Maximally nonlocal theories cannot be maximally random

Gonzalo de la Torre,1,‡ Matty J. Hoban,1∥ Chirag Dhara,1 Giuseppe Prettico,1 and Antonio Acín1,2

1ICFO–Institut de Ciencies Fotoniques, Mediterranean Technology Park, 08860 Castelldefels (Barcelona), Spain
2ICREA–Institució Catalana de Recerca i Estudis Avançats, E–08010 Barcelona, Spain

Correlations that violate a Bell Inequality are said to be nonlocal, i.e. they do not admit a local and deterministic explanation. Great effort has been devoted to study how the amount of nonlocality (as measured by a Bell inequality violation) serves to quantify the amount of randomness present in observed correlations. In this work we reverse this research program and ask what do the randomness certification capabilities of a theory tell us about the nonlocality of that theory. We find that, contrary to initial intuition, maximally nonlocal theories cannot allow maximal randomness certification. We go on and show that quantum theory, in contrast, permits certification of maximal randomness in all dichotomic scenarios. We hence pose the question of whether quantum theory is optimal for randomness, i.e. is it the most nonlocal theory that allows maximal randomness certification? We answer this question in the negative by identifying a larger-than-quantum set of correlations capable of this feat. Not only are these results relevant to understanding quantum mechanics’ fundamental features, but also put fundamental restrictions on device-independent protocols based on the no-signaling principle.

Since the early days of quantum theory, its seemingly ‘intrinsic’ unpredictability has been heavily debated \cite{1,2}. A great advance came when John Bell \cite{3} identified limitations on any theory founded on the following two basic physical principles: impossibility of instantaneous signaling (the no-signaling principle); and the existence of hidden variables that would restore determinism if known. Correlations among distant parties that satisfy both principles had to be constrained by a set of inequalities termed Bell inequalities. Correlations that do not violate any of those inequalities hence admit a deterministic explanation and are called “local”. Correlations violating any of these inequalities, such as some quantum correlations, are termed “nonlocal” and must possess some amount of intrinsic randomness so as not to violate the no-signalling principle.

In recent years, a full body of research has been devoted to studying the connection between nonlocality and randomness in a quantitative way \cite{4}. Beyond the evident fundamental motivation, this issue is relevant for applications. Indeed, nonlocality-certified randomness represents an information resource in the now well-established area of device-independent information processing. Celebrated examples of protocols exploiting nonlocal quantum correlations include randomness generation \cite{5,6}, randomness amplification \cite{7,8}, or key distribution \cite{9,10}. In a device-independent protocol, several parties receive physical systems in their laboratories, subject them to measurements and collect the resulting outcomes. By definition, no assumption is made about the precise inner-workings of the devices which are regarded as black boxes. However, key to this approach is the most general physical theory that is assumed to dictate the workings of these boxes. The kind of assumption made in device-independent protocols is whether the devices are quantum mechanical \cite{9} or just compatible with the no-signalling principle \cite{10}.

Quantum theory is not the only, or even the most, nonlocal theory respecting the no-signaling condition \cite{11}. We will call any theory capable of producing correlations as nonlocal as allowed by the no-signalling principle a maximally nonlocal theory. Given the eminent role of nonlocality for randomness certification, one could expect such theories to have powerful randomness certification capabilities and be more powerful than theories with limited nonlocality, such as quantum mechanics.

The main goal of our work is to understand the relationship between the nonlocality and randomness of a theory and, in particular, what the randomness capabilities of a theory tell us about the nonlocality allowed within that theory. The first result is to show that the previous intuition is wrong: were physical correlations solely restricted by the no-signaling principle, maximal randomness could not be certified in any scenario, i.e. maximally nonlocal theories cannot be maximally random. Secondly, we focus on quantum theory and provide, in contrast, scenarios with an arbitrary number of parties performing at least two dichotomic measurements where maximal randomness can be certified. This should be compared with other works that showed that if maximally nonlocal theories were permitted in Nature we would have unimaginable computational and communicating power \cite{12,13}. Here, being in a maximally nonlocal world limits our information processing capabilities. This observation leads us to ask if the nonlocality of quantum theory is in some sense optimal for randomness certification. That is, is quantum theory the most nonlocal theory capable of certifying maximal randomness? Our final result answers this question in the negative: we identify a set of supra-quantum non-signalling correlations where maximal randomness can also be certified.

**Boxes and Bell tests**—We use the scenario of a Bell test to study the correlations observed among space-like separated measurements within different physical theories. There are \( N \) distant parties and each party makes a choice of measurement upon their system and receives some outcome of this measurement. These processes are arranged so that they define space-like separated events. We then model these parties as black boxes with the measurement choice for the \( j \)th party (for \( j \in \{1,2,...,N\} \)) being an input \( x_j \in \{0,1,...,(M-1)\} \) where there are \( M \) possible choices; the measurement outcome for this \( j \)th party is then the output \( a_j \in \{0,1,...,(d-1)\} \) where
for every party there are \( d \) possible outcomes to the every measurement. Therefore, a string of “dits” is produced in each round of a Bell test; dits are just the generalization of bits where now we have a string of length \( N \) with each element taking \( d \) possible values. The Bell test is then labelled by the parameters \( (N, M, d) \). After a suitable number of uses of the boxes, the conditional probabilities \( p_{\text{obs}}(a|x) \) for all values of \( a = (a_1, \ldots, a_N) \) and \( x = (x_1, \ldots x_N) \) that describe the observed process are obtained. These conditional probabilities form a full distribution \( P_{\text{obs}} \) with elements \( p_{\text{obs}}(a|x) \). In general, we will use an upper-case \( P \) for a distribution and lower-case \( p \) for an element of that distribution.

Different physical theories predict for different correlations \( p(a|x) \in P \) that may or may not have been observed in our experiments though in principle they may be attainable. Let us call \( C, Q \) and \( NS \) the set of correlations allowed by classical, quantum and maximally nonlocal theories. We will call \( T \) the set of correlations allowed by a particular theory. As mentioned earlier, within any device-independent protocol we need to determine the kind of theory that dictates how our devices work – if not an exact description of how they work – thus the most general set of correlations that are achievable by our devices (and other devices held by, for example, an eavesdropper).

**Randomness Certification in Correlation Theories**—Given an observed conditional probability distribution \( p_{\text{obs}}(a|x) \) there are different notions of randomness to consider. Firstly, we call the observed predictability for a particular input \( x_0 \) the guessing probability of the most likely outcome, i.e. \( G_{\text{obs}}(x_0, P_{\text{obs}}) = \max_{a} p_{\text{obs}}(a|x_0) \). However, part of the observed randomness can arise from a lack of information about the preparation of the observed probability distribution. How much of the observed randomness is then intrinsic?

Assuming we allow \( T \) to be the most general set of correlations that can be generated by any experiment then we can define the intrinsic predictability for the theory generating \( T \). For some observed correlations \( P_{\text{obs}} \) for an input \( x_0 \) this predictability is the highest guessing probability for outcomes \( a \) for a party in possession of all the information about the preparation of these observed correlations. We consider the worst-case scenario where the preparation that would give the lowest certified randomness is utilized. This is then an optimization over all possible probabilistic decompositions in terms of allowed correlations in theory \( T \) compatible with the observed correlations \( P_{\text{obs}} \) of the form

\[
G_{\text{int}}^{T}(x_0, P_{\text{obs}}) = \max_{\{p(\cdot), p_{\text{ex}}^{\lambda}\}} \sum_{e} p(e) G_{\text{obs}}(x_0, P_{\text{ex}}^{\lambda})
\]

subject to:

\[
\sum_{e} p(e) p_{\text{ex}}^{\lambda}(a|x) = p_{\text{obs}}(a|x).
\]

(1)

Here \( P_{\text{ex}}^{\lambda} \) represents an extremal probability distribution in the theory, i.e. a distribution \( P \) that cannot be written as a convex combination of other distributions in the theory. \( G_{\text{obs}}(x, P_{\text{ex}}^{\lambda}) = \max_{a} p_{\text{ex}}^{\lambda}(a|x) \) is also the intrinsic predictability of \( P_{\text{ex}}^{\lambda} \), since intrinsic and observed predictabilities must coincide for extremal probability distributions of the allowed set of correlations \( T \). Note that we are explicitly assuming the independence between the preparation components labelled by \( \epsilon \) and the measurement settings \( x \). This is commonly known as the freedom of choice assumption. Recent work has shown that this assumption can even be relaxed by implementing randomness amplification protocols [7, 13]. We can also quantify both observed and intrinsic predictabilities in terms of dits of observed and intrinsic randomness, as measured by the min-entropy, through the simple equalities \( H_{\text{obs}}(x_0, P_{\text{obs}}) = -\log_{2} G_{\text{obs}}(x_0, P_{\text{obs}}) \) and \( H_{\text{int}}^{T}(x_0, P_{\text{obs}}) = -\log_{2} G_{\text{int}}^{T}(x_0, P_{\text{obs}}) \) respectively.

