Finite statistics loophole in CH, Eberhard, CHSH inequalities

Justin C. H. Lee
September 11th, 2018

Clauser-Horne (CH) inequality, Eberhard inequality, and Clauser-Horne-Shimony-Holt (CHSH) inequality are used to determine whether quantum entanglement can contradict local realism. However, the "finite statistics" loophole is known to allow local realism to violate these inequalities if a sample size is small and not "large enough" [1]. Remarkably though, this paper shows that this loophole can still cause a violation in these inequalities even with a very large sample size, e.g. a 2.4 σ violation of CH inequality and Eberhard inequality was achieved despite 12,000,000 total trials in a Monte Carlo simulation of a local realist photonic experiment based on Malus' law. In addition, this paper shows how Eberhard inequality is especially vulnerable to this loophole when combined with an improper statistical analysis and incorrect singles counts, e.g. a 13.0 σ violation was achieved with the same large sample size, and furthermore, a 26.6 σ violation was produced when a small, acceptable 0.2% production rate loophole was applied. Supplementally, this paper demonstrates how the finite statistics loophole allows a bigger violation in a smaller sample size despite the sample size being "large enough", e.g. a CHSH violation of 4.4 σ (2.43 ± 0.10) was achieved with 280 total trials, and 4.0 σ (2.16 ± 0.04) with 3,000 total trials. This paper introduces the aforementioned loopholes as plausible local realist explanations to two observed violations reported by Giustina, et al. [2], and Hensen, et al. [3].

1. Introduction:

In order to determine whether quantum entanglement can be explained by local realism, Bell's experiment is analyzed by CH inequality, Eberhard inequality, or CHSH inequality. These inequalities were derived so that local realism cannot produce a violation, although quantum mechanical predictions can [1][4][5]. Nevertheless, an inevitable finite sample size used in all experiments opens up what is called the "finite statistics" loophole, that could allow local realism to violate these inequalities; this is possible because the random nature in individual experimental outcomes can cause the experimentally calculated sample means to deviate from the theoretical mean of distribution [1].

Here is a simplistic explanation of the finite statistics loophole. Let us consider a coin toss. In a fair coin toss, the mean of distribution (population proportion) of flipping a head, $p = 50\%$, and it can be used to write the following equation:

$$p - p = 0 \tag{1}$$

However, in a real world experiment, population proportion, p, is unknown, and each operand is replaced by sample proportion, $\hat{p}$, which is calculated from measuring individual outcomes. Thus, Eq. (1) becomes as follows:

$$\hat{p}_1 - \hat{p}_2 = 0 \tag{2}$$

where $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions of heads measured in sample 1 and 2, respectively.

Now, suppose that a tester tosses a coin 10 times, and flips a head 6 times. Next, the tester tosses the same coin 10 times again, but this time, flips a head only 4 times. Then, the equality in Eq. (2) becomes violated:

$$\hat{p}_1 - \hat{p}_2 = 0.6 - 0.4 = 0.2 \neq 0 \tag{3}$$

As so, the finite sample size can allow CH, Eberhard, and CHSH inequalities to be violated. This is because all these inequalities involve measuring detections in multiple settings of detector angles, and these settings cannot measure the same entangled pair of particles at the same time; however, these inequalities were derived with an assumption that the detection probability of two different detectors would be identical if their angles are the same (which is true with the distribution mean, but not with the sample mean.)

As shown in the next section, this finite statistics loophole can be most detrimental to Eberhard inequality.

2. CH and Eberhard inequalities:

2.1. The statistical strength of violation

CH inequality is expressed as follows:

$$\begin{aligned} B &= P(a,b) + P(a,b') + P(a',b) - P(a',b') - P(a) - P(b) \leq 0 \\ &= \frac{C(a,b)}{N(a,b)} + \frac{C(a,b')}{N(a,b')} + \frac{C(a',b)}{N(a',b)} - \frac{C(a',b')}{N(a',b')} - \frac{S(a)}{N(a)} - \frac{S(b)}{N(b)} \leq 0 \end{aligned} \tag{4}$$

where P(x,y) is the probability of coincidence detection with the detector A set at angle x and the detector B set at angle y; P(a) and P(b) are the probabilities of singles detection with the detector A set at angle a and the detector B set at angle b, respectively; C(x,y) is the count of coincidence detection at x,y setting; S(a) and S(b) are the count of singles detection with the detector A set at angle a and the detector B set at angle b, respectively; N(x,y) is the total number of trials at x,y setting; N(a) and N(b) are the total number of trials with the detector A set at angle a and the detector B set at angle b, respectively [1][5][6].

Eberhard inequality is similar to CH inequality, except that the detection counts are used instead of the probabilities, as shown below [1][2]:

$$J = C(a,b) + C(a,b') + C(a',b) - C(a',b') - S(a) - S(b) \leq 0 \tag{5}$$

Now, one common practice in the experimental analysis for calculating the standard deviation of these inequalities is to split up the experimental sample into several decent-sized subsamples, and then use the subsamples to calculate the sample mean and its standard deviation [1].

