

## Article

# Optimal Statistical Analyses of Bell Experiments

Richard D. Gill 

Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands;  
gill@math.leidenuniv.nl

**Abstract:** We show how both smaller and more reliable  $p$ -values can be computed in Bell-type experiments by using statistical deviations from no-signalling equalities to reduce statistical noise in the estimation of Bell's  $S$  or Eberhard's  $J$ . Further improvement was obtained by using the Wilks likelihood ratio test based on the four tetranomially distributed vectors of counts of the four different outcome combinations, one 4-vector for each of the four setting combinations. The methodology was illustrated by application to the loophole-free Bell experiments of 2015 and 2016 performed in Delft and Munich, at NIST, and in Vienna, respectively, and also to the earlier (1998) Innsbruck experiment of Weihs et al. and the recent (2022) Munich experiment of Zhang et al., which investigates the use of a loophole-free Bell experiment as part of a protocol for device-independent quantum key distribution (DIQKD).

**Keywords:** Bell experiment; quantum entanglement; Bell-CHSH inequality; maximum likelihood estimation; testing statistical hypotheses; Wilks generalized likelihood ratio test

## 1. Introduction

In 2015 and 2016, four famous “loophole free Bell experiments” were performed, all of which produced statistically significant violations of Bell-CHSH (or related) inequalities. Their results are published in the papers by Hensen et al. (2015) [1], Rosenfeld et al. (2017) [2], Giustina et al. (2015) [3], and Shalm et al. (2015) [4]: the Delft, Munich, Vienna, and NIST experiments, respectively. The first two and the second two are strongly related. The last two, Vienna and NIST, both actually used the Eberhard inequality. The experiments in Delft and Munich had another special feature—the use of entanglement swapping for heralded entanglement generation—which we will mention later.

All four experiments have been criticized on various grounds, especially concerning an imperfect randomized choice of settings, and the drift of experimental parameters over time. The experimenters were themselves aware of these issues and used martingale based tests, instead of the traditional ones, to neutralize some of the problems. Anyway, later experiments have rectified many claimed defects. Nowadays, a loophole-free Bell test is part of the standard methodology for device-independent quantum key distribution (DIQKD). Naturally, more complex experiments have many more sore points for skeptics to point their fingers at; this is clearly just another episode in a never-ending story. In this paper, we do not enter into any of these discussions. Rather, we made the working assumption that each of the four experiments was performed in a close enough to ideal way that, for each sub-experiment corresponding to one of the four possible setting pairs, we had data which we may think of as being made up of independent and identically-distributed pairs of outcomes. We assumed that the experiment satisfied the (surface level) no-signalling property, namely that the probabilities of the two possible outcomes in each wing of the experiment, given the two settings applied in both wings, only depend on the setting in the wing under consideration.

This defines a simple statistical problem in which we have just four independent tetranomially distributed observations (multinomial with four categories), where four linear constraints (no-signalling) are known to hold on the sixteen probabilities parametrizing the



**Citation:** Gill, R.D. Optimal Statistical Analyses of Bell Experiments. *AppliedMath* **2023**, *3*, 446–460. <https://doi.org/10.3390/appliedmath3020023>

Received: 27 March 2023

Revised: 1 May 2023

Accepted: 5 May 2023

Published: 16 May 2023



**Copyright:** © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

four tetranomially distributed observations. Moreover, we wished to test a null-hypothesis of a further eight linear constraints: the hypothesis of local realism is, by Fine's theorem, equivalent to satisfaction of the eight one-sided Bell-CHSH inequalities. Now, had the constraints been linear constraints on the logarithms of the probabilities, statistical estimation and testing would have been computationally easy. However, linear constraints on the probabilities themselves forced us to undertake more work, and put us in a non-standard situation. Asymptotically optimal estimates and asymptotically-optimal tests of hypotheses cannot be written down in closed-form expressions, but numerical optimization turned out to be quite easy. A tricky point is that, if the experiment is a good one, estimates of the parameters assuming the null-hypothesis of local realism to be true will usually lie on the boundary of the parameter space. Wilks' statistic (twice the difference of the maximized log likelihoods) will not have the standard asymptotic null-hypothesis distribution. Instead of the chi-square (1) distribution, we will have, if the truth is indeed on the boundary, a 50–50 mixture of chi-square (1) and chi-square (0), because at such points, asymptotically, the optimal estimate of Bell's  $S$  without assuming local realism would, half the time be larger than 2 and half the time smaller than 2.

In this paper, we explore the relation between the various possible tests of local realism, showing that, in principle, under the standard assumptions, much better  $p$ -values could have been obtained in all four experiments with little extra computational effort. We solved the computational issues or, at least, avoided them, using a modern statistical methodology which physicists generally are not aware of. The  $p$ -values obtained with the Wilks test turn out to be smaller than those obtained by comparing an estimate to an estimated standard error, and there are good reasons to believe that they are actually more accurate, too.

The main point is that statistical deviations from no-signalling equalities are statistically correlated with statistical variation around the theoretical values of the Bell-CHSH statistic  $S$  or the Eberhard statistic  $J$ . Hence, one can improve the observed "naive" values of  $S$  or  $J$  by subtracting a prediction of the statistical error in the observed value, based on the *observed* statistical deviations from the four no-signalling equalities. This is not difficult to do, and moreover leads to other ways to improve the accuracy of the statistical estimates and tests.

We go on to look at one older and one much newer experiment: those of Weihs et al. (1998) [5] and of Zhang et al. (2022) [6]. We show that the control of the randomization of setting choices has reached an unparalleled perfection in the latest experiment. This makes statistical violations of no-signalling at the manifest level (correlation between Alice's setting and Bob's outcome) a thing of the past. We suggest that it was a spurious correlation caused by the hidden confounder "time". The physical parameters both of random setting generators and of source and transmission lines and of detectors can drift in time. With proper randomization of the settings, one is protected against drifts and jumps and correlation over time in the rest of the experiment. Instead of relying on assumptions which are unlikely to be true, we can design experiments whose statistical assumptions are guaranteed by the experimenter's own procedures.

