Kolmogorov Amplification from Bell Correlation

Ämin Baumeler[*†], Charles Alexandre Bédard[‡], Gilles Brassard[‡§] and Stefan Wolf[*†]
[*] Faculty of Informatics, Università della Svizzera italiana, 6900 Lugano, Switzerland
[†] Facoltà indipendente di Gandria, 6978 Gandria, Switzerland
[‡] Département d'informatique et de recherche opérationnelle,
Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal (QC), H3C 3J7 Canada
[§] Canadian Institute for Advanced Research

***Abstract*—It was first observed by John Bell that quantum theory predicts correlations between measurement outcomes that lie beyond the explanatory power of local hidden variable theories. These correlations have traditionally been studied extensively in the probabilistic framework. A drawback of this perspective is that one is then forced to use in a single argument the outcomes of mutually-exclusive measurements. One of us has initiated an alternative approach, invoking only data at hand, in order to circumvent this issue. In this factual view, which is based on Kolmogorov complexity, we introduce mechanisms such as *complexity amplification*. We establish that this functionality is realizable, just as its probabilistic counterpart, hereby underlining that Bell correlations are a precious information-processing resource.**

I. Motivation and results

Correlations pioneered by Bell [1], so-called non-local, challenge our classical conception of the world. Any attempt at explaining the correlations of entangled states by a local hidden variable theory runs into a dead end. The intrinsic-randomness feature of non-local correlations has become increasingly studied. For instance, violating a Bell inequality was shown to be a way to expand [2] and amplify [3] randomness when using untrusted quantum devices. But what does one mean by *randomness*? With probabilities at the very foundation of "Born-ruled" quantum theory, Shannon's notion of randomness — and of *information* — was a natural measure.

Ironically, the very same who formalized axiomatic probability theory suggested [4] a way to free randomness — and information — from its probabilistic context. *Kolmogorov complexity*, also called *algorithmic complexity*, relates to the data itself. It does not suppose any ensemble context nor probabilistic process for that data to have come about. The Kolmogorov complexity $K(s)$ of a bit string s is the length of the shortest program that outputs s on a fixed universal computing device. In this picture, a patternless bit string is said to be more random than one with structure that enables its compression even if, from a probabilistic perspective, the two strings have equal chances to be generated by the toss of a fair coin if they have the same length.

To boil down non-local correlations to their simplest expression, Popescu and Rohrlich [5] proposed a model of a non-signalling box that violates the CHSH inequality [6] maximally. Two parties, Alice and Bob, feed inputs a and b into the box, which responds with outputs x and y, respectively, where the condition $a \cdot b = x \oplus y$ holds (see Figure 1).

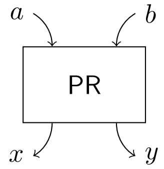

Figure 1. PR Box: Alice inputs some $a \in \{0,1\}$, and Bob inputs some $b \in \{0,1\}$. The output of the box $(x,y) \in \{0,1\}^2$, where Alice receives x and Bob receives y, is related to the inputs by $a \cdot b = x \oplus y$.

In a probabilistic analysis of the PR box, where A, B, X and Y are random variables taking value $(a,b,x,y) \in \{0,1\}^4$, it is well-known, under non-signalling assumptions, that if all input pairs have non-zero probability of occurring, then

$$P_{X|A=a,B=b}(x) = {}^1\!/_2 \quad \text{and} \quad P_{Y|A=a,B=b}(y) = {}^1\!/_2 \,.$$

So, the outputs must be completely random and independent from the inputs and hence fresh probabilistic randomness is created. Taking a factual and context-free position was a motivation for the first algorithmic analysis of the PR box [7], [8], where bit strings play the role of random variables.

This paper first improves on the previous results with better lower bounds on the complexity[1] of the outputs when the box is given incompressible inputs. Second, we show that each output of the PR box is algorithmically independent from the inputs, even under less restrictive conditions on the inputs. As mentioned before, this independence relation is also exhibited in the usual probabilistic analysis of the box. Third, our most fundamental result is the complexity amplification property, which informally states that the box must generate some *fresh* complex strings, even if given infinite ressources. Finally, we exhibit a case in which a compressible input pair is fed to the algorithmic PR box, and an incompressible output pair is returned.