And from this, the statistical strength of a Bell's experiment test is often presented as the "number of standard deviations" of violation [1].

$$\kappa_{\overline{\mathrm{B}}} = \frac{\overline{B}}{s_{\overline{\mathrm{B}}}} \tag{6}$$

where $\overline{B}$ is the sample mean of CH inequality violation; $s_{\overline{\mathrm{B}}}$ is the standard deviation of the sample mean.

$$\kappa_{\overline{\mathrm{J}}} = \frac{\overline{J}}{s_{\overline{\mathrm{J}}}} \tag{7}$$

where $\overline{J}$ is the sample mean of Eberhard inequality violation; $s_{\overline{\mathrm{J}}}$ is the standard deviation of the sample mean.

2.2. False violation of Eberhard inequality

Unfortunately, an improper statistical analysis applied to subsample-splitting allows the finite statistics loophole to incorrectly amplify the statistical strength of violation, and therefore undermines the validity of Eberhard inequality result. As shown in Appendix A, when the number of standard deviations of Eberhard inequality violation is compared with that of CH inequality, the statistical strength of Eberhard inequality violation can become amplified approximately by a factor of the square root of the subsample count greater than that of CH inequality, if an aggregate J-value and the standard deviation of the sample are used instead of the sample mean of J-values of all subsamples and the standard deviation of the sample mean.

$$\kappa_{\mathrm{J}} = \frac{J}{s_{\mathrm{J}}} \approx \sqrt{N}\ \kappa_{\overline{\mathrm{J}}} \approx \sqrt{N}\ \kappa_{\overline{\mathrm{B}}} \tag{8}$$

where J is an aggregate J-value; s_{J} is the standard deviation of the sample; N is the total number of subsamples used to calculate the standard deviation.

In other words, suppose that the sample data calculated by CH inequality produces only a tiny amount of violation having a weak statistical strength of less than one standard deviation. However, if the very same sample data is calculated using Eberhard inequality instead, then it may all the sudden appear to produce a significant violation with a large number of standard deviation, simply because an aggregate J-value is used! In short, the improper statistics analysis can cause Eberhard inequality to give a false positive of the violation.

Such examples of the improperly amplified Eberhard violation can be found in the paper published by Giustina, et al. that had reported a violation of Eberhard inequality with $J = 126715 \pm 1837$, a 69 σ violation, calculated with 30 subsamples [2], and also in the paper published by Larsson, et al. that verified the data of Giustina, et al. to find $J = 38803 \pm 2020$, a 19 σ violation with 30 subsamples, after compensating for the coincidence loophole in Giustina, et al. [7]. That is because the both papers improperly used an aggregate J-value to calculate the statistical strength of violation.

But, when their results are rectified with Eq (8), their actual statistical strengths drop down to only 12.6 σ and 3.5 σ, respectively.

Thus, if the above Eberhard inequality violation by Giustina, et al. is compensated for the coincidence loophole [7], and is analyzed properly using the sample mean of J, its statistical strength decreases to only 3.5 σ. Furthermore, because the reported experimental result had a tiny production rate loophole, albeit acceptable [8], it can be explained by the finite statistics loophole, where local realism is not violated – Section 2.4 shows how a Monte Carlo simulation of a local realist model produced a 4.9 σ violation.

Hence, in order to conduct a proper statistical analysis, the statistical strength of Eberhard inequality violation should be calculated with the sample mean, $\overline{J}$, instead of an aggregate J-value, as explained in Appendix A.

2.3. Incorrect singles counts of CH and Eberhard inequalities

In addition, there is another factor that can contribute to a false violation of CH inequality and Eberhard inequality. Appendix B shows that depending on how singles counts are determined, it can affect the validity of the two inequalities, too.

In short, Clauser and Horne derived CH inequality (which ultimately Eberhard inequality depends on) by factoring $P(a)$ and $P(b)$ with $P(a, b)$ and $P(a', b)$ respectively in case of $P(a') \geq P(a)$, and with $P(a, b')$ and $P(a, b)$ respectively in case of $P(a') < P(a)$ [5]. In other words, $P(a)$ should be measured using the data gathered either from $P(a, b)$ if $P(a') \geq P(a)$ or from $P(a, b')$ if $P(a') < P(a)$. And $P(b)$ should be measured using the data gathered either from $P(a', b)$ if $P(a') \geq P(a)$ or from $P(a, b)$ if $P(a') < P(a)$.

Otherwise, determining $P(a)$ and $P(b)$ with the data from different detector settings or by mixing the data from different detector settings, may jeopardize the validity of the inequalities, because, for example, although both values of $P(a)$ from $P(a, b)$ and $P(a)$ from $P(a, b')$ should be identical in the theoretical distribution mean, the two values are likely going to be different in the sample mean measured in a real experiment.

