Causality – Complexity – Consistency: Can Space-Time Be Based on Logic and Computation?

Åmin Baumeler\textsuperscript{1,2} and Stefan Wolf\textsuperscript{1,2}

\textsuperscript{1}Faculty of Informatics, Università della Svizzera italiana, Via G. Buffi 13, 6900 Lugano, Switzerland
\textsuperscript{2}Facoltà indipendente di Gandria, Lunga scala, 6978 Gandria, Switzerland

The difficulty of explaining non-local correlations in a fixed causal structure sheds new light on the old debate on whether space and time are to be seen as fundamental. Refraining from assuming space-time as given a priori has a number of consequences. First, the usual definitions of randomness depend on a causal structure and turn meaningless. So motivated, we propose an intrinsic, physically motivated measure for the randomness of a string of bits: its length minus its normalized work value, a quantity we closely relate to its Kolmogorov complexity (the length of the shortest program making a universal Turing machine output this string). We test this alternative concept of randomness for the example of non-local correlations, and we end up with a reasoning that leads to similar conclusions as, but is more direct than, in the probabilistic view since only the outcomes of measurements that can actually all be carried out together are put into relation to each other. In the same context-free spirit, we connect the logical reversibility of an evolution to the second law of thermodynamics and the arrow of time. Refining this, we end up with a speculation on the emergence of a space-time structure on bit strings in terms of data-compressibility relations. Finally, we show that logical consistency, by which we replace the abandoned causality, it strictly weaker a constraint than the latter in the multi-party case.

I. RANDOMNESS WITHOUT CAUSALITY

What is causality? — The notion has been defined in different ways and turned out to be highly problematic, both in Physics and Philosophy. This observation is not new, as is nicely shown by Bertrand Russell’s quote \cite{11} from more than a century ago:

“The law of causality [...] is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.”

Indeed, a number of attempts have been made to abandon causality and replace global by only local assumptions (see, e.g., \cite{35}). A particular motivation is given by the difficulty of explaining quantum non-local correlations according to Reichenbach’s principle \cite{39}. The latter states that in a given (space-time) causal structure, correlations stem from a common cause (in the common past) or a direct influence from one of the events to the other. In the case of violations of Bell’s inequalities, a number of results indicate that explanations through some mechanism as suggested by Reichenbach’s principle either fail to explain the correlations \cite{12} or are unsatisfactory since they require infinite speed \cite{33,20,5,4} or precision \cite{50}. All of this may serve as a motivation for dropping the assumption of a global causal structure in the first place.

Closely related to causality is the notion of randomness: In \cite{15}, a piece of information is called freely random if it is statistically independent from all other pieces of information except the ones in its future light cone. Clearly, when the assumption of an initially given causal structure is dropped, such a definition is not possible any longer. One may choose to consider freely random pieces of information as being more fundamental than a space-time structure — in fact, the latter can then be seen as emerging from the former: If a piece of information is free, then any piece correlated to it is in its causal future. But how can we define the randomness of an object purely intrinsically and independently of any context?

For further motivation, note that Colbeck and Renner’s definition of randomness \cite{18} is consistent with full determinism: A random variable with trivial distribution is independent of every other (even itself). How can we exclude this and additionally ask for the possibility in principle of a counterfactual outcome, i.e., that the random variable $X$ could have taken a value different from the one it actually took? Intuitively, this is a necessary condition for freeness. The question whether the universe (or a closed system) starting from a given state $A$ always ends up in the same state $B$ seems to be meaningless: Even if rewinding were possible, and two runs could be performed, the outcomes $B_1$ and $B_2$ that must be compared never exist in the same reality since rewinding erases the result of the rewound run \cite{40}: “$B_1 = B_2$?” is not a question which cannot be answered in principle, but that cannot even be formulated precisely. In summary, defining freeness of a choice or a random event, understood as the actual possibility of two (or more) well-distinguishable options, seems hard even when a causal structure is in place.$^2$

---

1. This change of perspective reflects the debate, three centuries ago, between Newton and Leibniz on the nature of space and time, in particular on as how fundamental this causal structure is to be considered.
2. In this context and as a reply to \cite{26}, we feel that the notion of a choice between different possible futures by an act of free will put forward there is not only hard to formalize but also
We look for an intrinsic definition of randomness that takes into account only the “factuality,” i.e., the state of the closed system in question. Clearly, such a definition is hard to imagine for a single bit, but it can be defined in a natural way for (long) strings of bits, namely its length minus the work value (normalized through dividing by $kT$) of a physical representation of the string with respect to some extraction device; we relate this quantity to the string’s best compression.

We test the alternative view of randomness for physical meaning. More specifically, we find it to be functional in the context of non-local correlations: A reasoning yielding a similar mechanism as in the probabilistic regime is realized which has the conceptual advantage not to require relating the outcomes of measurements that cannot all actually be carried out. That mechanism is: Random inputs to a non-local system plus no-signaling guarantee random outputs.

In the second half of this text, we consider consequences of abandoning (space-time) causality as being fundamental. In a nutshell, we put logical reversibility to the center of our attention here. We argue that if a computation on a Turing machine is logically reversible, then a “second law” emerges: The complexity of the tape’s content cannot decrease in time. This law holds without failure probability, in contrast to the “usual” second law, and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter. In the same spirit, we propose to define causal relations between physical points, modeled and implies the latter.

II. PRELIMINARIES

Let $U$ be a fixed universal Turing machine (TM). For a finite or infinite string $s$, the Kolmogorov complexity $K(s) = K_U(s)$ is the length of the shortest program for $U$ such that the machine outputs $s$. Note that $K(s)$ can be infinite if $s$ is.

Let $a = (a_1, a_2, \ldots)$ be an infinite string. Then

$$a_n := (a_1, \ldots, a_n, 0, \ldots).$$

We study the asymptotic behavior of $K(a_n) : \mathbb{N} \to \mathbb{N}$. For this function, we simply write $K(a)$, similarly $K(a | b)$ for $K(a_n | b_{n0})$, the latter being the length of the shortest program outputting $a_n$ upon input $b_{n0}$. We write

$$K(a) \approx n \iff \lim_{n \to \infty} \left( \frac{K(a_n)}{n} \right) = 1.$$

We call a string $a$ with this property incompressible. We also use $K(a_n) = \Theta(n)$, as well as

$$K(a) \approx 0 \iff \lim_{n \to \infty} \left( \frac{K(a_n)}{n} \right) = 0 \iff K(a_n) = o(n).$$

Note that computable strings $a$ satisfy $K(a) \approx 0$, and that incompressibility is, in this sense, the extreme case of uncomputability.

Generally, for functions $f(n)$ and $g(n) \neq 0$, we write $f \approx g$ if $f/g \to 1$. Independence of $a$ and $b$ is then

$$K(a | b) \approx K(a)$$

or, equivalently,

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

If we introduce

$$I_K(x; y) := K(x) - K(x \mid y) \approx K(y) - K(y \mid x),$$

independence of $a$ and $b$ is $I_K(a, b) \approx 0$.

In the same spirit, we can define conditional independence: We say that $a$ and $b$ are independent given $c$ if

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

or, equivalently,

$$K(a \mid b, c) \approx K(a \mid c),$$

or

$$I_K(a; b \mid c) := K(a \mid c) - K(a \mid b, c) \approx 0.$$

5 This is inspired by [17], where (joint) Kolmogorov complexity — or, in practice, any efficient compression method — is used to define a distance measure on sets of bit strings (such as literary texts of genetic information of living beings). The resulting structure in that case is a distance measure, and ultimately a clustering as a binary tree.
III. COMPLEXITY AS RANDOMNESS 1: WORK EXTRACTION

A. The Converse of Landauer’s Principle

In our search for an intrinsic notion of randomness — independent of probabilities or the existence of alternatives — expressed through the properties of the object in question, we must realize, first of all, that such a notion is impossible for single bits, since neither of the two possible values, 0 nor 1, is in any way more an argument for the “randomness” of that bit than not. The situation, however, changes for long strings of bits: No one would call the one-million-bit string 000\ldots0 random (even though, of course, it is not impossible that this string originates from a random process such as a million consecutive tosses of a fair coin). In the spirit of Rolf Landauer’s famous slogan “information is physical,” we may want to test our intuition physically: If the N bits in a string encode the position, being in the left (0) as opposed the right (1) half of some container, of the molecules of a gas, then the 0-string means that the gas is all concentrated in one half and, hence, allows for extracting work from the environmental heat; the amount is \(NkT\ln 2\) if \(k\) is Boltzmann’s constant and \(T\) is the temperature of the environment. This fact has also been called the converse of Landauer’s principle. Note that any other system which can be transformed by a reversible process into that maximally asymmetric gas has the same work value; an example is a physical representation of the first \(N\) bits of the binary expansion of \(\pi\) of the same length — although this string may look much more “random” at first sight. This reversible process is, according to the Church-Turing thesis\(^\text{6}\) imagined to be carried out by a Turing machine in such a way that every step is logically reversible (such as, e.g., a Toffoli gate) and can be uncomputed by the same device; the process is then also possible in principle in a thermodynamically reversible way: No heat is dissipated \(^\text{24}\). It is clear that most \(N\)-bit strings cannot have any work value provided there is no perpetual mobile of the second kind.