II. Complexity amplification, informally

Before going into the details of the model and the results, let us discuss how non-locality, studied with the PR-box model, once more challenges our classical conception of the world. This time, it strikes at the *classical computational models*.

If one gives a string s to a box, which outputs a string t such that $K(t \mid s)$ is very large, say N bits, then one concludes

[1] Throughout this paper, "complexity" stands for "Kolmogorov complexity".

that if the box is no more than a deterministic computational model, its program must be of at least N bits.

However, one could think of the box as a probabilistic computational model that has access to random bits. *Computation is physical*, and with a deterministic classical theory, the standard way to model probabilistic computation is via a string λ that is considered random. Let us suppose that λ is an *infinite* string that we supply to a deterministic computational device of program size m. We could think of λ as an encoding of an infinite random string, as well as arbitrarily long or even infinitely long programs. If $m \ll N$, this is no surprise, because to produce t from s, our deterministic machine can use arbitrarily many bits of λ. Now, call $K^{\lambda}(t \,|\, s)$ the length of the shortest program for a fixed universal machine to produce t from s, when it is given λ as oracle. In the previous understanding of a probabilistic computing device, we would expect $K^{\lambda}(t \,|\, s)$ to be no more than m, because all the complex procedure that permits to produce t from s is encoded in λ. However, our results turn out to show that, under certain conditions, the algorithmic PR box satisfies $K^{\lambda}(t \,|\, s) = K(t \,|\, s) = N \gg m$. How can deterministic computational models, even if provided with infinite ressources to make them probabilistic, *explain* this? What comes out of the box is fresh algorithmic randomness since it cannot be deterministically computed from any ressources that the box has at hand.

III. Model

A. Kolmogorov complexity

The Kolmogorov complexity $K(s)$ of a finite string s is defined as the length of the shortest program that outputs s [4], [9]. For a meaningful definition, we have to select an *additively optimal* universal Turing machine $\mathcal{U}$ to run all programs. Such a machine can simulate any other with a constant overhead. Thus, the complexity $K(s)$ is defined as the length of the shortest program for $\mathcal{U}$ to output s.

All strings discussed will be bit strings, *i.e.*, the alphabet is $\{0,1\}$. For an infinite string $a = a_1 a_2 \ldots$, we use $a_{[n]} = a_1 a_2 \ldots a_n$ to denote the first n bits of a. The Kolmogorov complexity $K(a)$ of an infinite string a is defined as a function

$$\begin{aligned} K(a)\colon \mathbb{N} &\to \mathbb{N} \\ n &\mapsto K(a_{[n]})\,. \end{aligned}$$

To each n, this function returns the length of the shortest program for $\mathcal{U}$ that outputs $a_{[n]}$. In order to describe the function $K(a)$ without caring about small oscillating patterns in the complexity of a, or in that of the number n itself, we equate $K(a)$ with any function f that differs from $K(a)$ by at most a logarithmic additive term. That is, for a $f\colon \mathbb{N} \to \mathbb{N}$ we say that $K(a)$ *is asymptotic to* f, and write $K(a) \approx f$, if

$$K(a_{[n]}) = f(n) \pm O(\log n)\,.$$

One motivation for using the asymptotic behaviour of infinite strings is the simplicity of the model; smaller order terms such as constants cropping up from the specification of $\mathcal{U}$ and even logarithmic terms can simply be ignored. This is useful to even out different definitions one could have chosen. For instance, $K(a_{[n]}) \approx K(a_{[n]} \,|\, n)$, because one could encode n in a $\log n$ long program. Also the difference between *plain* complexity and *prefix* complexity of $a_{[n]}$, being no more than $\log n$, makes this choice irrelevant. Note however that our "$\approx$" relation is meaningless when we consider strings with complexity smaller than logarithmic, but this paper studies strings in the linear-complexity regime.

An infinite string a is called *incompressible* if it satisfies $K(a) \approx n$. In that case, the length of the shortest program that outputs a differs from that of a program that simply recites a only by at most logarithmic terms. In the same spirit, a string a is called *computable* if $K(a_{[n]} \,|\, n)$ is $O(1)$, *i.e.*, there exists a constant-length program that outputs n bits of a if n is provided.