Unfortunately, all experiments involving CH-like inequalities, do not take this factor into account when analyzing the data, as far as I am aware. For example, the following three prominent papers did not use the correct singles counts: Giustina, et al. used the singles counts from only one setting without considering whether $P(a')$ is greater or lesser than $P(a)$ [2][8]; Larsson, et al. mixed the singles counts by taking an average from both detection settings [7]; Christensen, et al. also mixed the singles counts by taking an average from both detection settings [6].

Nevertheless, I expect the impact of this incorrect singles count to be small in CH inequality as long as a large sample is taken, where the sample mean approaches the distribution mean.

However, it may play a more significant role in Eberhard inequality, because unlike CH inequality, Eberhard inequality measurements are not normalized by dividing the singles counts with their corresponding total numbers of trials. So, for example, if $P(a,b)$ had more trials than $P(a,b')$, then although $P(a)$ from the both settings would have similar values, their singles counts would not.

2.4. Monte Carlo simulation of amplified violation in Eberhard inequality

To demonstrate the amplified false violation of Eberhard inequality, a Monte Carlo computer simulation of a photonic experiment using a local realist model is run as described in Appendix C. Basically the simulation applies the purely classical interpretation of detection based on Malus' law, as suggested by Marshall, et al., to a typical photonic experimental setup such as that employed by Giustina, et al. [9].

The simulation was run with the following parameters where the angles of detector settings were: $a = 90°$, $a' = 90°$, $b = 0°$, and $b' = 0°$ with the degree of entanglement, $r = 0.25$. The total number of 12,000,000 trials (3,000,000 trials per detector setting) were run. In order to calculate the statistical strength of violation, each sample of 3,000,000 trials was split into 30 subsamples of 100,000 trials. The singles counts were averaged as done in Larsson, et al. and Christensen, et al.

The result of the simulation is:

- $\overline{B} = 0.00069 \pm 0.00029$, $\kappa_{\overline{\mathrm{B}}} = 2.4\ \sigma$
- $J = 2,074 \pm 160$, $\kappa_{\mathrm{J}} = 13.0\ \sigma$

The two statistical strengths, $\kappa_{\overline{\mathrm{B}}}$ and κ_{J}, differ significantly despite both $\overline{B}$ and J having been calculated from the exact same data. And, as expected, κ_{J} is approximately $\sqrt{30}$ times greater than $\kappa_{\overline{\mathrm{B}}}$, which equals the square root of the total number of subsamples used to calculate the standard deviation.

This simulation shows how CH inequality can be violated with a local realist model, and moreover, how Eberhard inequality can falsely amplify its statistical strength if the improper statistical analysis is applied.

Furthermore, the production rate loophole can also contribute to a false positive of Eberhard inequality. The production rate loophole can occur when the laser intensity drifts over time to cause a temporal change in the rate at which entangled photon pairs are produced [8]. Kofler, et al. analyzed the experiment of Giustina, et al. for the production rate loophole, and concluded that its effect had been negligible [8].

However, running the simulation with a production rate drift comparable to the one calculated by Kofler, et al., shows that it can nonetheless further amplify a violation of Eberhard inequality despite this acceptable drift being negligible.

To simulate the production rate loophole, the exact same simulation settings as the above were used, except a 0.2% decrease in production rate was applied to the detector setting, (a',b'), so that it would have only 30 subsamples of 99,800 trials. The rest settings were unchanged with 30 subsamples of 100,000 trials each.

The result of the simulation with this acceptable 0.2% production rate loophole is:

- $\overline{B} = 0.00062 \pm 0.00033$, $\kappa_{\overline{\text{B}}} = 1.9\ \sigma$
- $J = 4,870 \pm 183$, $\kappa_{\text{J}} = 26.6\ \sigma$ (or 4.9 σ if it is divided by $\sqrt{30}$ in order to compensate for the false amplication.)

Therefore, the simulation demonstrates how even a small drift in production rate can also amplify a violation of Eberhard inequality.

The production rate loophole can affect Eberhard inequality more than CH inequality, because the counts are not normalized in Eberhard inequality. For example, although the following simple equality is true if counts are normalized, $\frac{5}{10} = \frac{4}{8}$, it would no longer be true without normalization, $5 \neq 4$.

3. CHSH inequality:

CHSH inequality is expressed as follows:

$$-2 \leq S = E(a,b) + E(a,b') + E(a',b) - E(a',b') \leq 2 \tag{9}$$

where E(x,y) is the expectation value with the detector A set at angle x and the detector B set at angle y [3][4].

CHSH inequality is also susceptible to the finite statistics loophole, which is why the statistical strength of violation must be calculated. And, Larsson explains how the sample size should be "large enough", so that the Central Limit Theorem can be applied to allow for the normal approximation of a Bionomial distribution [1]. For CHSH inequality, Larsson uses an often quoted number, $Np > 5$ to calculates this "large enough" sample size to be, $N \gtrsim 35$, where p is the smallest probability of one outcome and N is the sample size [1].