For a given string \(S\), its length minus the work value of a physical representation (divided by \(kT\)) may be regarded as an intrinsic measure for the randomness of \(S\). We address the question what in general the fuel value is of (a physical representation of) \(S\). Since (the reversible extraction of) the string \(0^N\) from \(S\) is equivalent to (the gain of) free energy of \(NkT\ln 2\), we have a first answer: Work extraction is data compression.

B. Free Energy and Data Compression

State of the art. Bennett\(^\text{13}\) claimed the fuel value of a string \(S\) to be its length minus \(K(S)\):

\[
W(S) = (\text{len}(S) - K(S))kT \ln 2.
\]

Bennett’s argument is that (the physical representation of) \(S\) can be — logically, hence, thermodynamically \(^\text{24}\) — reversibly transformed into the string \(P||000\ldots0\), where \(P\) is the shortest program for \(U\) generating \(S\) and the length of the generated 0-string is \(\text{len}(S) - K(S)\) (see Figure 1).

![Figure 1. Bennett’s argument.](image)

It was already pointed out by Zurek\(^\text{53}\) that whereas it is true that the reverse direction exists and is computable by a universal Turing machine, its forward direction, i.e., \(P\) from \(S\), is not. This means that the demon that can carry out the work-extraction computation on \(S\) from scratch does not physically exist if the Church-Turing hypothesis is true. We will see, however, that Bennett’s value is an upper bound on the fuel value of \(S\).

Dahlsten et al.\(^\text{21}\) follow Szilárd\(^\text{45}\) in putting the knowledge of the demon extracting the work to the center of their attention. More precisely, they claim

\[
W(S) = (\text{len}(S) - D(S))kT \ln 2,
\]

where the “defect” \(D(S)\) is bounded from above and below by a smooth Rényi entropy of the distribution of \(S\) from the demon’s viewpoint, modeling her ignorance. They do not consider the algorithmic aspects of the demon’s actions extracting the free energy, but the effect of the demon’s a priori knowledge on \(S\). If we model the demon as an algorithmic apparatus, then we should specify the form of that knowledge explicitly: Vanishing conditional entropy means that \(S\) is uniquely determined from the demon’s viewpoint. Does this mean that the demon possesses a copy of \(S\), or the ability to produce such a copy, or pieces of information that uniquely determine \(S\)? This question sits at the origin of the gap between the two described groups of results; it is maximal when the demon fully “knows” \(S\) which, however, still has maximal complexity even given her internal state (an example see below). In this case, the first result claims \(W(S)\) to be 0, whereas \(W(S) \approx \text{len}(S)\) according to the second. The
gap vanishes if “knowing $S$” is understood in a constructive — as opposed to entropic — sense, meaning that “the demon possesses or can produce a copy of $S$ represented in her internal state.” If that copy is included in Bennett’s reasoning, then his result reads

$$\frac{W(S,S)}{kT} \approx \text{len}(S,S) - K(S,S) \approx 2\text{len}(S) - K(S) \approx \text{len}(S).$$

In this case, knowledge has immediate work value.

**The model.** We assume the demon to be a universal Turing machine $\mathcal{U}$ the memory tape of which is sufficiently long for the tasks and inputs in question, but finite. The tape initially contains $S$, the string the fuel value of which is to be determined, $X$, a finite string modeling the demon’s knowledge about $S$, and 0’s for the rest of the tape. After the extraction computation, the tape contains, at the bit positions initially holding $S$, a (shorter) string $P$ plus 0\text{len}(S) - \text{len}(P)$, whereas the rest of the tape is (again) the same as before work extraction. The demon’s operations are logically reversible and can, hence, be carried out thermodynamically reversibly [24]. Logical reversibility in our model is the ability of the same demon to carry out the backward computation step by step, i.e., from $P\|X$ to $S\|X$. We denote by $E(S|X)$ the maximal amount of 0-bits extractable logically reversibly from $S$ given the knowledge $X$, i.e.,

$$E(S|X) := \text{len}(S) - \text{len}(P)$$

if $P$’s length is minimal (see Figure 2).

![Figure 2. The model.](image)

According to the above, the work value of any physical representation of $S$ for a demon knowing $X$ is

$$W(S|X) = E(S|X)kT\ln 2. $$

**Lower bound on the fuel value.** Let $C$ be a computable function

$$C : \{0,1\}^* \times \{0,1\}^* \rightarrow \{0,1\}^*$$

such that

$$(A, B) \mapsto (C(A, B), B)$$

is injective. We call $C$ a data-compression algorithm with helper. Then we have

$$E(S|X) \geq \text{len}(S) - \text{len}(C(S, X)).$$

This can be seen as follows. First, note that the function

$$A\|B \mapsto C(A, B)||\{\text{len}(A) - \text{len}(C(A, B))\}||B$$

is computable and bijective. Given the two (possibly irreversible) circuits computing the compression and its inverse, one can obtain a reversible circuit realizing the function and where no further input or output bits are involved. This can be achieved by first implementing all logical operations with Toffoli gates and uncomputing all junk [14] in both of the circuits. The resulting two circuits have now both still the property that the input is part of the output. As a second step, we can simply combine the two, where the first circuit’s first output becomes the second’s second input, and vice versa. Roughly speaking, the first circuit computes the compression and the second reversibly uncomputes the raw data. The combined circuit has only the compressed data (plus the 0’s) as output, on the bit positions carrying the input previously. (The depth of this circuit is roughly the sum of the depths of the two irreversible circuits for the compression and for the decompression, respectively.) We assume that circuit to be hard-wired in the demon’s head. A typical example for a compression algorithm that can be used is Ziv-Lempel [51].

**Upper bound on the fuel value.** We have the following upper bound on $E(S|X)$:

$$E(S|X) \leq \text{len}(S) - K_\mathcal{U}(S|X).$$

The reason is that the demon is only able to carry out the computation in question (logically, hence, thermodynamically) reversibly if she is able to carry out the reverse computation as well. Therefore, the string $P$ must be at least as long as the shortest program for $\mathcal{U}$ generating $S$ if $X$ is given.

Although the same is not true in general, this upper bound is tight if $K_\mathcal{U}(S|X) = 0$. The latter means that $X$ itself is a program for generating an additional copy of $S$. The demon can then bit-wisely XOR this new copy to the original $S$ on the tape, hereby producing $\{\text{len}(S)\}$ reversibly to replace the original $S$ (at the same time preserving the new one, as reversibility demands). When Bennett’s “uncomputing trick” is used — allowing for making any computation by a Turing machine logically reversible [14] —, then a history string $H$ is written to the tape during the computation of $S$ from $X$ such that after the XOR-ing, the demon can, going back step by step, uncompute...
the generated copy of $S$ and end up in the tape’s original state — except that the original $S$ is now replaced by $\Omega^{\text{len}(S)}$. This results in a maximal fuel value matching the (in this case trivial) upper bound. Note that this harmonizes with [16] if vanishing conditional entropy is so established.

**Discussion.** We contrast our bounds with the entropy-based results of [21]. According to the latter, a demon having complete knowledge of $S$ is able to extract maximal work: $E(S) \approx \text{len}(S)$. What means “knowing $S$?” (see Figure 3). We have seen that the results are in accordance with ours if the demon’s knowledge consists of (a) a copy of $S$, or at least of (b) the ability to algorithmically reconstruct $S$, based on a known program $P$, as discussed above. It is, however, possible (c) that the demon’s knowledge is of different nature, merely determining $S$ uniquely without providing the ability to build $S$. For instance, let the demon’s knowledge about $S$ be: “$S$ equals the first $N$ bits $\Omega_N$ of the binary expansion of $\Omega$.” Here, $\Omega$ is the so-called halting probability [19] of a fixed universal Turing machine (e.g., the demon $U$ itself). Although there is a short description of $S$ in this case, and $S$ is thus uniquely determined in an entropic sense, there is no set of instructions shorter than $S$ enabling the demon to generate $S$ — which would be required for work extraction from $S$ according to our upper bound. In short, this gap reflects the one between the “unique-description complexity”[8] and the Kolmogorov complexity.

