Talk proposal

Mapping Temporal Correlations to Contextuality Correlations

Amy Searle[1], Rui Soares Barbosa[2], and Samson Abramsky[3]

[1]Department of Physics, University of Oxford, U.K.
[2]INL – International Iberian Nanotechnology Laboratory, Braga, Portugal
[3]Department of Computer Science, University College London, U.K.

Overview

We are interested in a setup in which a sequence of measurements is performed on a single system, thereby producing a sequence of corresponding measurement outcomes. Measurements that can be performed on this system are collected into a set X, and the output of measurement m is taken to lie in the set O_m. Writing X^* for the set of all finite sequences of elements of X, we collect the sequences that can be performed on the system into a down-closed subset $\Sigma \subseteq X^*$ (with respect to the prefix order on X^*). We group this data together into a tuple $\mathcal{M} = \langle X, \Sigma, O \rangle$, which is called a **temporal measurement scenario**. A pair $\langle m, o \rangle$ with $m \in X$ and $o \in O_m$ is called a measurement event, and we write $\Delta_{\mathcal{M}}$ for the set of all measurement events associated with $\mathcal{M}$.

In contextuality scenarios, which are a spatial analogue of the current setup, there is a clear way in which to define *deterministic classicality* of a system: it corresponds to a measurement $m \in X$ being assigned the same outcome regardless of which other measurements are performed simultaneously; see e.g. the sheaf-theoretic treatment of contextuality [1]. For probabilistic systems, classicality in a sheaf approach sense is synonymous with the existence of a probabilistic local hidden variable model. That the assumptions of classicality hold experimentally can be guaranteed for Bell scenarios by appealing to special relativity, which has resulted in 'loophole-free' tests of Bell inequalities.

In the temporal case, assumptions of classicality are more subtle. The outcome of a measurement $m \in X$ may in principle depend on any of the previously performed measurements and their outcomes. So, a deterministic behaviour of the temporal system is instead described by a function

$$s : \Sigma \to O :: \sigma \mapsto o$$

which sends each sequence $\sigma \in \Sigma$ to an outcome $s(\sigma)$. The interpretation is that the last measurement of σ obtains the outcome $s(\sigma)$ when performed in the sequence σ, allowing for outcomes to depend on the whole history.[1] We refer to such functions as **global strategies**, and write $\mathcal{S}(\Sigma)$ for the set of all global strategies over Σ.

In practice, unbounded signalling of information from the past is not realistic and is limited by the **memory** of the system. For finite state machines—a paradigm of sequential computation—this limitation comes from the size of the state set. It therefore makes sense in this temporal setting to speak of classicality with respect to some memory bound. We consider a model of memory in which the system can only store a subset of the past measurements and their outcomes. We focus on three special cases, parametrised by for a fixed $k \in \mathbb{N}$:

1. the system remembers the k immediately preceding measurements;
2. the system remembers up to k of the previous measurements;
3. the system remembers the k immediately preceding measurement events (i.e. measurements and outcomes).

For example, given a sequence $\sigma = m_0 \dots m_n$, define $\mathsf{lookback}_k(\sigma) := m_{\max\{n-k+1,0\}} \dots m_n$ to be its suffix of length $k+1$ (if it exists). Item 1 requires that any strategy $s \in \mathcal{S}(\Sigma)$ satisfy

$$\mathsf{lookback}_k(\sigma) = \mathsf{lookback}_k(\sigma') \implies s(\sigma) = s(\sigma').$$

We say that a strategy is $\mathsf{lookback}_k$-consistent if it satisfies this property. We write $\mathcal{S}_{\mathsf{lookback}_k}(\Sigma)$ for the set of $\mathsf{lookback}_k$-consistent strategies on Σ, noting that for any k it holds that

$$\mathcal{S}_{\mathsf{lookback}_k}(\Sigma) \subseteq \mathcal{S}_{\mathsf{lookback}_{k+1}}(\Sigma).$$

[1]Here, outcomes only explicitly depend on prior measurement choices and not on their outcomes. However, adding dependence on previous outcomes for s would be pointless: if a system is responding deterministically, knowing the full history of measurements is enough to reconstruct all past outcomes, and so they need not be included explicitly.

We similarly define functions $\Theta_k \colon \Sigma \to X^*$ and $\mathsf{lookback}_k^{(e)} \colon \Sigma \times \mathcal{S}(\Sigma) \to \Delta^*_{\mathcal{M}}$ such that a strategy acting according to item 2 is Θ_k-consistent, and according to 3 is $\mathsf{lookback}_k^{(e)}$-consistent (the superscript indicating that it is the past k measurements *events*). We denote the associated subsets of strategies $\mathcal{S}_{\Theta_k}(\Sigma)$ and $\mathcal{S}_{\mathsf{lookback}_k^{(e)}}(\Sigma)$.

Now, given a measurement scenario $\mathcal{M}$, performing a particular sequence of measurements $\sigma \in \Sigma$ produces a probability distribution e_σ on corresponding output sequences, i.e. on assignments of outcomes to the set $\downarrow \sigma$ of prefix sequences of σ. The collection $e = (e_\sigma)_{\sigma\in\Sigma}$ is called an **empirical model**.

An empirical model which for each σ has support only on the $\mathsf{lookback}_k$-consistent strategies is said to be $\mathsf{lookback}_k$-consistent, and we write $\mathsf{EM}^{(\mathsf{lookback}_k)}(\mathcal{M})$ for the convex set of such models. We can make a similar definition based on Θ_k and $\mathsf{lookback}_k^{(e)}$. Note that Θ_k-consistent empirical models are restricted to use the available memory according to the fixed choice of function Θ_k. More generally we would like to describe systems which may respond with any strategy that stores at most k of the past measurements. To this end, we say that a model is L_k-consistent if every strategy in its support is Θ_k-consistent for some Θ.

So, in general, we write $\mathsf{EM}^{(F)}(\mathcal{M})$ for the set of F-consistent empirical models with $F \in \{\mathsf{lookback}_k, L_k, \mathsf{lookback}_k^{(e)}\}$. The function F introduces additional no-signalling constraints for the empirical model e, which go beyond the usual arrow of time constraints and are not captured by restrictions of the presheaf $\mathcal{D}_R \circ \mathcal{S}_F$. Such constraints reflect the fact that signalling from the past has been in some way restricted.

While in the spatial case classicality is synonymous with the existence of a local hidden variable theory, here we take classicality of an F-consistent empirical model to mean realisability by a classical machine E_F that produces F-consistent strategies according to some probability distribution on global strategies $h \in \mathcal{D}_R \circ \mathcal{S}_F(\Sigma)$. We say that e is **E_F-classical.**[2]Thus for e E_F-nonclassical, *local* probability distributions are consistent with those that E_F produces, but there is no extension of these to a consistent *global* distribution (as with contextuality). Note that defining classicality with respect to memory bounds on a classical machine, as captured by a restriction on strategies F, avoids hypothesising about what it means for temporal correlations to be classical (as in Leggett and Garg's macrorealistic assumptions, which suppose that noninvasiveness should hold). This circumvents a debatable philosophical issue by appealing to resource-theoretic notions, where in this case the resource of interest is memory.

An advantage of casting the setup in a sheaf-theoretic framework is that when the system stores only measurements in memory, i.e. in the cases $F = \mathsf{lookback}_k$ or $F = L_k$, we are able to map temporal measurement scenarios to a particular type of contextuality setup, in which correlations are no longer temporal but spatial. Nonclassical temporal empirical models are then the pullback of contextual empirical models on this image scenario. We first show that strategies $f \in \mathcal{S}_F(U)$ are in one-to-one correspondence with sections $s \in \mathcal{E}(\mathcal{C}_F(U))$, where

$$\mathcal{C}_F(U) := \{F(\sigma) | \sigma \in U\}.$$

Calling this bijection α, then given an empirical model $\{w_C\}_{C\in\Sigma_{\mathcal{C}_F(\mathcal{M})}}$ we define a temporal empirical model

$$e_C(f) := w_{\mathcal{C}_F(C)}(\alpha(f)).$$

This forms the first theorem of the paper, stated below.

Theorem 1. *For $F \in \{\mathsf{lookback}_k, L_k\}$ there is a map $\mathcal{C}_F$ from temporal measurement scenarios to contextuality measurement scenarios, such that empirical models $w \in \mathsf{EM}(\mathcal{C}_F(\mathcal{M}))$ can be pulled back via $\mathcal{C}_F^*$ to F-consistent empirical models $\mathcal{C}_F^* w \in \mathsf{EM}^{(F)}(\mathcal{M})$ on the temporal measurement scenario $\mathcal{M}$. This map preserves and reflects nonclasicality, meaning that an empirical model w on $\mathcal{C}_F(\mathcal{M})$ is contextual if and only if $\mathcal{C}_F^* w$ is E_F-nonclassical.*

Vorob'ev's theorem and quantum advantage

The advantage of mapping an empirical model of the temporal type to one of the contextual type is that the latter have been extensively studied in the sheaf-theoretic approach. Therefore, there are a number of well-developed tools we can readily utilise on contextuality scenarios. Crucially, as the map both preserves and reflects contextuality, it can be used to transfer results, allowing us to say something in turn about the classicality of empirical models in the original temporal scenario.

With this in mind, we recall an application of Vorob'ev's Theorem, originating in game theory [3], to contextuality measurement scenarios, which was studied in [4]. This theorem characterises the contextuality measurement scenarios that admit contextual empirical models.

[2]An E_F machine is for example a finite state machine which uses its state set to store past measurements and outcomes. Note that contextuality and finite state machines has been studied too in [2].

Theorem 2 ([4, 3])**.** *Let $\mathcal{M} = \langle X_\mathcal{M}, \Sigma_\mathcal{M}, O_\mathcal{M}\rangle$ be a contextuality measurement scenario where Σ is a simplicial complex representing the compatibility relation of elements in X. Then all empirical models defined on $\mathcal{M}$ are non-contextual if and only if $\Sigma_\mathcal{M}$ is acyclic.*

Acyclicity here means that one can remove the measurements from the scenario one by one in such a way that the removed measurement at each stage belongs to a single maximal context, i.e. all the measurements compatible with it are jointly compatible.

The following corollary follows straightforwardly from Theorems 1 and 2.

Corollary 3. *Let $\mathcal{M}$ be a temporal measurement scenario. Every F-consistent empirical model on $\mathcal{M}$ is E_F-classical if and only if the corresponding contextual measurement scenario $\mathcal{C}_F(\mathcal{M})$ has acyclic simplicial complex $\Sigma_{\mathcal{C}_F(\mathcal{M})}$.*

Note that we obtain both directions of Vorob'ev's Theorem, so that a non-acyclic $\Sigma_{\mathcal{C}_F(\mathcal{M})}$ implies that there exists an E_F-nonclassical empirical model $e \in \mathsf{EM}^{(F)}(\mathcal{M})$.

Although the constructed map breaks down when memory storage of outputs is allowed, we are nevertheless able to show that $\Sigma_{\mathcal{C}_{\mathsf{lookback}_k}(\mathcal{M})}$ not acyclic also implies the existence of an empirical model that is $\mathsf{lookback}_k^{(e)}$-consistent but $\mathsf{E}_{\mathsf{lookback}_k^{(e)}}$-nonclassical. This is done by constructing an empirical model which is deterministic at certain measurements, so that allowing for strategies which store outputs of these measurements is not more advantageous. This extends one direction of Corollary 3 to $\mathsf{lookback}_k^{(e)}-$consistent strategies.

Theorem 4. *Let $\mathcal{M}$ be a temporal measurement scenario. If the measurement scenario $\mathcal{C}_{\mathsf{lookback}_k}(\mathcal{M})$ has a non-acyclic simplicial complex, then there exists a $\mathsf{lookback}_k^{(e)}$-consistent empirical model on $\mathcal{M}$ which cannot be generated by a classical machine $\mathsf{E}_{\mathsf{lookback}_k^{(e)}}$.*

Conclusion and Future Work

The assumption underlying the F-consistent models we consider is that the system at any given time can only remember some subset of the past measurements that have been performed and outputs obtained (as specified by F). In our paper we show that

1. When only measurements are stored, we can map temporal scenarios to contextuality scenarios via $\mathcal{C}_F$ in way that preserves and reflects nonclassicality. This allows the identification of empirical models which have *local* support on deterministic strategies that a classical machine E_F can produce, but which are inconsistent with any *global* distribution on such strategies. We call this behaviour E_F-nonclassial.
2. This map does not work when outputs are stored. Nevertheless,
   (a) An empirical model generated by a classical machine $\mathsf{E}'_{F'}$ that can store past measurements and their outcomes can be realised by a machine E_F that stores only measurements if one increases the allowed number of stored measurements. We analyse this trade-off for certain examples of empirical models.
   (b) We can for any measurement scenario which admits $\mathsf{E}_{\mathsf{lookback}_k}$-nonclassical empirical models find at least one empirical model which is also $\mathsf{E}'_{\mathsf{lookback}_k^{(e)}}-$nonclassical.