The Kolmogorov complexity $K(a, b, \ldots, z)$ for multiple *finite* bit strings $a, b, \ldots, z$ is the length of the shortest program for $\mathcal{U}$ that outputs an encoding[2] $\langle a, b, \ldots, z\rangle$ of *all* strings. For infinite strings, $K(a, b, \ldots, z)$ is again a function defined by

$$K(a, b, \ldots, z)(n) \overset{\text{def}}{=} K(a_{[n]}, b_{[n]}, \ldots, z_{[n]})\,.$$

The *conditional* Kolmogorov complexity $K(a \,|\, b)$ is, for all n, defined as the length of the shortest program for $\mathcal{U}$ that outputs $a_{[n]}$ when given $b_{[n]}$:

$$K(a \,|\, b)(n) \overset{\text{def}}{=} K\left(a_{[n]} \,|\, b_{[n]}\right)\,.$$

A bit string a is called *pseudo-probabilistic* if the only way to compress it is by some coding of the alphabet, *i.e.*, there is no more structure to exploit for compression besides the frequency bias of the bits appearing in the string. This only makes sense if that frequency exists. Therefore, a pseudo-probabilistic bit string a is such that

$$\lim_{n\to\infty} \frac{\#1(a_{[n]})}{n} = \alpha \quad \text{and} \quad K(a) \approx h(\alpha) n\,, \tag{PP}$$

provided $0 < \alpha < 1$, where $\#1(a_{[n]})$ is the number of 1s in the bit string $a_{[n]}$ and $h(\alpha) = -\alpha \log(\alpha) - (1-\alpha)\log(1-\alpha)$ is the binary entropy of α. Pseudo-probabilistic strings can be seen as a *relaxation of incompressible strings* corresponding to a typical sequence of a memoryless (possibly unfair) coin flip of fixed bias. An incompressible bit string can thus be seen as a special pseudo-probabilistic string with $\alpha = {}^1\!/_2$. Pseudo-probabilistic strings always have complexity of linear order because we impose $0 < \alpha < 1$.

It is well known [9], [10] that the Kolmogorov complexity satisfies a *chain rule*. For n-bit strings,

$$K(a_{[n]}, b_{[n]}) = K(b_{[n]}) + K(a_{[n]} \mid b_{[n]}) \pm O(\log n)\,,$$

which means that the trivial recipe to compute the pair $(a_{[n]}, b_{[n]})$ by first computing one string (here, $b_{[n]}$) and then

[2] One can first define $\langle\cdot,\cdot\rangle : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}^*$ as an arbitrary bijective computable mapping. Then, for an encoding of multiple strings, one simply iterates the encoding. Since the encoding is computable, its specification adds only a constant term to Kolmogorov complexity.

the other from the first is optimal up to logarithmic terms. For infinite stings, the chain rule simplifies to

$$K(a,b) \approx K(b) + K(a \mid b)\,.$$

Two bit strings a and b are called *independent* when

$$K(a\,|\,b) \approx K(a)\,, \tag{IND}$$

which informally states that knowing b does not help to compress a. Thanks to the chain rule, the independence can easily be seen to be symmetric.

Finally, we define the *algorithmic mutual information*

$$I_K(a:b) \overset{\text{def}}{=} K(b) - K(b\,|\,a)\,,$$

which is positive since knowing a can only help in compressing b, and symmetric up to logarithmic terms, *i.e.*,

$$I_K(a:b) \approx I_K(b:a)\,.$$

B. Assumptions

Our model is about an algorithmic flavour of the PR box. We consider quintuples (a, b, x, y, λ) of infinite bit strings, where the strings within the quintuples satisfy some conditions. The goal is then to make statements about such quintuples. We call this the *facts-only* view, since no choices and no probabilities are involved; only the data. We refer to a, b as the *inputs*, and to x, y as the *outputs* of Alice and Bob respectively. First, we assume all (a, b, x, y) to satisfy the *bit-by-bit PR condition*:

$$a \cdot b = x \oplus y\,, \tag{PR}$$

which means that for all i: $a_i \cdot b_i = x_i \oplus y_i$.

Alice's and Bob's boxes should be seen as computing devices, which share an *a priori* resource $\lambda \in \{0,1\}^\omega$ (see Figure 2). This λ could be an infinite random string, it could be an infinitely long program for each of them, possibly correlated to coordinate their strategies. It could also be an oracle that allows their devices to go higher on the arithmetic hierarchy. Thanks to Hilbert's hotel, it could even be an encoding of all of these.