## 2. Background: The Physics Story

For those not familiar with the physics background, here is a potted history of Bell's theorem and the notable experiments on what is now called "quantum non-locality", which led to the 2022 Nobel prize in physics for John Clauser, Alain Aspect, and Anton Zeilinger. John Bell himself, the star of the story, unfortunately died quite young and unexpectedly in 1990. The main purpose of this section is to put a number of key papers into the bibliography in order to help the reader who is blissfully ignorant of this backstory but would like to know more to orient themselves. The author has written one survey paper on statistical issues in Bell experiments, Gill (2014) [7], written shortly before the miraculous year of 2015, with the first successful loophole-free Bell experiment, immediately followed by three more. I do not go into any nitty gritty of the quantum mechanics (QM) framework of states, observables, measurements, and time evolution. The point to remember is

that QM does not explain what actually happens when quantum systems are measured. It only tells the physicist what the statistics will be of repeating the same preparation and measurement many times. Since the birth of quantum mechanics, this has been a deep cause of discomfort, mystery, and debate; today, some physicists still search for an underlying theory of a more classical nature which would actually explain the randomness in observations of quantum systems as merely the reflection of deterministic processes with initial conditions which cannot be controlled in any way. One path to the unification of relativity theory and quantum mechanics would be a description of quantum mechanics as a collection of emergent phenomena arising from a deeper hidden level where more classical physical rules are followed. Others believe that determinism must give way and relativity theory will need adjustment. Many other standpoints are possible.

The story may start with the paper by Einstein, Podolsky, and Rosen (1935) [8], in which it was argued that quantum mechanics was either wrong or incomplete. A thought experiment involving the measurement of either the position or of the momentum of two particles in a so-called singlet state showed that each particle possessed definite values of both properties while, according to quantum mechanics, a particle only received a definite value of either property, after it was measured in an appropriate measurement set-up. The assumption in the EPR argument was a locality assumption: measuring one particle could not have any influence on another, distant, particle. Discussion of the foundations of quantum mechanics subsided, under the influence of its enormous success and Feynman's famous dictum, "shut up and calculate". A few stubborn individuals did continue to think and to question. David Bohm converted the EPR thought experiment into an experiment concerning the spin of two entangled spin-half particles; one can measure the spin of such particles in any chosen direction but the outcome of the measurement is binary: the particle, as it were, chooses either the direction set by the experimenter or the opposite direction. Then came Bell's famous (1964) [9] paper, taking the EPR-B model and now adding a new twist: instead of only the same two possible measurements on each particle, he considered several different possible measurements on each. EPR had concluded, assuming local realism, that QM is either wrong or incomplete. Bell's conclusion was the more shocking: QM is either wrong or non local.

Bell's thought experiment was, in 1964, far from being experimentally feasible. A few wild spirits became interested and started working towards experimental testing. Their actual expectation was that quantum entanglement would rapidly decay as particles moved further apart. They did not expect to see the signature of quantum entanglement in the measurements of particles widely separated in space. In order to get closer to experimental test, the EPR-B model was transposed from the spin of spin-half particles to the polarization of photons (each offered the choice between two perpendicular 2D orientations, instead of the choice between two opposite 2D directions). Bell's original inequality was also generalized to what is now called the CHSH inequality, after Clauser, Horne, Shimony, and Holt (1969) [10]. A first experiment in which the violation of Bell inequalities was observed was performed by Freedman and Clauser (1972) [11]. A big defect was that the settings of the polarizers were kept fixed for many consecutive photon pairs. Thus, each photon had plenty of time to know how both were going to be measured. In a now world-famous experiment, Aspect et al. (1982) [12] managed to observe a statistically significant violation of Bell-CHSH inequalities while measurement settings were chosen while the photons were in flight. Further experiments made further refinements. For a while, the most impressive was Weihs et al. (1998) [5]. However, a big defect in all these experiments was what is called the detection loophole. In Weihs' experiment, it appeared that only one in twenty photons made it from source to detection, so only one in four hundred emitted photon pairs resulted in measurements of both their polarizations. Already, Pearle (1970) [13] had shown that quantum correlations could be faked using a classical and local mechanism if enough particles did not show up at the detectors, and Garg and Mermin (1987) [14] showed that the critical detection rate is 83% in an Aspect- or Weihs-type experiment with maximally-entangled photon polarizations. That is a very long way to go from 5%.

The subsequent decades were spent on working towards so called loophole-free Bell experiments in which statistics were obtained, predicted by QM, and impossible to be explained by a classical physical mechanism without recourse to superluminal messaging or even more outlandish explanations. The big breakthrough came in 2015 with four experiments carried out in Delft, Munich, Vienna, and at NIST (Boulder, CO, USA). The Delft and Munich experiments used a novel technology called entanglement swapping, developed by Zeilinger and others in preceding decennia. These experiments had no “no shows” at all, but rather small sample sizes. The Vienna and NIST experiments used an alternative to Bell’s inequality called the Eberhard (1993) [15] inequality, which used the clever device of detecting particles polarized in one orientation only, merging all “no shows” with particles which with the perpendicular polarization, thus resulting in guaranteed binary outcomes. Eberhard had discovered paradoxically that, using less than maximally entangled photons, one could get away with a 67% detection rate, so detector efficiency need not be as high as for the old-style (Aspect, Weihs) experiments.

The 2015 experiments were not perfect and various defects need to be honed away, but the net impact of four resounding confirmations of Bell’s genial discovery was enough for the 2022 Nobel prize committee. Research continues on using an embedded loophole-free Bell experiment as part of a protocol for creating shared secret random keys at two distant locations while communicating over public communication channels. The most promising technology is that based on the Delft and Munich experiments. Here, two distant “solid state” stationary qubits are brought into quantum entanglement by having both emit a photon which meet one another and interfere at a third intermediate location, where a third collaborator, Charlie, measures the two photons after they have interfered and reports his findings to Alice and Bob, who at the same time were measuring their qubit in one of several ways. Alice and Bob study the statistics of the measurement outcomes and settings corresponding just to those occasions when Charlie obtained a certain measurement outcome. Just as one can generate statistical dependence between originally independent random variables  $X$  and  $Y$  by conditioning on a function of  $X$  and  $Y$ , it is also possible to generate quantum entanglement between quantum systems which have never physically interacted with one another by conditioning on a measurement outcome of two emitted particles which have interacted at a third location.

A complicated protocol now, in principle, allows them either to determine that there has been no interference in their communications and to distil some number of secret shared random bits, or to detect interference or imperfection and abort the process. Input into all these experiments consists, preferably, of independent, local, and completely random setting choices. In many experiments, physical random number generators have been used, whose properties tend to slowly drift as time goes by (the experiment might last several days), and occasionally moreover jump when shocks occur (e.g., a lorry crosses the campus). At the same time, the same external processes are causing drifts and jumps in the physics of source, transmission lines, and detectors. The result can be a spurious correlation between, for instance, Alice’s settings and Bob’s outcome, even though Alice’s settings could not have reached Bob’s apparatus in time to influence the measurement outcome. Both are influenced by a hidden confounder: time. Plenty of techniques are available to discount a certain amount of deviation from complete randomness in setting choices, but this leads to less transparent results, depending moreover on assumed limits on the amount of bias.