In what follows, we provide bounds on how much randomness can be certified within a given theory \( T \), that is, we establish bounds on the solution to the optimization problem (1) over all correlations in \( T \). We also say that a theory \( T \) allows maximal randomness certification when if it is possible to obtain correlations \( P_{\text{obs}} \) in \( T \) for \( N \) parties and \( d \)-output measurements such that for a particular input \( x_0 \) the solution to (1) defines \( N \) dits of randomness (clearly, maximal randomness certification for all inputs is impossible). While solving this problem over all possible correlations of a theory seems at first hopeless, we manage to derive a non-trivial bound on randomness certification in maximally nonlocal theories, valid for arbitrary Bell test scenarios. This bound proves that maximally randomness certification is impossible in these theories. Moving to quantum correlations, we prove that maximal randomness certification is possible by providing correlations among \( N \) observers that perform two-measurement outputs such that the solution to (1) defines \( N \) random bits.

**Classical Correlations**—Before proving these results we briefly comment on classical correlations. How do these definitions of randomness apply to the set of classical correlations \( C \)? Classical correlations are those that can admit a local and deterministic explanation, i.e.

\[
p_{\text{obs}}(a|x) = \int_{\Lambda} p(\lambda) d\lambda \prod_{j=1}^{N} \delta_{f_j(x_j, \lambda)}(a_j),
\]

(2)

where \( p(\lambda) \geq 0 \) is a valid probability distribution over values of \( \lambda \in \Lambda \) such that \( \int_{\Lambda} p(\lambda) d\lambda = 1 \) and \( f_j \) is a deterministic function that takes the input \( x_j \) and \( \lambda \) and returns a value \( a_j \). It is easy to see that although all possible values of observed randomness \( H_{\text{obs}} \) can be engineered within this class of models, the intrinsic randomness \( H_{\text{int}}^{C} \) is always equal to zero. We now turn our attention towards sets of correlations where randomness can be certified.

**No-signaling Correlations**—We now formally define the set of maximally nonlocal correlations \( NS \). Here we permit any valid normalized probability distribution \( P \) with all elements satisfying \( 1 \geq p(a|x) \geq 0 \) where marginals are well-defined. That is, the probabilities (correlations) satisfy \( \sum_{a} p(a|x) = 1 \) and for all bipartitions of \( N \) parties into \( m \)
parties and \((N - m)\) parties then
\[
\sum_{a_j \in \{1, \ldots, m\}} p(a_1, a_2, \ldots, a_N | x_1, x_2, \ldots, x_N) = p(a_{m+1}, a_{m+2}, \ldots, a_N | x_{m+1}, x_{m+2}, \ldots, x_N),
\]
for all \(x_j\) such that \(j \in \{1, 2, \ldots, m\}\). Moreover, this condition is also applied to the probability distribution \(p(a|x)\) for all permutations of the \(N\) parties. Now this set is defined we present our first result.

**Result 1:** Given an arbitrary Bell test scenario \((N, M, d)\), assuming solely the no-signaling principle the randomness \(G^{NS}_{int}(x_0, P_{obs})\) that can be certified satisfies the bound
\[
G^{NS}_{int}(x_0, P_{obs}) \geq \frac{1}{d^N - (d - 1)^N},
\]
for all inputs \(x_0\).

To prove this result we only need to consider the randomness of the extreme points \(P^{es}\) of \(NS\) as indicated by \((1)\). Our proof is based on the simple observation that if elements \(p^{es}_x(a|x_0)\) of \(P^{es}\) are equal to zero then this severely inhibits the randomness of the distribution. That is, if for a particular \(x_0\), \(n\) values of \(p^{es}_x(a|x_0)\) are equal to zero then \(\max_{a} p^{es}_x(a|x_0) \geq \frac{1}{n}\). Result 1 then follows from the following Theorem, which provides a bound on the number of non-zero entries in extreme non-signalling correlations.

**Theorem:** Let \(p(a|x)\) be an extreme probability distribution in the set \(NS\) in an arbitrary Bell test scenario \((N, M, d)\). For a given combination of settings \(x_0\), denote by \(n(x_0)\) the number of probabilities \(p(a|x_0)\) that are equal to zero and define \(n = \min_{x_0} n(x_0)\). Then, \(n \geq (d - 1)^N\).

**Proof:** The proof of the result follows from a relatively simple counting argument. First we introduce some useful notation to describe the marginals of a probability distribution. If we have a set \(P\) with elements \(p(a|x)\) and we have a set \(J \subseteq \{1, 2, \ldots, N\}\) of the \(N\) parties then the probability distribution only over these parties in \(J\) is \(p(a^J|x^J) = \sum_{a_j \in \{1, \ldots, N\} \setminus J} p(a^J|x^J)\) where \(a^J\) and \(x^J\) consist of only elements \(a_j\) and \(x_j\) respectively for all \(j \in J\). Following a simple generalization of Ref. \([10]\), a probability distribution (for any input \(x_0\)) satisfying the no-signalling principle can be parametrized by \(p(a^J|x^J)\) for all possible sets \(J\). What is more, due to normalization we only consider \((d - 1)\) outputs for each party in all of these distributions. Therefore the probability \(p(a|x)\) is a function of \(p(a^J|x^J)\) for all \(J\) but the elements \(a_j\) of \(a^J\) only range over \((d - 1)\) values. Apart from when \(J\) contains all \(N\) parties, every other marginal probability \(p(a^J|x^J)\) will result from another probability distribution \(p(a'|x')\) for \(x' \neq x_0\) by summing over outputs of the appropriate parties as per \((3)\). Therefore the values of these marginals are fixed by probabilities for inputs \(x' \neq x_0\) and the only free parameters defining \(p(a|x_0)\) are the \((d - 1)^N\) probabilities when \(J\) contains all \(N\) parties.

Clearly, the space of this \(d^N\)-outcome probability distributions \(p(a|x_0)\) is convex. Moreover, if \(p(a|x_0)\) is not an extreme point in this space, neither are the original correlations \(p(a|x)\) in the original non-signalling space. As mentioned, when restricted to the specific setting \(x_0\), there are \((d - 1)^N\) free parameters. Now, the hyperplanes defining this convex space correspond to the positivity constraints defined by the \(d^N\) probabilities \(p(a|x_0)\). An extreme point in this space of dimension \((d - 1)^N\) should then be defined by the intersection of \((d - 1)^N\) hyperplanes. This implies that a necessary condition for the correlations \(p(a|x)\) to be extreme is that at least \((d - 1)^N\) probabilities \(p(a|x_0)\) are zero for each value of \(x_0\). This completes the proof. \(\square\)

At this point, it is worth mentioning two facts. First, this result indicates an important limitation on maximally nonlocal theories. In fact, the gap between the maximal randomness in principle attainable in a general Bell test scenario \(H^{max}_{int}\) and that achievable in maximally nonlocal theories \(H^{NS}_{int}\) is unbounded, as \(\lim_{d \to \infty} H^{NS}_{int} - H^{max}_{int} \to \infty\). Second, the derived bound is in general not tight. For instance, all extreme non-signalling correlations in Bell test scenarios \((2, M, 2)\) were obtained in \([17, 18]\). In this case, the number of zeros satisfies \(n \geq 2\), and thus one has at most one bit of certifiable randomness, while the previous bound predicts only one zero. Interestingly, the same gap of one zero appears when considering the extreme points of the \((3, 2, 2)\) scenario, classified in \([19]\). However, in the asymptotic limit of \(d \to \infty\) our bound gives \(\frac{1}{O(d^{-2})}\), which can be shown to be tight by comparing it with the results in \([20]\). We now move to randomness certification in quantum theory.