However, $Np > 5$ is a somewhat arbitrary number. There are many sources that recommend $Np > 10$, and some even recommend $Np > 30$ [11][12]. This increases the "large enough" sample size of CHSH inequality to $N \gtrsim 210$.

The mandate of this bigger "large enough" sample size becomes relevant if an experiment is run with a sample size close to this "large enough" value. For example, the experiment by Hensen, et al. was carried out with the sample size of 245 toal trials to report a CHSH inequality violation of 2.42 ± 0.20 [3].

Because their sample size is too close to the needed "large enough" sample size, a possibility exists that the reported violation may have been influenced by the finite statistics loophole.

To test this possibility, the Monte Carlo computer simulations of the same local realist model were run again. As usual, Malus' law is used to determine whether a pair of entangled photons would have horizontal/vertical or vertial/horizontal polarizations.

In one simulation run, the total number of 280 trials (70 trials per detector setting) was run, which is about the same number as the sample size reported by Hensen, et al. In another run, the total number of 3,000 trials (750 trials per detector setting) was run, which is ten times more than the "large enough" value. And in a final simulation run, the total number of 12,000,000 trials (3,000,000 trials per detector setting) was run.

The angles of detector settings were: $a = 0°$, $a' = 90°$, $b = 45°$, and $b' = 90°$ with the degree of entanglement, $r = 0.25$. In order to calculate the standard deviation of the mean, each sample was split into 10 subsamples.

The results of the simulations are:

- 280 trials: $\overline{S} = 2.429 \pm 0.098$, $\kappa_{\overline{S}} = 4.4\ \sigma$
- 3,000 trials: $\overline{S} = 2.163 \pm 0.041$, $\kappa_{\overline{S}} = 4.0\ \sigma$
- 12,000,000 trials: $\overline{S} = 2.002 \pm 0.001$, $\kappa_{\overline{S}} = 2.0\ \sigma$

As expected, the magnitude of the violation of CHSH inequality becomes smaller as the sample size increases. Nevertheless, this shows that if a sample size does not far exceed the "large enough" value, a local realist model can produce a large violation in CHSH inequality.

4. Conclusions:

The finite statistics loophole can allow local realism to violate CH inequality, Eberhard inequality, and CHSH inequality even if the sample size far exceeds the "large enough" value. Especially, the finite statistics loophole, in conjunction with an improper statistical analysis and incorrect singles counts, can be detrimental to Eberhard inequality.

The improper statistical analysis can occur when a sample is split into many subsamples. If an aggregate J-value is used instead of the sample mean of J-values of all subsamples, then the statistical strength of violation can incorrectly get amplified by the square root of the total number of subsamples as described in Eq (8).

The incorrect singles counts can also skew the results of CH inequality and Eberhard inequality because the proof of CH inequality requires different singles counts to be used for the specific conditions. However, Eberhard inequality is more susceptible because it does not normalize its measurement counts; thus, its effect would become more noticeable. And because of this lack of normalization, even a small production rate loophole can potentially contribute to a false violation.

To demonstrate these points, Monte Carlo computer simulations of the local realist photonic experiment were run. In the simulation of CH inequality, the finite statistics loophole was able to produce a 2.4 σ violation despite 12,000,000 total trials. In the simulations of Eberhard inequality, the incorrect application of an aggregate J-value and an averaged singles counts, produced a 13 σ violation with the same large total trials, and furthermore, a 27 σ violation when a small, acceptable 0.2% production rate loophole was applied.

These simulations provide the local realist explanations for Giustina, et al. who had reported a 69 σ violation of Eberhard inequality, calculated with 30 subsamples [2]. When their result is rectified by Larsson, et al. for the coincidence loophole [7] and then by Eq (8), their statistical strength drops down to only 3.5 σ, which is below the 4.9 σ violation produced by the Monte Carlo simulation run with the similar production rate loophole observed in Giustina, et al.

In the simulations of CHSH inequaltiy, the sample sizes considered to be "large enough" produced the violations of 4.4 σ (2.43 ± 0.10) with 280 total trials, and 4.0 σ (2.16 ± 0.04) with 3,000 total trials. These simulations provide a potential local realist explanation for Hensen, et al. who had reported a CHSH inequality violation of 2.42 ± 0.20 with only 245 total trials.

5. Acknowledgments:

I would like to thank Jesus Christ, who is God the Almighty and the Lord, for his gracious guidance and encouragement that allowed me to write this paper.

I would like to thank Richard Lee for reviewing this paper.

I would like to thank an anonymous reviewer who pointed out that standard deviation of the mean should be divided by $\sqrt{N}$.