In future work we hope to extend these notions to causal scenarios studied in [5], including the causal Bell scenarios studied in [6].

References

[1] S. Abramsky and A. Brandenburger, "The sheaf-theoretic structure of non-locality and contextuality," *New Journal of Physics*, vol. 13, p. 113036, Nov. 2011.

[2] C. Budroni, "Contextuality, memory cost and non-classicality for sequential measurements," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 377, no. 2157, p. 20190141, 2019.

[3] N. N. Vorob'ev, "Consistent Families of Measures and Their Extensions," *Theory of Probability & Its Applications*, vol. 7, no. 2, pp. 147–163, 1962.

[4] R. Barbosa, "Contextuality in quantum mechanics and beyond," 2015.

[5] S. Abramsky, R. S. Barbosa, and A. Searle, "Combining contextuality and causality: a game semantics approach," 2023.

[6] S. Gogioso and N. Pinzani, “The sheaf-theoretic structure of definite causality,” *Electronic Proceedings in Theoretical Computer Science*, vol. 343, pp. 301–324, sep 2021.

Mapping Temporal Correlations to Contextuality Correlations

Amy Searle[1], **Rui Soares Barbosa**[2] **and Samson Abramsky**[3]

1. Department of Physics, University of Oxford, Oxford, U.K.
2. INL – International Iberian Nanotechnology Laboratory, Braga, Portugal
3. Department of Computer Science, University College London, London, U.K.

E-mail: `amy.searle@physics.ox.ac.uk`

Abstract. We study temporal correlations in a simple setup using a sheaf-theoretic framework in the spirit of Abramsky and Brandenburger's sheaf-theoretic approach to contextuality [3]; the setup consists of a system and a set of measurement sequences which are performed in turn on the system, thereby producing outcomes with a particular probability. Nonclassicality then coincides with obstructions to the existence of a global section—or of a global strategy, as we will prefer to call them—as in the case of contextuality. We limit the amount of information that can be signalled; in particular, when a deterministic system is measured the outcome can only depend upon k of the previous measurement choices (which for Bell scenarios becomes k previous measurement settings). This can be thought of as a restricted type of memory, and we show that in the case $k = 0$ we recover the macrorealistic correlations which were advocated, for example, by Leggett and Garg, and that the case $k = \infty$ coincides with a system with unbounded memory as considered by Gogioso and Pinzani. We then show that for each k, there exists a map from the temporal setup to a contextuality setup, such that every nonclassical temporal empirical model corresponds to a contextual empirical model. Doing so allows us to utilise a result from Vorob'ev', originating in the realm of game theory, in order to say, for any measurement scenario and choice of k, whether nonclassical correlations can arise. We show moreover that this last result holds also when allowing a system to store k previous measurements *and* their outcomes.

1. Introduction

Contextuality is an important feature of quantum systems which distinguishes them from their classical counterparts. While distributions on measurement outcomes produced by a classical system are consistent with there being a fixed outcome for observables defined prior to measurement, with randomness arising due to imperfect state preparation, in quantum systems this does not hold. In the latter, measurement contexts are important and distributions on measurement outcomes become context specific, albeit in a way that the no-signalling principle is still satisfied. Such an observation has created a fundamental shift in the way quantum systems are to be viewed, but has also led to important discoveries concerning the rôle of contextuality in information processing tasks— not least of all, contextuality has been shown to be necessary for increasing the computational power in certain models of computation [6, 18]. The sheaf theoretic approach to contextuality has contributed to understanding this role by importing tools from sheaf theory; this had led to a number of important developments in the field, including amongst others the contextual fraction as a measure of contextuality [1], resource-theoretic notions and definitions [15, 5, 8], and a logical basis for Bell inequalities [4].

Nevertheless, the correlations which arise in contextuality scenarios are by assumption spatial, which limits the reach of applicability of these tools. Temporal correlations describe correlations which are time-like separated, such as those which arise with the repeated measurement of a single system at different time steps. While in the spatial case all correlations are required to satisfy no-signaling (or more generally no-disturbance), in temporal setups correlations instead obey the arrow of time (AoT) constraints, which allow for signaling from the past to the future but not in the opposite direction.

The consideration of temporal correlations is not new, and, in analogue with the spatial case in which contextuality inequalities are ubiquitous[16], inequalities can be derived which separate classical and quantum correlations in such setups, with violations of these inequalities corresponding to a quantum advantage in certain information processing tasks [9, 22]. However, although there are many similarities with results on spatial correlations, the assumptions of classicality are different, and of a more subtle kind. Most notably, Leggett and Garg supposed that macro-realism was a valid axiom of classicality in the temporal regime, however loophole free violation of inequalities derived under these assumptions have so far not been possible. The subtlety arises because in principle a system responding to measurements sequentially can store an unbounded amount of information from the past, making it important to carefully consider different models of classicality and the classical resources they require. In short, in the temporal case quantum advantage in certain information processing tasks is always relative to some bounds on the classical system, such as memory bounds [12, 11, 20, 10].

Here, we cast into the sheaf theoretic framework a setup in which sequential measurements are performed on a single system. We study how placing restrictions on the amount of information which can be signaled from the past alters the correlation polytope, by allowing a signaling window of specified depth k. In the deterministic case this amounts to allowing the system to store a subset of past measurements, so that measurement outcomes may depend on some of the information from the past but not all. In the two opposite extreme limits, corresponding to the most and least restrictive cases of no and full forward signaling, we show that the produced correlations coincide with two cases already studied in the literautre. Namely, the former limit corresponds to Leggett and Garg's macrorealistic correlations [17] and the latter to those correlations studied by Gogioso and Pinzani in [14].

We map the temporal setups under consideration to contextuality setups, where the contextuality setup obtained will depend explicitly on the choice of signalling depth k.

Doing so allows us to invoke a Theorem due to Vorob'ev [21], originating in the field of game theory, to understand whether there exists any separation between the classical and AoT correlation polytopes which the temporal setup at hand produces—note that both are required to satisfy the additional k-dependent no-signaling conditions. The separation between the two depends explicitly on the combinatorial structure of Σ and choice of k, and forms the main Theorem of the paper. This allows us to deduce that if maximal signaling from the past is allowed every temporal correlation can arise classically. Moreover, for temporal Bell setups we give a sufficient and necessary condition, phrased in terms of a combinatorial condition relating the number of parties N in the setup, k, and the distance between agents with more than two inputs, for nonclassicality to arise.

Next, we relate classicality as defined in sheaf-theoretic terms to realisability by input-output machines, of which finite state machines are a particular type. Finite state machines (FSMs) are a paradigm of sequential computation, and we are able to derive an upper bound on the size of the state set of an FSM needed to generate all correlations corresponding to a particular setup.

Lastly, we consider more general models of memory. Since the signaling restrictions come about by requiring that any measurement in a sequence depend only on k of the previous measurements, there is no dependency on outcomes. Nevertheless we (i) discuss how allowing the system to store more measurements can offset the advantage of storing outcomes and (ii) show that our main Theorem extends to the case where there is dependency on k previous measurements and outputs.

2. The Setting

We study scenarios in which a sequence of measurements is performed on a single system, thereby producing a sequence of corresponding measurement outcomes. Measurements that can be performed on the system are drawn from a finite set X, and the output of each measurement is assumed to lie in a finite set O. The choice of sequence may vary but is not allowed to be arbitrary. We specify the allowed sequences by collecting them in a prefix-closed subset $\Sigma \subseteq X^*$, so that each element $\sigma \in \Sigma$ is a sequence of measurements from X. Here, X^* denotes the set of all finite sequences which can be constructed from X, and we order elements in X^* via the prefix relation; that is, if τ is a prefix sequence of σ, then $\tau \leq \sigma$ in X^*. We indicate the concatenation of two sequences $\sigma_1, \sigma_2 \in X^*$ by $\sigma_1\sigma_2$ and let $\sigma[i]$ and $\sigma[-i]$ be the ith measurement in σ indexing from the beginning and end of the sequence respectively. Finally, each measurement is taken to have output lying in the set O, which for now we take to be the same for every measurement.

We summarise this by grouping data together into a tuple $\mathcal{M} = \langle X, \Sigma, O\rangle$, which is called a **temporal measurement scenario**.

A **measurement event** is a pair (x, o), where $x \in X$ is a choice of measurement and $o \in O$ the outcome obtained upon measurement. Therefore in every round of measurement on the system $\mathcal{T}$, a sequence of measurement events is obtained. Concretely, to a temporal measurement scenario $\mathcal{M}$ we associate a **set of histories** $\mathcal{H}(\mathcal{M})$, defines as follows

$$\mathcal{H}(\mathcal{M}) := \{(m_0, o_0)(m_1, o_1)...(m_n, o_n)|m_0 m_1 ... m_n \in \Sigma \ \wedge \ \forall i.\, o_i \in O_i\}. \tag{1}$$

An element $h \in \mathcal{H}(\mathcal{M})$ consists of a sequence of measurement events which could be observed in a round of measurement.

A subset of measurement scenarios are produced by considering agents which sequentially perform a measurement on the system, with each agent only being allowed to perform a measurement from a specified set. This is described by a set of agents, $\mathcal{A}$, together with a linear order on that set, and for each agent $a \in \mathcal{A}$ an input set I_i and an output set O_i. We call such scenarios **temporal Bell scenarios**, to match the terminology in the spatial case ‡.

Example 2.1 (Temporal Bell Scenarios)**.** *Consider a linearly ordered set of agents $\mathcal{A}$. In each round of experiment, agent $a \in \mathcal{A}$ performs the measurement indicated by their input $i \in I_a$. We therefore define the set of measurements*

$$X := \{(a, i)|a \in \mathcal{A} \wedge i \in I_a\}. \tag{2}$$

Numbering the agents from 1 to N according to the linear order, the set of sequences which can be performed is

$$\Sigma := \{(1, i_1)(2, i_2)...., (N, i_N)|\forall j \in [1, N]\, .i_j \in I_j\}. \tag{3}$$

The scenario for two agents $\mathcal{A} = \{A, B\}$ with $A < B$ is known as the temporal CHSH setup and has been considered in detail in [13].

The scenarios we have so far considered are theory independent. Usually, however, we are interested in quantum scenarios as a special case; in this case measurements are taken to be projectors on a Hilbert space.

Example 2.2 (Quantum Temporal Measurement Scenario)**.** *Let $\mathbf{M}$ be a list of Projective Valued Measurements, or PVMs for short. Then a triple $\langle \mathbf{M}, \Sigma, O\rangle$ is a quantum temporal measurement scenario, with Σ denoting which sequences of PVMs can be performed.*

We can for example consider the temporal CHSH setup [13] in this framework as follows.

The set of PVMs consists of four measurements,

$$\mathbf{M} = \{M_{(a,0)}, M_{(a,1)}, M_{(b,0)}, M_{(b,1)}\}. \tag{4}$$

Sequences which can be performed are given by

$$\Sigma = \{M_{(a,0)}, M_{(a,0)}M_{(b,0)}, M_{(a,0)}M_{(b,1)}, M_{(a,1)}, M_{(a,1)}M_{(b,0)}, M_{(a,1)}M_{(b,1)}\}, \tag{5}$$

which note is a prefix-closed subset of $\boldsymbol{M}^*$.