**Quantum Correlations**—For \(\rho\) being some positive semi-definite Hermitian matrix with unit trace and \(O_{a_j}\) being a positive operator valued measurement (POVM) for input \(a_j\) and output \(a_j\) such that \(\sum_{a_j} O_{a_j} = I\) (the identity matrix), all quantum correlations result from the following expression:

\[
p(a|x) = \text{tr}(\rho \bigotimes_{j=1}^N O_{a_j}),
\]

Characterizing the set of correlations achievable in this way is a great open problem in quantum information theory. Therefore, in what follows, rather than solving the optimization problem \((\text{I})\), we consider a relaxation of it where randomness is certified only by an observed Bell violation.

Given that a Bell inequality is just a (real) linear combination of probabilities \(p(a|x)\) over all inputs \(a\) and outputs \(x\), by the linearity of the trace, the quantum value of a Bell expression is \(\text{tr}(B)\) where \(B = \sum_{a, x} \beta_{ax} O_{a_j}^x\) and \(\beta_{ax}\) are real coefficients. Finding the guessing probability \(G^Q(x_0, P)\) certified by the maximal violation of a Bell inequality for the set of quantum correlations for any particular input \(x_0\) then
amounts to solving the following optimization \cite{5} for all $a$

$$\text{maximize } p(a|x_0)$$

subject to $\text{tr}(B \rho) = q_{\max}$, \quad $p(a|x_0) = \text{tr}(\rho \otimes O_{ax}^j)$,

where $q_{\max}$ is the maximal quantum violation of the Bell inequality described by the coefficients $\beta_{ax}$.

In \cite{21} a method was provided to detect when the maximal quantum violation of a Bell inequality certifies that the outputs are maximally random. The method has the advantage that it can be easily applied, but unfortunately it only works under the assumption that the maximal quantum violation of the inequality is unique. The uniqueness of the maximal quantum violation is in general hard to prove. However, in what follows, we consider Bell inequalities for which the uniqueness of the maximal violation can be proven using the results of Refs. \cite{22 23}. This then allows us to apply the simple method introduced in \cite{21} and prove the following result.

**Result 2:** For Bell test scenarios $(N, M, 2)$ for $N \geq 3$, with quantum mechanical correlations we can obtain $N$ bits of maximal randomness, i.e. $G^Q_{\text{int}}(x_0, P_{\text{obs}}) = \frac{1}{2^N}$.

We prove this result in the Appendix by generalizing the results of \cite{21} to all $N$ via a Bell inequality introduced in \cite{24}. We actually prove Result 2 for the $(N, 2, 2)$ scenario but this trivially applies to the $(N, M, 2)$ since we can always ignore $(M - 2)$ of the inputs for each party. While our proof does not apply to the case of two parties, it has been shown in \cite{22} that for the $(2, 2, 2)$ scenario an amount of randomness arbitrarily close to the maximum of 2 random bits can be certified in some limit. Additionally, numerical and analytical evidence indicates that exactly 2 bits of maximal randomness can be obtained in the $(2, 3, 2)$ scenario \cite{21}. All of this serves to show that quantum correlations certify maximal randomness even if maximally nonlocal theories cannot always do this.

We have outlined generic properties of two sets of correlations: the maximally nonlocal set and the quantum set. These properties contrast with each other in a concrete fashion and is the first time this contrast has been studied and highlighted. This observation allows us to posit that this highlights the **uniqueness** of quantum correlations: not only can we certify the generation of randomness (something impossible in classical physics) but we can certify **maximal** randomness (something impossible in maximally nonlocal theories). Perhaps the certification of maximal randomness could be an information principle that allows us to recover quantum mechanical correlations from the set of all correlations that satisfy no-signaling. Other examples of information principles include Information Causality \cite{12}, Non-trivial Communication Complexity \cite{13} and Local Orthogonality \cite{24}. Is it possible to recover quantum correlations utilising the principle that a theory should certify maximal randomness? In other words, are there other sets of correlations that allow for maximal randomness certification? We now address this question.

**Supra-quantum Correlations**—Navascués, Pironio and Acín introduced a means to approximate the set of quantum correlations which was an infinite hierarchy of semi-definite programs \cite{27}. This hierarchy has an infinite number of levels where each level defines a set of correlations defined in terms of a positive semi-definite matrix. For example, the first non-trivial level of this hierarchy is $Q^1$ and this set is provably larger than the set of quantum correlations $Q$. Already in the work of Pironio et al in Ref. \cite{5} these first few levels in the hierarchy was used to lower bound the amount of randomness that can be certified for quantum correlations. Since the first few levels of the hierarchy produce supra-quantum correlations, we would expect these sets of correlations not to give maximal randomness. We give our third main result that this intuition is incorrect by introducing a set of correlations that can produce maximal randomness.

For simplicity, we first describe the largest set of correlations called $Q^1$. In this case we have a quantum state $|\psi\rangle$ and associated with every measurement outcome $a_j$ for input $x_j$ for party $j$ we have a projector $E_j^{a_j,x_j}$. These projectors satisfy $\sum_j E_j^{a_j, x_j} = I$, $E_j^{a_j, x_j} E_j^{a'_j, x_j} = 0$ and $(E_j^{a_j, x_j})^\dagger = E_j^{a_j, x_j}$. We take the set $E = \{ E_j^{a_j, x_j} \}$ to be the set of all of these projectors (plus the identity). Correlations in the set $Q^1$ are obtained from all matrices $\Gamma$ with elements $\Gamma_{ij} = \langle \psi | \Pi_i \Pi_j | \psi \rangle$ where $\Pi_i, \Pi_j$ are linear combinations (over the complex numbers) of projectors in the set $E$. We can define subsets of $Q^1$ (and subsequently better approximations of $Q$) by allowing the set $E$ to contain products (or sequences) of projectors, e.g. $E_1^{a_1, x_1} E_2^{a_2, x_2}$.

Going further into the hierarchy, we now introduce the set of correlations $Q^{1+ABC}$ for three parties. At this level, $E$ is supplemented with the set $\{ E_1^{a_1, x_1} E_2^{a_2, x_2} E_3^{a_3, x_3} \}$ of sequences of three projectors where each projector corresponds to a different party. The set $Q^{1+ABC}$ is then defined in exactly the same way as $Q^1$ but with this new set of projectors $E \cup \{ E_1^{a_1, x_1} E_2^{a_2, x_2} E_3^{a_3, x_3} \}$. We are now in a position to present our final main result.

**Result 3:** For Bell test scenarios $(3, M, 2)$ for all $M$, with correlations belonging to the set $Q^{1+ABC}$, we can obtain 3 bits of maximal randomness, i.e. $G^Q_{\text{int}}(x_0, P_{\text{obs}}) = \frac{1}{8}$.

The proof of this result is presented in the Appendix. The crucial element in this proof is showing that there is only one probability distribution in the set $Q^{1+ABC}$ that maximally violates the Mermin inequality \cite{28}. We can then invoke again the results in Ref. \cite{21}, where it was shown that if there is only one probability distribution violating the Mermin inequality, then the maximal violation of the inequality certifies a maximal randomness (of 3 bits) in a particular input. It is perhaps worth noting that the set $Q^1$ does not permit maximal randomness certification utilising the Mermin inequality.

At first, this result may seem disappointing but there are
other examples of limitations to recovering quantum correlations from information principles. For example it is known that we need truly multipartite information theoretic principles \cite{24}. It has also been shown that other information principles will never recover quantum mechanical correlations \cite{25} and our work fits squarely within this foundational research program.