IV. COMPLEXITY AS RANDOMNESS 2: NON-LOCALITY

A. Non-Locality from Counterfactual Definiteness

Non-local correlations [12] are a fascinating feature of quantum theory. The conceptually challenging aspect is the difficulty of explaining the correlations’ origin causally, i.e., according to Reichenbach’s principle, stating that a correlation between two space-time events can stem from a common cause (in the common past) or a direct influence from one event to the other [39]. More specifically, the difficulty manifests itself when alternatives — hence, counterfactuals — are taken into account: The argument leading up to a Bell inequality relates outcomes of alternative measurements — only one of which can actually be realized. Does this mean that if we drop the assumption of counterfactual definiteness [52], i.e., the requirement to consistently understand counterfactual events, the paradox or strangeness disappears? The answer is no: Even in the “factual-only view,” the joint properties — in terms of mutual compressibility — of the involved (now: fixed) pieces of information are such that consequences of non-local correlations, as understood in a common probability-calculus, persist: An example is the significant complexity forced upon the output given the input’s maximal complexity plus some natural translocation of no-signaling to the static scenario (see Figure 4).

In the traditional, probabilistic view, a Popescu-Rohrlich (PR) box [37] gives rise to a mechanism of the following kind: Let $A$ and $B$ the respective input bits to the box and $X$ and $Y$ the output bits; the (classical) bits satisfy

$$X \oplus Y = A \cdot B.$$  \hfill (1)

This system is no-signaling, i.e., the joint input-output behavior is useless for message transmission. Interestingly, on the other hand, the non-locality of the correlation means that classically speaking, signaling would be required to explain the behavior since shared classical information is insufficient.) According to a result by Fine [23], the non-locality of the system (i.e., conditional distribution) $P_{XY|AB}$, which means that it cannot be written as a convex combination of products $P_{X|A} \cdot P_{Y|B}$, is equivalent to the fact that there exists no “roof distribution” $P'_{XY|AB}$ such that

$$P'_{XY} = P_{XY|A=i, B=j}$$

for all $(i, j) \in \{0, 1\}^2$. In this view, non-locality means that the outputs to alternative inputs cannot consistently coexist. The counterfactual nature of this reasoning has already been pointed out by Specker [12]: “In einem gewissen Sinne gehören aber auch die scholastischen Spekulationen über die Infutasurablen hierher, das heisst die Frage, ob sich die göttlichen Allwissenheit auch

---

[8] A diagonal argument, called Berry paradox, shows that the notion of “description complexity” cannot be defined generally for all strings.
Figure 4. The traditional (a) vs. the new (b) view: Non-locality à la Popescu/Rohrlich (PR) plus no-signaling leads to the output inheriting randomness (a) or complexity (b), respectively, from the input.

auf Ereignisse erstrecke, die eingetreten wären, falls etwas geschehen wäre, war nicht geschehen ist.” — “In some sense, this is also related to the scholastic speculations on the infuturabilis, i.e., the question whether divine omniscience even extends to what would have happened if something had happened that did not happen.”

We intend to challenge this view. Let us first restate in more precise terms the counterfactual reasoning. Such reasoning is intrinsically assuming or concluding statements of the kind that that some piece of classical information, such as a bit \( U \), exists or does not exist. What does this mean? Classicality of information is an idealized notion implying that it can be measured without disturbance and that the outcome of a measurement is always the same (which makes it clear this is an idealized notion requiring the classical bit to be represented in a redundantly extended way over an infinite number of degrees of freedom). It makes thus sense to say that a classical bit \( U \) exists, i.e., has taken a definite value.

In this way of speaking, Fine’s theorem \([23]\) reads: “The outputs cannot exist before the inputs do.” Let us make this qualitative statement more precise. We assume a perfect PR box, i.e., a system always satisfying \( X \oplus Y = A \cdot B \). Note that this equation alone does not uniquely determine \( P_{XY|AB} \) since the marginal of \( X \), for instance, is not determined. If, however, we additionally require no-signaling, then the marginals, such as \( P_{X|A=0} \) or \( P_{Y|B=0} \), must be perfectly unbiased under the assumption that all four \((X,Y)\)-combinations, i.e., \((0,0),(0,1),(1,0)\), and \((1,1)\), are possible. To see this, assume on the contrary that \( P_{X|A=0,B=0}(0) > 1/2 \). By the PR condition \([1]\), we can conclude the same for \( Y \): \( P_{Y|A=0,B=0}(0) > 1/2 \). By no-signaling, we also have \( P_{X|A=0,B=1}(0) > 1/2 \). Using symmetry, and no-signaling again, we obtain both \( P_{X|A=1,B=0}(0) > 1/2 \) and \( P_{Y|A=1,B=1}(0) > 1/2 \). This contradicts the PR condition \([1]\) since two bits which are both biased towards 0 cannot differ with certainty. Therefore, our original assumption was wrong: The outputs must be perfectly unbiased. Altogether, this means that \( X \) as well as \( Y \) cannot exist (i.e., take a definite value — actually, there cannot even exist a classical value arbitrarily weakly correlated with one of them) before for some nontrivial deterministic function \( f : \{0,1\}^2 \rightarrow \{0,1\} \), the classical bit \( f(A,B) \) exists. The paradoxical aspect of non-locality — at least if a causal structure is in place — now consists of the fact that fresh pieces of information come to existence in a spacelike-separated way but that are nonetheless perfectly correlated.

B. Non-Locality without Counterfactual Definiteness

We propose an understanding of non-locality that refrains from using counterfactual definiteness but invokes solely the data at hand, i.e., existing in a single reality \([18]\).

Uncomputability of the outputs of a PR box. Let first \((a,b,x,y)\) be infinite binary strings with

\[
x_i \oplus y_i = a_i \cdot b_i .
\]

Obviously, the intuition is that the strings stand for the inputs and outputs of a PR box. Yet, no dynamic meaning is attached to the strings anymore (or to the “box,” for that matter) since there is no free choice of an input — i.e., a choice that “could also have been different” (a notion we discussed and suspect to be hard to define precisely in the first place) — and no generation of an output in function of an input; all we have are four fixed strings satisfying the PR condition \([2]\). However, nothing prevents us from defining this (static) situation to be no-signaling:

\[
K(x | a) \approx K(x | ab) \text{ and } K(y | b) \approx K(y | ab) .
\]

Recall the mechanism which the maximal non-locality displayed by the PR box enables: If the inputs are not entirely fixed, then the outputs must be completely unbiased as soon as the system is no-signaling. We can now draw a similar conclusion, yet entirely within actual — and without having to refer to counterfactual — data:

If the inputs are incompressible and independent, and no-signaling holds, then the outputs must be uncomputable.
For a proof of this, let \((a, b, x, y) \in \{(0, 1)^N\}^4\) with \(x \oplus y = a \cdot b\) (bit-wisely), no-signaling [3], and
\[
K(a, b) \approx 2n ,
\]
i.e., the “input” pair is incompressible. We conclude
\[
K(a \cdot b \mid b) \approx n/2 .
\]
Note first that \(b_i = 0\) implies \(a_i \cdot b_i = 0\), and second that any further compression of \(a \cdot b\), given \(b\), would lead to “structure in \((a, b)\)” i.e., a possibility of describing (programming) \(a\) given \(b\) in shorter than \(n\) and, hence, \((a, b)\) in shorter than \(2n\). Observe now
\[
K(x \mid b) + K(y \mid b) \geq K(a \cdot b \mid b) ,
\]
which implies
\[
K(y \mid b) \geq K(a \cdot b \mid b) - K(x \mid b) \gtrsim n/2 - K(x) .
\]  
(4)
On the other hand,
\[
K(y \mid a, b) \approx K(x \mid a, b) \leq K(x) .
\]  
(5)
Now, no-signaling [3] together with [4] and [5] implies
\[
n/2 - K(x) \lesssim K(x) ,
\]
and
\[
K(x) \geq n/4 = \Theta(n) : \text{The string } x \text{ must be uncomputable.}
\]
We have seen that if the pair of inputs \((a, b)\) is maximally incompressible, then the outputs \(x\) and \(y\) must at least be uncomputable. This observation raises a number of natural questions: Does a similar result hold with respect to the conditional complexities \(K(x \mid a)\) and \(K(y \mid b)\)? With respect to quantum non-local correlations? Can we give a suitable general definition of non-locality and does a similar result as the above hold with respect to any non-local correlation? Can we strengthen and tighten our arguments to show, for instance, that uncomputable inputs plus no-signaling and maximal non-locality leads to incompressibility of the outputs? What results might turn out to be incompressibility-amplification methods. Let us address these questions.