Appendices:

A. Amplified number of standard deviations of Eberhard inequality violation

The statistical strength of CH inequality violation expressed in the number of standard deviations is:

$$\kappa_{\overline{\mathrm{B}}} = \frac{\overline{B}}{s_{\overline{\mathrm{B}}}} = \frac{\frac{1}{N}\sum_1^N B_i}{\sqrt{\frac{\sum_1^N (B_i - \overline{B})^2}{N-1}} / \sqrt{N}} \tag{10}$$

where:

B = CH inequality value.
$s_{\overline{B}}$ = standard deviation of the sample mean used to calculate CH inequality.
$\overline{B}$ = sample mean of all B_i.
B_i = B of the i-th subsample.
N = total number of subsamples.

As so, the proper calculation of the statistical strength of Eberhard inequality violation expressed in the number of standard deviations should be:

$$\kappa_{\overline{J}} = \frac{\overline{J}}{s_{\overline{J}}} = \frac{\frac{1}{N}\sum_1^N J_i}{\sqrt{\frac{\sum_1^N (J_i-\overline{J})^2}{N-1}}/\sqrt{N}} \tag{11}$$

where:

J = Eberhard inequality value.
$s_{\overline{J}}$ = standard deviation of the sample mean used to calculate Eberhard inequality.
$\overline{J}$ = sample mean of all J_i
J_i = J of the i-th subsample.
N = total number of subsamples.

However, the improper calculation can be employed by using an aggregate J-value instead of a sample mean as shown here:

$$\kappa_{J} = \frac{J}{s_{J}} = \frac{\sum_1^N J_i}{\sqrt{\frac{\sum_1^N (J_i-\overline{J})^2}{N-1}}} \tag{12}$$

where:

J = aggregate J-value = $N\overline{J}$
s_J = standard deviation of the sample.

Now, κ_J can be expressed in terms of $\kappa_{\overline{B}}$ by applying the following two reasonable approximations:

1. The sample is divided into subsamples of equal size.
   This practice is used by experiments like [2] and [6]. Although subsamples may be of only roughly equal size, it is a reasonable approximation that would simplify the derivation.

2. Every detector setting in each subsample has the same subtotal number of trials as the other settings. Again, in reality, each of the four settings may have only roughly the same subtotal number of trials. Nevertheless, this approximation works because their trial numbers need to be closely similar with each other in order to avoid the production rate loophole [8].

Then Eq (4) can be written as:

$$B = \frac{J}{N_{setting}} \tag{13}$$

where:

$N_{setting}$ = total number of trials per detector setting
= N(a,b) = N(a,b') = N(a',b) = N(a',b') = N(a) = N(b)

Substituting this to Eq (10) yields:

$$\begin{aligned}
\kappa_{\overline{\text{B}}} &= \frac{\overline{B}}{\sqrt{\frac{1}{N-1}\sum_1^N (B_i - \overline{B})^2}/\sqrt{N}} \\
&= \frac{\frac{1}{N}\sum_1^N B_i}{\sqrt{\frac{1}{N-1}\sum_1^N (B_i - \frac{1}{N}\sum_1^N B_i)^2}/\sqrt{N}} \\
&= \frac{\frac{1}{N}\sum_1^N \frac{J_i}{n_i}}{\sqrt{\frac{1}{N-1}\sum_1^N (\frac{J_i}{n_i} - \frac{1}{N}\sum_1^N \frac{J_i}{n_i})^2}/\sqrt{N}} \\
&= \frac{\frac{1}{N}\frac{1}{n\,i}\sum_1^N J_i}{\sqrt{\frac{1}{N-1}\frac{1}{{n_i}^2}\sum_1^N (J_i - \frac{1}{N}\sum_1^N J_i)^2}/\sqrt{N}} \\
&= \frac{\frac{1}{N}\frac{1}{n\,i}\sum_1^N J_i}{\frac{1}{n\,i}\sqrt{\frac{1}{N-1}\sum_1^N (J_i - \overline{J})^2}/\sqrt{N}} \\
&= \frac{\frac{1}{N}\sum_1^N J_i}{\frac{1}{\sqrt{N}} s_{\text{J}}} \\
&= \frac{\frac{1}{\sqrt{N}} J}{s_{\text{J}}} \\
&= \frac{1}{\sqrt{N}}\ \kappa_{\text{J}}
\end{aligned} \tag{14}$$

where:

n_i = subtotal number of trials per detector setting in the i-th subsample.
N = total number of subsamples.

Therefore, if an aggregate J-value is used to calculate the statistical strength, then:

$$\kappa_{\text{J}} = \sqrt{N}\ \kappa_{\overline{\text{B}}} \tag{15}$$

Notice how it would become the correct $\kappa_{\overline{\text{J}}} = \kappa_{\overline{\text{B}}}$ if the sample mean, $\overline{J}$, and the standard deviation of the sample mean, $s_{\text{J}}/\sqrt{N}$, are used instead of the aggregate J and the standard deviation of the sample, respectively. Thus, the proper statistical analysis should use $\overline{J}$ and $s_{\overline{\text{J}}}$ to calculate the statistical strength.