Discussion—We have shown that correlations in maximally nonlocal theories and quantum theory have drastically different consequences for randomness certification. Therefore, if we assume Nature does not abide by quantum mechanics but some more general theory, paradoxically we could have more predictive power than quantum theory allows. In fact, were the correlations only limited by the no-signalling principle, we could always have some non-trivial predictability in the outputs observed in a correlation experiment. These results are not only of foundational interest but have application in randomness extraction, certification and amplification. For example, in Ref. \cite{26} a lower bound on certifiable randomness was obtained using only the no-signaling principle, and this bound has found applications in other protocols (e.g. in Ref. \cite{27}). Not only do we show that maximal randomness is impossible for maximally nonlocal theories but we also give a bound on the randomness that is possible. Hopefully this bound will be useful in the design and analysis of future protocols. As mentioned above, the derived bound is not tight in general. An interesting follow-up question is to determine the exact maximum randomness allowed just by the no-signalling principle, a fundamental number providing a quantitative link between randomness and no-signalling.

In contrast, we have shown that quantum theory allows for maximal randomness certification where parties perform dichotomic measurements. However, we conjecture that quantum correlations can produce certifiable maximal randomness for all scenarios. In fact, another interesting follow-up question is to identify scenarios for maximal quantum randomness certification beyond the two-output case. To finish, we have also shown that maximal randomness certification is not a property that only quantum theory holds. We have given an example of a more nonlocal theory with the same ability in the case of three parties performing dichotomic measurements. This last result indicates that quantum theory is not so special from an information theoretic perspective (cf. Ref. \cite{30}). Also the set of quantum correlations is notoriously difficult to define whereas the set of maximally nonlocal correlations is defined by linear inequalities therefore there is an apparent trade off in the complexity of the set of correlations with the randomness that can be obtained from it. The fact that there exists a set of correlations that has a relatively simple description but facilitates maximal randomness certification provides a “third way” for the design and analysis of future protocols.

Acknowledgements—We acknowledge financial support from the ERC Consolidator Grant QITBOX, the grant “Intrinsic randomness in the quantum world” from the John Templeton Foundation, an SGR from the Generalitat de Catalunya, and the Spanish MINECO, through an FPI grant and projects Intrinqra and Chist-Era DIQIP.

* Electronic address: gonzalo.delatorre@icfo.es
† Electronic address: matthew.hoban@icfo.es

[1] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. \textbf{47}, 777-780, (1935).
[2] N. Bohr, Phys. Rev. \textbf{48}, 696-702, (1935).
[3] J. S. Bell, \textit{Physics} \textbf{1}: 195200 (1964).
[4] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, S. Wehner, \textit{arXiv}:1303.2849.
[5] Pironio et al, \textit{Nature} \textbf{464}, 1021 (2010).
[6] R. Colbeck, PhD Thesis, University of Cambridge (2006).
[7] R. Colbeck and R. Renner, \textit{Nat Phys}. \textbf{8}(6), 450-454, (2012).
[8] R. Gallego et al, \textit{Nature Communications} \textbf{4}, 2654 (2013).
[9] Acín et al, \textit{Phys. Rev. Lett.} \textbf{98}, 230501 (2007); B. W. Reichardt, F. Unger and U. Vazirani, \textit{Nature} \textbf{496}, 456640 (2013).
[10] J. Barrett, L. Hardy and A. Kent, \textit{Phys. Rev. Lett.} \textbf{95}, 010503 (2005).
[11] S. Popescu and D. Rohrlich, \textit{Found. Phys.} \textbf{24}, 379 (1994).
[12] M. Pawlowski et al, \textit{Nature} \textbf{461}, 1101 (2009).
[13] G. Brassard et al, \textit{Phys. Rev. Lett.} \textbf{96}: 250401, (2006).
[14] C. Dhara, G. de la Torre and A. Acín, \textit{Phys. Rev. Lett.} \textbf{112}, 100402 (2014).
[15] L. P. Thinh, L. Sheridan, V. Scarani, \textit{Phys. Rev. A} \textbf{87}, 062121 (2013).
[16] D. Collins and N. Gisin, \textit{J. Phys. A: Math. Gen.} \textbf{37} 1775 (2004).
[17] N. S. Jones, Ll. Masanes \textit{Phys. Rev. A} \textbf{72}, 052312 (2005).
[18] J. Barrett and S. Pironio, Phys. Rev. Lett. \textbf{95}, 140401 (2005).
[19] S. Pironio, J-D. Bancal, V. Scarani, \textit{J. Phys. A: Math. Theor.} \textbf{44}, 065303 (2011)
[20] L. Aolita, R. Gallego, A. Cabello, A. Acín, \textit{Phys. Rev. Lett.} \textbf{108}, 100401 (2012).
[21] C. Dhara, G. Prettico and A. Acín, \textit{Phys. Rev. A} \textbf{88}, 052116 (2013).
[22] T. Franz, F. Furrer, R.F. Werner, \textit{Phys. Rev. Lett.} \textbf{106}, 250502 (2011).
[23] C. A. Miller and Y. Shi, \textit{8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013)} \textbf{22}, 254-262 (2013).
[24] M. J. Hoban et al, \textit{New J. Physics} \textbf{13}, 023014 (2011).
[25] A. Acín, S. Massar and S. Pironio, \textit{Phys. Rev. Lett.} \textbf{108}, 100402 (2012).
[26] T. Fritz et al, \textit{Nature Communications} \textbf{4}, 2263 (2013).
[27] M. Navascués, S. Pironio, A. Acín, \textit{Phys. Rev. Lett.} \textbf{98}, 010401 (2007); M. Navascués, S. Pironio, A. Acín, \textit{New J. Phys.} \textbf{10}, 073014 (2008).
[28] D. Mermin, \textit{Phys. Rev. Lett.} \textbf{65}, 1838-1840, (1990).
[29] R. Gallego, L. E. Würflinger, A. Acín, M. Navascués, \textit{Phys. Rev. Lett.} \textbf{107}, 210403 (2011).
[30] M. Navascués, Y. Guruyanova, M. J. Hoban, A. Acín, \textit{in preparation}.
[31] U. Vazirani and T. Vidick, \textit{arXiv}:1210.1810v2 [quant-ph] (2012).
APPENDIX

Proof of Result 2

In this section we show that it is possible to obtain \( N \) bits of global randomness for all \( N \). In Ref. [21] it was shown how to achieve and certify \( N \) bits of global randomness for all odd \( N \). Here, we show that there is a Bell inequality in the \((N, 2, 2)\) setting for all \( N \) (which is a generalization of the Mermin inequality first studied in [24]) which if maximally violated, gives \( N \) bits of global randomness. We use the tools developed in Ref. [21] to obtain maximal randomness based on the symmetries of the inequality we use.

First, we need to introduce some notation. As standard in the literature, we introduce the \( n \)-party correlators where \( n \leq N \). Take a subset \( J_n \subseteq \{1, 2, ..., N\} \) of \( n \) parties from all \( N \) parties. Then associated with this subset and a string of inputs \( x_{J_n} = \{x_j | j \in J_n\} \), a string of outputs \( a_{J_n} = \{a_j | j \in J_n\} \) and a marginal probability distribution \( p(a_{J_n} | x_{J_n}) \). We then define the correlators to be

\[
\langle x_{J_n} \rangle := 2 \left( \sum_{a_{J_n}} \alpha p(a_{J_n} | x_{J_n}) \right) - 1, \tag{6}
\]

with \( \alpha = 1 + \sum_{k \in J_n} a_k \text{ mod } 2 \). We can define the full joint probabilities in terms of these correlators as

\[
p(a|x) = \frac{1}{2^N} \sum_{J_n} (-1)^{\sum_{k \in J_n} a_k} \langle x_{J_n} \rangle, \tag{7}
\]

where we take a sum over all \( 2^N \) subsets of \( N \) parties (including the empty set).

The Bell inequality that will concern is the following inequality discussed in Ref. [24]:

\[
\sum_x (-1)^{f(x)} \delta_0^{g(x)} \langle x_{J_n} \rangle \leq \epsilon < 2^{N-1}, \tag{8}
\]

where \( f(x) = \sum_{j=1}^{N-1} x_j \left( \sum_{k=j+1}^{N} x_k \right) \text{ mod } 2 \) and \( g(x) = \sum_{j=1}^{N} x_j \text{ mod } 2 \). As indicated the upper-bound for local hidden variables \( \epsilon \) is strictly less than \( 2^{N-1} \), the number of terms in the sum. Crucially, quantum mechanics can violate this inequality and achieve the algebraic upper bound of \( 2^{N-1} \) as shown in Ref. [24] using a Greenberger-Horne-Zeilinger state. Also, there is only one probability distribution that maximally violates this inequality as can be shown by applying the techniques of Ref. [22] or by a self-testing argument due to [23]. In Ref. [21] this property of uniqueness of a probability distribution maximally violating an inequality was used to prove that global randomness can be generated from a Bell test.