Conditional uncomputability of the outputs of a PR box. With respect to the same assumptions as in the previous section, we now consider the quantities \(K(x \mid a)\) and \(K(y \mid b)\), respectively. Note first
\[
K(x \mid a) \approx 0 \Leftrightarrow K(x \mid ab) \approx K(y \mid ab) \approx 0 \Leftrightarrow K(y \mid b) \approx 0 ,
\]
i.e., the two expressions vanish simultaneously. We show that, in fact, they both fail to be of order \(o(n)\). In order to see this, assume \(K(x \mid a) \approx 0\) and \(K(y \mid b) \approx 0\). Hence, there exist programs \(P_n\) and \(Q_n\) (both of length \(o(n)\)) for functions \(f_n\) and \(g_n\) with
\[
f_n(a_n) \oplus g_n(b_n) = a_n \cdot b_n .
\]  
(6)
For fixed (families of) functions \(f_n\) and \(g_n\), asymptotically how many \((a_n, b_n)\) can at most exist that satisfy (6)?

The question boils down to a parallel-repetition analysis of the PR game: A result by Raz [25] implies that the number is of order \((2 - \Theta(1))^2n\). Therefore, the two programs \(P_n\) and \(Q_n\) together with the index, of length
\[
o(n) + (1 - \Theta(1))2n ,
\]
of the correct pair \((a, b)\) within the list of length \((2 - \Theta(1))^2n\) lead to a program, generating \((a, b)\), that has length

\[
o(n) + (1 - \Theta(1))2n ,
\]
in contradiction to the assumption of incompressibility of \((a, b)\).

Conditional uncomputability from quantum correlations. In the “traditional view” on non-locality, the PR box is an idealization unachievable by the behavior of any quantum state. If it did exist, on the other hand, it would be a most precious resource, e.g., for cryptography or randomness amplification. The reason is that — as we have discussed above — under the minimal assumption that the inputs are not completely determined, the outputs are perfectly random, even given the inputs.

Perfect PR boxes are not predicted by quantum theory, but sometimes, the best approximations to PR boxes that are quantum physically achievable (~ 85%) can be used for information-processing tasks, such as key agreement [27]. For our application here, however, we found this not to be the case. On the other hand, it has been shown [6, 15, 19] that correlations which are achievable in the laboratory [13] allow for similar applications; they are based on the chained Bell inequality instead of perfect PR-type non-locality. We show the same to hold here.

To the chained Bell inequality belongs the following idealized system: Let \(A, B \in \{1, \ldots, m\}\) be the inputs. We assume the “promise” that \(B\) is congruent to \(A\) or to \(A + 1\) modulo \(m\). Given this promise, the outputs \(X, Y \in \{0, 1\}\) must satisfy
\[
X \oplus Y = \chi_{A=m, B=1} ,
\]  
(7)
where \(\chi_{A=m, B=1}\) is the characteristic function of the event \(\{A = m, B = 1\}\).

Barrett, Hardy, and Kent [6] showed that if \(A\) and \(B\) are random, then \(X\) and \(Y\) must be perfectly unbiased if the system is no-signaling. More precisely, they were even able to show such a statement from the gap between the error probabilities of the best classical — \(\Theta(1/m)\) — and quantum — \(\Theta(1/m^2)\) — strategies for winning this game.
In our framework, we show the following statement. Let \((a, b, x, y) \in (\{1, \ldots, m\}^n)^2 \times (\{0, 1\}^n)^2\) be such that the promise holds, and such that
\[
K(a, b) \approx (\log m + 1) \cdot n ,
\]
i.e., the string \(a|b\) is maximally incompressible given the promise; the system is no-signaling if otherwise there would be the possibility of compressing \((a, b)\).

These positions, the string due to the string’s incompressibility. If we condition on the fractions of 1’s in \(b\) from position 1 to \(n\), is to generate \(x[n]\) and \(y[n]\) as well as the string indicating the positions where \((7)\) is violated, the complexity of the latter being at most
\[
K(x = \Theta(n))
\]

Let us prove this statement. First, \((a, b)\) being maximal implies
\[
K(\chi_{a=m,b=1} | b) \approx \frac{n}{m} \quad (8)
\]
The fractions of 1’s in \(b\) must, asymptotically, be \(1/m\) due to the string’s incompressibility. If we condition on these positions, the string \(\chi_{a=m,b=1}\) is incompressible, since otherwise there would be the possibility of compressing \((a, b)\).

Now, we have
\[
K(x | b) + K(y | b) + h(\Theta(1/m^2)) n \geq K(\chi_{a=m,b=1} | b)
\]
since one possibility for “generating” the string \(\chi_{a=m,b=1}\), from position 1 to \(n\), is to generate \(x[n]\) and \(y[n]\) as well as the string indicating the positions where \((7)\) is violated, the complexity of the latter being at most
\[
\log \left( \frac{n}{\Theta(1/m^2)n} \right) \approx h(\Theta(1/m^2)) n .
\]

Let us compare this with \(1/m\): Although the binary entropy function has slope \(\infty\) in 0, we have
\[
h(\Theta(1/m^2)) < 1/(3m)
\]
if \(m\) is sufficiently large. To see this, observe first that the dominant term of \(h(x)\) for small \(x\) is \(−x \log x\), and second that
\[
c(1/m) \log(m^2/c) < 1/3
\]
for \(m\) sufficiently large.

Together with \((8)\), we now get
\[
K(y | b) \geq \frac{2n}{3m} - K(x) \quad (9)
\]
if \(m\) is chosen sufficiently large. On the other hand,
\[
K(y | ab) \leq K(x | ab) + h(\Theta(1/m^2)) n
\]
\[
\leq K(x) + \frac{n}{3m} . \quad (10)
\]

Now, \((9), (9)\), and \((10)\) together imply
\[
K(x) \geq \frac{n}{6m} = \Theta(n) ;
\]
in particular, \(x\) must be uncomputable.

For any non-local behavior characterizable by a condition that is always satisfiable with entanglement, but not without this resource — so called “pseudo-telepathy” games —, the application of Raz’ parallel-repetition theorem shows that incompressibility of the inputs leads to uncomputability of at least one of the two outputs even given the respective input, i.e.,
\[
K(x | a) \neq 0 \text{ or } K(y | b) \neq 0 .
\]

We illustrate the argument with the example of the magic-square game [3]: Let \((a, b, x, y) \in (\{1, 2, 3\}^N)^2 \times (\{1, 2, 3, 4\}^{N^2})\) be the quadruple of the inputs and outputs, respectively, and assume that the pair \((a, b)\) is incompressible as well as \(K(x | a) \approx 0 \approx K(y | b)\). Then there exist \(o(n)\)-length programs \(P_a, Q_a\) such that \(x[n] = P_a(a[n])\) and \(y[n] = Q_a(b[n])\). The parallel-repetition theorem [38] implies that the length of a program generating \((a[n], b[n])\) is, including the employed sub-routines \(P_a\) and \(Q_a\) of order \(1 - \Theta(1)\) (\(\log(\Theta(1)\) — in contradiction to the incompressibility of \((a, b)\).

An all-or-nothing flavor to the Church-Turing hypothesis. Our lower bound on \(K(x | a)\) or on \(K(y | b)\) means that if the experimenters are given access to an incompressible number (such as \(\Omega\)) for choosing their measurement bases, then the measured photon (in at least one of the two labs) is forced to generate an uncomputable number as well, even given the string determining its basis choices. Roughly speaking, there is either no incompressibility at all in the world, or it is full of it. We can interpret that as an all-or-nothing flavor attached to the Church-Turing hypothesis: Either no physical system at all can carry out “beyond-Turing” computations, or even a single photon can.

General definition of (non-)locality without counterfactuality. We propose the following definition of when a no-signaling quadruple \((a, b, x, y) \in (\{0, 1\}^N)^4\) (where \(a, b\) are the “inputs” and \(x, y\) the outputs) is local. There must exist \(\lambda \in \{0, 1\}^N\) such that
\[
K(a, b, \lambda) \approx K(a, b) + K(\lambda) , \quad (11)
\]
\[
K(x | a\lambda) \approx 0 , \quad \text{and}
\]
\[
K(y | b\lambda) \approx 0 .
\]

Sufficient conditions for locality are then
\[
K(a, b) \approx 0 \quad \text{or} \quad K(x, y) \approx 0 ,
\]
since we can set \(\lambda := (x, y)\). At the other end of the scale, we expect that for any non-local “system,” the fact that

\footnotesize
\[\text{Here, } h \text{ is the binary entropy } h(x) = -p \log p - (1 - p) \log(1 - p) \quad .\]

\[\text{Usually, } p \text{ is a probability, but } h \text{ is invoked here merely as an approximation for binomial coefficients.}\]
$K(a, b)$ is maximal implies that $x$ or $y$ is conditionally uncomputable, given $a$ and $b$, respectively.