### 3. The Statistical Model

In a standard ideal Bell experiment, two separated experimenters, Alice and Bob, each repeatedly insert a binary setting into some apparatus and a short time after observe a binary outcome. Alice and Bob work in a carefully synchronized way, such that each setting of Alice could not reach Bob’s lab before Bob’s outcome was registered, even if travelling at the speed of light and vice-versa. Alice and Bob might be inserting settings which were somehow generated “on demand” by some auxiliary randomization procedure, whether

physical or algorithmic. Alternatively, they might be reading off settings one at a time from a pre-generated database. All of these possibilities have advantages and disadvantages which we do not discuss here. We will use the word “trial” to denote one set of four binary outcomes, namely a setting for each of Alice and Bob, and an outcome for each of Alice and Bob. In the first instance, the experiment generates an  $N \times 4$  spreadsheet of settings  $a, b$ , taking values, say, in the set  $\{1, 2\}$  and outcomes  $x, y$ , taking values, say, in the set  $\{-1, +1\}$ .

Given the pair of settings  $(a, b)$  used in just one trial, we consider the pair of outcomes as being the realizations of two Rademacher (i.e.,  $\pm 1$  valued) random variables with a joint probability distribution which depends only on  $(a, b)$ ; given all the settings, we considered all the pairs of Rademachers as being independent of one another. With sufficiency, we may reduce the data to the sixteen counts  $N(x, y | a, b)$  of trials with outcomes  $(x, y)$  and settings  $(a, b)$ . Grouping these according to the settings, we have four realizations of four independent tetranomially-distributed random vectors  $(N(x, y | a, b) : x, y \in \{-1, +1\})$ , for  $(a, b)$  in  $\{1, 2\} \times \{1, 2\}$ . By definition of the multinomial distribution and by conditioning the settings of all the trials, the sums  $n_{ab} = N_{ab} := \sum_{x,y} N(x, y | a, b)$  are fixed. For each  $(a, b)$ , the probability distribution of the 4-vector  $(N(x, y | a, b) : x, y \in \{-1, +1\})$  is the multinomial distribution with number of cells = 4, number of trials =  $n_{ab}$ , and multinomial probabilities  $p_{ab} = (p(xy | ab), x, y = -1, +1)$ ; a vector of four probabilities adding to one.

We did expect a number of constraints to hold on the 16 probabilities  $p(xy | a, b)$ . Obviously, they add up to +1 in groups of four. These constraints are called the normalization constraints. Less obviously, we have the *no-signalling* constraints. For a well-conducted experiment, we believe and we will moreover assume that, given all settings, the marginal probability distribution of Alice’s outcomes does not depend on Bob’s setting and vice versa. Using a “+” to denote addition over all values of a given argument, we assume that  $p_a(x) := p(x, + | a, b) = p(x, + | a, b')$  for all  $a$  and all  $b \neq b'$ , and  $q_b(y) := p(+, y | a, b) = p(+, y | a', b)$  for all  $b$  and all  $a \neq a'$ . These equations are the so-called no-signalling equalities. A little thought shows that, because of the normalization constraints (probabilities add up to +1) and the no-signalling constraints, our 16 probabilities  $p(x, y | a, b)$  depend on just eight free parameters. We can take them as the four marginal probabilities of the outcome +1 given the local setting on each side of the experiment  $p_1(+1), p_2(+1), q_1(+1), q_2(+1)$ , and the four correlations

$$\rho_{ab} := p(+1, +1 | a, b) + p(-1, -1 | a, b) - p(+1, -1 | a, b) - p(-1, +1 | a, b).$$

To be specific,

$$p(x, y | a, b) = \frac{1}{4} + \frac{1}{2}(p_a(x) - \frac{1}{2}) + \frac{1}{2}(q_b(y) - \frac{1}{2}) \pm \frac{\rho_{ab}}{4}$$

where the ‘ $\pm$ ’ sign is ‘+’ if  $a = b$  and ‘−’ if  $a \neq b$ , and  $p_a(-1) := 1 - p_a(+1)$ ,  $q_b(-1) := 1 - q_b(+1)$ . The eight parameters vary freely in the sense that they vary in an eight-dimensional closed convex polytope with non-empty interior, bounded by the hyperplanes determined by the non-negativity of the  $p(x, y | a, b)$ . Another way to say this is that the vector of all 16 probabilities  $p(x, y | a, b)$  lies in a closed convex polytope in an eight-dimensional affine subspace of  $\mathbb{R}^{16}$ , with non-empty relative interior. It is called the no-signalling polytope.

As is well known, according to quantum mechanics, the possibilities are limited to a strictly smaller closed convex subset called the quantum body and, according to local realism, they are limited even further to a polytope called the local realism polytope. The two smaller sets are both full, having a non-empty relative interior relative to the eight-dimensional affine subspace in which all are constrained to lie. For instance, the point where all sixteen probabilities equal  $\frac{1}{4}$  lies in all of their relative interiors.

Each of these three convex sets are the convex hull of their boundary, and the boundaries of the two polytopes are finite sets of points.

This has defined a nice statistical model for four independent tetranomially-distributed random vectors; the only unusual feature (relative to standard statistical theory) is the no-signalling constraints on the mean vectors of the four observations. A further non-standard feature is that we are interested in testing the null hypothesis of local realism against the alternative of quantum mechanics. We have non-standard estimation problems and a non-standard testing problem.

#### 4. The Methodology

As we have seen, a standard Bell-type experiment with

- two parties,
- two measurement settings per party,
- two possible outcomes per measurement setting per party,

generates a vector of  $16 = 4 \times 4$  numbers of outcome combinations per setting combination.

As is now well known, this can be applied to two-channel experiments without a detection loophole, but also to one-channel experiments and (equivalently) to two-channel experiments with  $-1$  and “no-detection” combined, as long as the experimental units are “time-slots”. The four sets of four counts can be thought of as four observations each of a tetranomially-distributed vector over four categories.