We need to use an input \( x' \) that does not appear in the left-hand-side of (8), since the outputs of measurements for inputs in (8) will be highly correlated, and thus not random. We need to show that for input \( x' \), the probability \( p(a|x') \) in (7) is equal to \( \frac{1}{2^N} \) for all \( a \). This occurs if all correlators satisfy \( \langle x_{J_n} \rangle = 0 \) for all (non-empty) subsets \( J_n \) and \( \langle x_{\{\}} \rangle = 1 \) for \( J_n = \{\} \), the empty set when \( n = 0 \). The aim of this section is to show that this is true.

To do this, we utilize the tools in Ref. [21] where we perform transformations on the data obtained in a Bell test that do not affect the correlators that appear in the Bell inequality of (8). These transformations affect correlators that do appear in the inequality. If we take the unique probability distribution that maximally violates (8) then, under these transformations, it still violates the same inequality maximally. If we call the original probability distribution \( P \) with elements \( p(a|x) \) and the transformed distribution \( P' \), then \( P = P' \), and so all correlators resulting from these two distributions must be equal as well. If one of these symmetry transformations is to flip an outcome of a measurement depending on the choice of input then this can alter correlators, e.g. \( a_1 \rightarrow a_1 \oplus x_1 \), then all correlators \( \langle x_{J_n} \rangle \) that contain \( x_1 = 1 \) have their sign flipped as \( \alpha = 1 + \sum_{k \in J_n} a_k \text{ mod } 2 \rightarrow 2 + \sum_{k \in J_n} a_k \text{ mod } 2 \). However, due to uniqueness of quantum violation this implies that the correlators before and after the transformation are equal, so in the case that the transformation flips the sign of the correlator then \( \langle x_{J_n} \rangle = -\langle x_{J_n} \rangle = 0 \). This thus demonstrates a way to show that correlators are zero for particular distributions.

For clarity we introduce the notation to show when a correlator’s sign is flipped. Our symmetry operations are captured by an \( n \)-length bit-string \( s \), where if the \( j \)-th element \( s_j \) is zero, then we flip \( a_j \) for the choice of input \( x_j = 0 \), and if \( s_j = 1 \), then we flip \( a_j \) for choice of input \( x_j = 1 \). Then the correlator \( \langle x_{J_n} \rangle \) under the symmetry transformation described by \( s \) is mapped to \( (-1)^{N-H(x,s)} \langle x_{J_n} \rangle \) where \( H(x, s) \) is the Hamming distance between the bit-strings \( x, s \): the number of times \( x_j \neq s_j \) for bit-strings \( x, s \). Another way of writing the Hamming distance is

\[
H(x, s) = \sum_{j=1}^{N} x_j + s_j \text{ mod } 2 = \sum_{j=1}^{N} x_j + s_j - 2s_jx_j, \tag{9}
\]
We now return to the Bell inequality in (8) and focus on even $N$. We apply $N$ transformations described by the bit-strings $s$: $(0,0,...,0)$ (the all-zeroes bit-string) and the $(N-1)$ bit-strings $s$ that all have $s_N = 1$, and only one other element being equal to one, e.g. $(1,0,0,...,1)$ or $(0,1,0,...,1)$. All of these bit-strings have an even number of ones, therefore $\sum_{j=1}^{N} s_j = 2k$ for $k \in \{0,1\}$. For correlators in the inequality of (8), the inputs $x$ satisfy $\sum_{j=1}^{N} x_j \mod 2 = 0$, so that $\sum_{j=1}^{N} x_j = 2k'$ for some integer $k'$. Therefore all the correlators $\langle x \rangle$ that appear in (8) are mapped to $(-1)^{2(k+k'-\sum_{j=1}^{N} s_j x_j)} \langle x \rangle = \langle x \rangle$ and thus the transformation does not alter the inequality.

To obtain $N$ bits of global randomness for even $N$, we choose the input $x' = (1,1,...,1,0)$, the bit-string of all-ones except $x_N' = 0$. This input does not appear in (8) and indeed $\sum_{j=1}^{N} x_j = 2k'+1$ for some integer $k'$, therefore the above transformations map $\langle x' \rangle$ to $(-1)^{1+2(k+k'-\sum_{j=1}^{N} s_j x_j)} \langle x' \rangle = -\langle x' \rangle$. Due to the uniqueness of the probability distribution maximally violating the Bell inequality, $\langle x' \rangle = -\langle x' \rangle = 0$.

We now need to show that all correlators $\langle x_{J_n} \rangle$ where $x_{J_n}$ is the string of $n < N$ elements from $x' = (1,1,...,1,0)$ for a subset $J_n$. We consider the Hamming distance $H(x_{J_n}, s_{J_n})$ between $x_{J_n}$ and the corresponding string $s_{J_n}$ of elements of $s$ where $s_j$ is in $s_{J_n}$ if $j \in J_n$. Immediately we see that for at least one string $s_{J_n}$, the Hamming distance is $H(x_{J_n}, s_{J_n}) = (n - 1)$. Therefore, there is at least one transformation $s$ that maps $\langle x_{J_n} \rangle$ to $(-1)^{n-(n-1)} \langle x_{J_n} \rangle = -\langle x_{J_n} \rangle$ for all $J_n$. Again, given that all correlators should be equal after the transformation we have that $\langle x_{J_n} \rangle = -\langle x_{J_n} \rangle = 0$.

To summarize, we have shown that all correlators that appear in (21) for the input $x' = (1,1,...,1,0)$ equal to zero if the probability distribution that produces them maximally violates the inequality in (8). This therefore implies that $p(a|x') = \frac{1}{2^n}$ for all $a$ and for all even $N$. For this input $x'$ we obtain $N$ bits of global randomness. We can use another inequality to obtain $N$ bits of global randomness for odd $N$ as shown in Ref. [27]. Therefore, we can obtain $N$ bits of global randomness for all $N$. We have used the fact that there is a unique quantum violation of the inequality in (8). However, for more general theories this may not be the case.

**Proof of Result 3**

We present a proof that it is possible to certify 3 bits of global randomness for a set of correlations that is strictly larger than the quantum set $Q$. We call this set $Q^{1+ABC}$ in the terminology of the multipartite generalization of the Navascués-Pironio-Acin hierarchy of correlations that can be characterized through semi-definite programming [27]. We prove this result utilising the tripartite Mermin inequality [28], so we are therefore in the $(3,2,2)$ scenario.

We first recall from Ref. [27] that correlations $p(a_1,a_2,a_3|x_1,x_2,x_3)$ are contained in the set $Q^{1+ABC}$ if there exists a pure quantum state $|\psi\rangle$, and projectors $\{E_{x_1}^{a_1}, F_{x_1}^{a_1}, G_{x_2}^{a_3}\}$ labelled by inputs $x_j \in \{0,1\}$ and outputs $a_j \in \{0,1\}$, such that

1. **(Hermiticity)** $- (E_{x_1}^{a_1})^\dagger = E_{x_1}^{a_1}, (F_{x_1}^{a_1})^\dagger = F_{x_1}^{a_1}, (G_{x_2}^{a_3})^\dagger = G_{x_2}^{a_3}$ for all $x_j$ and $a_j$

2. **(Normalization)** $\sum_{a_1} E_{x_1}^{a_1} = 1, \sum_{a_1} F_{x_1}^{a_1} = 1, \sum_{a_3} G_{x_2}^{a_3} = 1$ for all $x_j$