It is a natural question whether the given definition harmonizes with the probabilistic understanding. Indeed, the latter can be seen as a special case of the former: If the (fixed) strings are typical sequences of a stochastic process, our non-locality definition implies non-locality of the corresponding conditional distribution. The reason is that a hidden variable of the distribution immediately gives rise, through sampling, to a $\lambda$ in the sense of [11]. Note, however, that our formalism is strictly more general since asymptotically, almost all strings fail to be typical sequences of such a process.

V. DROPPING CAUSALITY 1: OBJECTIVE THERMODYNAMICS AND THE SECOND LAW

It has already been observed that the notion of Kolmogorov complexity can allow, in principle, for thermodynamics independent of probabilities or ensembles: Zurek [53] defines physical entropy $H_p$ to be

$$H_p(S) := K(M) + H(S | M),$$

where $M$ stands for the collected data at hand while $H(S | M)$ is the remaining conditional Shannon entropy of the macrostate $S$ given $M$. That definition of a macrostate is subjective since it depends on the available data. How instead can the macrostate — and entropy, for that matter — be defined objectively? We propose to use the Kolmogorov sufficient statistics [25] of the microstate: For any $k \in \mathbb{N}$, let $M_k$ be the smallest set such that $S \in M_k$ and $K(M_k) \leq k$ hold. Let further $k_0$ be the value of $k$ at which the function $\log |M_k|$ becomes linear with slope $-1$. Intuitively speaking, $k_0$ is the point beyond which there is no more “structure” to exploit for describing $S$ within $M_{k_0}$. $S$ is a “typical element” of the set $M_{k_0}$. We define $M(S) := M_{k_0}$ to be $S$’s macrostate. It yields a program generating $S$ of minimal length

$$K(S) = k_0 + \log |M_{k_0}| = K(M(S)) + \log |M(S)|.$$

The fuel value (as discussed in Section [11]) of a string $S \in \{0, 1\}^N$ is now related to the macrostate $M(S) \supseteq S$ by

$$E(S) \leq N - K(M(S)) - \log |M(S)|$$

(see Figure 5): Decisive is neither the complexity of the macrostate nor its log-size alone, but their sum.

A notion defined in a related way is the sophistication or interestingness as discussed by Aaronson [1] investigating the process where milk is poured into coffee (see Figure 6). Whereas the initial and final states are “simple” and “uninteresting,” the intermediate (non-equilibrium) states display a rich structure; here, the sophistication — and also $K(M)$ for our macrostate $M$ — becomes maximal.

During the process under consideration, neither the macrostate’s complexity nor its size is monotonic in time:

$$E(S) \leq N - K(M(S)) - \log |M(S)|$$

(see Figure 5): Decisive is neither the complexity of the macrostate nor its log-size alone, but their sum.

A notion defined in a related way is the sophistication or interestingness as discussed by Aaronson [1] investigating the process where milk is poured into coffee (see Figure 6). Whereas the initial and final states are “simple” and “uninteresting,” the intermediate (non-equilibrium) states display a rich structure; here, the sophistication — and also $K(M)$ for our macrostate $M$ — becomes maximal.

$$E(S) \leq N - K(M(S)) - \log |M(S)|$$

On the other hand, the complexity of the microstate,

$$K(S) = K(M) + \log |M|,$$

is a candidate for a (essentially) monotonically nondecreasing quantity: Is this the second law of thermodynamics in that view? This law, which claims a certain quantity to be (essentially) monotonic in time, is by many believed to be the origin of our ability to distinguish the future from the past.

The second law, traditional view. Let a closed system be in a thermodynamical equilibrium state of entropy $S_1$ at time $t_1$. Assume that the system evolves to another equilibrium state, of entropy $S_2$, at some fixed later time $t_2 > t_1$. Then, for $s > 0$,

$$\text{Prob}[S_1 - S_2 \geq s k \ln 2] = 2^{-s}.$$
It is a rare example — outside quantum theory — of a physical “law” holding only with some probability.

Is there an underlying fact in the form of a property of an evolution holding with certainty and also for all intermediate states?

Clearly, that fact would not talk about the coarse-grained behavior of the system, which we have seen in the discussed example to be non-monotonic in time. If, however, we consider the microstate, then logical reversibility — meaning that the past can be computed step by step from the future (not necessarily vice versa) — is a good candidate: Indeed, also Landauer’s principle links the second law to logical irreversibility. A logically reversible evolution is potentially asymmetric in time if the backward direction is not logically reversible.

In the spirit of the Church-Turing hypothesis, we see the state of a closed system as a finite binary string and its evolution (through discretized time) as being computed by a universal Turing machine.

The second law, revisited. The evolution of a closed system is logically reversible and the past at time $t_1$ can be computed from the future at time $t_2$ ($> t_1$) by a constant-length program on a Turing machine.

It is somewhat ironic that this view of the second law puts forward the reversibility of the computation, whereas the law is usually linked to the opposite: irreversibility. A consequence of the law is that the decrease of the Kolmogorov complexity of the string encoding the system’s state is limited.

Consequence of the second law, revisited. Let $x_1$ and $x_2$ be the contents of a reversible Turing machine’s tape at times $t_1 < t_2$. Then

$$K(x_1) \leq K(x_2) + \Theta(\log(t_2 - t_1)).$$

If the Turing machine is deterministic, the complexity increases at most logarithmically in time. On the other hand, this growth can of course be arbitrarily faster for probabilistic machines. Turned around, Kolmogorov complexity can yield an intrinsic criterion for the distinction between determinism and indeterminism (see Figure 8). In the case of randomness, a strong asymmetry and an objective arrow of time can arise. A context-free definition of randomness (or free will for that matter) has the advantage not to depend on the “possibility that something could have been different from how it was,” a metaphysical condition we came to prefer to avoid.

The traditional second law from complexity increase. It is natural to ask what the connection between logical reversibility and complexity on one side and the traditional second law on the other is. We show that the latter emerges from increasing complexity — including the exponential error probabilities.

Let $x_1$ and $x_2$ be the microstates of a closed system at times $t_1 < t_2$ with $K(x_2) \geq K(x_1)$. If the macrostates $M_1$ and $M_2$ of $x_1$ and $x_2$, respectively, have small Kolmogorov complexity (such as traditional thermodynamical equilibrium states characterized by global parameters like volume, temperature, pressure, etc.), then

$$|M_1| \lesssim |M_2|.$$

If the macrostates are simple, then their size is non-decreasing. Note that this law is still compatible with the exponentially small error probability ($2^{-N}$) in the traditional view of the second law for a spontaneous immediate drop of entropy by $\Theta(n)$: The gap opens when the simple thermodynamical equilibrium macrostate of a given microstate differs from our macrostate defined through the Kolmogorov statistics. This can occur if, say, the positions and momenta of the molecules of some (innocent-, i.e., general-looking) gas encode, e.g., $\pi$ and have essentially zero complexity.

We can now finish up by closing a logical circle. We have started from the converse of Landauer’s principle, went through work extraction and ended up with a complexity-theoretic view of the second law: We have returned back to our starting point.

Landauer’s principle, revisited. The (immediate) transformation of a string $S$ to the 0-string of the same length requires free energy at least

$$K(S)kT \ln 2,$$

which is then dissipated as heat to the environment. For every concrete lossless compression algorithm $C$,

$$\text{len}(C(S))kT \ln 2 + \Theta(1),$$

is, on the other hand, an upper bound on the required free energy.

Finally, Landauer’s principle can be combined with its converse and generalized as follows.

![Figure 8. Randomness vs. determinism.](image-url)
Generalized Landauer’s principle. Let \( A \) and \( B \) two bit strings of the same length. The (immediate) transformation from \( A \) to \( B \) costs at least
\[
(K(A) - K(B)) kT \ln 2
\]
free energy, or it releases at most the absolute value of \([12]\) if this is negative.

If the Turing machine is a closed physical system, then this principle reduces to the complexity-non-decrease stated above. This suggests that the physical system possibly simulated by the machine — in the spirit of the Church-Turing hypothesis — also follows the second law (e.g., since it is a closed system as well). The fading boundaries between what the machine is and what is simulated by it are in accordance with Wheeler’s [46] “it from bit.” Every “it” — every particle, every field of force, even the spacetime continuum itself — derives its function, its meaning, its very existence entirely [...] from the apparatus-elicited answers to yes or no questions, binary choices, “bits.” If we try to follow the lines of such a view further, we may model the environment as a binary string \( R \) as well. The goal is a unified discourse avoiding to speak about complexity with respect to one system and about free energy, heat, and temperature to the other. The transformation addressed by Landauer’s principle and its converse then looks as in Figure 9: The transformation from inside would like to find a causal structure arising from inside, i.e., from the properties of, and relations between, these strings. The intuition is that an \( x \in \mathcal{C} \) encodes the totality of momentary local physical reality in a “point,” i.e., parameters such as mass, charge, electric and magnetic field density.