‡ These are the scenarios considered in Gogioso and Pinzani in [14] where the poset on agents is taken to be a linear order

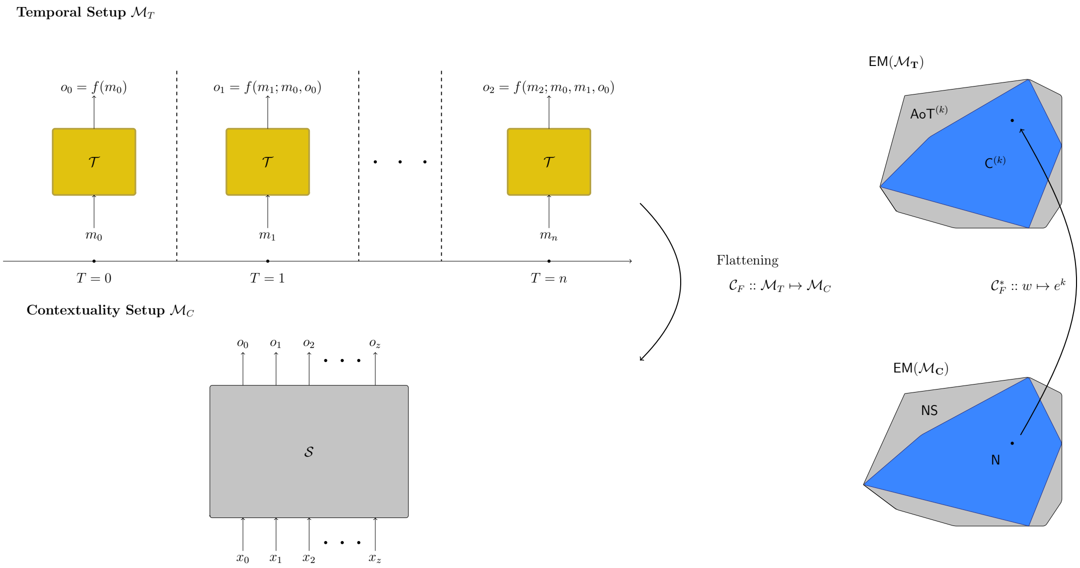


Figure 1: *Upper left.* A system $\mathcal{T}$ undergoing a sequence of measurements. At time step i measurement m_i is performed, thereby producing an outcome o_i which is a function of the measurement and any previous measurements and outcomes. Dependence on previous measurement events requires that $\mathcal{T}$ have some form of memory it can access, as shown by the inset box. *Lower left.* A commuting set of measurements x_1 through to x_z are performed at a single instant in time on the system $\mathcal{T}$, thereby producing a corresponding set of outcomes. The outcome o_i is a function only of measurement m_i due to measurements being performed simultaneously. Temporal setups with certain types of memory can be 'flattened' to contextuality setups under a map of measurement scenarios $\mathcal{C}_\mathsf{F}$. *Right.* Given a temporal measurement scenario $\mathcal{M}$, correlations on the flattened scenario $\mathcal{C}_\mathsf{F}(\mathcal{M})$ can be pulled back to correlations on $\mathcal{M}$. In particular, all correlations in the contextual (noncontextual) polytope of $\mathcal{C}_\mathsf{F}(\mathcal{M})$ can be pulled back to nonclassical (classical) correlations on $\mathcal{M}$.

2.1. Classicality in the Setup

The measurement correlations that arise from a measurement scenario $\mathcal{M}$ together with a choice of system $\mathcal{S}$ will depend on whether the type of system of hand— that is, whether is it acting in accord with the laws of classical physics, quantum physics, or, more generally, any theory which obeys the appropriate no-signalling constraints.

First think about how a deterministic system behaves: in a single run of the experiment, a sequence of measurements $\sigma = m_0 m_1 ... m_i ... m_n$ is mapped to a sequence of outcomes.

To this end define a function

$$f : \Sigma \to O :: \sigma \mapsto o \tag{6}$$

such that if one inputs the sequence $\sigma = m_0 ... m_n$ to the system $\mathcal{T}$ then the system returns the sequence of outputs $f(m_0) f(m_0 m_1) ... f(m_0 ... m_n)$. Note that here the function is defined so that $f(\sigma)$ returns an outcome only for the final measurement in σ. Therefore, measurements are mapped to outcomes conditional on the previous measurements which have been made.

Given f, we define a function f^* on $\sigma = m_0...m_n$ in Σ to be

$$f^*(m_0...m_n) := f(m_0)f(m_0m_1)...f(m_0...m_n), \tag{7}$$

so that f^* instead returns for a choice of sequence of measurements a corresponding sequence of outcomes. In certain cases it will be convenient to use this representation.

Let $\mathcal{S}(\Sigma)$ denote the set of all possible functions as in Equation 6, which can be thought of the set of all (deterministic) behaviours of the system.

Then for any $U \in \mathcal{P}(\Sigma)$ § we can let $\mathcal{S}(U)$ be the set of all such functions on U i.e. any $f \in \mathcal{E}(U)$ is a function

$$f : U \to O :: (\sigma \in U) \mapsto o. \tag{8}$$

We refer to elements of $\mathcal{S}(U)$ as **strategies over U**, so that the set $\mathcal{S}(\Sigma)$ contains the **global strategies**.

Naturally, strategies over a set $U \in \mathcal{P}(\Sigma)$ can be restricted to a subset $U' \subset U$ with $U' \in \mathcal{P}(\Sigma)$, because if one knows how sequences are mapped in U then one knows how they are mapped in only a subset of those sequences. We will write this as $\mathcal{S}(U)|_{U'}$, where each $f \in \mathcal{S}(U)$ is restricted as $f|_{U'} :: (\sigma \in U') \mapsto (o \in O)$ where $f|_{U'}(\sigma) := f(\sigma)$—note that $\sigma \in U'$ implies that $\sigma \in U$, so this is a well-defined restriction map.

This restriction map satisfies two properties which make $\mathcal{S}$ into a type of functor known as a presheaf. These are that (1) For all $f \in \mathcal{S}(U)$, $f|_U = f$ and (2) For sets of sequences where $V \subset W \subset U$ it always holds that $f|_W|_V = f|_V$.

In order to classify $\mathcal{S}$ as a sheaf—a more well behaved type of functor—an additional property must be checked. Take a subset $U \in \mathcal{P}(\Sigma)$ and an indexed set $\{U_i\}_{i\in I}$ which cover U so that $\bigcup U_i = U$. $\mathcal{S}$ is a sheaf if the two conditions below are met.

(i) *Existence of gluing.* Given a strategy $f_i \in \mathcal{S}(U_i)$ for all $i \in I$, such that $f_i|_{U_i\cap U_j} = f_j|_{U_i\cap U_j}$ for all $i, j \in I$, there exists a *global* strategy $f \in \mathcal{S}(U)$ with, for all $i \in I$, $f|_{U_i} = f_i$.

(ii) *Uniqueness of gluing.* If there exists two global strategies $f, g \in \mathcal{S}(U)$ such that $f|_{U_i} = g|_{U_i}$ for all $i \in I$, then $f = g$.

Since $\mathcal{S}$ does satisfy these properties, we may refer to $\mathcal{S}$ as a sheaf.

So far we have allowed for the system to have a lot of flexibility in the way in can respond to a measurement. By allowing a measurement to have different outcomes depending on **any** previous measurement made, we are allowing the system to store an arbitrary number of previous measurements— in other words, we are allowing the system to have an unbounded memory of a particular kind. However, as would usually be the case in real systems, we can suppose that this is restricted—for example, the

§ Here $\mathcal{P}$ denotes the powerset of Σ.

system might only be capable of storing the previous k measurements in memory at any given time.

To capture this, we introduce a function $\mathsf{lookback}_k : \Sigma \to X^*$ such that given a sequence $\sigma = m_0...m_n$ it is mapped under this function as $\mathsf{lookback}_k(m_0 m_1, ..., m_n) = m_{\max(0,n-k)}...m_n$.

The k-lookback therefore either returns the $k+1$-length tail of σ, or the entirety of σ, depending on which is longer. Note that the string that is returned is of length $k+1$ and not k—this is because we would like the lookback to refer to the k measurements *before* the measurement being performed in that timestep.

We introduce a property which we refer to as $\mathsf{lookback}_k$-consistency which a strategy $f \in \mathcal{S}$ may satisfy. A $\mathsf{lookback}_k$-consistent section is one which maps measurements which have had the k same measurement events before them to the same outcome.

Definition 2.1. *A strategy $f \in \mathcal{S}(U)$ is said to be $\mathsf{lookback}_k$-consistent if for every pair of sequences $\sigma, \tau \in \Sigma$ it holds that*

$$\mathsf{lookback}_k(\sigma) = \mathsf{lookback}_k(\tau) \implies f(\sigma) = f(\tau). \tag{9}$$

Requiring a system have support only on the $\mathsf{lookback}_k$-consistent strategies places a specific type of memory restriction on the system. We may think of the memory as having k 'spaceholders', which correspond to the k previous measurement made. Once a new measurement is made, a measurement is kicked off the 0 placeholder and each measurement moves down one space to make space for the latest measurement. This, at the moment, seems like a rather tight restriction to place on a system with memory, but we will see in what follows how this special case turns out to be useful for understanding more general cases.

Now for each set $U \in \mathcal{P}(\Sigma)$, we denote as $\mathcal{S}_k(U)$ the subset of $\mathcal{S}(U)$ containing only strategies which are $\mathsf{lookback}_k$-consistent. We obtain the following order of inclusions

$$\mathcal{S}_0 \subseteq \mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq .... \subseteq \mathcal{S}_\infty = \mathcal{S}. \tag{10}$$

This reflects that the smaller the value of k the less strategies there are which satisfy the k-consistency property, and in particular that if a strategy σ satisfies k-consistency then it automatically satisfies $(k+1)$- consistency. When we study distributions on these strategies, the extreme ends of k capture notions of classicality already studied in the literature: when k is maximal we recover the notion of classicality of Gogioso and Pinzani [14] (instantiated to a special case of linear causal order). On the other hand, when k is minimal, a deterministic classical system is a map from each measurement to an outcome, without any consideration of which measurements had been performed before. In this case, we capture the correlations produced by macro-realistic models, first introduced by Leggett and Garg [17] and subsequently studied in other contexts [13]. We give a more precise statement of these observations in a later section, but for now it serves as an initial justification of the k-lookback function; the lookback can be regarded as an interpolation between the two extreme ends of classicality which are studied in the literature.

2.2. Probabilistic Classical Systems and Empirical Models

We would like to study probabilistic systems, or equivalently allow a system to only probabilistically respond with a particular strategy. To deal with this, we precompose $\mathcal{S}_k$ with the distribution functor $\mathcal{D}_R$.

The distribution functor sends a set S to the set $\mathcal{D}_R(S)$ of probability distributions over the objects in S. For example, if $S = \{1, 2, 3, 4\}$, then an element $p \in \mathcal{D}_R(S)$ would be written

$$p = p_1 |1\rangle + p_2 |2\rangle + p_3 |3\rangle + p_4 |4\rangle \,, \tag{11}$$

such that $p_1 + p_2 + p_3 + p_4 = 1$. The set $\mathcal{D}_R(\mathcal{S}(U))$ therefore contains probability distributions on all strategies over the measurement set U.

By precomposing the $\mathsf{lookback}_k$-consistent sheaf of strategies $\mathcal{S}_k$ with $\mathcal{D}_R$, we obtain, for each set $U \subseteq \Sigma$ the set of all normalised probability distributions on the set of $\mathsf{lookback}_k$-consistent strategies on U. Moreover, we can restrict from a set U to a set V with $V \subseteq U$, to obtain a normalised set of probability distributions on V. Given a distribution $p \in \mathcal{D}_R \circ \mathcal{S}_k(U)$, we define, for each $f \in \mathcal{S}_k(V)$

$$p|V(f) := \sum_{g \in \mathcal{S}_k(U)} p(g). \tag{12}$$

2.3. Measurement Covers and Empirical Models

Contextuality is an essential feature of quantum systems. It relies on the noncommutativity of certain observables, thereby placing a restriction on which sets of observables, or measurements, can be observed in a single round of experiment. Analogously, in the current temporal setup only subsets of measurements which occur as a sequence in Σ can be observed in a single run. While the commutation relation is a fundamental constraint from Nature, here the constraint is experimenter imposed ‖. Nevertheless, just as in the spatial case, temporal inequalities can be derived which separate the classical and quantum correlation polytopes [11].

An empirical model is exactly the probability distributions which we observe upon measurement. Analogous to the case in contextuality setups, in order to define the empirical models we need a cover of $\mathcal{P}(\Sigma)$. We take $\{C_i\}_{i \in I}$ where each C_i consists of causally consistent sequences of measurements.

Definition 2.2 (Causally consistent sequences)**.** *Two sequences $\sigma, \tau \in \Sigma$ are said to be causally consistent if one is a prefix sequence of the other, so that either $\sigma \leq \mu$ or $\mu \leq \sigma$.*

A **$\mathsf{lookback}_k$-consistent empirical model** is one in which the observed probability distributions over each set C_i have support only in $\mathcal{S}_k(C_i)$. Note that this is only relevant when Σ contains sequences with repeated measurements—otherwise, the $\mathsf{lookback}_k$-consistency requirement becomes trivial on a single sequence $\sigma \in \Sigma$.

‖ In the temporal CHSH setup considered in [13], the quantum realisation corresponds to a single qubit undergoing a sequence of measurements, which do not need to commute.