3. **(Orthogonality)** $-E_{x_1}^{a_1} F_{x_1}^{a_1} = \delta_{a_1} E_{x_1}^{a_1}, F_{x_1}^{a_1} G_{x_2}^{a_2} = \delta_{a_2} F_{x_1}^{a_1}, G_{x_2}^{a_2} G_{x_3}^{a_3} = \delta_{a_3} G_{x_2}^{a_3} G_{x_3}^{a_3}$ for all $x_j$,

such that probabilities are $p(a_1,a_2,a_3|x_1,x_2,x_3) = \langle \psi | E_{x_1}^{a_1} F_{x_1}^{a_1} G_{x_2}^{a_3} | \psi \rangle$. In addition to these general constraints, linear combinations of these probabilities are elements of a positive semidefinite matrix $\Gamma_{1+ABC} \succeq 0$. We choose a specific positive semidefinite matrix with elements $[\Gamma_{1+ABC}]_{ij} = \langle \psi | O_i^\dagger O_j | \psi \rangle$ where $O_i \in \{I,\{A_i\},\{B_i\},\{C_i\},\{A_i,B_i,C_i\}\}$ for $A_i = E_{x_1}^{a_1} - F_{x_1}^{a_1}, B_j = F_{x_1}^{a_1} - F_{x_1}^{b_1}$ and $C_k = G_{x_2}^{a_3} - G_{x_2}^{b_3}$. Therefore the matrix $\Gamma_{1+ABC}$ is a 15-by-15 matrix with each $O_i$ labelling a row or column. We can now make several observations: $O_i O_I = I$ for all $O_I$ therefore $(O_I)^\dagger = O_I$; $\langle x_1 x_2 x_3 \rangle = \langle x_1 x_2 x_3 | A_i B_j C_k | \psi \rangle$; $\langle x_1 x_2 \rangle = \langle x_1 x_2 | A_i C_k | \psi \rangle$; $\langle x_1 x_3 \rangle = \langle x_1 x_3 | A_i B_j | \psi \rangle$; $\langle x_2 x_3 \rangle = \langle x_2 x_3 | B_j C_k | \psi \rangle$; $\langle x_1 \rangle = \langle x_1 | A_i | \psi \rangle$; $\langle x_2 \rangle = \langle x_2 | B_j | \psi \rangle$; $\langle x_3 \rangle = \langle x_3 | C_k | \psi \rangle$. Here we utilized the notation introduced in the previous section. Finally, the set of quantum correlations $Q$ is a subset of $Q^{1+ABC}$ since the former can be recovered from the latter by imposing more constraints on the projectors. It can also be shown that $Q$ is a strict subset of $Q^{1+ABC}$ for all possible scenarios ($N,M,d$).

Now that we have defined the set $Q^{1+ABC}$ of correlations that concerns us, we return to the issue of randomness certification. We wish to show that for correlations in this set that maximally violate the tripartite Mermin inequality [28]

$$\langle 001 \rangle + \langle 010 \rangle + \langle 100 \rangle - \langle 111 \rangle \leq 2,$$

$p(a|x_0) = \frac{1}{2}$ for all $a$ for a particular input $x_0$. We choose this input to be $x_0 = (0,0,0)$ but it will turn out that we could choose any input $x$ that does not appear in the Mermin inequality. The maximal violation of the Mermin inequality is 4 and because this violation is achievable with quantum mechanics [28] and $Q \subseteq Q^{1+ABC}$, it is achievable in $Q^{1+ABC}$ also. Therefore,
ascertaining the maximal probability \( p(\mathbf{a}|000) \) compatible with this violation and for correlations in \( Q^{1+ABC} \) is an optimization of the form:

\[
\begin{align*}
\text{maximize} & \quad p(\mathbf{a}|000) \\
\text{subject to} & \quad \langle 001 \rangle + \langle 010 \rangle + \langle 100 \rangle - \langle 111 \rangle = 4, \\
& \quad p(\mathbf{a}|000) \in Q^{1+ABC}.
\end{align*}
\]

(11)

Given our construction of correlations in \( Q^{1+ABC} \), we can rephrase this optimization in terms of a semidefinite program:

\[
\begin{align*}
\text{maximize} & \quad \frac{1}{2}\text{tr}(M \Gamma_{1+ABC}) \\
\text{subject to} & \quad \frac{1}{2}\text{tr}(B \Gamma_{1+ABC}) = 4, \\
& \quad \Gamma_{1+ABC} \succeq 0, \\
& \quad \frac{1}{2}\text{tr}(D_i \Gamma_{1+ABC}) = 0, i \in \{1, 2, ..., m\},
\end{align*}
\]

(12)

where \( M, B \) and \( D_i \) are real, symmetric 15-by-15 matrices such that \( \frac{1}{2}\text{tr}(M \Gamma_{1+ABC}) = p(\mathbf{a}|000) \) and \( \frac{1}{2}\text{tr}(B \Gamma_{1+ABC}) = \langle 001 \rangle + \langle 010 \rangle + \langle 100 \rangle - \langle 111 \rangle \). Due to (7), we can impose the former equality on \( 4\text{tr}(M \Gamma_{1+ABC}) \). The \( m \) matrices \( D_i \) just impose constraints on elements of \( \Gamma_{1+ABC} \) such that they are compatible with \( Q^{1+ABC} \).

We now fix the particular representation of \( \Gamma_{1+ABC} \) with elements \([\Gamma_{1+ABC}]_{ij} = \langle \psi|O_i O_j|\psi \rangle \) such that for both rows \( i \) and columns \( j \) we write the ordered vector of operators \( O_i \) with \( i \) increasing from left to right:

\[(O_1, ..., O_{15}) = (I, A_0, A_1, B_0, B_1, C_0, C_1, A_0 B_1 C_1, A_1 B_0 C_1, A_1 B_1 C_0, A_1 B_0 C_0, A_0 B_0 C_0, A_0 B_1 C_0, A_0 B_0 C_1, A_1 B_1 C_1).
\]

(13)

Immediately we observe that the diagonal elements of the matrix \([\Gamma_{1+ABC}]_{ii} = 1 \) and thus the magnitude of all elements of the matrix \([|\Gamma_{1+ABC}]_{ij}| \leq 1 \) are bounded if the matrix is positive semidefinite. For example, given this representation \( B = C + C^\dagger \) where \( C = \left( \begin{array}{c} 0 \\ \vdots \\ 0 \end{array} \right) \) for \( 0 \) being a 14-by-15 matrix of zeroes and \( w = (0, 0, ..., 0, 1, 1, -1) \).

There is a unique solution to the problem in (12) if instead of the probability distribution being in \( Q^{1+ABC} \) it is constrained to be in \( Q \). As mentioned \( Q \subseteq Q^{1+ABC} \), so we can write this solution as a matrix of the form \( \Gamma_{1+ABC} \), and we call this solution matrix \( \Gamma_M \) and define it as follows:

**Definition 1.** The only solution matrix \( \Gamma_M \) to (12) that can be realized in quantum theory has elements

1. \([\Gamma_M]_{ij} = 1 \) if \( i = j \), \([\Gamma_M]_{ij} \in \{(001), (010), (100)\}\) and \([\Gamma_M]_{ij} \in \{\langle \psi|P(P')^\dagger|\psi \rangle, \langle \psi|(P')^\dagger P|\psi \rangle\} \) where \( P, P' \in \{A_0, A_1, B_0 B_1, C_0 C_1\} \) and \( P \neq P' \);
2. \([\Gamma_M]_{ij} = -1 \) if \([\Gamma_M]_{ij} = \langle 111 \rangle \) and \([\Gamma_M]_{ij} \in \{\langle \psi|PP'|\psi \rangle, \langle \psi|P'P|\psi \rangle\} \) where \( P, P' \in \{A_0 A_1, B_0 B_1, C_0 C_1\} \) and \( P \neq P' \);
3. \([\Gamma_M]_{ij} = 0 \) otherwise.

We now present the main theorem of this section.

**Theorem 1.** The only possible solution matrix \( \Gamma_{1+ABC} \) to the semidefinite program in (12) is \( \Gamma_M \).

This immediately leads to the following corollary that is relevant for randomness certification. That is, since the solution to the semidefinite program in (12) is the quantum solution we inherit the result of Dhara et al [21] that shows that we obtain three random bits if we maximally violate the Mermin inequality [28]. We state this result more formally in the following corollary.