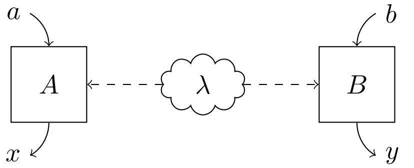

Figure 2. Model of an algorithmic PR box represented by Alice's and Bob's computing devices A and B, to which is supplied the oracle λ as a shared ressource.

For a finite string s, we define $K^\lambda(s)$ to be the length of the shortest program for $\mathcal{U}$ that produces s when it has access to λ as an oracle. For an infinite string a, $K^\lambda(a)$ becomes a function $\mathbb{N} \to \mathbb{N}$ defined by $K^\lambda(a)(n) = K^\lambda(a_{[n]})$. Notice that we write $K^\lambda(\cdot)$ rather than $K(\cdot\,|\,\lambda)$ because for each n, the *whole* λ is given to the computing device as opposed to only $\lambda_{[n]}$.

We impose that no matter how powerful the shared ressource λ is, it does not permit a shorter description of the input pair a, b. Therefore, the inputs are *oracle-independent* of λ in the sense that

$$K^\lambda(a,b) \approx K(a,b)\,. \tag{λ-IND}$$

Another condition the strings (a, b, x, y, λ) have to satisfy is *no-signalling*:

$$\begin{aligned} K^\lambda(x\,|\,a,b) &\approx K^\lambda(x\,|\,a)\,,\\ K^\lambda(y\,|\,a,b) &\approx K^\lambda(y\,|\,b)\,. \end{aligned} \tag{NS}$$

Inasmuch as no-signalling in a probabilistic theory means that knowing Bob's input cannot help Alice to predict her output better; in the factual case, one simply changes "predict" by "compute", and understands that "better" is measured in program length. Furthermore, we consider only those quintuples for which the strings a and b are *independent* (IND), and a, b are *pseudo-probabilistic* (PP) with an asymptotic frequency of 1s of α and β, respectively. We define S as the set of objects (a, b, x, y, λ) such that (PR), (λ-IND), (NS), (IND) and (PP) hold. In the following, we shall make statements about quintuples from the set S.

IV. Complexity Amplification, formally

In this section, we state and discuss two lemmas, whose proofs are relegated to Section V. Then, we state and prove our main theorem on complexity amplification. Finally, we mention a consequence of the theorem.

Lemma 1. *For all* $(a, b, x, y, \lambda) \in S$,

$$K^\lambda(y) \approx I_{K^\lambda}(x:y \mid a) + K^\lambda(a \cdot b \mid a)\,.$$

This lemma improves on the result of ref. [8], in which it was shown that if the box was provided completely random and independent strings a and b, *i.e.*, with $K(a,b) \approx 2n$, then $K(y) \gtrsim n/4$. Indeed, in this case Lemma 1 implies that

$$K(y) \gtrsim K(a \cdot b \mid a) \approx \frac{n}{2}\,.$$

To see this, observe that $(a \cdot b)_i = 0$ whenever $a_i = 0$ and $(a \cdot b)_i = b_i$ whenever $a_i = 1$. This last case leads to an incompressible substring of b, even given a. Since a is incompressible, each case happens on an asymptotic proportion of $^1/_2$. The former case leaves us no complexity, but the latter case leaves us with maximal complexity. Note that (PP) is not needed to prove this lemma.

Lemma 2. *For all* $(a, b, x, y, \lambda) \in S$,

$$K^\lambda(y) \approx K^\lambda(y \mid b)\,.$$

Lemma 2 links with the probabilistic study of the PR box, in which the distribution of the output does not depend on the input. Here, we have an algorithmic analogue: The length of a program for the output string y does not change if it is given its input b.

We say that a process *amplifies complexity by an asymptotic term* $t(n)$ if the joint input-output complexity is larger than the complexity of the inputs by a $t(n)$ term. By the chain rule, this is equivalent to $K(\mathrm{Out}|\mathrm{In}) \approx t(n)$. We now show that the non-local box amplifies complexity by an asymptotic linear term, when it is given pseudo-probabilistic and independent inputs. It does so even if supplied with an arbitrary λ that does not provide signalling capabilities. What comes out of the box cannot be produced even by an infinitely long shared program (encoded in λ) run on the inputs.