Definition 2.3 ($\mathsf{lookback}_k$-Consistent Temporal Empirical Model)**.** *Given a temporal measurement scenario $\mathcal{M}$, a $\mathsf{lookback}_k$-consistent temporal empirical model consists of a choice $e_i^{(k)} \in \mathcal{D}_{\mathbb{R}} \circ \mathcal{S}_k(C_i)$ for each set C_i in the causally consistent cover $\{C_i\}_{i\in I}$.*

The Arrow of Time constraints, which forbid signalling from the future to the past, are automatically satisfied due to $\mathcal{D}_{\mathbb{R}} \circ \mathcal{S}_k$ being a presheaf. Concretely, for any sets C, C' in the cover with $C' \subseteq C$,

$$e_C^{(k)}|_{C'} = e_{C'}^{(k)}. \tag{13}$$

We now consider what it means for an empirical model generated in this temporal fashion to be nonclassical. In the spatial case noncontextuality is synonymous with the existence of a hidden variable model. Here, we do not wish to make any hypotheses about the nature in which physical systems produce outcomes when sequentially measured—quantum or otherwise. Macrorealism assumptions, for example, hypothesise about how a classical system should behave under measurement; namely that it should exist in a definite state at all times and measurements should be noninvasive. Here, instead, we wish to characterise quantum and AoT correlations, as captured by an empirical model, with regards to the types of classical machines which could produce them.

Let $\mathcal{M} = \langle X, \Sigma, O\rangle$ be a measurement scenario. We define a classical machine E to be a quadruple $\langle X, O, F, h\rangle$, where

(i) X is the input set

(ii) O is the output set

(iii) $F \subseteq \{f :: (\sigma \in \Sigma) \mapsto \prod_{m\in\sigma} O\}$, is the set of possible input-output functions over the set Σ of input strings

(iv) h is a probability distribution on the set F

The set F may be unconstrained, in which case the machine can produce, for input string $i_0 i_1 ... i_n$, with each $i_k \in X$, any corresponding output string $o_0 o_1 ... o_n$ with $o_i \in O$. Usually this set F is constrained in some way. For finite state machines, for example, in which the output is dependent only on the current input and current state, the size of the state set has an influence on which strings of outputs are producible.

A function $f :: (\sigma \in \Sigma) \mapsto \prod_{m\in\sigma} O$, corresponds to the starred version s^* of a strategy $s \in \mathcal{S}(\Sigma)$. Thus, we can define F to be the set of functions satisfying $\mathsf{lookback}_k$-consistency. In particular, if $f \in F$ maps $i_0 i_1 ... i_n \mapsto o_0 o_1 ... o_n$ then the $\mathsf{lookback}_k$-constraint corresponds to requiring that

$$o_k = g(i_k, i_{k-1}, ..., i_{\max(n-(k+1),0)}), \tag{14}$$

so that any output in the sequence is only a function of the previous k inputs as well as the current input.

We will write $\mathsf{E}_{\mathsf{lookback}_k}$ to represent an E-machine with F constrained in this way.

Definition 2.4 ($\mathsf{E}_{\mathsf{lookback}_k}$-classical)**.** *Given a measurement scenario $\mathcal{M}$, an empirical model $e \in \mathsf{EM}^{(k)}(\mathcal{M})$ is said to be $\mathsf{E}_{\mathsf{lookback}_k}$-classical if there exists an E-machine $\mathsf{E} = \langle X, O, F, h\rangle$ such that $h|C = e_C$ for all causally consistent subsets $C \subseteq \Sigma$.*

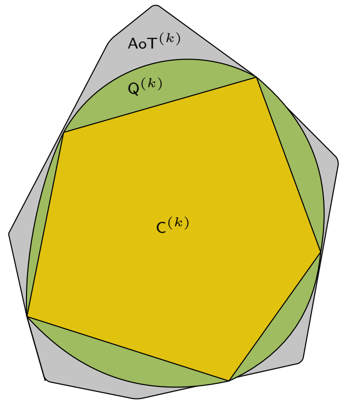


Figure 2: The convex sets of empirical models which are $\mathsf{lookback}_k$-consistent and $\mathsf{E}_{\mathsf{lookback}_k}$-classical (denoted $\mathsf{C}^{(k)}$), or $\mathsf{E}_{\mathsf{lookback}_k}$-nonclassical (denoted $\mathsf{Q}^{(k)}$ for the quantum realisable, while $\mathsf{AoT}^{(k)}$ denotes those correlations obeying AoT in addition to the k-lookback window constraints)

Note that all strategies in the support of e_C for each $C \in \mathscr{C}$ can be produced in a single run of $\mathsf{E}_{\mathsf{lookback}_k}$; however, this does not imply that e is $\mathsf{E}_{\mathsf{lookback}_k}$-classical.

In Figure **??** we give an abstract representation of the convex set of correlations which can be achieved under different assumptions. The largest set, the grey set labelled $\mathsf{AoT}^{(k)}$, represents those correlations which obey the AoT no-signaling constraints as well as the additional ones imposed by $\mathsf{lookback}_k$-consistency. The green set $\mathsf{Q}^{(k)}$ represents the subset of these which are quantum realisable. Finally, the yellow subset $\mathsf{C}^{(k)}$ is the set of $\mathsf{E}_{\mathsf{lookback}_k}$-classical correlations.

3. Contextuality and Vorob'ev's Theorem

We now give a brief recap of the sheaf theoretic approach to contextuality and non-locality introduced in [3]. It will be needed in order to introduce Vorob'ev's Theorem for contextuality setups.

In contextuality setups we instead have **contextuality measurement scenarios** $\mathcal{M} = \langle X, \Sigma, O\rangle$, where X again specifies the set of available measurements and O the set of outcomes. Here there is no dependence on time and Σ instead specifies the commutation relation of of the measurements in X, so that a face $\sigma \in X$ indicates a set of commuting measurements, which can be thought of as being performed at a single instant in time.

In this setting, a deterministic behaviour s over a set $U \subseteq X$ is simply a function $s : U \to O$, indicating that outcome $s(m)$ was observed when the measurement $m \in U$ is performed. These functions are referred to as **sections**. We write $\mathcal{E}(U)$ for the set of sections associated with the measurement set U. We will refer to $\mathcal{E}$ as the event sheaf, and its sheaf properties can be relatively easily checked.

An empirical model consists of a collection of probability distributions $\{e_\sigma\}_{\sigma\in\Sigma}$, where $e_\sigma \in \mathcal{D}_\mathbb{R}(\mathcal{E}(\sigma))$. Importantly, here distributions are defined on sets of measurements, while in the temporal case it was sets of sequences.

A non-contextual empirical model e is one for which there exists a global section $h \in \mathcal{D}_\mathbb{R}(\mathcal{E}(X))$ which marginalises to the empirical model at each measurement context:

$$e_\sigma(s) = \sum_{r \in \mathcal{E}(X), r|\sigma = s} h(r). \tag{15}$$

3.1. Vorob'ev's Theorem

The central result of this paper is using Vorob'ev's Theorem, a result originating in the field of game semantics [21] and which was first applied to the area of Quantum Information in [7, 19], to understand which measurement scenarios can result in empirical models in which there is some separation between classical and no-signalling correlation. Central to the Theorem is the compatibility structure of measurements as encoded by Σ— certain compatibility structures are able to host contextual empirical models, while it turns out that others cannot.

Vorob'ev's result, however, applied directly to temporal measurement scenarios, only gives the corresponding result for full lookback empirical models. This is because no information about the lookback depth is contained in Σ. We can get around this however by mapping $\mathcal{M}$ to a series of contextuality measurement scenarios indexed by k, such that Vorob'ev's result for empirical models can be extended to $\mathsf{lookback}_k$-consistent empirical models, too. We therefore recall the theorem for contextuality measurement scenarios $\mathcal{M}$.

We begin by introducing some terminology pertaining to simplicial complexes. A simplicial complex may posses a certain property called acyclicty, which can be thought of as the generalisation of acyclicty of graphs where it can intuitively be understood as the absence of cycles in the graph. It is useful to bear in mind that in our case K represents the compatibility of measurements, however Vorob'ev's Theorem applies more generally to any presheaf which is defined over a simplicial complex K.

We will give the algorithmic definition of acyclicity, in which a simplicical complex is acyclic if and only if we can remove vertices from it in a certain way so as to produce the zero complex $\Delta_0 := \{\emptyset\}$, which contains no vertices.

If vertices can be removed from the simplicial complex it is said to be Graham reducible.

Definition 3.1 (Graham reducible)**.** *We say that a simplicial complex K is Graham reducible if there exists a face $\Sigma \in MaxK$ which has vertex $v \in K$ such that the vertex v is not in any other maximal faces.*

The removal of the vertex v to form the simplicial complex $K' := \{\sigma | (\sigma \cup \{v\} \in K) \vee (\sigma \in K)\}$ is called a Graham reduction step. We write $K \rightsquigarrow K'$.

Definition 3.2 (Acyclic Simplicical Complex)**.** *A simplicial complex K is said to be acyclic if there is a series of Graham reduction steps $K \rightsquigarrow K' \rightsquigarrow ... \rightsquigarrow \Delta_0$, ending in the zero complex Δ_0.*

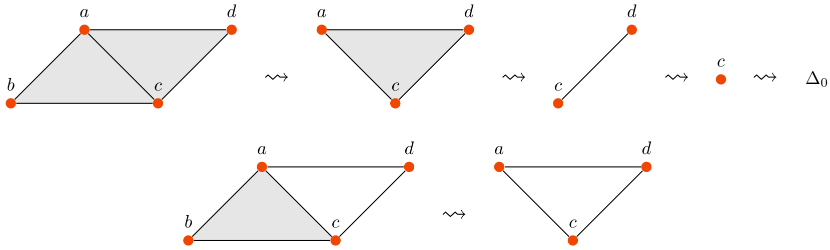


Figure 3: *Top.*Successful Graham reduction of an acyclic simplicical complex. *Bottom.* A simplicical complex which cannot be Graham reduced to Δ_0 and is therefore not acyclic.

3.1.1. Examples of Graham Reduction Some simple examples are useful to illustrate what is meant by acyclicity. First consider the acyclic simplicial complex shown in the top half of Figure 3 together with its Graham reduction to the complex Δ_0. Looking first at the leftmost complex we see that vertices a and d both occur in a single maximal face of the simplicial complex. We begin by removing a, which leaves us with the next simplicial complex in that row, where $\rightsquigarrow$ is used to indicate the process of Graham reduction. After this, every vertex occurs in one maximal face which is simply the face $\{b, c, d\}$, and so any one of them can be removed, which is chosen to be b resulting in the next simplicial complex. In this manner, we are left with the simplicial complex $\{d\}$ containing only the single vertex d. This can clearly be reduced to Δ_0, and thus this simplicial complex can be seen to be acyclic.

On the other hand, the simplicial complex shown in the bottom half of Figure 3 is not acyclic. This is because although we can initially reduce it as before by removing the vertex a, after this each vertex occurs in *two* maximal faces, and therefore none qualify to be removed in a Graham reduction step.

Vorob''s Theorem can then be stated quite compactly. The version we give here is from [7].

Theorem 3.1 (Vorob'ev [21])**.** *Given a measurement scenario $\mathcal{M} = \langle X, \Sigma, O \rangle$, any empirical model $\{e_\sigma\}$ defined over $\mathcal{M}$ is non-contextual if and only if Σ is acyclic.*

Theorem 3.1 says not only that acyclicity of Σ guarantees all empirical models in $\mathsf{EM}(\mathcal{M})$ are noncontextual, but moreover that if Σ is *not* acyclic then a contextual empirical model can always be constructed.

4. Mapping Temporal Setups to Conextuality Setups

We map a temporal measurement scenario $\mathcal{M}$ to a set of measurement scenarios $\{\mathcal{M}'_k\}_{k \in I}$ indexed by an integer k, such that the empirical models on these latter measurements scenarios are in one-to-one correspondence with the $\mathsf{lookback}_k$-consistent empirical models on $\mathcal{M}$. The upshot of doing so is that Vorob'ev's Theorem applied to these new scenarios describes, for each choice of k, whether there are non-classical $\mathsf{lookback}_k$-consistent empirical models on $\mathcal{M}$. Crucially, since the simpicial complex of $\mathcal{M}'_k$ is different for different values of k, this will depend on the choice of lookback.

Theorem 4.1. *For all lookback depths k there is a map $\mathcal{C}_k$ from temporal measurement scenarios to contextuality measurement scenarios, such that empirical models $w \in \mathsf{EM}(\mathcal{C}_k(\mathcal{M}))$ can be pulled back via $\mathcal{C}_k^*$ to $\mathsf{lookback}_k$-consistent empirical models $e^{(k)} \in \mathsf{EM}(\mathcal{M})$ on the temporal measurement scenario $\mathcal{M}$. This map preserves and reflects nonclasicality, meaning that an empirical model w on $\mathcal{M}'_k$ which is in the domain of $\mathcal{C}_k$ is contextual if and only if $e^{(k)}$ is a $\mathsf{lookback}_k$-nonclassical model.*

We begin by defining the map $\mathcal{C}_k$ which 'flattens' the temporal measurement scenario to a contextuality measurement scenario. Given any subset $U \subseteq \Sigma$ of measurement sequences, define

$$\mathcal{C}_k(U) := \{\mathsf{lookback}_k(\sigma) | \sigma \in U\}. \tag{16}$$

Given a temporal measurement scenario $\mathcal{M} = \langle X, \Sigma, O\rangle$ with cover $\mathcal{C}$, define a contextuality scenario $\mathcal{M}'_k = \langle X'_k, \Sigma'_k, O'_k\rangle$ as

(i) $X'_k := \mathcal{C}_k(\Sigma)$,
(ii) $\Sigma'_k := \{\mathcal{C}_k(C) | C \in \mathcal{C}\}$,
(iii) $(O'_k)_{\mathsf{lookback}_k\sigma} := O_m$ where $\sigma[-1] = m$.