**Corollary.** The maximal value of the objective function \( \frac{1}{2}\text{tr}(M \Gamma_{1+ABC}) = p(\mathbf{a}|000) \) in the semidefinite program (12) is equal to \( \frac{1}{5} \) for all \( \mathbf{a} \).

**Proof** – First we observe that, as obtained from the definition of matrix \( \Gamma_M \), \( \langle 0 \rangle = 0 \) for every party’s single-body correlator, and equally \( \langle 00 \rangle = 0 \) for all two-body correlators between the three parties, and \( \langle 000 \rangle = 0 \). Substituting these values into (7), we then obtain \( p(\mathbf{a}|000) = \frac{1}{5} \) for all \( \mathbf{a} \). Since \( \Gamma_M \) is the only possible solution to (12), this is the only possible probability distribution over \( \mathbf{a} \). \( \square \)

To prove theorem 1 we require two lemmas that will be introduced and proved in the sequel. The first lemma describes the structure of the feasible matrices \( \Gamma_{1+ABC} \), i.e. the matrices that satisfy all of the constraints in (12). The second lemma just says that matrices \( \Gamma_{1+ABC} \) of this form are positive semidefinite if and only if they are equal to \( \Gamma_M \). We now present and prove these lemmas. For simplicity we utilize the notation for the notation \( \langle O \rangle = \langle \psi|O|\psi \rangle \) with \( O \in \{A_j, B_j, C_j| j \in \{0, 1\}\} \).
Lemma 1. Matrices $\Gamma_{1+ABC}$ are feasible (satisfy all constraints therein) for the semidefinite program [12] if and only if they are of the form:

$$\Gamma_{1+ABC} = \begin{pmatrix}
1 & q_1 & q_2 & q_3 \\
q_1^T & X & Y & Z \\
q_2^T & \mathbb{X} & \mathbb{Y} & \mathbb{Z} \\
q_3^T & \mathbb{Y} & \mathbb{Z} & \mathbb{D}
\end{pmatrix},$$

with

$$q_1 = \left(\langle A_0\rangle, \langle A_1\rangle, \langle B_0\rangle, \langle B_1\rangle, \langle C_0\rangle, \langle C_1\rangle\right),$$

$$q_2 = (0, 0, 0, 0),$$

$$q_3 = (1, 1, 1, -1),$$

$$\mathbb{W} = \begin{pmatrix}
\mathbb{I} & C & B \\
C^T & \mathbb{I} & A \\
B^T & A^T & \mathbb{I}
\end{pmatrix},$$

$$\mathbb{X} = \begin{pmatrix}
\langle A_1 \rangle - \langle A_1 \rangle & \langle A_1 \rangle - \langle A_1 \rangle \\
\langle A_0 \rangle - \langle A_0 \rangle & \langle A_0 \rangle - \langle A_0 \rangle \\
\langle B_1 \rangle - \langle B_1 \rangle & \langle B_1 \rangle - \langle B_1 \rangle \\
\langle B_0 \rangle - \langle B_0 \rangle & \langle B_0 \rangle - \langle B_0 \rangle \\
\langle C_1 \rangle - \langle C_1 \rangle & \langle C_1 \rangle - \langle C_1 \rangle \\
\langle C_0 \rangle - \langle C_0 \rangle & \langle C_0 \rangle - \langle C_0 \rangle
\end{pmatrix},$$

$$\mathbb{Y} = \begin{pmatrix}
\langle A_0 \rangle - \langle A_0 \rangle & \langle A_0 \rangle - \langle A_0 \rangle \\
\langle A_1 \rangle - \langle A_1 \rangle & \langle A_1 \rangle - \langle A_1 \rangle \\
\langle B_0 \rangle - \langle B_0 \rangle & \langle B_0 \rangle - \langle B_0 \rangle \\
\langle B_1 \rangle - \langle B_1 \rangle & \langle B_1 \rangle - \langle B_1 \rangle \\
\langle C_0 \rangle - \langle C_0 \rangle & \langle C_0 \rangle - \langle C_0 \rangle \\
\langle C_1 \rangle - \langle C_1 \rangle & \langle C_1 \rangle - \langle C_1 \rangle
\end{pmatrix},$$

$$\mathbb{D} = \begin{pmatrix}
1 & 1 & 1 & -1 \\
1 & 1 & 1 & -1 \\
1 & 1 & 1 & -1 \\
-1 & -1 & -1 & 1
\end{pmatrix},$$

where $\mathbb{D}$ is a 4-by-4 matrix of all-zeroes, with $A = \left(\langle A_1 \rangle - \langle A_1 \rangle, B = \left(\langle B_1 \rangle - \langle B_1 \rangle, C = \left(\langle C_1 \rangle - \langle C_1 \rangle \right)$ and $\mathbb{I} = \begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix}$.

Proof – Vectors $q_1$ and $q_3$ are trivially obtained if the constraints in [12] are satisfied.

We now use the observation that for all feasible matrices $\Gamma_{1+ABC}$, the elements $[\Gamma_{1+ABC}]_{ij} \in \{\langle 001\rangle, \langle 010\rangle, \langle 100\rangle\}$ are all equal to 1 and when $[\Gamma_{1+ABC}]_{ij} = \langle 111\rangle$ the element is equal to -1. This is due to the fact that this is the only combination of values compatible with maximal violation of the Mermin inequality. This fact implies that $\langle \psi | R | \psi \rangle = \langle \psi | \psi \rangle$ for $R \in \{A_0B_0C_1, A_0B_1C_0, A_1B_0C_0\}$ and $\langle \psi | A_1B_1C_1 | \psi \rangle = -\langle \psi | \psi \rangle$ by normalization,

$$A_0B_0C_1 | \psi \rangle = | \psi \rangle,$$

$$A_0B_1C_0 | \psi \rangle = \langle \psi |,$$

$$A_1B_0C_0 | \psi \rangle = | \psi \rangle,$$

$$A_1B_1C_1 | \psi \rangle = -\langle \psi |.$$

This implies that $\langle \psi | P R | \psi \rangle = \langle \psi | P | \psi \rangle$ for $R \in \{A_0B_0C_1, A_0B_1C_0, A_1B_0C_0\}$ and $\langle \psi | P A_1B_1C_1 | \psi \rangle = -\langle \psi | P | \psi \rangle$ where $P$ is any $O_j$ as described above for $j \in \{1, 2, ..., 15\}$. Utilising this observation we obtain the sub-matrix $\mathbb{D}$ in [13] if $P$ is equal to any of the $R$ described above. Also for $P \in \{A_i, B_j, C_k\}$ for all $i, j, k$, we again utilize this observation to obtain $\mathbb{Y}$ and certain elements of $\mathbb{X}$. The elements of $\mathbb{X}$ that are obtained via this observation are those where $\langle \psi | O_j O_j | \psi \rangle = \langle \psi | P R | \psi \rangle$ with $P$ and $R$ being described above.

To obtain the remaining elements of $\mathbb{X}$ that do not satisfy the above conditions, we utilized another consequence of the conditions of [15]. That is, since $O_j O_i = \mathbb{I}$, any element of $\Gamma_{1+ABC}$ equal to $\langle \psi | S | \psi \rangle$ for $S \in \{A_i B_j, A_i C_k, B_j C_k\}$ is equal to $\pm \langle \psi | S' | \psi \rangle$ for $S' \in \{A_0B_0C_1, A_0B_1C_0, A_1B_0C_0, A_1B_1C_1\}$. The sign in front of $\langle \psi | S' | \psi \rangle$ is determined by the product $SS'$. We also use this observation to obtain matrices $A, B$ and $C$. 
It remains to be shown how the vector $\mathbf{q}_2$, the matrix $\mathcal{O}$ and the submatrices $\mathbb{I}$ in $\mathcal{W}$ are obtained. We first observe that $\mathbf{q}_2 = (w, x, y, z)$ where $w = \langle \psi | A_0 B_1 C_1 | \psi \rangle$, $x = \langle \psi | A_1 B_0 C_1 | \psi \rangle$, $y = \langle \psi | A_1 B_1 C_0 | \psi \rangle$, and $z = \langle \psi | A_0 B_0 C_0 | \psi \rangle$. Utilising the relations in (15), we obtain
\[
\mathcal{O} = \begin{pmatrix}
w & w & w & -w \\
x & x & x & -x \\
y & y & y & -y \\
z & z & z & -z \\
\end{pmatrix}.
\] (16)