Theorem 1 (Complexity amplification)**.**
For all $(a, b, x, y, \lambda) \in S$,

$$K^\lambda(x, y \mid a, b) \gtrsim h(\alpha\beta)n\,.$$

Proof. Thanks to the chain rule,

$$K^\lambda(x, y \mid a, b) \approx K^\lambda(y \mid a, b) + K^\lambda(x \mid a, b, y)\,.$$

Since x can be computed by a, b and y, the last term vanishes, and by (NS) the first term can be reduced to $K^\lambda(y \mid b)$. The amplification of the box is therefore as much as $K^\lambda(y \mid b)$ and the statement of the theorem boils down to bounding this term by $h(\alpha\beta)n$. From Lemma 2, $K^\lambda(y\,|\,b) \approx K^\lambda(y)$. We shall now show that $K^\lambda(y\,|\,x) \approx h(\alpha\beta)n$ and the conclusion will follow from $K^\lambda(y) \gtrsim K^\lambda(y \mid x)$.
Observe that

$$\begin{aligned} K^\lambda(y \mid x) &\approx K^\lambda(x \oplus y \mid x) \\ &\approx K^\lambda(x \mid x \oplus y) + K^\lambda(x \oplus y) - K^\lambda(x) \\ &= K^\lambda(x \mid a \cdot b) + K^\lambda(a \cdot b) - K^\lambda(x)\,. \end{aligned} \qquad (1)$$

But, by acquiring information (twice), no-signalling and the symmetric statement of Lemma 2,

$$K^\lambda(x) \gtrsim K^\lambda(x \mid a \cdot b) \gtrsim K^\lambda(x \mid a, b) \approx K^\lambda(x \mid a) \approx K^\lambda(x)\,,$$

so the first and third terms of Eq. (1) cancel out. Since a and b are independent and pseudo-probabilistic, the $a \cdot b$ string is also pseudo-probabilistic with an asymptotic proportion of 1s of $\alpha\beta$. Therefore,

$$K^\lambda(y \mid x) \approx K^\lambda(a \cdot b) \approx h(\alpha\beta)n\,. \qquad \square$$

We conclude this section with a special case in which a non-maximally complex input pair gets amplified into a maximally complex output pair. Take α and β such that $\alpha\beta = {}^1\!/_2$ and $h(\alpha) + h(\beta) = 1 + \varepsilon$ for arbitrarily small ε by choosing α or β sufficiently close to 1. In this case, the complexity of the pair of inputs is $K(a, b) \approx h(\alpha)n + h(\beta)n \approx (1 + \varepsilon)n$. But then, the complexity of the outputs is

$$\begin{aligned} K(x, y) &\approx K(x) + K(y \mid x) \\ &\gtrsim K(x \mid y) + K(y \mid x) \\ &\approx h(\alpha\beta)n + h(\alpha\beta)n \\ &\approx 2n\,. \end{aligned}$$

So we find explicit cases in which the PR box is given just a little more than n complexity, and spits out $2n$ complexity. In sharp contrast, a maximally complex output pair can*not* be obtained from a maximally complex input pair (*i.e.*, $K(a, b) \approx 2n$ and hence $\alpha = \beta = {}^1\!/_2$). Indeed,

$$\begin{aligned} K(x, y) &\approx K(x) + K(y \,|\, x) \\ &\approx K(x) + h(\alpha\beta)n \\ &\lesssim (1 + h({}^1\!/_4))n\,, \end{aligned}$$

is strictly smaller than $2n$.

V. Proofs

A. Proof of Lemma 1

Observe first that (NS) and (PR) imply

$$K^\lambda(x \mid a) \approx K^\lambda(x \mid a, b) \approx K^\lambda(y \mid a, b) \approx K^\lambda(y \mid b)\,.$$

On the one hand, acquiring information leads to

$$\begin{aligned} K^\lambda(y \mid b) &\approx K^\lambda(x \mid a) \\ &= K^\lambda(a \cdot b \oplus y \mid a) \\ &\gtrsim K^\lambda(a \cdot b \oplus y \mid a \cdot b, a) \\ &\approx K^\lambda(y \mid a \cdot b, a)\,; \end{aligned}$$

but on the other hand, no-signalling and forgetting information imply

$$\begin{aligned} K^\lambda(y \mid b) &\approx K^\lambda(y \mid a, b) \\ &\lesssim K^\lambda(y \mid a \cdot b, a)\,. \end{aligned}$$