We will rewrite what we discussed in the previous section in a different notation, more convenient for converting formulas into programming code in the language *R* or any other modern programming language. Write  $X_{ij}$  for the number of times outcome combination  $j$  was observed, when setting combination  $i$  was in force. Let  $n_i$  be the total number of trials with the  $i$ th setting combination. The four random vectors  $\vec{X}_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})$ ,  $i = 1, 2, 3, 4$ , are independent each with a multinomial  $(n_i; \vec{p}_i)$  distribution, where  $\vec{p}_i = (p_{i1}, p_{i2}, p_{i3}, p_{i4})$ .

The 16 probabilities  $p_{ij}$  can be estimated by relative frequencies  $\hat{p}_{ij} = X_{ij}/n_i$ , which have the following variances and covariances:

$$\begin{aligned}\text{var}(\hat{p}_{ij}) &= p_{ij}(1 - p_{ij})/n_i, \\ \text{cov}(\hat{p}_{ij}, \hat{p}_{i'j'}) &= -p_{ij}p_{i'j'}/n_i \quad \text{for } j \neq j', \\ \text{cov}(\hat{p}_{ij}, \hat{p}_{i'j'}) &= 0 \quad \text{for } i \neq i'.\end{aligned}$$

These variances and covariances can be arranged in a  $16 \times 16$  block diagonal matrix  $\Sigma$  of four  $4 \times 4$  diagonal blocks of non-zero elements.

Arrange the 16 estimated probabilities and their true values correspondingly in (column) vectors of length 16. I will denote these simply by  $\hat{p}$  and  $p$ , respectively. We have  $E(\hat{p}) = p \in \mathbb{R}^{16}$  and  $\text{cov}(\hat{p}) = \Sigma \in \mathbb{R}^{16 \times 16}$ .

We are interested in the value of one particular linear combination of  $p_{ij}$ ; let us denote it by  $\theta = a^\top p$ . The vector  $a$  might specify the CHSH quantity  $S$ , or Eberhard's  $J$ . We know that four other particular linear combinations are identically equal to zero: the so-called no-signalling conditions. This can be expressed as  $B^\top p = 0$ , where the  $16 \times 4$  matrix  $B$  contains, as its four columns, the coefficients of the four linear combinations. We can sensibly estimate  $\theta$  by  $\hat{\theta} = a^\top \hat{p} - c^\top B^\top \hat{p}$ , where  $c$  is any vector of dimension 4. For whatever choice we make,  $E\hat{\theta} = \theta$ . We propose to choose  $c$  so as to minimize the variance of the estimator. This minimization problem is an elementary problem from statistics and linear algebra (“least squares”). Define

$$\begin{aligned}\text{var}(a^\top \hat{p}) &= a^\top \Sigma a = \Sigma_{aa}, \\ \text{cov}(a^\top \hat{p}, B^\top \hat{p}) &= a^\top \Sigma B = \Sigma_{aB}, \\ \text{var}(B^\top \hat{p}) &= B^\top \Sigma B = \Sigma_{BB};\end{aligned}$$

then the optimal choice for  $c$  is

$$c_{\text{opt}} = \Sigma_{aB} \Sigma_{BB}^{-1}$$



leading to the optimal variance

$$\Sigma_{aa} - \Sigma_{aB} \Sigma_{BB}^{-1} \Sigma_{Ba}.$$

In the experimental situation, we do not know  $p$  in advance, hence we also do not know  $\Sigma$  in advance. However, we can estimate it in the obvious way (“plug-in”), and for  $n_i \rightarrow \infty$  we will have an asymptotic normal distribution for our “approximately best” Bell inequality estimate, with an asymptotic variance which can be estimated using the natural “plug-in” procedure, leading again to asymptotic confidence intervals, estimated standard errors, and so on. The asymptotic width of this confidence interval is the smallest possible and, correspondingly, the number of standard errors deviation from “local realism” the largest possible. The fact that  $c$  is not known in advance does not harm these results.

The methodology is called “generalized least squares”.

One can go further. It is sensible to use these estimates as the starting point of Newton–Raphson iterations searching for the maximum over the two polytopes of interest for the multinomial log likelihood instead of the quadratic loss function. Subsequently, we may compute the Wilk’s generalized log likelihood ratio test, evaluated through its asymptotic chi-square distribution (actually, because of boundary issues, a mixture of chi-square distributions with different numbers of degrees of freedom). In fact, as far as asymptotic results are concerned, just a single Newton–Raphson iteration should produce asymptotically optimal estimates. Switching from generalized least squares to minimize a variance to one-step Newton–Raphson on the log likelihood maximum likelihood and then to true maximum likelihood typically gives a better approximation, at each step, of the asymptotic distribution. Asymptotically, all three are equivalent. We investigated what happens when we indeed try to obtain more reliable estimates and tests in this way.

## 5. The Results

We first performed these computations on the data sets of the four famous loophole-free Bell tests of 2015. The analyses were performed using scripts written in the *R* language [16], and published on the website <https://rpubs.com/gill1109> (accessed on 4 May 2023) using the IDE *RStudio*. For each of the four experiments, we performed one analysis computing the CHSH quantity  $S$  and one computing Eberhard’s  $J$ ; we computed standard errors and  $p$ -values using the asymptotic normality of the multinomial distribution; then, we computed an optimized version of  $S$  and  $J$  by subtracting the linear combination of the four observed deviations from the no-signalling equalities, which maximally reduced the (estimated) variance, thus leading to a maximal  $p$ -value. As must be the case, the optimized  $S$  and  $J$  were related by the theoretical identity  $S = 2 + 4J$ , and the resulting standard errors differed by a factor of 4; the  $p$ -values based on asymptotic normality were identical. The  $p$ -values were all approximate, being based on large sample multivariate normal approximations to the distribution of the 16 raw counts and estimated covariance matrices.

After that, we attempted to estimate  $S$  and  $J$  by maximum likelihood for four independent tetranomially-distributed vectors of counts; tests of the hypotheses of interest were then computed using the generalized likelihood ratio test and its asymptotic distribution: a 50–50 mixture of chi-squared (1) and chi-squared (0) distributions. Thus, we avoided the initial step of approximation of multinomials by multivariate normals. In the case of NIST and Vienna, this led to numerical problems: the numbers of trials were so large that numerical optimization starting at the earlier obtained failed to be able to improve the initial value of the function being maximized. From the numerical point of view, the earlier obtained estimates were already indistinguishable from maximum likelihood estimates. The numerical issues reported by the software was in fact a non-problem. Indeed, the sample sizes were so large in these two experiments that systematic violation of model assumptions was much more important than the statistical variation of observed counts. Our assumptions of constant physical parameters throughout the whole run are possibly wrong. Time variation in the physics and, in particular, time drifts in the physical random number generation of settings, means that, to a small extent, “no signalling” is violated:

from observing her local statistics, Alice could, in principle, to a tiny extent, guess Bob's settings better than assuming independence and constant probabilities. This can, to a large extent, be taken care of by using martingale-based tests based on assumptions about the randomness of the settings, instead of tests based on assumptions that trials (under the same pair of settings) are i.i.d. We will return to that option at the end of the paper when we discuss the Zhang et al. (2022) experiment.