The measurement set of the kth image scenario $\mathcal{M}'_k$ is the set of k-lookbacks of the sequences in Σ. Thus, for $k = 0$ we obtain a set of single measurements (analogous to the usual contextuality case) and in the case $k = \infty$ we obtain the set of sequences Σ. Next, a set of k-lookbacks forms a context $K \in \Sigma'_k$ only when their preimage under the k-lookback function is a valid context in $\mathcal{C}$. Finally, the output set of the string of measurements $\mathsf{lookback}_k\sigma$ describing the k-lookback of the sequence σ is just given by the output set O_m of the terminal measurement m in that sequence. We consider an example.

Example 4.1. *Consider the mapping $\mathcal{C}_k$ of temporal Bell scenario with three agents, $\mathcal{A} = \{A, B, C\}$ where $A < B < C$, and in which each agent has two inputs, $I_A = I_B = I_C = \{0, 1\}$. The measurement scenario is therefore*

(i) $X = \{(a, 0), (a, 1), (b, 0), (b, 1), (c, 0), (c, 1)\}$
(ii)

$$\sigma \in \Sigma \implies \sigma = \begin{cases} (a,i)(b,j)(c,k) & i,j,k \in \{0,1\} \\ (a,i)(b,j) & i,j \in \{0,1\} \\ (a,i) & i \in \{0,1\} \end{cases}$$

(iii) $\forall m \in X. O_m := \{0, 1\}$.

The cover $\mathcal{C}$ has as elements causally consistent subsets of Σ. For instance $\{(a,i)(b,j)(c,k), (a,i)(b,j), (a,i)\} \in \mathcal{C}$ since $(a,i) \leq (a,i)(b,j) \leq (a,i)(b,j)(c,k)$. There are three possible mappings in this case: $\mathcal{C}_0, \mathcal{C}_1$ and $\mathcal{C}_2$, since all mappings $\mathcal{C}_k$ for $k > 2$ are identical to the $k = 2$ case.

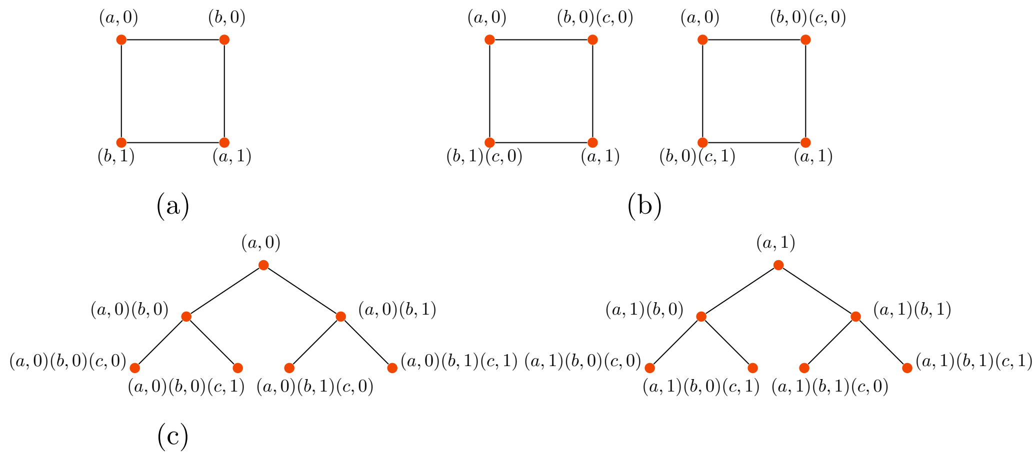


Figure 4: The image simplicial complexes, or subcomplexes thereof, under the map $\mathcal{C}_k$ for the three party temporal Bell scenario, with binary inputs for agents. (a) image complex for lookback depth $k = 0$, (b) subcomplex of the image complex for lookback depth $k = 1$ (c) image complex for lookback depth $k = 2$.

(i) <u>*The $\mathcal{C}_0$ mapping*</u>

- $X'_0 = \{m | m \in X\}$
- *Σ'_0 is shown in Figure 4(a). Two important points are that (i) the simplicical complex is not Graham reducible and (ii) It is identical to the simplicial complex Σ for the spatial 2-party Bell scenario.*
- $\forall m \in X'.O_m \in \{0, 1\}$.

(ii) <u>*The $\mathcal{C}_1$ mapping*</u>

- $$\begin{aligned} X'_1 \;=\; \{&(a,0),(a,1),(a,0)(b,0),(a,0)(b,1),(a,1)(b,0),(a,1)(b,1),\\ &(b,0)(c,0),(b,0)(c,1),(b,1)(c,0),(b,1)(c,1)\} \end{aligned}$$
- *The simplicial complex is too large to be informative when displayed, but it contains the induced subcomplexes shown in Figure 4(b). Note that the left subcomplex encodes the incompatibility of lookbacks: the outcome of $(c, 0)$ can only be obtained either after $(b, 0)$ or after $(b, 1)$— not both. On the other hand, the right simplicical complex encodes the commutation relation of measurements: $(c, 1)$ and $(c, 0)$ cannot be simultaneously measured.*

<u>*The $\mathcal{C}_2$ mapping*</u>

- $X'_3 := \Sigma$
- *$\Sigma'_3 := \mathcal{C}$, which is shown in Figure 4(c).*
- $O_{...m} := O_m$

As is clear from the above example, for different values of k the mapping produces a different cover for the sheaf (as encoded by Σ'_k). For $k \geq 2$, the image scenario is actually identical to $\mathcal{M}$.

The next part of the theorem describes how the pullback of this map, denoted $\mathcal{C}_k^*$, can produce from an empirical model $w^{(k)} \in \mathsf{EM}(\mathcal{M}'_k)$ a valid k-foregetful empirical model $e^{(k)} \in \mathsf{EM}(\mathcal{M})$.

4.1. Mapping Empirical Models

We will begin by showing that $\mathsf{lookback}_k$-consistent strategies on any subset $U \subseteq \Sigma$ are in bijection with sections on the subset $U' := \mathcal{C}_k(U)$.

We define a function

$$\alpha : \mathcal{S}_k(U) \to \mathcal{E}(U') :: f \mapsto s \tag{17}$$

which sends a $k-$ forgetful strategy $f \in \mathcal{S}_k(U)$ to a section $s \in \mathcal{E}(U')$. Define s to be the function

$$s(l) := f(\sigma) \tag{18}$$

where l is the k-lookback of the sequence σ. Note that although we could have σ, σ' with $\mathsf{lookback}_k(\sigma) = \mathsf{lookback}_k(\sigma')$ so that the above seems ambiguous, both give the same definition of s due to f being $\mathsf{lookback}_k$-consistent.

Proposition 4.1. *The function α is a bijection.*

Proof. It has an inverse α^{-1} which sends $s \in \mathcal{E}(U')$ to $f \in \mathcal{S}_k(U)$ where $f(\sigma) := s(\mathsf{lookback}_k(\sigma))$. The result is a map which $\mathsf{lookback}_k$-consistent. □

For any $U \subseteq \Sigma$, using the pullback $\mathcal{C}_k^*\mathcal{E}(U) := \mathcal{E}(\mathcal{C}_k(U))$ we obtain sections on the k-lookback of sequences in U directly, and these are in bijection with those on $\mathcal{S}_k(U)$. This bijection extends also to the probability distributions on these sections, so that we obtain

$$\mathcal{D}_{\mathbb{R}} \circ \mathcal{C}_k^*\mathcal{E}(U) \cong^{\alpha} \mathcal{D}_{\mathbb{R}} \circ \mathcal{S}_k(U). \tag{19}$$

Take an empirical model $w \in \mathsf{EM}(\mathcal{M}'_k)$. Recall that this consists of a probability distribution $w_L \in \mathcal{D}_{\mathbb{R}} \circ \mathcal{E}(L)$ for each $L \in \Sigma'_k$. Now define an empirical model $e^{(k)}$ to be

$$e^{(k)}_C(f) = w_{\mathcal{C}_k(C)}(\alpha(f)). \tag{20}$$

The final part of the Theorem is to show that this map reflects and preserves classicality.

Lemma 4.1. *The induced map between empirical models $w \mapsto e^{(k)}$ given in Equation 20 preserves and reflects classicality.*

Proof. We start by showing that classicality is preserved. Assume $e^{(k)}$ is classical, so that there exists a global section $h \in \mathcal{D}_{\mathbb{R}} \circ \mathcal{S}_k(\Sigma)$ which marginalises to $e^{(k)}$ on the cover $\mathcal{C}$. Let $\delta_s(s')$ be equal to 1 iff $s = s'$, and 0 otherwise. Here, $s, s' \in \mathcal{F}(U)$ where $\mathcal{F} \in \{\mathcal{E}, \mathcal{S}_k\}$

and $U \subseteq X'_k$ for $\mathcal{F} = \mathcal{E}$, and $U \subseteq \Sigma$ for $\mathcal{F} = \mathcal{S}_k$. Then we have that

$$\begin{aligned} w &= e^{(k)}_{\mathcal{C}_k(C)}(\alpha(f)) \\ &= \sum_{s \in \mathcal{S}_k(\Sigma)} \delta_{s|\mathcal{C}_k(C)}(\alpha(f))h(s) \\ &= \sum_{\alpha^{-1}(s) \in \mathcal{E}(X'_k)} \delta_{\alpha^{-1}(s)|C}(f)\alpha^{-1}(h)(\alpha^{-1}(s)). \end{aligned}$$

where the first line follows by definition (Equation 20), the second from $e^{(k)}$ being a classical empirical model and therefore the existence of a global section $s \in \mathcal{S}_k(\Sigma)$, and the last can be seen to be true by checking equality of each term (the bijection ensures this holds). Thus $\alpha^{-1}(h)$ is a valid global section for w.

Showing that classicality is reflected follows similarly by first assuming that w is classical. $\square$

4.2. No-Signalling Constraints beyond the AoT Constraints

The AoT constraints contained in Equation 13 express that the probability of a measurement obtaining a particular outcome should be independent of the choice of measurements at a later time. This can be implemented in a loophole free way by ensuring measurements are time-like separated.

However, the k-lookback constraints impose additional no-signalling constraints which we have so far not considered. Such additional constraints on an empirical model $e^{(k)} \in \mathsf{EM}^k(\mathcal{M})$ become clear when considering the image scenario $\mathcal{M}'_k := \mathcal{C}_k(\mathcal{M})$ and its corresponding empirical models $\mathsf{EM}(\mathcal{M}'_k)$.

Take $w^{(k)} \in \mathsf{EM}(\mathcal{M}'_k)$. Note that we can have two sets in the causally compatible cover $C_1, C_2 \in \mathcal{C}$ which before had empty intersection, but which under the mapping $\mathcal{C}_k$ obtain non-empty intersection. Therefore, we obtain the no-signalling constraint on $w^{(k)}$:

$$w^{(k)}_{\mathcal{C}_k(C_1)}|_{\mathcal{C}_k(C_1)\cap\mathcal{C}_k(C_2)} = w^{(k)}_{\mathcal{C}_k(C_2)}|_{\mathcal{C}_k(C_1)\cap\mathcal{C}_k(C_2)}. \tag{21}$$

Example 4.2. *Consider again the measurement scenario of the three party temporal Bell setup, and consider the set of empirical models $w^{(1)} \in \mathsf{EM}(\mathcal{M}'_k)$ on the image scenario $\mathcal{M}'_k := \mathcal{C}_1(\mathcal{M})$. Consider the causally consistent sets $C_1 = \{(a,0),(a,0)(b,0),(a,0)(b,0)(c,0)\}$ and $C_2 = \{(a,1),(a,1)(b,0),(a,1)(b,0)(c,0)\}$. Under $\mathcal{C}_1$ these become*

$$\mathcal{C}_1(C_1) = \{(a,0),(a,0)(b,0),(b,0)(c,0)\} \tag{22}$$

$$\mathcal{C}_1(C_2) = \{(a,1),(a,1)(b,0),(b,0)(c,0)\} \tag{23}$$

which have nonempty intersection $C'_1 \cap C'_2 = \{(b,0)(c,0)\}$. Therefore, we obtain the no-signalling constraint

$$w_{C'_1}|_{C'_1\cap C'_2} = w_{C'_2}|_{C'_1\cap C'_2} \tag{24}$$

or

$$w_{\{(a,0),(a,0)(b,0),(b,0)(c,0)\}}|_{\{(b,0)(c,0)\}} = w_{\{(a,1),(a,1)(b,0),(b,0)(c,0)\}}|_{\{(b,0)(c,0)\}}. \tag{25}$$

What do these new no-signalling equations express? They express that there can be no signalling of the measurement setting outside of the k-lookback window. For k maximal, such no-signalling constraints coincide with the AoT constraints imposed by the arrow of time. For k minimal and equal to zero, the no-singalling constraints are identical to those that one would get in a spatially separated system.