Therefore, $K^\lambda(y \mid b) \approx K^\lambda(y \mid a \cdot b, a)$. This is useful in

$$\begin{aligned} K^\lambda(x, y \mid a) &\approx K^\lambda(x \oplus y, y \mid a) \\ &\approx K^\lambda(x \oplus y \mid a) + K^\lambda(y \mid a, x \oplus y) \\ &= K^\lambda(a \cdot b \mid a) + K^\lambda(y \mid a, a \cdot b) \\ &\approx K^\lambda(a \cdot b \mid a) + K^\lambda(y \mid b) \\ &\approx K^\lambda(a \cdot b \mid a) + K^\lambda(x \mid a)\,. \end{aligned}$$

Hence,

$$K^\lambda(y \mid a, x) \approx K^\lambda(x, y \mid a) - K^\lambda(x \mid a) \approx K^\lambda(a \cdot b \mid a)\,.$$

Also, expressing $K^\lambda(a, b, y)$ in two different ways, namely,

$$\begin{aligned} K^\lambda(a, b, y) &\approx K^\lambda(a, b) + K^\lambda(y \mid a, b) \\ &\approx K^\lambda(a) + K^\lambda(b) + K^\lambda(y \mid b) \end{aligned}$$

and

$$K^\lambda(a, b, y) \approx K^\lambda(b) + K^\lambda(y \mid b) + K^\lambda(a \mid y, b)\,,$$

we conclude that $K^\lambda(a \mid y, b) \approx K^\lambda(a)$. This implies in particular that $K^\lambda(a \mid y) \approx K^\lambda(a)$ and by symmetry of the independence relation, that $K^\lambda(y \mid a) \approx K^\lambda(y)$.

We complete the proof of the lemma by observing that

$$\begin{aligned} K^\lambda(y) &\approx K^\lambda(y \mid a) \\ &\approx K^\lambda(y \mid a) - K^\lambda(y \mid a, x) + K^\lambda(a \cdot b \mid a) \\ &\approx I_{K^\lambda}(x : y \mid a) + K^\lambda(a \cdot b \mid a)\,. \end{aligned} \qquad \square$$

Notice that until now, the pseudo-probabilistic property of the inputs has not been used.

B. *Proof of Lemma 2*

We develop $I_{K^\lambda}(x : y \mid a)$ in a different expression and start using the (PP) hypothesis:

$$\begin{aligned} I_{K^\lambda}(x : y \mid a) &\approx K^\lambda(x \mid a) - K^\lambda(x \mid a, y) \\ &\approx K^\lambda(y \mid b) - K^\lambda(x \oplus y \mid a, y) \\ &= K^\lambda(y \mid b) - K^\lambda(a \cdot b \mid a, y)\,. \end{aligned}$$

Now, we know $(a \cdot b)_i = 0$ whenever $a_i = 0$, which happens on an asymptotic proportion of $1-\alpha$ of the bits, and we know on which because we are given a. But whenever $a_i = 1$, which happens on an asymptotic proportion of α of the bits, we know $(a \cdot b)_i = b_i$. Therefore, the second term is asymptotic to $\alpha K^\lambda(b \mid a, y)$.

$$\begin{aligned} K^\lambda(b \mid a, y) &\approx K^\lambda(a, b, y) - K^\lambda(a, y) \\ &\approx K^\lambda(a) + K^\lambda(b) + K^\lambda(y \mid a, b) \\ &\qquad -K^\lambda(a) - K^\lambda(y) \\ &\approx K^\lambda(b) + K^\lambda(y \mid b) - K^\lambda(y)\,. \end{aligned}$$

We shall now put this together with Lemma 1. For the same reason as before, $K^\lambda(a \cdot b \mid a) \approx \alpha K^\lambda(b \mid a)$ and by independence and the pseudo-probabilistic property of b, this is asymptotic to $\alpha h(\beta) n$. Therefore,

$$\begin{aligned} K^\lambda(y) &\approx I_{K^\lambda}(x : y \mid a) + \alpha h(\beta) n \\ &\approx K^\lambda(y \mid b) - \alpha(K^\lambda(b) + K^\lambda(y \mid b) - K^\lambda(y)) \\ &\qquad +\alpha h(\beta) n\,. \end{aligned}$$