We published the results of running eight *R* scripts back in 2019. They can be found on the following web pages (accessed: 4 May 2023):

[https://rpubs.com/gill1109/OptimisedDelft\\_2](https://rpubs.com/gill1109/OptimisedDelft_2);  
<https://rpubs.com/gill1109/OptimizedMunich>;  
<https://rpubs.com/gill1109/OptimizedNIST>;  
<https://rpubs.com/gill1109/OptimizedVienna>;  
<https://rpubs.com/gill1109/AdvancedDelft>;  
<https://rpubs.com/gill1109/AdvancedMunich>;  
<https://rpubs.com/gill1109/AdvancedNIST>;  
<https://rpubs.com/gill1109/AdvancedVienna>.

The addition of “underscore 2” means that an original document from 2019 has been improved in 2022 with some minor editing. We will next discuss our findings for the Delft experiment at some length, and then point out any notable features of the analyses of the other three experiments.

We also added similar analyses of the Weihs et al. (1998) [5] experiment in Innsbruck and of the Zhang et al. (2022) [6] experiment (on DIQKD) in Munich (websites accessed: 4 May 2023):

<https://rpubs.com/gill1109/OptimizedWeihs>;  
<https://rpubs.com/gill1109/AdvancedWeihs>;  
<https://rpubs.com/gill1109/OptimizedDIQKD>;  
<https://rpubs.com/gill1109/AdvancedDIQKD>.

## 6. The Experiments

The appendix of the paper contains, for each experiment discussed, tables with the observed frequencies (counts) and relative frequencies. That means, per experiment, both for the raw counts and for the relative frequencies, four  $2 \times 2$  tables. Displayed this way, one can see the nice patterns of symmetries in the relative frequencies, as well as other notable features. The reader who wants to do further analyses with this data is advised to take it from the *R* scripts listed above.

### 6.1. Delft

We proceed to read off some of the statistical results of our analysis [https://rpubs.com/gill1109/OptimisedDelft\\_2](https://rpubs.com/gill1109/OptimisedDelft_2) (accessed on 4 May 2023) of the Delft experiment, see Table A1. One observes  $S = 2.4225$  with an estimated standard error of 0.2038266, giving a *z*-value of 2.07284 and an approximate *p*-value of 0.0190936. In round numbers,  $S = 2.42(0.21)$ , giving us a *p*-value of 0.02. We can slightly reduce the standard error and the *p*-value by optimally subtracting noise, by assuming that in reality no-signalling is true (the probabilities of Alice's outcomes do not depend on Bob's setting and vice-versa). The relative frequencies do vary slightly with the settings on the other side. This is (under our model assumptions) pure noise, but it is pure noise which is correlated with the noise in the estimate of  $S$ . The slightly better estimate (in the sense of lower variance) is  $S = 2.462658$  or in round numbers  $S = 2.46(0.20)$  and the *p*-value is now 0.01.

All *p*-values here are approximate: they assume that the asymptotic normal distributions of the statistics give a good approximation to the actual distribution, and the asymptotic theory is conditional on our assumption of four independent tetranomially-distributed observed count vectors.

Lovers of the Eberhard test might prefer to look at Eberhard's  $J$ , which takes the value 0.1195162. Under our assumptions,  $E(S) = 2 + 4E(J)$ , so the observed values of  $J$  and  $S$



do correspond nicely. However, for these data,  $J$  has a much higher estimated variance than  $S/4$ . Its estimated standard error is 0.09475703, leading to a  $z$ -value of just larger than 1 and a  $p$ -value of about 10%, which is terrible compared to that of  $S$ , namely, about 1%. When we improve  $J$  using the same noise reduction strategy as we applied to  $S$ , we end up with an improved estimate of  $J$  exactly equal to the improved estimate of  $S$ , divided by 4; and the same  $p$ -value.

In all cases, the experiments all exhibit a violation of the one-sided Bell-CHSH inequality in which three correlations were added and one was subtracted. By recoding outcome labels if necessary, we have arranged that the exceptional correlation (large, negative) corresponds to the setting pair (2, 2); the other three correlations are large and positive.

Next, we took a look at the script <https://rpubs.com/gill1109/AdvancedDelft> (accessed on 4 May 2023). Here, we stuck to multinomial distributions. We assumed no-signalling and estimated the 16 probabilities corresponding to the 16 relative frequencies by maximizing the likelihood (a) without any restriction and (b) assuming local realism.

The two numerical optimizations gave no problems. We took as initial estimates the estimates we obtained before, using approximate normality. Without assuming local realism, there are eight free parameters: each of the four tables has its own correlation; then there are the marginal probabilities of Alice's outcome + under each of Alice's settings, and similarly for Bob. We estimated  $S$  as the sum of three of the correlations minus the fourth;  $J$  was estimated by  $(S - 2)/4$ . Under local realism, the maximum likelihood estimate of  $S$  (corresponding to adding the first three correlations and subtracting the fourth) exactly equals 2: the data quite strongly violate the corresponding one-sided CHSH inequality. The fourth correlation is an affine function of the other three. We had to optimize over seven free parameters, forcing  $S = 2$ .

We tested the hypothesis of local realism by comparing the maximized log likelihood in the two situations. Large sample theory (Wilks' test) tells us that, under the null hypothesis, twice the difference should have a mixture of 50–50 of a chi-square distribution with one degree of freedom, and 50–50 identically equal to zero. The latter case occurs when the unconstrained estimate of the eight parameters is a point inside the null-hypothesis. This means that, in our case, our  $p$ -value according to the Wilks test is  $0.04704162/2 = 0.02352081$ , or in round numbers, a  $p$ -value of 2.4%.