Let \( \mathcal{C} \subseteq \{0,1\}^N \) be finite. We define the following order relation on \( \mathcal{C} \):
\[
x \preceq y \iff K(x|y) \approx 0 .
\]
We say that \( x \) is a cause of \( y \), and that \( y \) is an effect of \( x \). So, \( y \) is in \( x \)’s future exactly if \( y \) contains the entire information about \( x \); no information is ever lost. The intuition that any “change” in the cause affects each one of its effects — if sufficient precision is taken into account. We write \( x \prec y \) if \( x \preceq y \) as well as \( x \succeq y \) hold. If \( x \preceq y \) and \( x \not\preceq y \), we write \( x \not\preceq y \) and call \( x \) and \( y \) spacelike separated. We call the pair \((\mathcal{C}, \preceq)\) a causal structure.

For a set \( \{x_i\} \subseteq \mathcal{C} \) and \( y \in \mathcal{C} \), we say that \( y \) is the first common effect of the \( x_i \) if it is the least upper bound: \( x_i \preceq y \) holds for all \( x_i \), and for any \( z \) with \( x_i \preceq z \) for all \( x_i \), also \( y \preceq z \) holds. The notion of last common cause is

\[\text{Figure 9. Work extraction and Landauer’s principle in the view of “the church of the larger bit string.”}\]

VI. DROPPING CAUSALITY 2: SPACE-TIME FROM COMPLEXITY

A. Information and Space-Time

If, motivated by the above, we choose to regard information as being more fundamental than space and time, how can the latter be imagined to emerge from the former? Can such a causal structure be understood to be of logical rather than physical nature? In other words, is it more accurate to imagine causal relations to be a proper-ty of logical rather than physical spaces [47]? We address these questions here, continuing to avoid speaking about “what could have been different,” i.e., the counterfactual viewpoint.

In Section [V] an arrow of time has emerged under the assumption of (uni-directional) logical reversibility. Here, we refine the same idea in an attempt to derive a causal structure based on the principle that any point carries complete information about its space-time past.

B. Causal Structures

Let us start with a finite set \( \mathcal{C} \) of strings on which we would like to find a causal structure arising from inside, i.e., from the properties of, and relations between, these strings. The intuition is that an \( x \in \mathcal{C} \) encodes the totality of momentary local physical reality in a “point,” i.e., parameters such as mass, charge, electric and magnetic field density.

Let \( \mathcal{C} \subseteq \{0,1\}^N \) be finite. We define the following order relation on \( \mathcal{C} \):
\[
x \preceq y \iff K(x|y) \approx 0 .
\]
We say that \( x \) is a cause of \( y \), and that \( y \) is an effect of \( x \). So, \( y \) is in \( x \)’s future exactly if \( y \) contains the entire information about \( x \); no information is ever lost. The intuition that any “change” in the cause affects each one of its effects — if sufficient precision is taken into account. We write \( x \prec y \) if \( x \preceq y \) as well as \( x \succeq y \) hold. If \( x \preceq y \) and \( x \not\preceq y \), we write \( x \not\preceq y \) and call \( x \) and \( y \) spacelike separated. We call the pair \((\mathcal{C}, \preceq)\) a causal structure.

For a set \( \{x_i\} \subseteq \mathcal{C} \) and \( y \in \mathcal{C} \), we say that \( y \) is the first common effect of the \( x_i \) if it is the least upper bound: \( x_i \preceq y \) holds for all \( x_i \), and for any \( z \) with \( x_i \preceq z \) for all \( x_i \), also \( y \preceq z \) holds. The notion of last common cause is 10

10 In this section, conditional complexities are understood as follows: In \( K(x|y) \), for instance, the condition \( y \) is assumed to be the full (infinite) string, whereas the asymptotic process runs over \( x_{[n]} \). The reason is that very insignificant bits of \( y \) (intu- itively: the present) can be in relation to bits of \( x \) (the past) of much higher significance. The past does not disappear, but it fades.
defined analogously. A minimum (maximum) of \((C, \preceq)\) is called without cause (without effect). If \(C\) has a smallest (greatest) element, this is called big bang (big crunch).

We call a causal structure deterministic if, intuitively, every \(y\) which is not without cause is completely determined by all its causes. Formally, for some \(y \in C\), let \(\{z_i\}\) be the set of all \(x_i \in C\) such that \(x_i \preceq y\) holds. Then we must have
\[
K(y \mid x_1, x_2, \ldots) \approx 0 .
\]
Otherwise, \(C\) is called probabilistic.

**C. The Emergence of Space-Time**

Observe first that every deterministic causal structure which has a big bang is trivial: We have
\[
x \preceq y \text{ for all } x, y \in C .
\]
This can be seen as follows. Let \(b\) be the big bang, i.e., \(b \preceq x\) for all \(x \in C\). On the other hand, \(K(x \mid z_1) \approx 0\) if \(\{z_i\}\) is the set of predecessors of \(x\). Since the same is true for each of the \(z_i\), we can continue this process and, ultimately, end up with only \(b\): \(K(x \mid b) \approx 0\), i.e., \(x \preceq b\), and thus \(b \approx x\) for all \(x \in C\). In this case, we obviously cannot expect to be able to explain space-time.

(Note, however, that there can still exist deterministic \(C\)'s — without big bang — with non-trivial structure.) However, the world as it presents itself to us — with both big bang and arrow of time — seems to direct us away from determinism (in support of [23]).

The situation is very different in probabilistic causal structures: Here, the partial order relation \(\preceq\) gives rise to a non-trivial picture of causal relations and, ideally, causal space-time including the arrow of time. Obviously, the resulting structure depends crucially on the set \(C\). Challenging open problems are to understand the relationship between sets of strings and causal structures: Can every partially ordered set be implemented by a suitable set of strings? What is the property of a set of strings that gives rise to the “usual” space-time of relativistic light-cones?

Is it helpful to introduce a metric instead of just an order relation? As a first step, it appears natural to define \(K(y \mid x)\) as the distance of \(x\) from the set of effects of \(y\). In case \(y\) is an effect of \(x\), this quantity intuitively measures the time by which \(x\) happens before \(y\).

Generally in such a model, what is a “second law,” and under what condition does it hold? Can it — and the arrow of time — be compatible even with determinism (as long as there is no big bang)?

What singles out the sets displaying quantum non-local correlations as observed in the lab? (What is the significance of Tsirelson’s bound in the picture?)

**VII. DROPPING CAUSALITY 3: PRESERVING LOGICAL CONSISTENCY**

A recent framework for quantum [35] and classical [10] correlations without causal order is based on local assumptions only. These are the local validity of quantum or classical probability theory, that laboratories are closed (parties can only interact through the environment), and that the probabilities of the outcomes are linear in the choice of local operation. The global assumption of a fixed global causal order is replaced by the assumption of logical consistency: All probabilities must be non-negative and sum up to 1. Some correlations — termed non-causal — that can be obtained in this picture cannot arise from global quantum or classical probability theory. Similarly to the discovery of non-local correlations that showed the existence of a world between the local and the signaling; in a similar sense, we discuss here a territory that lies between what is causal and what is logically inconsistent: It is not empty.

In the spirit of Section [IV] where we studied the consequences of non-locality, we show that the results from non-causal correlations carry over to the picture of (conditional) compressibility of bit strings, where we do not employ probabilities, but consider actual data only. In that sense, these are the non-counterfactual versions of results on non-causal correlations.

**A. Operational Definition of Causal Relations**

We define causal relations operationally, where we use the notion of parties. A party can be thought of as a system, laboratory, or an experimenter, performing an operation. In the traditional view, the choice of operation is represented by randomness, and thus by a probability distribution. Here, in contrast, we refrain from this counterfactual approach (probabilities), and consider actual — as opposed to potential — choices only. The traditional view with probabilities has a dynamic character: Systems undergo (randomized) evolutions. Like in Section [IV] we obtain a static situation if we consider actual data only. All statements are formulated with bit strings and relations between these strings modeling the “operations.”

A party \(A\) is modeled by two bit strings \(A_I\) and \(A_O\). We restrict ourselves to pairs of bit strings that satisfy some relation \(A\). Within a party, we assume a fixed causal structure \((A_I\text{ precedes }A_O)\) (see Figure [10]). The relation \(A\) is called local operation of \(A\), the string \(A_I\) is called input to \(A\), and \(A_O\) is \(A\)'s output. If we have more than one party, we consider only those input and output bit strings that satisfy some global relation. These relations are, as in Section [IV], to be understood to act locally on the involved strings: A relation involves only a finite number of instances (bit positions), and it is repeated \(n (\to \infty)\) times for obtaining the global relation.