We obtain the two following propositions, which are expanded upon and proven in Appendix B and Appendix C.

Proposition 4.2. *Take a measurement scenario $\mathcal{M}$. If a $\mathsf{lookback}_0$-consistent empirical model $e \in \mathsf{EM}^{(0)}(\mathcal{M})$ is non-classical, or equivalently extendable, then this implies that the system obeys macroscopic realism.*

Proposition 4.3. *Take a temporal Bell scenario $\mathcal{M}$. The set of $\mathsf{lookback}_k-$-consistent empirical models $\mathsf{EM}^{(k)}(\mathcal{M})$ for $k \geq N-1$, where N is the number of agents in the setup is in bijection with the set of empirical models on $\mathcal{M}$ as defined by Gogioso and Pinzani in [14].*

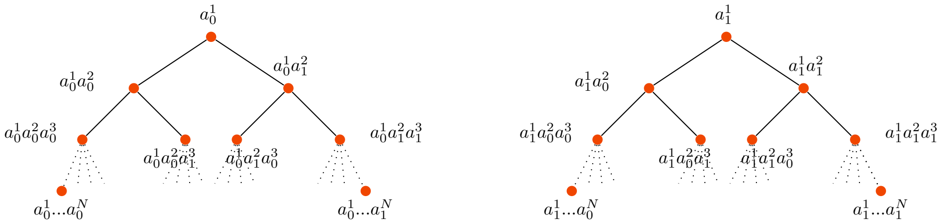
4.3. Combinatorial Conditions for Classicality

Combining Vorob'ev's Theorem with the Mapping Theorem above results in the following corollary.

Corollary 4.1. *Let $\mathcal{M}$ be a temporal measurement scenario. Every empirical model on $\mathcal{M}$ is $\mathsf{lookback}_k$-classical iff the corresponding contextuality measurement scenario $\mathcal{C}_k(\mathcal{M})$ has acyclic simplicial complex Σ'_k.*

Proof. Suppose first that the simplicial complex Σ'_k associated with the measurement scenario $\mathcal{C}_k(\mathcal{M})$ is acyclic. From Theorem 4.1, every empirical model on top of Σ'_k is noncontextual. Each of these corresponds to an empirical model on $\mathcal{M}$, which, together with the property that $\mathcal{C}_k$ preserves and reflects classicality of e, is enough to say that every empirical model on $\mathcal{M}$ must, too, be k-lookback classical. On the other hand, suppose that every empirical model on $\mathcal{M}$ is $\mathsf{lookback}_k$-classical. Since every empirical model w on $\mathcal{C}_k(\mathcal{M})$ pullsback to give an empirical model on $\mathcal{M}$, and moreover this map preserves and reflects classicality, it must be that every empirical model $w \in \mathsf{EM}(\mathcal{C}_k(\mathcal{M}))$ is noncontextual. From 3.1, we therefore know that the simplicical complex Σ'_k must be acyclic. □

One immediately sees that all simplicial complexes of a measurement scenario $\mathcal{M}$ arising under the mapping $\mathcal{C}_\infty$ are tree-like and, most importantly, acyclic. We display the general from of Σ'_∞ for temporal Bell scenarios below (to save space, we switch notation letting a_i^m denote the m-th measurement of agent i).



We therefore arrive at the following proposition.

Proposition 4.4. *Let $\mathcal{M} = \langle X, \Sigma, O \rangle$ be a temporal measurement scenario, and let N be the length of the longest sequence in Σ, $N := \max_{\sigma \in \Sigma} |\sigma|$. Then for all $k \geq N-1$, every empirical model $e \in \mathsf{EM}^{(k)}(\mathcal{M})$ is classical.*

Proof. Immediate. □

For the special case of temporal Bell scenarios, N is equal to the number of agents in the setups, $N := |\mathcal{A}|$. We might wonder, then, what combinatorial conditions on Σ ensure that Σ'_k is not acyclic, thereby guaranteeing that an empirical model that is $\mathsf{E}_{\mathsf{lookback}_k}$-nonclassical exists.

We give a condition on the agents and their inputs which is both sufficient and necessary for Σ'_k to be irreducible.

Proposition 4.5. *Let $\mathcal{M} = \langle X, \Sigma, O \rangle$ be a temporal Bell scenario with N linearly ordered agents $\mathcal{A}$ with I_a denoting the input set of an agent $a \in \mathcal{A}$. The simplicial complex Σ'_k is Graham irreducible if and only if there exists at least two agents i and j (w.l.o.g. take $i < j$) with $|I_i| \geq 2$ and $|I_j| \geq 2$ such that either*

(i) $j - i > k$ or

(ii) There exists an $i < j < l$ with $l - i > k$ and $l - j \leq k$.

The proof is contained in Appendix D. Key to the proof of sufficiency is the fact that if either (i) or (ii) holds then the image complex $\mathcal{C}_k(\Sigma)$ will be Graham irreducible due to the presence of a square. This then allows us to invoke Vorob'ev's Theorem. Necessity on the other hand follows by considering all cases in which we do not have at least two agents with at least two inputs, and showing that this always results in an acyclic simplicical complex.

It is informative to study the subcomplexes displayed in Figure 4(b), which correspond to a 3 agent temporal Bell scenario with $k = 1$. The left-hand subcomplex corresponds to the case in which $i = 0$, which corresponds to agent a, and $j = 1$ which corresponds to agent b. Then since for $l = 2$, we have $|l - i| = |2 - 0| = 2 > 1$ and $|l - j| = |2 - 1| = 1 \leq 1$ we get a square subcomplex which is acyclic, so that the entire simplcial complex must be acyclic.

On the other hand, the right-hand subcomplex corresponds to the case in which $i = 0$ and $j = 2$ corresponding to agents a and c respectively. Then since $|j - i| = 2 > 1$ there exists a subcomplex of $\mathcal{C}_1(\Sigma)$ which is a square. As such, we expect to be able to construct an empirical model $e \in \mathsf{EM}^{(1)}(\mathcal{M})$ which is nonclassical; we do this below simplifying to one input for B (so that we consider acyclicity which arises due to (i)).

Example 4.3 (3-party temporal Bell)**.** *We now construct a non-classical empirical model for the* 3*-party temporal Bell scenario, with the input at agent* B *taken to be the singleton* $\{0\}$ *and the inputs at* A *and* B *to be* $\{0, 1\}$*.*

	000	001	010	011	100	101	110	111
$(a,0)\,(b,0)\,(c,0)$	1/4	0	1/4	0	0	1/4	1/4	0
$(a,0)\,(b,0)\,(c,1)$	1/4	0	1/4	0	0	1/4	1/4	0
$(a,1)\,(b,0)\,(c,0)$	1/4	0	1/4	0	0	1/4	1/4	0
$(a,1)\,(b,0)\,(c,1)$	0	1/4	0	1/4	1/4	0	1/4	0

Table 1

It can be checked that it satisfies the required no-signalling constraints:

$$e_{\{(a,i),(a,i)(b,j),(a,i)(b,j)(c,k)\}\}}(s)|_{\{(a,i)\}} = e_{\{(a,i),(a,i)(b,j),(a,i)(b,j)(c,k)\}}(s)|_{\{(a,i)\}}$$
$$e_{\{(a,i),(a,i)(b,j),(a,i)(b,j)(c,k)\}\}}(s)|_{\{(a,i)(b,j)(c,k)\}} = e_{\{(a,i'),(a,i')(b,j),(a,i')(b,j)(c,k)\}}(s)|_{\{(a,i)(b,j)(c,k)\}}.$$

It is therefore a valid $\mathsf{lookback}_1$-consistent empirical model, so we have an empirical model $e^{(1)} \in \mathsf{EM}^{(1)}(\mathcal{M})$. We will show that the $e^{(1)}$ is non-classical by showing that there is no 1-lookback global strategy which is consistent with the support of this empirical model.

Proposition 4.6. *The empirical model defined in table 1 is a non-classical* $\mathsf{lookback}_1$*-consistent empirical model.*

Proof. Consider the image scenario $\mathcal{C}_1(\mathcal{M})$ and the empirical model $w \in \mathsf{EM}(\mathcal{C}_1(\mathcal{M}))$ of which $e^{(1)}$ is the pullback. We will prove that w must be non-classical by contradiction, and so begin by assuming that there exists a probability distribution h on global strategies $\mathcal{E}(X'_1)$ consistent with w. As a weaker condition, we would like to find a single section $s \in \mathcal{E}(X'_1)$ which is consistent with the support of w. This section will be of the following form

$$f :: \{(a,0) \mapsto o_1, (a,0)(b,0) \mapsto o_2, (b,0)(c,0) \mapsto o_3, (b,0)(c,1) \mapsto o_4, (a,1) \mapsto o_5, (a,1)(b,0) \mapsto o_6\}.$$

For example, looking at table 1 we might take $o_1 = 0$ and $o_2 = 0$, then we must take $o_3 = 0$ (it is the only way that this section will have support on entries which occur with non-zero probability) and $o_4 = 0$. No matter the choice of o_5 and o_6, there will be a probability in the table which is zero but which f contains in its support. This is true regardless of how $o_1 - o_6$ are defined, as can be checked. Thus, w is non-classical and therefore $e^{(1)}$ is non-classical (from Theorem 4.1). $\square$

5. Finite State Machines

We have defined a machine to be a tuple $\mathsf{E} = \langle X, O, F, h \rangle$. Oftentimes, rather than define a subset F of realisable input-output functions, one gives the machine a finite number of states and refers to it as a **finite state machine**. A finite state machine, given its internal state and a choice of input produces an output. This process, of deterministically producing outputs given the machines current state and the choice of input, is described jointly via a transition function and output function.

Definition 5.1 (Finite State Machine)**.** *A finite state machine* E *is a* 6*-tuple* $\langle S, S_0, X, O, T, Q \rangle$ *consisting of*

- *A finite set of states* S
- *An initial state* $S_0 \in S$
- *A finite input alphabet* χ
- *A finite output alphabet* Λ
- *A transition function* $T : S \times \chi \to S$
- *An output function* $Q : S \times \chi \to \Lambda$

A strategy will be defined in terms of the output function Q of the mealy machine via

$$f(m_0 m_1 ... m_n) := Q(s_0, m_0) Q(m_1, s_1) ... Q(m_n, s_n). \tag{26}$$

We can clearly then generate the set F of realisable input-output functions over some choice of $\Sigma \subseteq X^*$.

We first show that for Q and T can be defined such that the strategies realised by E are $\mathsf{lookback}_k$-consistent.

Proposition 5.1. *For appropriately defined function* Q *and* T *and internal states* S*, the FSM* E *generates only* $\mathsf{lookback}_k$*-consistent input-output functions.*

Proof. Take an arbitrary $U \subseteq \Sigma$ and consider the set of $\mathsf{lookback}_k$-consistent strategies $\mathcal{S}_k(U)$. Recall that such strategies are in bijection with the set of sections $\mathcal{E}(\mathcal{C}_k(U))$, where $\mathcal{C}_k$ is the map of measurement scenarios for lookback depth k. Take the internal states of the system to be $S = \{1, 2, ..., d\}$. We require that, for a fixed measurement $m \in X$, the set of internal states be large enough to allow the k-looback for all sequences ending in m to be stored. To this end define

$$L(m) := \{\mathsf{lookback}_k(\sigma) | \sigma \in \Sigma, \sigma[-1] = m\} \tag{27}$$

and let

$$d := \max_{m \in X} |L(m)| \,. \tag{28}$$

For non-local setups *without* repeating measurements we require $d = |I|^k$, where I is the set of inputs of an agent.