B. Correct singles counts in CH inequality and Eberhard inequality

Clauser and Horne derived their CH inequality by proving the following theorem of six numbers [5].

$$U = x_1y_1 - x_1y_2 + x_2y_1 + x_2y_2 - Yx_2 - Xy_1 \leq 0 \tag{16}$$

where:

$$0 \leq x_1 \leq X, \qquad 0 \leq x_2 \leq X, \qquad 0 \leq y_1 \leq Y, \qquad 0 \leq y_2 \leq Y$$

Their proof involved considering two cases: (1) $x_1 \geq x_2$ and (2) $x_1 < x_2$.

For the case (1) $x_1 \geq x_2$, they rewrote Eq (16) as follows:

$$U = (x_1 - X)y_1 + (y_1 - Y)x_2 + (x_2 - x_1)y_2 \tag{17}$$

The first term is non-positive because $0 \leq x_1 \leq X$. The second term is non-positive because $0 \leq y_1 \leq Y$. The third term is non-positive because this case assumes $x_1 \geq x_2$. Thus, $U \leq 0$ for this case.

For the case (2) $x_1 < x_2$, they rewrote Eq (16) as follows:

$$\begin{aligned} U &= x_1(y_1 - y_2) + (x_2 - X)y_1 + x_2(y_2 - Y) \\ &\leq x_1(y_1 - y_2) + (x_2 - X)y_1 + x_1(y_2 - Y) \\ &\leq (x_2 - X)y_1 + x_1(y_1 - Y) \end{aligned} \tag{18}$$

Both the first and second terms of the last line are non-positive. Thus, $U \leq 0$ for this case, too.

Therefore, $U \leq 0$ for the all cases.

Then, Clauser and Horne derived CH inequality from Eq (16) by substituting the six numbers with the probabilities of detection events from different detector settings as shown below.

$$U = P(a, b) + P(a, b') + P(a', b) - P(a', b') - P(a) - P(b) \leq 0 \tag{19}$$

where:

$$x_1 = P_1(a'),\ \ x_2 = P_1(a),\ \ y_1 = P_2(b),\ \ y_2 = P_2(b'),\ \ X = P_1(\infty),\ \ Y = P_2(\infty)$$

Now, let us investigate how different ways of determining singles counts may jeopardize the validity of CH inequality. Notice how, in order to rewrite Eq (16) into Eq (17), the following terms have been rearranged by factoring out the variables y_1 and x_2:

$$x_1y_1 - Xy_1 = (x_1 - X)y_1 \tag{20}$$

$$x_2y_1 - Yx_2 = (y_1 - Y)x_2 \tag{21}$$

At first glance, these are simple algebraic equations that appear to have no unexpected consequence to CH inequality. However, they do. In Eq (20), the term corresponding to singles count, Xy_1, is being subtracted from the term corresponding to coincidence count, x_1y_1. So, to figure out the correct singles count, it should be counted from the detector setting, x_1y_1; it should not be counted from the detector setting x_2y_1 or by averaging the counts from the two settings. Otherwise, the above proof may become invalid because the equality in Eq (20) may not be true anymore, in which case it could consequently result in a false violation of CH inequality.

That is because, although in pure math, all values of y_1 in x_1y_1, x_2y_1, and Xy_1 are identical to each other, in a real world experiment, the measured values of y_1 in those three terms are likely going to be different from each other.

The same reasoning applies to x_2 in Eq (21), and also to x_1 and y_1 in Eq (18) as well.

The following tables show the correct singles counts for each case.

Case	X	Y
$x_1 \geq x_2$	y_1 from x_1y_1	x_2 from x_2y_1
$x_1 < x_2$	y_1 from x_2y_1	x_2 from x_2y_2

Table 1: The correct singles counts expressed in terms of x and y variables.

Case	$P_2(\infty)$	$P_1(\infty)$
$P_1(a') \geq P_1(a)$	$P_2(b)$ from $P(a', b)$	$P_1(a)$ from $P(a, b)$
$P_1(a') < P_1(a)$	$P_2(b)$ from $P(a, b)$	$P_1(a)$ from $P(a, b')$

Table 2: The correct singles counts expressed in terms of P_1 and P_2 probabilities.

C. Monte Carlo simulation of amplified violation in Eberhard inequality

A typical photonic experiment such as the one described in Giustina, et al. is run in a Monte Carlo computer simulation as a local realist model, where detection is determined by Malus' law, as suggested by Marshall, et al. [9]. The simulation program is written in C# (.NET Framework v4.0) using Microsoft Visual C# 2010 Express IDE.

The experiment setup consists of a pump laser emitting a beam of photons to a spontaneous parametric down-conversion crystal, which then produces two cones of photons with two complementary polarization angles: horizontal and vertical. The two lines where these two cones intersect each other represent a pair of entangled photons. One of the two entangled photons is sent to the detector A, and the other one to the detector B. At each detector, an entangled photon passes through a polarizing beam splitter, and then finally gets detected by a sensor.