For two parties \(A\) and \(B\), we say that \(A\) is in the causal
past of $B$, $A \preceq B$, if and only if
\begin{equation}
K(B_1 | A_O) \neq K(B_1) \neq 0.
\end{equation}

Intuitively, $A$ is in the causal past of $B$ if and only if $B$'s input is uncomputable — otherwise $B$ could simply obtain it herself — and better compressible with $A$’s output than without, i.e., the two strings depend on each other. Expressed according to intuitive dynamic thinking, the definition means that $A$ is in the causal past of $B$ if and only if $B$ learns parts of an incompressible string from $A$. The causal relation among parties defined here is extended straight-forwardly to the scenario where one or more parties are in the causal past of one or more parties.

This definition is different from the one proposed in Section VI. There, a string is said to be the cause of another string $y$ (the effect) if and only if $K(x | y) \approx 0$. The intuition there is logical reversibility: Future events contain all information about past events, no information is ever lost, and $x$ and $y$ are understood to encode complete physical reality in some space-time point. In contrast, the definition here only relates pieces of information chosen and processed by the parties: If one party’s input depends on another’s output, then she is in the causal future of the latter. (Reversibility, the central notion in Section VI does not play a role here.) Since the strings now just correspond to the pieces of information manipulated by the parties, we cannot simply define freeness as an attribute of complexity. Instead, we postulate the output strings to be free.

The rationale of Definition (13) above is similar to the one we propose in §4 for the probability picture. There, $A \preceq B$ holds if and only if both random variables $A$ and $B$ are correlated and $A$ is postulated free. The motivation is to define causal relations based on freeness, and not the other way around (see Figure 11). Intuitively, if you flip a switch that is correlated to a light bulb, then flipping the switch is in the causal past of the light turning on or off — the definition of a causal relation relies on what we call free (the switch in this case). Such a definition based on postulated freeness is similar to the interventionist’s approach to causality, see e.g., [49]. In the approach studied here, the analog to correlation is dependence. The distinction between free and not free variables is done in the same way by distinguishing between input and output bit strings.

B. Causal Scenario

Causal scenarios describe input and output strings of the parties where the resulting causal relations reflect a partial ordering of the parties (see Figure 12a). In the most general case, the partial ordering among the parties of a set $S$, who are all in the causal future of some other party $A \notin S$, i.e., for all $B \in S : A \preceq B$, can depend (i.e., satisfy some relation with) the bit strings of $A$ [8]. [36]. A causal scenario, in particular, implies that at least one party is not in the causal future of some other parties. If no partial ordering of the parties arises, then the scenario is called causal (see Figure 12b).

A trivial example of a causal scenario is a communication channel over which a bit is perfectly transmitted from a party to another. This channel, formulated as a global relation, is $f(x,y) = (0,x)$, with $x,y \in \{0,1\}$, and where the first bit belongs to $A$ (sender) and the second to $B$ (receiver) (see Figure 13a). Consider the $n (\to \infty)$-fold sequential repetition of this global relation, and assume that both output bit strings are incompressible and independent: $K(A_O, B_O) \approx 2n$. The bit string $A_I$ is $(0,0,0,\ldots)$ according to the global relation. In contrast, $B_I$ is equal to $A_O$. Since $K(B_I) \approx n$ and $K(B_I | A_O) \approx 0$, the causal relation $A \preceq B$ holds, restating that $A$ is in the causal past of $B$. Conversely, $K(A_I) \approx 0$ and, therefore, $A \not\preceq B$: The receiver is not in the causal future of the sender.

11 Transitivity arises from the assumption of a fixed causal structure within a party, where the input is causally prior to the output.
C. Non-Causal Scenario

Consider the global relation
\[
g(x,y) = (y,x),
\]
(14)
which describes a two-way channel: A’s output is equal to B’s input and B’s output is equal to A’s input (see Figure 13a). This global relation can describe a non-causal scenario. If \( K(O_A, O_B) \approx 2n \), then indeed, the causal relations that we obtain are \( A \preceq B \) and \( B \preceq A \). What we want to underline here is that for this particular choice of local operations of the parties, input bit strings that are consistent with the relation \( (\ref{eq:14}) \) exist. In stark contrast, if we fix the local operations of the parties to be \( A_O = A_I \) and \( B_O = B_I \), then no choice of inputs \( A_I \) and \( B_I \) satisfies the desired global relation \( (\ref{eq:14}) \). This inconsistency is also known as the grandfather antinomy. If no satisfying input and output strings exist, then we say that the global relation is inconsistent with respect to the local operations. Otherwise, the global relation is consistent with respect to the local operations.

For studying bit-wise global relations, i.e., global relations that relate single output bits with single input bits, that are consistent regardless the local operations, we set the local operation to incorporate all possible operations on bits. These are the constants 0 and 1 as well as the identity and bit-flip operations. The parties additionally hold incompressible and independent strings that define which of these four relations is in place at a given bit position. For party \( P \), let this additional bit string be \( P_C \). Formally, if we have \( k \) parties \( A, B, C, \ldots \), then
\[
K(A_C, B_C, C_C, \ldots) \approx kn.
\]

The local operation of a party \( P \) is
\[
P_O^{(i)} = \begin{cases} 
0 & \text{if } (P_C^{(2i)}, P_C^{(2i+1)}) = (0,0), \\
1 & \text{if } (P_C^{(2i)}, P_C^{(2i+1)}) = (1,1), \\
P_I^{(i)} & \text{if } (P_C^{(2i)}, P_C^{(2i+1)}) = (0,1), \\
P_I^{(i)} \oplus 1 & \text{if } (P_C^{(2i)}, P_C^{(2i+1)}) = (1,0), 
\end{cases}
\]
where a superscript \((i)\) selects the \( i \)-th bit of a string. Depending on pairs of bits on \( P_C \), the relation \( (\ref{eq:15}) \) states that a given output bit is either equal to 0 or 1 or equal to or different from the corresponding input bit. An example is presented in Figure 14. Since all pairs of bits appear equally often in \( P_C \) (asymptotically speaking), in half the cases bits of the output string are identical to bits of the (incompressible) string \( P_C \) in the respective positions. Thus, the output satisfies \( K(P_O) = \Theta(n) \). We call the local operation of Eq. \( (\ref{eq:15}) \) of a party universal local operation. If a global relation is consistent with respect to universal local operations, then we call it logically consistent. If we consider all (bit-wise) operations, the global relation \( (\ref{eq:14}) \) becomes inconsistent: No input and output strings exist that satisfy the desired global relation \( (\ref{eq:14}) \). To see this, note that since we are in the asymptotic case, there exist positions \( i \) where the relation of \( A \) states that the \( i \)-th output bit is equal to the \( i \)-th input bit, and the relation of \( B \) states that the \( i \)-th output bit is equal to the negated \( i \)-th input bit, which results in a contradiction — the global relation cannot be satisfied. In more detail, there exists an \( i \) such that the bit string \( A_C \) contains the pair \((0,1)\) at position \( 2i \), and such that the bit string \( B_C \) contains the pair \((1,0)\) at the same position:
\[
A_C = (\ldots,0,1,\ldots), \\
B_C = (\ldots,1,0,\ldots).
\]
On the one hand, the input to \( A \) has a value \( a \) on the \( i \)-th position, and, because of \( A_C \), the same value is on the \( i \)-th position of \( A_O \):
\[
A_I = (\ldots,a,\ldots), \\
A_O = (\ldots,a,\ldots).
\]
The input and output bit strings of \( B \), on the other hand, must, due to \( B_C \), have opposite bits on the \( i \)-th position:
\[
B_I = (\ldots,b,\ldots), \\
B_O = (\ldots,b \oplus 1,\ldots).
\]

Figure 13. (a) The global relation \( f \) describes a channel from \( A \) to \( B \). (b) The input to party \( A \) is, as defined by the global relation \( g \), identical to the output from party \( B \), and the input to party \( B \) is identical to the output from party \( A \).
A contradiction arises: No choice of bits $a$ and $b$ exist that satisfy the global relation (15). The global relation (14), which is depicted in Figure 13, is logically inconsistent.