We now observe that $\mathcal{O}$ can be defined in an equivalent way since $\mathcal{O}_i \mathcal{O}_i = \mathbb{I}$ for all $\mathcal{O}_i$. Using this observation and $\langle \psi | \mathcal{O}_i \mathcal{O}_j | \psi \rangle = \langle \psi | \mathcal{O}_j \mathcal{O}_i | \psi \rangle$ for $\mathcal{O}_i, \mathcal{O}_j \in \{A_i, B_j, C_k\}$ and $\mathcal{O}_i \neq \mathcal{O}_j$, we obtain
\[
\mathcal{O} = \begin{pmatrix}
w & \langle C_1 | C_1 \rangle & \langle B_0 B_1 \rangle & \langle A_0 A_1 \rangle \\
\langle C_1 | C_1 \rangle & x & \langle A_0 A_1 \rangle & \langle B_0 B_1 \rangle \\
\langle B_0 B_1 \rangle & \langle A_0 A_1 \rangle & y & \langle C_1 | C_1 \rangle \\
\langle A_0 A_1 \rangle & \langle B_0 B_1 \rangle & \langle C_1 | C_1 \rangle & -z \\
\end{pmatrix},
\] (17)

where again we are using the notation $\langle \psi | \mathcal{O}_i \mathcal{O}_j | \psi \rangle = \langle \mathcal{O}_j \mathcal{O}_i | \psi \rangle$ for brevity. Since the matrix in (16) and (17) have to be equal to each other, the only possible solution is that $\mathcal{O}$ is a 4-by-4 matrix of zeroes. This also implies that $\mathbf{q}_2 = (0, 0, 0, 0)$ and $\langle A_0 A_1 \rangle = \langle B_0 B_1 \rangle = \langle C_0 C_1 \rangle = 0$, thus completing the matrix $\mathcal{W}$. This also completes our proof. $\square$

We now present our final lemma that will complete the proof of theorem [1].

**Lemma 2.** The matrix $\Gamma_{1+ABC}$ described by (1) is positive semidefinite if and only if $\Gamma_{1+ABC} = \Gamma_M$.

**Proof** – We can use the Schur complement of $\Gamma_{1+ABC}$ in (1) and that $\mathbb{D} = \mathbf{q}_3^T \cdot \mathbf{q}_3$ and $\mathbb{Y} = \mathbf{q}_1^T \cdot \mathbf{q}_3$ to show that $\Gamma_{1+ABC}$ is positive semidefinite if and only if
\[
\begin{pmatrix}
\mathbb{W}' & \mathbf{X} \\
\mathbf{X}^T & \mathbb{D} \\
\end{pmatrix} \succeq 0,
\] (18)

where $\mathbb{W}' = \mathbb{W} - \mathbf{q}_1^T \cdot \mathbf{q}_1$. For example, for the matrix $\Gamma_M$, the corresponding submatrix $\Gamma'_M$ from (18) is
\[
\Gamma'_M = \begin{pmatrix}
\mathbb{I} & 0 \\
0^T & \mathbb{D} \\
\end{pmatrix},
\] (19)

where $\mathbb{I}$ is the 6-by-6 identity matrix and $0$ is a 6-by-4 matrix of zeroes. This submatrix of $\Gamma_M$ is positive semidefinite if and only if $\mathbb{D} \succeq 0$ which is indeed true.

Since the space of positive semi-definite matrices is convex, the set of feasible matrices $\Gamma_{1+ABC}$ for the semidefinite program (12) is a convex set. Therefore, if there is a submatrix $\Gamma_1$ of the form (18), we can obtain another submatrix $\Gamma_2$ of the form (18) that is a convex combination of $\Gamma_1$ and $\Gamma'_M$. We assume that $\Gamma_1$ has elements corresponding to some non-zero values $\{\langle A_i \rangle, \langle B_j \rangle, \langle C_k \rangle\}$, therefore completely unlike $\Gamma_M$. We now show that there exist matrices of the form $\Gamma_2$ that are not positive semidefinite which implies that any matrix $\Gamma_1$ as described is not positive semidefinite. This in turn implies that the only positive semidefinite matrix of the form (18) is $\Gamma_M$.

We choose $\Gamma_2$ such that $\sum_{j=0}^3 |\langle A_j \rangle| + |\langle B_j \rangle| + |\langle C_j \rangle| \ll 1$ but at least one of the elements of the set $\{\langle A_0 \rangle, \langle B_0 \rangle, \langle C_0 \rangle\}$ is non-zero. As mentioned before, since the space of solution matrices $\Gamma_{1+ABC}$ is convex we can always choose such a matrix without loss of generality. Therefore, the matrix in (18) is positive semidefinite if and only if
\[
\begin{pmatrix}
\mathbb{W}' & \mathbf{X} \\
\mathbf{X}^T & \mathbb{D} \\
\end{pmatrix} - \frac{1}{(1 - \langle A_0 \rangle^2)} \begin{pmatrix}
\mathbf{s} \mathbf{r}^T \\
\mathbf{r} \mathbf{s}^T \\
\end{pmatrix} \succeq 0,
\] (20)

where $\mathbf{X} = \begin{pmatrix} \mathbf{X} \end{pmatrix}$ where $\mathbf{r} = (-\langle A_1 \rangle, \langle A_1 \rangle, \langle A_1 \rangle, \langle A_1 \rangle)$ is the first row of $\mathbf{X}$, $\mathbb{W}'$ is $\mathbb{W}$ without the first column and first row, and $\mathbf{s}$ is the first row of $\mathbb{W}'$ excluding the element $[\mathbb{W}]_{11}$. Since every diagonal element of $\mathbb{W}' - \frac{1}{(1 - \langle A_0 \rangle^2)} \mathbf{s} \mathbf{r}^T \cdot \mathbf{s}$ is positive by construction, then the matrix in (20) is positive semidefinite if and only if $\mathbb{E} = \mathbb{D} - \frac{1}{(1 - \langle A_0 \rangle^2)} \mathbf{r} \mathbf{r}^T \cdot \mathbf{r} \succeq 0$. Note that every diagonal element of $\mathbb{E}$ is equal to $1 - \frac{1}{(1 - \langle A_0 \rangle^2)} \langle A_1 \rangle^2$. However, the element $[\mathbb{E}]_{12} = 1 + \frac{1}{(1 - \langle A_0 \rangle^2)} \langle A_1 \rangle^2$. For a matrix to be positive semidefinite off-diagonal elements have a magnitude that is bounded by the diagonal terms, therefore for $\mathbb{E}$ to be positive semidefinite we must satisfy $\langle A_1 \rangle = 0$.

We can now repeatedly apply the same analysis to subsequent bottom-left submatrices of (20), where the matrix in (18) is positive semidefinite if and only if $\mathbb{E}' = \mathbb{D}' - \frac{1}{\alpha} \mathbf{r}' \mathbf{r}'^T \cdot \mathbf{r}' \succeq 0$ where $\alpha < 1$ is some positive real number and $\mathbf{r}'$ is any row of $\mathbf{X}$.
For every matrix $E'$ the diagonal elements are $1 - \frac{1}{n} \langle P \rangle^2$ where $P \in \{A_i, B_j, C_k\}$ but there are off-diagonal terms in $E'$ that take the value $1 + \frac{1}{n} \langle P \rangle^2$. Therefore for all $\Gamma_{1 + ABC}$ described by $1$, $\langle P \rangle = 0$ for $P \in \{A_i, B_j, C_k\}$ for all $i, j, k$. This matrix thus corresponds to $\Gamma_M$ and completes our proof. □

Combining the two lemmas above we then obtain our proof of Theorem 1. This concludes our observation that maximally random numbers can be certified within a set of correlations that is not the quantum set. Our proof is analytic and makes concrete the numerical observations in Dhara et al [21]. It would be interesting to extend this proof to other scenarios even though we have used a lot of the structure of the Mermin inequality and the $(3, 2, 2)$ scenario.