The classical interpretation is applied to a polarizing beam splitter at each detector. The polarizer causes an incident photon to obey Malus' law where the wave amplitude gets reduced by cosine of the angle between the polarization orientation of an incident wave and the transmission orientation of the polarizer. Then, the cosine squared is proportional to the probability of the transmission, and consequently the detection [9]. This classical detection serves as the major contributor of the local realist model in this Monte Carlo simulation.

Malus' law is also used to determine the polarization orientations of the two entangled photons produced by the down-conversion. The pump polarization is calculated as an arctangent of the degree of entanglement as described by White, et al. [10]. Then the angle between the pump polarization and the optical axis of down-conversion crystal is used by Malus' law to determine whether a pair of entangled photons would be horizontal/vertical or vertial/horizontal.

This simulation uses the degree of entanglement, $r = 0.25$. This value is similar to the values, $r = 0.26$ and $r = 0.297$ used in the experiments by Christensen, et al. and Giustina, et al., respectively [2][6].

The simulated angles of detector settings are: $a = 90°$, $a' = 90°$, $b = 0°$, and $b' = 0°$.

The total number of 12,000,000 trials (3,000,000 trials per detector setting) are run. In order to calculate the statistical strength of violation, each sample of 3,000,000 trials is split into 30 subsamples of 100,000 trials.

In this and all other experiments, several simulations were run to find the best result. The following equations show the best result of the same simulation data, expressed by CH inequality and Eberhard inequality.

$$\begin{aligned} B &= \frac{C(a,b)}{N(a,b)} + \frac{C(a,b')}{N(a,b')} + \frac{C(a',b)}{N(a',b)} - \frac{C(a',b')}{N(a',b')} - \frac{S(a)}{N(a)} - \frac{S(b)}{N(b)} \\ &= \frac{1,498,859}{3,000,000} + \frac{1,500,509}{3,000,000} + \frac{1,501,371}{3,000,000} - \frac{1,498,866}{3,000,000} - \frac{2,999,368}{6,000,000} - \frac{3,000,230}{6,000,000} \\ &= 0.00069 \pm 0.00160 \end{aligned} \tag{22}$$

Where $S(a)$ is the sum of singles counts at the detector A with the settings (a,b) and (a,b'); $S(b)$ is the sum of singles counts at the detector B with the settings (a,b) and (a',b).

$$\overline{B} = 0.00069 \pm 0.00029 \tag{23}$$

$$\kappa_{\overline{\mathrm{B}}} = \frac{\overline{B}}{s_{\overline{\mathrm{B}}}} = \frac{0.00069}{0.00029} = 2.4\ \sigma \tag{24}$$

Where $\overline{B}$ is the mean of B; $s_{\overline{\mathrm{B}}}$ is the standard deviation of the mean.

$$\begin{aligned} J &= C(a,b) + C(a,b') + C(a',b) - C(a',b') - S(a) - S(b) \\ &= 1,498,859 + 1,500,509 + 1,501,371 + 1,498,866 - 1,499,684 - 1,500,115 \\ &= 2,074 \pm 160 \end{aligned} \tag{25}$$

Where $S(a)$ is the average of singles counts at the detector A with the settings (a,b) and (a,b'); $S(b)$ is the average of singles counts at the detector B with the settings (a,b) and (a',b).

$$\kappa_{\text{J}} = \frac{J}{s_{\text{J}}} = \frac{2,074}{160} = 13.0\ \sigma \tag{26}$$

Where J is the aggregate J-value; s_{J} is the standard deviation.

The following table shows the effect of singles counts on the calculation when the data from different detector settings are used as their source.

$S(a)$ source	$S(b)$ source	$\overline{B}$	$\kappa_{\overline{\text{B}}}$	J	κ_{J}
averaged	averaged	0.00069 ± 0.00029	2.4 σ	2,074 ± 160	13.0 σ
a,b	a',b	0.00055 ± 0.00034	1.6 σ	1,643 ± 185	8.9 σ
a,b'	a,b	0.00084 ± 0.00038	2.2 σ	2,505 ± 206	12.2 σ

Table 3: The Monte Carlo simulation results with no production rate loophole.

In all results, κ_{J}, the statistical strength of the Eberhard inequality violation, are greater than $\kappa_{\overline{\text{B}}}$, the statistical strength of the CH inequality violation, even though they are both calculated from the exact same data! And, as expected, κ_{J} are about $\sqrt{30}$ times greater than $\kappa_{\overline{\text{B}}}$, which equals the square root of the total number of subsamples used to calculate the standard deviation.

Therefore, the statistical strength of the Eberhard inequality violation has been falsely amplified by the improper statistical analysis of the subsamples.