Simplifying and rewriting, we get

$$(1-\alpha)K^\lambda(y) \quad \approx \quad (1-\alpha)K^\lambda(y \mid b)\,.$$

The conclusion follows from $\alpha \neq 1$. □

VI. Conclusions

Non-local correlations, such as those arising from the PR box, have traditionally been studied in the *probabilistic* realm. There, probability distributions are assigned to inputs, processes and outputs. This has lead to a series of interesting statements on non-local correlations from a cryptographic as well as foundational point of view, *e.g.*, randomness amplification and expansion were shown to be possible. This probabilistic view has also been employed to study *quantum* correlations obtained by two or more parties measuring an entangled quantum state. However, the necessary settings of the measurement apparatuses to exhibit these correlations from such states represent non-commuting observables. Now, if one describes the settings required to arrive at such correlations by a probability distribution, then one ends up with statements where non-commuting observables appear *simultaneously* — yet, this is prohibited by quantum theory itself. This means that the results drawn from the probabilistic view are based on "infuturabili" or counterfactuals: The statements talk about "what would have happened if something had happened that did not happen" [11].

Contrary to this counterfactual view, we advocate the *facts-only perspective*: Data from a potential experiment are considered without using any probability distributions — and we aim at talking about *what actually happened*, without referring to any alternatives. In this view, we model series of measurement settings and results by bit strings. Hence, we can make statements about the relations among such bit strings. In the study performed here, we look at tuples of bit-strings that are possible from a PR-box setup and conclude (under appropriate conditions) that the PR box *must* amplify Kolmogorov complexity.

An open question is how this result can be extended beyond the PR box to settings that arise in nature: quantum correlations cropping up from measurements of entangled states. One approach to adapt this work to the quantum world would be by considering chained Bell-inequalities or a pseudo-telepathy game. Another direction that should be investigated is to tighten the analysis with a relation "$\approx$" defined as differing by an $O(1)$ term. For this, prefix Kolmogorov complexity will be necessary.

Acknowledgements

The work of ÄB and SW is supported by the Swiss National Science Foundation (SNF), the National Centre of Competence in Research "Quantum Science and Technology" (QSIT), and the COST action on Fundamental Problems in Quantum Physics. The work of CAB is supported by the public society Hydro-Québec, as well as the Fonds de recherche du Québec — Nature et technologies (FRQNT). The work of GB is supported in part by the Canadian Institute for Advanced Research (CIFAR), the Canada Research Chair program, Canada's Natural Sciences and Engineering Research Council (NSERC) and Québec's Institut transdisciplinaire d'information quantique (INTRIQ).

References

[1] J. S. Bell, "On the Einstein Podolsky Rosen paradox," *Physics*, vol. 1, no. 3, pp. 195–200, 1964.

[2] R. Colbeck and A. Kent, "Private randomness expansion with untrusted devices," *Journal of Physics A: Mathematical and Theoretical*, vol. 44, no. 9, 095305, 2011.

[3] R. Colbeck and R. Renner, "Free randomness can be amplified," *Nature Physics*, vol. 8, no. 6, pp. 450–454, 2012.

[4] A. N. Kolmogorov, "Three approaches to the quantitative definition of information," *Problemy Peredachi Informatsii*, vol. 1, no. 1, pp. 3–11, 1965.

[5] S. Popescu and D. Rohrlich, "Quantum nonlocality as an axiom," *Foundations of Physics*, vol. 24, no. 3, pp. 379–385, 1994.

[6] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Physical Review Letters*, vol. 23, no. 15, pp. 880–884, 1969.

[7] S. Wolf, "Nonlocality without counterfactual reasoning," *Physical Review A*, vol. 92, no. 5, 052102, 2015.

[8] Ä. Baumeler and S. Wolf, "Causality – Complexity – Consistency: Can space-time be based on logic and computation?" *preprint arXiv:1602.06987 [quant-ph]*, 2016.

[9] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and its Applications*. Springer, New York, 2008.

[10] G. J. Chaitin, "A theory of program size formally identical to information theory," *Journal of the ACM*, vol. 22, no. 3, pp. 329–340, 1975.

[11] E. Specker, "Die Logik nicht gleichzeitig entscheidbarer Aussagen," *Dialectica*, vol. 14, no. 2-3, pp. 239–246, 1960.