Now define a bijection for each $m \in X$ as $\gamma_m : L(m) \to S$ which assigns to each lookback for m a state in S. This lets us in turn define the transition and output functions T and Q in terms of a section $f \in \mathcal{C}_k^* \mathcal{E}(U)$ as

$$Q(m_i, s_i) = f(\gamma_{m_i}^{-1}(s_i)) \tag{29}$$

$$T(m_i, s_i) = \gamma_{m_{i+1}}(conc(\gamma_{m_i}^{-1}(s_i), m_{i+1})). \tag{30}$$

The output function uses the lookback for measurement m_i corresponding to the state s_i to define the output in terms of the $\mathsf{lookback}_k$-consistent section f. The transition function simply updates the lookback to include the next measurement to be performed m_{i+1} and maps this back to the corresponding internal state. □

Since a $\mathsf{E}_{\mathsf{lookback}_k}$consistent empirical models can be realised by an FSM of sufficient internal dimension, which uses its internal states to store the lookback of measurements, we can automatically extend Proposition 4.4 to derive a lower bound on the number of states needed to realise any empirical model for a given setup.

Proposition 5.2. *Let $\mathcal{M} = \langle X, \Sigma, O \rangle$ be a temporal measurement scenario. Every temporal Bell scenario $e \in \mathsf{EM}(\mathcal{M})$ can be produced by a FSM with with internal dimension $d \geq \max_{m \in X} |L_k(m)|$ where $k = N-1$. Here the set $L_k(m)$ for each measurement $m \in X$ is defined as the total number of inequivalent k-lookbacks for a measurement $m \in X$,*

$$L_k(m) := \{\mathit{lookback}_k(\sigma) | \sigma \in \Sigma \wedge \sigma[-1] = m\}. \tag{31}$$

Proof. Take the FSM to have output and transition functions as in Equations 29-30. Since this is a $\mathsf{lookback}_k$-consistent model with $k \geq N - 1$, from Proposition 4.4 it is classical. □

6. Θ-functions as generalisations of $\mathsf{lookback}_k$

We now consider the case in which a measurement can depend on up to k previous measurements, although these need not be the k preceding measurements. This can be envisioned by, at time step T, having the memory register storing k measurements which occurred previously. At time step $T + 1$ a measurement m is performed and the system can either (i) kick off one of the k measurements and replace it with m or (ii) keep it as is.

We first briefly introduce some additional terminology concerning sequences: we write $\tau \preceq \sigma$ when τ is a subsequence of σ, meaning it can be obtained from σ by deleting certain measurements but not changing the order of the remaining measurements. Moreover if $\tau \leq \sigma$ we write $\sigma - \tau$ to denote the tail of the sequence of σ not contained in τ.

We then define a Θ-function to be a function $\Theta : \Sigma \to X^*$ which satisfies the following three properties

(i) $\Theta(\sigma) \preceq \sigma$,

(ii) $\Theta(\sigma)[-1] = \sigma[-1]$ for all $\sigma \in \Sigma$, and

(iii) for $\sigma_1 \leq \sigma_2$, $\Theta(\sigma_2) = s_1 s_2$ where $s_1 \preceq \Theta_k(\sigma_1)$ and $s_2 \preceq \sigma_2 - \sigma_1$.

The first requirement says that Θ should map any sequence σ to a subsequence of σ. The second requirement is that the final value in the measurement sequence $\Theta(\sigma)$ be the terminal measurement of σ. Finally, the third requirement is that given two sequences σ_1, σ_2 with σ_1 a prefix sequence of σ_2, the value of Θ at σ_2 should be a concatenation of a subsequence of Θ at σ_1 and a subsequence of the tail of σ_2 not in σ_1.

It is of particular interest whether an empirical model has support on Θ-functions which are restricted to having in their codomain subsequences of length $k+1$. This amounts to considering functions which store arbitrary k measurements from the past, as well as the terminal measurement m on σ. For example, the $\mathsf{lookback}_k$ function is for each k an example of such a function.

Consider the Θ-functions which satisfy for fixed k and all $\sigma \in \Sigma$

$$\Theta(\sigma) \leq k+1, \tag{32}$$

and write L_k for the set of Θ-functions which satisfy this property, bearing in mind that the actual memory occupies just the first k space holders. For $U \subseteq \Sigma$ we define the set of strategies

$$\bar{\mathcal{S}}_k(U) := \bigcup_{\Theta \in L_k} \mathcal{S}_\Theta(U). \tag{33}$$

A strategy $s \in \bar{\mathcal{S}}_k(U)$ is said to be L_k-consistent. Definitions of L_k-consistent empirical models and E_{L_k}-classicality are then straightforward generalisations of those which we had before, replacing the presheaves $\mathcal{S}_k$ with $\bar{\mathcal{S}}_k$.

Moreover, Theorem 4.1 can be extended to L_k-consistent empirical models. Each function $\Theta \in L_k$ results in a unique image scenario $\mathcal{C}_\Theta(\mathcal{M})$, so that every L_k-consistent empirical model $e \in \mathsf{EM}^{(L_k)}$ is the pullback of an empirical model $w_1 \,\&\, w_2 \,\&\, ...\, w_{|L_k|}$ where each $w_i \in \mathsf{EM}(C_{\Theta_i}(\mathcal{M}))$. Here, $e_1 \,\&\, e_2$ indicates the controlled choice operation applied to empirical models e_1 and e_2 (see for example [5]).

Whether one can construct an L_k-consistent empirical model which is E_{L_k}-nonclassical requires that for at least one $\Theta \in L_k$, the image $\mathcal{C}_\Theta$ is Graham irreducible.

We now turn to a question we have left lingering, which is whether one can say anything interesting in the current framework about the case in which strategies can be dependent on past measurements *and* outcomes.

7. Allowing outputs to be stored in memory

Our first observation is that allowing for the system to store previous outputs, rather than measurements, still allows for the system $\mathcal{T}$ to keep track of previous measurements by encoding these in the outputs. The result, however, is that the strategies are constrained to use outputs solely for keeping track of previous measurements. In the extreme case, where we allow measurements to have unbounded output sets, for every empirical model

$e \in \mathsf{EM}(\mathcal{M})$ there exists an empirical model $e' \in \mathsf{EM}(\mathcal{M})$ whose output can be projected onto the first m values to give an empirical model identical to e, but which is classical.

The idea is that the second empirical model uses the output set to store, in addition to the output of the first empirical model, all measurements which have previously occurred. Much like Hilbert's Hotel, in which a fully occupied hotel with infinitely many rooms may still accommodate additional guests, our output sets, which we take to be infinitely big, can always store additional information. As such, we can store the full lookbacks in the output register.

This observation makes it clear that the size of the output set is important in determining to what extend including the outputs is important.

In this section we justify the use of $\mathsf{lookback}_k$, and more generally Θ-functions, by showing

(i) The strategies of a system storing outcomes can be realised by storing more measurements. We show that for certain examples the number of extra measurements required is linear in the size of the output set.

(ii) That we can for any measurement scenario which admits $\mathsf{lookback}_k$-consistent empirical models which cannot be produced by a classical machine E_{lookback_k} (i.e. $\mathsf{E}_{\mathsf{lookback}_k}$-nonclassical) find at least one empirical model which cannot be produced by a machine E' which can store the past k measurements **and** outcomes.

7.1. Parity Machines and Related Examples

A standard example for classical sequential machines, such as finite state machines, is the parity checker. The parity checker for a string of inputs $x_0 ... x_l$ will output $o_l = 0$ if $x_0 ... x_l$ is of even parity (i.e. it contains an even number of 1's), and $o_l = 1$ if it is of odd parity (or contains an odd number of 1's). Here we take inputs x_i to be choices of measurement and outputs o_i to be the outcome of this measurement.

A system which is able to store the previous outcome can implement the parity checker deterministically. The table below defines given o_{l-1} and x_l what the output o_l will be, by simply performing an XOR operation on the two bits. Clearly, this implements

o_{l-1}	x_l	$o_{l-1} \oplus x_l$
0	0	0
0	1	1
1	0	1
1	1	0

the parity checker. Now we ask what the minimal, in the sense of the number of inputs contained in its codomain, Θ-function is that can produce the parity checker. Clearly, if the entire history of inputs is stored the parity can be checked by computing $x_0 \oplus x_1 \oplus ... \oplus x_l$. Note however that for any bitstring, we can find a substring of length at most 2 with parity identical to the original. In fact, any bitstring which has odd parity always contains 1 as a substring, so we define $\Theta(x_0x_1...x_l) = 1x_l$ if $x_0x_1...x_{l-1}$ has odd parity. On the other hand, a bit string can contain only 1's and have even parity. Thus we define $\Theta(x_0x_1...x_l) = 0x_l$ if $x_0x_1...x_{l-1}$ has even parity and contains at least one 0, or if the bitstring contains no 0's then $\Theta(x_0x_1...x_l) = 11x_l$. Thus $\Theta \in L_3$, so that while only one outcome needs to be stored in memory we need two measurements.

This is not such a bad trade-off. Consider now a system with a larger output set $|O| = z$, and suppose we have the empirical model in which the output o_l is the number of 1's in the bitstring $x_0...x_l$, and so implements the parity checker when $z = 2$. In this case it can similarly be shown that the minimal Θ-function is in the set L_{z+1}. In this case therefore, the number of extra measurements which need to be stored scales as the size of the output set when the size of the input set is kept fixed.

7.2. Nonclassicality and Memory with Outputs

We denote as $\mathsf{lookback}_k^{(e)} : \mathcal{S}(\Sigma) \times \Sigma \to \Delta^*_{\mathcal{M}}$ the map which sends a tuple (f, σ) consisting of a strategy f and a sequence σ to the $k+1$ measurement events $\left\langle m_{\max(n-k,0)}, o_{\max(n-k,0)} \right\rangle ... \left\langle m_{n-i}, o_{n-i} \right\rangle \left\langle m_n, \perp \right\rangle$ where $o_i := f(m_0...m_{i-1}m_i)$. Note that this is strategy dependent, unlike $\mathsf{lookback}_k$.

Nevertheless, as with Θ-functions, it is straightforward to extend previous definitions of $\mathsf{lookback}_k$-consistent empirical models and their classicality to $\mathsf{lookback}_k^{(e)}$. Unfortunately, it is not clear how Theorem 4.1 can be extended, since inclusion of outputs in the mapping complicates matters. In effect, empirical models on the image scenario which pull-back to give an empirical model on the temporal scenario are not well-defined; specifically they do not satisfy the required no-signalling constraints.

We can show nevertheless that one direction of Corollary 4.1 extends also to empirical models which are $\mathsf{lookback}_k^{(e)}$-consistent.

Theorem 7.1. *Let $\mathcal{M}$ be a temporal measurement scenario. If the measurement scenario $\mathcal{C}_k(\mathcal{M})$ has simplicial complex Σ'_k which is not acyclic, then there is at least one $\mathsf{lookback}_k^{(e)}-$ consistent empirical model on $\mathcal{M}$ which cannot be generated by a classical machine $E_{\mathsf{lookback}_k^{(e)}}$.*

Proof of the Theorem utilises *simulations*, and what it means for one empirical model to simulate another. This was first studied in [15] and [5], which should be consulted for the relevant definitions.

Essentially, Theorem 7.1 holds because we can define an empirical model d on $\mathcal{C}_k(\mathcal{M})$ which simulates an empirical model on the restricted scenario $\mathcal{C}_k(\mathcal{M})|_Z$, where Z corresponds to the vertices of any square subcomplex of Σ'_k—Proposition 4.5 guarantees that such a restriction always exists. Using the fact that one can always define a contextual empirical model on this restricted scenario, we show that the empirical model d which simulates e (i) must also be contextual and (ii) can be defined in such a way that it pulls back to give an $\mathsf{lookback}_k^{(e)}$-consistent model and is $\mathsf{E}_{\mathsf{lookback}_k^{(e)}}$-nonclassical.

We prove this Theorem in **??**.

8. Conclusion and Outlook

We have considered the sequential analogue of a constextuality setup, in which nonlocality constraints are relaxed to the AoT, or Arrow of Time, constraints. By utilising a sheaf theoretic approach, we modeled a setup in which the amount of information a system can store from the past is bounded. Starting with storage of just the previous k measurements, we showed that, given a particular setup captured by a temporal measurement scenario $\mathcal{M}$, there exist nonclassical correlations with support on $\mathsf{lookback}_k$-consistent strategies only if the simplicial complex Σ_k' is *not* acyclic. This simplicical complex is the one associated with a contextuality measurement scenario $\mathcal{C}_k(\mathcal{M})$ which is the image of $\mathcal{M}$ under the map $\mathcal{C}_k$. Moreover, we showed that for temporal Bell scenarios, there are certain conditions on the order of agents and their input sets that, which, when met, guarantee that nonclassicality can arise.

We next showed how this results generalises when replacing the $\mathsf{lookback}_k$ functions with Θ-functions, which allow for storage or arbitrary measurements from the past. We considered the set of strategies which are Θ-consistent for some Θ-function which stores at most k measurements. Again, nonclassicality of Θ-consistent empirical models on a temporal scenario $\mathcal{M}$ relies on an image complex $\mathcal{C}_{L_k}$ being acyclic.