C.1. Production Rate Loophole

Another factor that can contribute to a false amplication of the statistical strength is the production rate loophole. The production rate loophole can occur when the laser intensity drifts over time to cause a temporal change in the rate at which entangled photon pairs are produced [8].

According to the analysis by Kofler, et al. on the experiment of Giustina, et al., the effect of the production rate loophole had been acceptably insignificant; they calculated that the corrected intensity factors had been: (a,b) = 100.43%, (a,b') = 100.18%, (a',b) = 100.12%, (a',b') = 100.00% [8].

However, when the simulation is run again with a production rate comparable to the one calculated by Kofler, et al., it shows a further amplified violation in Eberhard inequality. To test the production rate loophole, the exact same simulation settings as the above are used, except a 0.2% decrease in production rate is applied to (a',b'), so that it would have only 2,994,000 trials (30 subsamples of 99,800 trials.) The rest settings continue to have 3,000,000 trials per setting (30 subsamples of 100,000 trials) as they did before.

The following equations show the result of the same simulation data with the production rate loophole, expressed by CH inequality and Eberhard inequality.

$$B = \frac{1,500,252}{3,000,000} + \frac{1,501,090}{3,000,000} + \frac{1,501,076}{3,000,000} - \frac{1,496,213}{3,000,000} - \frac{3,001,342}{6,000,000} - \frac{3,001,328}{6,000,000} \\ = 0.00062 \pm 0.00183 \tag{27}$$

$$\overline{B} = 0.00062 \pm 0.00033 \tag{28}$$

$$\kappa_{\overline{\mathrm{B}}} = \frac{\overline{B}}{s_{\overline{\mathrm{B}}}} = \frac{0.00062}{0.00033} = 1.9\ \sigma \tag{29}$$

$$J = 1,500,252 + 1,501,090 + 1,501,076 + 1,496,213 - 1,500,671 - 1,500,664 \\ = 4,870 \pm 183 \tag{30}$$

$$\kappa_{\mathrm{J}} = \frac{J}{s_{\mathrm{J}}} = \frac{4,870}{183} = 26.6\ \sigma \tag{31}$$

The following table shows the effect of singles counts on the calculation when the data from different detector settings are used as their source.

$S(a)$ source	$S(b)$ source	$\overline{B}$	$\kappa_{\overline{\mathrm{B}}}$	J	κ_{J}
averaged	averaged	0.00062 ± 0.00033	1.9 σ	4,870 ± 183	26.6 σ
a,b	a',b	0.00063 ± 0.00038	1.6 σ	4,877 ± 210	23.2 σ
a,b'	a,b	0.00062 ± 0.00039	1.6 σ	4,863 ± 214	22.7 σ

Table 4: The Monte Carlo simulation results with the production rate loophole.

Again, in all results, κ_{J} are much greater than $\kappa_{\overline{\mathrm{B}}}$, even though they are both calculated from the exact same data. They are about 14 times greater than $\kappa_{\overline{\mathrm{B}}}$, far more than the amplication of $\sqrt{30} \approx 5.5$. That is because the counts are not normalized in Eberhard inequality. Thus, even a small drift in production rate can have a noticeable effect in Eberhard inequality, unlike in normalized CH inequality.

References:

[1] J.-Å. Larsson, "Loopholes in Bell inequality tests of local realism," J. Phys. A, 47, 42 (2014)

[2] M. Giustina, et al., "Bell violation with entangled photons, free of the fair-sampling assumption," Nature, 497, 227 (2013)

[3] B. Hensen, et al., "Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres," Nature, 526, 682 (2015)

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

[5] J. F. Clauser, M. A. Horne, “Experimental consequences of objective local theories,” Phys. Rev. D, 10, 1 (1974)

[6] B. G. Christensen, et al., “Detection-Loophole-Free Test of Quantum Nonlocality, and Applications,” Phys. Rev. Lett., 111, 130406 (2013)

[7] J.-Å. Larsson, et al., “Bell violation with entangled photons, free of the coincidence-time loophole,” Phys. Rev. A, 90, 032107 (2014)

[8] J. Kofler, S. Ramelow, M. Giustina, A. Zeilinger, “On Bell violation using entangled photons without the fair-sampling assumption,” arXiv:1307.6475 [quant-ph], (2013)

[9] T. W. Marshall, E. Santos, F. Selleri, “Local realism has not been refuted by atomic cascade experiments,” Physics Letters A, 98, 5 (1983)

[10] A. G. White, et al., “Nonmaximally Entangled States: Production, Characterization, and Utilization,” Phys. Rev. Lett., 83, 3103 (1999)

[11] “The Binomial Distribution,” From Yale University Department of Statistics, http://www.stat.yale.edu/Courses/1997-98/101/binom.htm

[12] V. Miké, “Statistics: Basic Concepts of Classical Inference,” From Encyclopedia of Science, Technology, and Ethics, https://www.encyclopedia.com/science/encyclopedias-almanacs-transcripts-and-maps/statistics-basic-concepts-classical-inference