This seems less attractive than the previously obtained 1%; however, experience shows that the  $p$ -value obtained from the log likelihood ratio-based Wilks test is a better asymptotic approximation than the  $p$ -value based on a  $z$ -value based on the approximate normal distribution of an estimator. Certainly, a shorter chain of approximations is being created when we stick closer to the underlying multinomial distributions. In the asymptotic theory, the asymptotic chi-square (1) distribution does correspond to squaring an asymptotically standard, normally distributed quantity, so everything is still based on the central limit theorem and the law of large numbers. Still, there are sound theoretical explanations for the aforementioned practical experience, based on higher order asymptotic theory, where one also looks at the rate of convergence to asymptotic normal distributions. Essentially, the Wilks approach eliminates a second order term due to skewness. The statistical approach automatically eliminates a possible asymmetry leading to a nonzero coefficient of skewness in the estimators. The Wilks' approach is invariant under reparametrization; whereas many unfortunate parametrizations lead to skewness.

## 6.2. Munich

In this experiment, see Table A2, the observed value of  $S$  was the impressive 2.609047, with estimated standard error 0.2484456. In round numbers,  $S = 2.61(0.25)$ . This resulted in a very nice  $p$ -value of 0.007114475, or in a round number, 0.7%. The optimized value of  $S$  was 2.582261, so slightly less. As must be the case, its estimated standard error was slightly smaller too, resulting in much the same  $p$ -value, 0.008782296, which one should report in a round number as 0.9%

Maximum likelihood estimation based on the multinomial likelihood worked without a hitch. The Wilks test gave a  $p$ -value of  $0.04104834/2$ , which we can fairly report as 2%. Again, theory and experience suggest that this is a rather more reliable number than the  $p$ -values based on a  $z$ -statistic.

### 6.3. Nist

Now, see Table A3, we find  $S = 2.000092$  with a standard error of  $1.572689 \times 10^{-5}$ . This gives us a  $z$ -value of  $92/15.7 = 5.859873$ . The  $p$ -value is astronomically small  $2.062969 \times 10^{-9}$ . Eberhard's  $J$  would be preferred by many for this situation. It gives a  $z$ -value of 4.778576 and a  $p$ -value of  $8.827054 \times 10^{-7}$ . Optimizing the variance gives  $S = 2.000051$ , so closer to 2, but of course also a smaller standard error, and in this case a much smaller standard error, resulting in a  $z$ -value of 7.637903 and a  $p$ -value of  $1.110193 \times 10^{-14}$ . However, all these  $p$ -values need to be taken with a very large pinch of salt, since the convergence to asymptotic normality is worse and worse, in terms of relative error in true and approximated tail values, the further in the tail we are. However, certainly, it would have been nice in the published paper to talk about 7.6 standard errors deviation from local realism rather than 4.8.

Maximum likelihood based on multinomial distributions resulted in warnings being issued by the numerical procedure. Essentially, the optimization of the log likelihood over seven or eight parameters does not succeed in moving the estimates away from values found by generalized least squares after switching to multivariate normal distributions, so a standard numerical optimization program just gives up after a while, after uttering a lot of complaints; it cannot improve on the initial estimates by an amount greater than expected numerical accuracy. No matter. Still, the Wilks test provides a perhaps slightly more reliable  $p$ -value than the  $z$ -tests we already discussed. The Wilks test statistic comes out as 57.19689 corresponding to a  $z$ -value of the square root of that number, 7.562863; so the message is 7.6 standard errors, and a  $p$ -value of  $1.971474 \times 10^{-14}$ .

### 6.4. Vienna

This experiment, see Table A4, had about a 10 times larger sample size than NIST. We find  $S = 2.000028$ , so the difference with the local realism bound of 2 is three times smaller than for NIST. The standard error is  $3.283419 \times 10^{-14}$ , so the  $z$ -value is 8.527696. Optimizing the variance reduces it by a factor of 2 while leaving the estimate of  $S$  essentially unaltered, so finally we get a square root of 2 times larger  $z$ -value of just above 12.

Maximum likelihood using the original multinomial has exactly the same numerical issues (or if you like – non-issues) as in the case of NIST. It comes up with an even larger  $z$ -value than that obtained in the method based on optimizing the estimate of  $S$  by minimizing its variance, through using the statistical deviations from no-signalling to reduce the error in  $S$ . In fact, we now get a  $z$ -value of 17.5, and experience and theory says this is more reliable than the  $z$ -values mentioned so far.

### 6.5. Weihs et al.

Just for fun, we also carried out our analyses for this earlier (1998) and quite famous experiment, see Table A5. It was for a long time the definitive Bell type experiment, having a large enough distance between Alice and Bob that the “locality loophole” was closed. However, there remained a very serious “detection loophole”. Thinking in terms of photon pairs emitted from the source, only one in 20 of the photons resulted in a detection event, and only 1 in 400 emissions of a photon pair resulted in two detections. In order to conclude that local realism has been disproved by this experiment, one must make the “fair sampling assumption” that detection of photons is independent of those hidden variables and independent of the settings. This experiment has  $S = 2.73$  and estimating it optimally hardly changes it, resulting in  $S = 2.71$ . This is more than  $z = 25$  estimated standard deviations away from the local realism bound  $S \leq 2$ .

### 6.6. Zhang et al.

The paper [6] was an attempt to show the feasibility of using a loophole-free Bell experiment as a component in a procedure for “device independent quantum key distribution”. For the data, see Table A6, Alice and Bob actually choose randomly between three settings, in such a way that some of the trials are performed with equal (or equal and opposite) settings. The experiment was actually a three-party experiment similar to the earlier experiments at Delft and Munich. One studies correlations between Alice and Bob’s outcomes conditional on a particular outcome having been obtained by Charlie. This is sometimes mistaken for post-selection, but it is not. Always one studies experimental data after the experiment is completed, so if, for instance, one looks at the correlation between Alice and Bob’s outcomes when their settings are, say, (1,1), one is only looking at selected trials. In order for these three party experiments to be loophole-free there must be no locality loophole concerning the three parties. In particular, Charlie’s outcomes must not be able to influence Alice’s or Bob’s measurement apparatus during the pre-allocated time-slots of the three parties. In the experiment [6], Alice and Charlie shared a lab, in fact, they shared a lot of the electronics, so this was not actually a loophole-free experiment.

What is rather nice about this experiment is that it seems that the random generation of settings has become very stable and close to unbiased. Our procedures for optimizing CHSH led to almost no improvement in  $p$ -value. Because of the near perfect experimental symmetry and the lack of drifts in physical parameters over time, the statistical deviations from no-signalling are hardly correlated with  $S$  at all. The experiment has  $S = 2.58$  with  $z \approx 7$ . That gives, of course, a tiny  $p$ -value of  $3 \times 10^{-12}$  but this should not be taken too seriously. Since  $7^2 = 49 \approx 50$ , a conservative but much more reliable  $p$ -value would be 0.01.