We show that there exist logically consistent global relations that are non-causal [7]. Suppose we are given three parties $A$, $B$, and $C$ with universal local operations. There exist global relations where the input to any party is a function of the outputs from the remaining two parties. An example [10] of such a global relation is

$$x = \neg b \land c, \quad y = a \land \neg c, \quad z = \neg a \land b,$$

(16)

where all variables represent bits, and where $x, y, z$ is the input to $A, B, C$ and $a, b, c$ is the output from $A, B, C$, respectively. This global relation can be understood as follows: Depending on the majority of the output bits, the relation either describes the identity channel from $A$ to $B$ to $C$, and back to $A$, or it describes the bit-flip channel from $A$ to $C$ to $B$, and back to $A$ (see Figure 15). We

study the causal relations that emerge from $n (\rightarrow \infty)$ sequential repetitions of this global relation, i.e., infinite strings that satisfy the global relation (16). The input to party $A$ is uncomputable, even $K(A) \not= 0$, because some bit positions of the outputs from $B$ and $C$ are uncomputable. Yet, the outputs from $B$ and $C$ completely determine the input to $A$, i.e., $K(A|B, C) \approx 0$. Therefore, the causal relation $(B, C) \preceq A$ holds. Due to symmetry, the causal relations $(A, C) \preceq B$ and $(A, B) \preceq C$ hold as well. All together imply that every party is in the causal future of some other parties — the scenario is non-causal. On the other hand, it is logically consistent: There exist input and output bit strings that satisfy the global relation (16) at every bit-position.

In the probability view, there exists an example of a randomized process that results in non-causal correlations [7], shown in Figure 16. In every run, the process models with probability $1/2$ the clockwise identity channel or the clockwise bit-flip channel. In the probabilistic view, both channels appear with equal probability. This leads to every party’s inability to influence its past. For instance, if parties $B$ and $C$ copy the input to the output and party $A$ has $a$ on the output, then $A$ has a random bit on the input — party $A$ cannot influence its past, and the grandfather antinomy does not arise. If, however, the probabilities of the mixture are altered slightly, a contradiction arises [10]. Furthermore, the process from Figure 16 cannot be embedded into a process with more inputs and outputs such that the larger process becomes deterministic and remains logically consistent [8]. Since in the view studied here, we look at single runs without probabilities — either of the global relations from the left or from the right channel must hold. Thus, if all parties use the universal local operation, a contradiction always arises, showing the inconsistency of the process.

Discussion. In Section I we saw that one consequence of dropping the notion of an a priori causal structure is that randomness becomes hard to define. Thus, we are forced to take the “factual-only view”: No probabilities are involved. Here, we show two facts that we formulate without considering counterfactuals. The first is that causal relations among parties can be derived by considering fixed bit strings only, without the use of the probability language. These causal relations are an inherent property of the bit strings of the parties. In other words, these strings are understood to be logically prior to the causal relations (just as in Section VI). The second consequence is that the causal relations that stem from certain strings can describe non-causal scenarios. This means that logical consistency does not imply a causal scenario: Causality is strictly stronger than logical consistency.

VIII. CONCLUSIONS

Whereas for Parmenides of Elea, time was a mere illusion — “No was nor will, all past and future null” — Heraclitus saw space-time as the pre-set stage on which his play of permanent change starts and ends. The follow-up debate — two millennia later and three centuries ago — between Newton and Leibniz about as how fundamental space and time, hence, causality, are to be seen was decided by the course of science in favor of Newton: In this view, space and time can be imagined as fundamental and given a priori. (This applies also to relativity theory, where space and time get intertwined and dynamic but remain fundamental instead of becoming purely relational in the sense of Mach’s principle.) Today, we have more reason to question a fundamental causal structure — such as the difficulty of explain-
ing quantum non-local correlations according to Reichenbach’s principle. So motivated, we care to test refraining from assuming space-time as initially given; this has a number of consequences and implications, some of which we address in this text.

When causality is dropped, the usual definitions of randomness stop making sense. Motivated by this, we test the use of intrinsic, context-independent “randomness” measures such as a string’s length minus its (normalized) fuel value. We show that under the Church-Turing hypothesis, Kolmogorov complexity relates to this value. We argue that with respect to quantum non-locality, complexity allows for a reasoning that avoids comparing results of different measurements that cannot all be actually carried out, i.e., that is not counterfactual. Some may see this as a conceptual simplification. It also leads to an all-or-nothing flavor of the Church-Turing hypothesis: Either no physical system can generate uncomputable sequences, or even a single photon can. Finally, it is asked whether logical reversibility is connected to the second law of thermodynamics — interpreted here in complexities and independent of any context expressed through probabilities or ensembles — and potentially to the arrow of time, past and future. Finally, we have speculated that if a causal structure is not fundamental, how it may emerge from data-compressibility relations.

When causality is dropped, one risks antinomies. We show, in the complexity-based view, that sticking to logical consistency does not restore causality but is strictly weaker. This observation has recently been extended to computational complexity [11]: Circuits solely avoiding antinomies are strictly stronger than causal circuits.

ACKNOWLEDGMENTS

This text is based on a presentation at the “Workshop on Time in Physics,” organized by Sandra Rankovic, Daniela Frauchiger, and Renato Renner at ETH Zurich in Summer 2015.

The authors thank Mateus Araújo, Veronika Baumann, Charles Bédard, Gilles Brassard, Harvey Brown, Caslav Brukner, Harry Buhrman, Matthias Christandl, Sandro Coretti, Fabio Costa, Bora Dakic, Frédéric Dupuis, Paul Erker, Adrien Feix, Jürg Fröhlich, Nicolas Gisin, Esther Hänggi, Arne Hansen, Marcus Huber, Lorenzo Maccone, Alberto Montina, Samuel Ranellucci, Paul Raymond-Robichaud, Louis Salvail, L. Benno Salwey, Andreas Winter, and Magdalena Zych for inspiring discussions, and the Einstein Kaffee as well as the Reitschule Bern for their inspiring atmosphere. — Grazie mille!

The authors thank Claude Crépeau for his kind invitation to present this work, among others, at the 2016 Bellairs Workshop, McGill Research Centre, Barbados.

Our work was supported by the Swiss National Science Foundation (SNF), the National Centre of Competence in Research “Quantum Science and Technology” (QSIT), the COST action on Fundamental Problems in Quantum Physics, and the Hasler Foundation.

[1] S. Aaronson, [http://www.scottaaronson.com/blog/?p=762], 2012.
[2] A. A. Abbott, C. S. Calude, J. Conder, K. Svozil, Strong Kochen-Specker theorem and incomputability of quantum randomness, Physical Review A, Vol. 86, pp. 062109, 2012.
[3] P. K. Aravind, Bell’s theorem without inequalities and only two distant observers, Foundations of Physics Letters, Vol. 15, No. 4, pp. 397-405, 2002.
[4] T. J. Barnea, J.-D. Bancal, Y.-C. Liang, N. Gisin, Tripartite quantum state violating the hidden influence constraints, Physical Review A, Vol. 88, pp. 022123, 2013.
[5] J.-D. Bancal, S. Pironio, A. Acín, Y.-C. Liang, V. Scarani, N. Gisin, Quantum non-locality based on finite-speed causal influences leads to superluminal signalling, Nature Physics, Vol. 8, pp. 867-870, 2012.
[6] J. Barrett, L. Hardy, A. Kent, No-signalling and quantum key distribution, Physical Review Letters, Vol. 95, pp. 010503, 2005.
[7] A. Baumeler, A. Feix, S. Wolf, Maximal incompatibility of locally classical behavior and global causal order in multi-party scenarios, Physical Review A, Vol. 90, pp. 042106, 2014.
[8] A. Baumeler, S. Wolf, Perfect signaling among three parties violating predefined causal order, Proceedings of IEEE International Symposium on Information Theory 2014, pp. 526–530, IEEE, Piscataway, 2014.
[9] A. Baumeler, F. Costa, T. Ralph, S. Wolf, in preparation.
[10] A. Baumeler, S. Wolf, The space of logically consistent classical processes without causal order, New Journal of Physics, Vol. 18, pp. 013036, 2016.
[11] A. Baumeler, S. Wolf, Non-causal computation avoiding the grandfather and information antinomies, arXiv preprint, arXiv:1601.06522 [quant-ph], 2016; accepted for publication in New Journal of Physics, 2016.
[12] J. S. Bell, On the Einstein-Podolsky-Rosen paradox, Physics, Vol. 1, pp. 195–200, 1964.
[13] C. H. Bennett, The thermodynamics of computation, International Journal of Theoretical Physics, Vol. 21, No. 12, pp. 905–940, 1982.
[14] C. H. Bennett, Logical reversibility of computation, IBM Journal of Research and Development, Vol. 17, No. 6, pp. 525–532.
[15] G. Brassard, A. Broadbent, and A. Tapp, Quantum pseudo-telepathy, arXiv preprint, arXiv:quant-ph/0407221, 2004.
[16] G. Chaitin, A theory of program size formally identical to information theory, Journal of the ACM, Vol. 22, pp. 329-340, 1975.
[17] R. Cilibrasi and P. Vitányi, Clustering by compression, IEEE Transactions on Information Theory Vol. 51, No. 4.