An alternative statistical analysis, if we assume the settings are chosen again and again by independent fair coin tosses, is based on a so-called martingale test, also known as the Bell game. One says that Alice and Bob have won each separate trial (in a game they play against nature) if their outcomes are equal and neither chose setting “2”, or their outcomes are opposite and both chose “2”. In this experiment, the number of wins was 1357 out of 1649 trials, notice that  $1357/1649 \approx 0.82$ . Under local realism (per trial, conditional on the past) the number of wins cannot have larger tail probabilities than those of the binomial distribution with  $N = 1649$  and  $p = 0.75$ . Under quantum mechanics, one could theoretically achieve  $p = 0.85$ , corresponding to Tsirelson’s bound  $S \leq 2\sqrt{2}$ . Thus, we can also obtain a  $p$ -value which is robust against violation of the independence and identical distributions assumption needed to reduce the data to multinomial counts. In this case, the Bell game  $p$ -value is the probability that a  $\text{Bin}(1649, 3/4)$  distributed random variable exceeds 1356; and that turns out to equal  $5 \times 10^{-13}$ . (The Bell-game test was used by the experimenters in Delft and Munich; the Delft group had refined and simplified results found by the present author, 20 years earlier. See the “supplementary material” of Hensen et al. (2015) [1].

We find this result rather exciting. Provided one has taken care of really good randomization of measurement settings, the martingale test is hardly different from the test based on a conventional calculation of  $z$ -values using approximate normality. The former provides, moreover, security against trends or jumps in the physical parameters of detectors. We suggest that past observations of locality violations at the manifest level of apparent correlations between Alice’s setting and Bob’s outcome, or vice versa, were simply manifestations of the statistical phenomenon of spurious correlations being caused by hidden confounders; the hidden confounder simply being time.

## 7. Conclusions

In a Bell test, if we are confident that there have not been shifts or jumps in the physics of the systems being studied during the course of the experiment, improved estimates of  $S$  or  $J$  are not difficult to obtain, and more reliable  $p$ -values can be found without much difficulty either. This can lead to big improvement of the results of experiments of the 2015 Vienna and NIST types: the Eberhard inequality is not the best test of local realism, by

a long way (though not surprisingly, it is better than CHSH). Of course, many scientists will be more convinced by a very simple and very robust estimation method. Rutherford said “if you need statistics you did the wrong experiment”. Well, some experiments do need statistics anyway. In that case one should process the data in the most efficient way possible using time honoured methods completely familiar to applied statisticians working in all fields of science. There is no excuse for the experimental physicist not to use the best tools available.

We remark that those astronomically small  $p$ -values in the Vienna and NIST experiments need to be taken with more than just a grain of salt. They are meaningless. The absolute error in the normal approximation must be huge compared to either actual or nominal (according to the normal distribution) value of these tail probabilities. In fact, we suggest that a meaningful, probably conservative,  $p$ -value is obtained from Chebyshev’s inequality. A  $z$ -value of 17.5 would correspond to a  $p$ -value of  $1/17.5^2 = 0.0033$  or 3 pro mille. On the other hand, the most recent and clearly best controlled Bell experiment to date, that of Zhang et al. (2022) [6], has essentially the same  $p$ -value independently of whether one uses a conventional estimation of standard error based on multinomial distributed counts and normal approximation to that distribution, or uses a martingale based statistic following the idea of the Bell game, which should insure the user against jumps and trends in the physical systems being studied over time. The close likeness of all the statistical test results suggests that there were hardly any shifts in time.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The author declares no conflict of interest.

## Appendix A

The appendix contains the summary statistics of the 2015 experiments of Delft, Munich (actually only published in 2017), NIST, Vienna, as well as that of Weihs et al. (1998) [5] at Innsbruck, and Zhang et al. (2022) [6] in Munich again. In each case we present the raw counts and the relative frequencies, per setting combination. Rows denote Alice’s outcomes, columns denote Bob’s outcomes; they are conventionally taken to be  $\pm 1$  in some of the experiments, and “d” and “n” standing for “detection” and “nondetection” in others. The settings have sometimes been reordered so that in cases, setting pairs (1,1), (1,2), (2,1) produce large counts (and large relative frequencies) on the diagonals, small off-diagonal; setting pair (2,2) has large numbers (and large relative frequencies) off-diagonal, small on the diagonal. In all cases, using a common notation, the sample value of the one-sided CHSH statistic  $S = \rho_{11} + \rho_{12} + \rho_{22} - \rho_{21}$  exceeds 2.

**Table A1.** Hensen et al.’s (2015) Experiment at Delft [1].

Raw counts					
	Settings (1,1)			Settings (1,2)	
	+	−		+	−
+	23	3	+	33	11
−	4	23	−	5	30
	Settings (2,1)			Settings (2,2)	
	+	−		+	−
+	22	10	+	4	20
−	6	24	−	21	6

**Table A1.** *Cont.*

Relative frequencies					
Settings (1,1)			Settings (1,2)		
	+	−		+	−
+	0.43	0.06	+	0.42	0.14
−	0.08	0.43	−	0.06	0.38
Settings (2,1)			Settings (2,2)		
	+	−		+	−
+	0.35	0.16	+	0.08	0.39
−	0.10	0.39	−	0.41	0.12

**Table A2.** Rosenfeld et al.'s (2017) Experiment at Munich [2].

Raw counts					
Settings (1,1)			Settings (1,2)		
	+	−		+	−
+	16	4	+	11	4
−	3	13	−	2	17
Settings (2,1)			Settings (2,2)		
	+	−		+	−
+	19	4	+	4	22
−	3	16	−	10	2
Relative frequencies					
Settings (1,1)			Settings (1,2)		
	+	−		+	−
+	0.44	0.11	+	0.32	0.12
−	0.08	0.36	−	0.06	0.50
Settings (2,1)			Settings (2,2)		
	+	−		+	−
+	0.45	0.10	+	0.11	0.58
−	0.07	0.38	−	0.26	0.05

**Table A3.** Shalm et al.'s (2015) Experiment at NIST (Boulder, Colorado) [4].

Raw counts					
Settings (1,1)			Settings (1,2)		
	d	n		d	n
d	6378	3282	d	6794	2821
n	3189	43,897,356	n	23,243	43,276,943
Settings (2,1)			Settings (2,2)		
	d	n		d	n
d	6486	21,334	d	106	27,539
n	2843	43,338,281	n	30,040	4,250,278
Relative frequencies $\times 10^6$					
Settings (1,1)			Settings (1,2)		
	d	n		d	n
d	145.25	74.74	d	156.87	65.14
n	72.63	999,707.38	n	536.67	999,241.33
Settings (2,1)			Settings (2,2)		
	d	n		d	n
d	149.55	491.92	d	2.49	647.06
n	65.55	999,292.97	n	705.82	998,644.63

The bottom right cell in all four  $2 \times 2$  tables is close to  $10^6$ , the difference is negligible.

**Table A4.** Giustina et al.'s (2015) Experiment at Vienna [3].

Raw counts					
	Settings (1,1)			Settings (1,2)	
	d	n		d	n
d	141,439	73,391	d	146,831	67,941
n	76,224	875,392,736	n	326,768	874,976,534
	Settings (2,1)			Settings (2,2)	
	d	n		d	n
d	158,338	425,067	d	8392	576,445
n	58,742	875,239,860	n	463,985	874,651,457
Relative frequencies $\times 10^6$					
	Settings (1,1)			Settings (1,2)	
	d	n		d	n
d	161.52	83.81	d	167.71	77.60
n	87.05	999,667.63	n	373.23	999,381.46
	Settings (2,1)			Settings (2,2)	
	d	n		d	n
d	180.78	485.30	d	9.58	658.27
n	67.07	999,266.86	n	529.84	998,802.30

The bottom right cell in all four  $2 \times 2$  tables is close to  $10^6$ , the difference is negligible.

**Table A5.** Weihs et al. (1998) Innsbruck Experiment [5].

Raw counts					
	Settings (1,1)			Settings (1,2)	
	+	−		+	−
+	1683	418	+	1100	269
−	361	1578	−	156	1386
	Settings (2,1)			Settings (2,2)	
	+	−		+	−
+	1728	313	+	179	1636
−	351	1978	−	1143	294
Relative frequencies					
	Settings (1,1)			Settings (1,2)	
	+	−		+	−
+	0.42	0.10	+	0.38	0.09
−	0.09	0.39	−	0.05	0.48
	Settings (2,1)			Settings (2,2)	
	+	−		+	−
+	0.40	0.07	+	0.06	0.50
−	0.08	0.45	−	0.35	0.09

**Table A6.** Zhang et al.'s (2022) Experiment at Munich [6].

Raw counts					
	Settings (1,1)			Settings (1,2)	
	+	−		+	−
+	178	44	+	199	36
−	29	183	−	28	160
	Settings (2,1)			Settings (2,2)	
	+	−		+	−
+	160	47	+	38	160
−	31	151	−	166	39



Table A6. Cont.

Relative frequencies					
	Settings (1,1)			Settings (1,2)	
	+	−		+	−
+	0.41	0.10	+	0.47	0.09
−	0.07	0.42	−	0.07	0.38
	Settings (2,1)			Settings (2,2)	
	+	−		+	−
+	0.41	0.12	+	0.09	0.40
−	0.08	0.39	−	0.41	0.10

## References

1. Hensen, B.; Bernien, H.; Dréau, A.E.; Reiserer, A.; Kalb, N.; Blok, M.S.; Ruitenber, J.; Vermeulen, R.F.L.; Schouten, R.N.; Abellán, C.; et al. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **2015**, *682*, 526. [CrossRef] [PubMed]
2. Rosenfeld, W.; Burchardt, D.; Garthoff, R.; Redeker, K.; Ortegel, N.; Rau, M.; Weinfurter, H. Event-ready Bell test using entangled atoms simultaneously closing detection and locality loopholes. *Phys. Rev. Lett.* **2017**, *119*, 010402. [CrossRef] [PubMed]
3. Giustina, M.; Versteegh, M.A.; Wengerowsky, S.; Handsteiner, J.; Hochrainer, A.; Phelan, K.; Steinlechner, F.; Kofler, J.; Larsson, J.Å.; Abellán, C.; et al. Significant-loophole-free test of Bell’s theorem with entangled photons. *Phys. Rev. Lett.* **2015**, *115*, 250401. [CrossRef] [PubMed]
4. Shalm, L.K.; Meyer-Scott, E.; Christensen, B.G.; Bierhorst, P.; Wayne, M.A.; Stevens, M.J.; Gerrits, T.; Glancy, S.; Hamel, D.R.; Allman, M.S.; et al. Strong Loophole-Free Test of Local Realism. *Phys. Rev. Lett.* **2015**, *115*, 250402. [CrossRef] [PubMed]
5. Weihs, G.; Jennewein, T.; Simon, C.; Weinfurter, H.; Zeilinger, A. Violation of Bell’s inequality under strict Einstein locality conditions. *Phys. Rev. Lett.* **1998**, *81*, 5039–5043. [CrossRef]
6. Zhang, W.; van Leent, T.; Redeker, K.; Garthoff, R.; Schwonnek, R.; Fertig, F.; Eppelt, S.; Rosenfeld, W.; Scarani, V.; Lim, C.C.-W.; et al. A device-independent quantum key distribution system for distant users. *Nature* **2022**, *607*, 687–691. [CrossRef] [PubMed]
7. Gill, R.D. Statistics, Causality and Bell’s Theorem. *Statist. Sci.* **2014**, *29*, 512–528. [CrossRef]
8. Einstein, A.; Podolsky, B.; Rosen, N. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.* **1935**, *47*, 777–780. [CrossRef]
9. Bell, J.S. On the Einstein Podolsky Rosen parodox. *Physics* **1964**, *1*, 195–290. [CrossRef]
10. Clauser, J.F.; Horne, M.A.; Shimony, A.; Holt, R.A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **1969**, *23*, 880–884. [CrossRef]
11. Freedman, S.J.; Clauser, J.F. Experimental Test of Local Hidden-Variable Theories. *Phys. Rev. Lett.* **1972**, *28*, 938–941. [CrossRef]
12. Aspect, A.; Dalibard, J.; Roger, G. Experimental Test of Bell’s Inequalities Using Time-Varying Analyzers. *Phys. Rev. Lett.* **1982**, *49*, 1804–1807. [CrossRef]
13. Pearle, P. Hidden-Variable Example Based upon Data Rejection. *Phys. Rev. D* **1970**, *2*, 1418–1425. [CrossRef]
14. Garg, A.; Mermin, N.D. Correlation Inequalities and Hidden Variables. *Phys. Rev. D* **1987**, *35*, 3831–3835. [CrossRef]
15. Eberhard, P.H. Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment. *Phys. Rev. A* **1993**, *47*, R747–R750. [CrossRef]
16. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021. Available online: <https://www.R-project.org/> (accessed on 4 May 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.