Finally, we discussed the case in which measurements and outputs are stored, concluding with Theorem 7.1 which gives a Vorob'ev-type result (albeit only one direction) for the case in which the k immediate measurements events are stored by the system.

There are a few possible directions for future work. Extending the results developed in this manuscript to causal measurement scenarios, as defined in [2] and where we have both temporal and spatial correlation, would be informative. Doing so, even just for the case $\mathcal{C}_\infty$ and for Bell scenarios, would shed light on which arrangement of agents can give rise to nonclassical empirical models. Are the linearly ordered agents considered here, for example, the only case in which one can never achieve non-classical AoT correlations? While we have mostly considered examples in which the measurement strings in σ are finite, oftentimes in sequential computation one allows the system to run indefinitely, thereby generating strings of outcomes of arbitrary length. It might therefore be informative to specialise to such cases in order to understand separations between quantum and classical correlations in these setups.

References

[1] Samson Abramsky, Rui Soares Barbosa, and Shane Mansfield. "Contextual Fraction as a Measure of Contextuality". In: **Physical Review Letters** 119.5 (Aug. 2017). ISSN: 1079-7114. DOI: `10.1103/physrevlett.119.050504`. URL: `http://dx.doi.org/10.1103/PhysRevLett.119.050504`.

[2] Samson Abramsky, Rui Soares Barbosa, and Amy Searle. **Combining contextuality and causality: a game semantics approach**. 2023. arXiv: `2307.04786 [quant-ph]`.

[3] Samson Abramsky and Adam Brandenburger. "The sheaf-theoretic structure of non-locality and contextuality". In: **New Journal of Physics** 13.11, 113036 (Nov. 2011), p. 113036. DOI: 10.1088/1367-2630/13/11/113036. arXiv: 1102.0264 [quant-ph].

[4] Samson Abramsky and Lucien Hardy. "Logical Bell inequalities". In: **Physical Review A** 85.6 (June 2012). ISSN: 1094-1622. DOI: 10.1103/physreva.85.062114. URL: http://dx.doi.org/10.1103/PhysRevA.85.062114.

[5] Samson Abramsky et al. "A comonadic view of simulation and quantum resources". In: **2019 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)**. IEEE, June 2019. DOI: 10.1109/lics.2019.8785677. URL: http://dx.doi.org/10.1109/LICS.2019.8785677.

[6] Janet Anders and Dan E. Browne. "Computational Power of Correlations". In: **Physical Review Letters** 102.5 (Feb. 2009). ISSN: 1079-7114. DOI: 10.1103/physrevlett.102.050502. URL: http://dx.doi.org/10.1103/PhysRevLett.102.050502.

[7] Rui Barbosa. "Contextuality in quantum mechanics and beyond". 2015.

[8] Rui Soares Barbosa, Martti Karvonen, and Shane Mansfield. "Closing Bell Boxing Black Box Simulations in the Resource Theory of Contextuality". In: **Outstanding Contributions to Logic**. Springer International Publishing, 2023, 475–529. ISBN: 9783031241178. DOI: 10.1007/978-3-031-24117-8_13. URL: http://dx.doi.org/10.1007/978-3-031-24117-8_13.

[9] Caslav Brukner et al. **Quantum Entanglement in Time**. 2004. arXiv: quant-ph/0402127 [quant-ph].

[10] Costantino Budroni. "Contextuality, memory cost and non-classicality for sequential measurements". In: **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences** 377.2157 (2019), p. 20190141. DOI: 10.1098/rsta.2019.0141. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2019.0141. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2019.0141.

[11] Costantino Budroni and Clive Emary. "Temporal Quantum Correlations and Leggett-Garg Inequalities in Multilevel Systems". In: **Physical Review Letters** 113.5 (July 2014). ISSN: 1079-7114. DOI: 10.1103/physrevlett.113.050401. URL: http://dx.doi.org/10.1103/PhysRevLett.113.050401.

[12] Costantino Budroni, Gabriel Fagundes, and Matthias Kleinmann. "Memory cost of temporal correlations". In: **New Journal of Physics** 21.9 (2019), p. 093018. DOI: 10.1088/1367-2630/ab3cb4. URL: https://doi.org/10.1088%2F1367-2630%2Fab3cb4.

[13] Tobias Fritz. "Quantum correlations in the temporal Clauser-Horne-Shimony-Holt (CHSH) scenario". In: **New Journal of Physics** 12.8, 083055 (Aug. 2010), p. 083055. DOI: 10.1088/1367-2630/12/8/083055. arXiv: 1005.3421 [quant-ph].

[14] Stefano Gogioso and Nicola Pinzani. "The Sheaf-Theoretic Structure of Definite Causality". In: **Electronic Proceedings in Theoretical Computer Science** 343 (2021), pp. 301–324. DOI: `10.4204/eptcs.343.13`. URL: `https://doi.org/10.4204%2Feptcs.343.13`.

[15] Martti Karvonen. "Categories of Empirical Models". In: **Electronic Proceedings in Theoretical Computer Science** 287 (Jan. 2019), 239–252. ISSN: 2075-2180. DOI: `10.4204/eptcs.287.14`. URL: `http://dx.doi.org/10.4204/EPTCS.287.14`.

[16] Alexander A. Klyachko et al. "Simple Test for Hidden Variables in Spin-1 Systems". In: **Phys. Rev. Lett.** 101 (2 2008), p. 020403. DOI: `10.1103/PhysRevLett.101.020403`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.101.020403`.

[17] A. J. Leggett and Anupam Garg. "Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks?" In: **Phys. Rev. Lett.** 54 (9 1985), pp. 857–860. DOI: `10.1103/PhysRevLett.54.857`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.54.857`.

[18] Robert Raussendorf. "Contextuality in measurement-based quantum computation". In: **Physical Review A** 88.2 (Aug. 2013). ISSN: 1094-1622. DOI: `10.1103/physreva.88.022322`. URL: `http://dx.doi.org/10.1103/PhysRevA.88.022322`.

[19] Rui Soares Barbosa. "On monogamy of non-locality and macroscopic averages: examples and preliminary results". In: **Electronic Proceedings in Theoretical Computer Science** 172 (Dec. 2014), 36–55. ISSN: 2075-2180. DOI: `10.4204/eptcs.172.4`. URL: `http://dx.doi.org/10.4204/EPTCS.172.4`.

[20] Lucas B. Vieira and Costantino Budroni. "Temporal correlations in the simplest measurement sequences". In: **Quantum** 6 (2022), p. 623. DOI: `10.22331/q-2022-01-18-623`. URL: `https://doi.org/10.22331%2Fq-2022-01-18-623`.

[21] N. N. Vorob'ev. "Consistent Families of Measures and Their Extensions". In: **Theory of Probability Its Applications** 7.2 (1962), pp. 147–163. DOI: `10.1137/1107014`.

[22] Tian Zhang, Oscar Dahlsten, and Vlatko Vedral. **Quantum correlations in time**. 2020. arXiv: `2002.10448 [quant-ph]`.

Appendix A. The $\mathcal{S}_k$ Subpresheaves

Proposition Appendix A.1. *The functors $\mathcal{S}_k$ are, for every k, subsheaves of the sheaf of strategies $\mathcal{S}$.*

Proof. Recall that a presheaf $\mathcal{F}$ is called a subpresheaf of a presheaf $\mathcal{G}$ if $\mathcal{F}(U) \subset \mathcal{G}(U)$ for all open sets $U \subset X$ such that the restriction maps of $\mathcal{G}$ induce the restriction maps of $\mathcal{F}$. If equality holds, so that $\mathcal{F}(U) \subset \mathcal{G}(U)$ for all U, then we obtain the original sheaf.

For all k and $U \subseteq \Sigma$ it clearly holds that either $\mathcal{S}_k(U) \subset \mathcal{S}(U)$ or, for large enough k, $\mathcal{S}_k(U) = \mathcal{S}(U)$. In the latter case, $\mathcal{S}_k$ retains all restriction maps of $\mathcal{S}$ and therefore is equivalent to $\mathcal{S}$. In the former case, we consider a restriction map of $\mathcal{S}_k$ from a subset U to a subset V:

$$\rho_V^{(k)U} : \mathcal{S}_k(U) \to \mathcal{S}_k(V) :: f_k \mapsto f_k|_V. \tag{A.1}$$

Where $f_k|_V(\sigma) = f(\sigma)$. Now consider the restriction map from $\mathcal{S}(U)$ to $\mathcal{S}(V)$,

$$\rho_V^U : \mathcal{S}(U) \to \mathcal{S}(V) :: f \mapsto f|_V \tag{A.2}$$

where $f|_V(\sigma) = f(\sigma)$. However, any $f \in \mathcal{S}_k(U)$ is also in $\mathcal{S}(U)$, and taking the restriction of f using the restriction maps of the full $\mathcal{S}$ is identical to doing it with the functor $\mathcal{S}_k$. □

While it holds that each $\mathcal{S}_k$ is a subpresheaf of $\mathcal{S}$, it is not the case that $\mathcal{S}_k$ is, for general k, a sheaf.

For demonstration of how this fails, take the simple example in which $X = x, y, z$, $\Sigma = \{x, xy, z, zy\}$ and $O = \{0, 1\}$. Consider the subsets and associated $\mathsf{lookback}_0$-consistent strategies

$$U_1 = \{x, xy\} \qquad f_1(x) = 0, f_1(xy) = 0 U_2 = \{z, zy\} \qquad f_2(z) = 0, f_2(zy) = 1. \tag{A.3}$$

The strategies f_1 and f_2 agree on their intersection (their intersection is empty), however, there is no way to form a 0-lookback strategy $f \in \mathcal{S}(U_1 \cup U_2)$ such that $f|_{U_1} = f_1$ and $f|_{U_2} = f_2$. If we define f in the expected way

$$f :: \{x \mapsto 0, xy \mapsto 0, z \mapsto 0, zy \mapsto 1\} \tag{A.4}$$

then f does not obey the 0-lookback constraint: $\mathbf{lookback}_0(zy) = \mathbf{lookback}_0(xy)$ but $f(zy) \neq f(xy)$. While this might seem to be an artefact of this particular example, it occurs in general and so $\mathcal{S}_k$ is not a sheaf in general, because the gluing property fails.

Appendix B. Proof of proposition 4.2

In this section we prove proposition 4.2. To this end, for a fixed scenario $\mathcal{M} = \langle X, \Sigma, O \rangle$ with causally consistent sequences $\mathcal{C}$ we consider strategies $s \in \mathcal{S}_0(\Sigma)$ and empirical modes $e \in \mathsf{EM}^{(0)}(\mathcal{M})$. We will show that the correlations arising assuming 0-lookback are identical to those which arise assuming that the system is macrorealistic, thereby providing another manner in which to interpret the assumptions of macrorealism.

Since $k = 0$, we consider the image scenario $\mathcal{M}' := \mathcal{C}_0(\mathcal{M})$. Recalling the definition of $\mathcal{C}_k$, in the specific case $k = 0$ it holds that $\mathcal{M}' = \langle X', \Sigma', O' \rangle$, where we have left off the 0 subindex for clarity of presentation, where

(i) $X' = \{\sigma[-1] | \sigma \in \Sigma\}$,

(ii) $\Sigma' = \{\{\sigma[-1] | \sigma \in C\} | C \in \mathcal{C}\} = \{\tau | \tau \subseteq \sigma, \sigma \in \Sigma\}$

(iii) $O'_{\sigma[-1]} = O_\sigma$.

Importantly, $\Sigma' = \{\tau | \tau \subseteq \sigma, \sigma \in \Sigma\}$ (point (ii)) , which is the set of all partial sequences which can be performed. It simply follows from a causally consistent subset $C \in \mathcal{C}$ consisting of subsets of some maximal sequence $\sigma \in \max(\Sigma)$, so that the terminal measurements of C are a subsets of σ.

Another way to look at it is to note that we can form a poset $(\mathcal{P}(X), \subseteq)$ from the powerset of the measurement set and ordering via subset inclusion. Then, for some $A \in \mathcal{P}(X)$, define the lower set generated by A to be

$$A{\downarrow} = \{y \mid \exists x, x \in A \wedge y \subseteq x\}. \tag{B.1}$$

Thus we simply have $\Sigma' = \Sigma{\downarrow}_\subseteq$. ¶

A section $f \in \mathcal{E}(U)$ for $U \subseteq X'$ is a function

$$f : U \to O. \tag{B.2}$$

The section f therefore specifies a single output $f(m)$ for each measurement in U.

Furthermore, an empirical model is a probability distribution $w_C \in \mathcal{D}_\mathbb{R}(\mathcal{E}(\sigma))$ for each $\sigma \in \Sigma{\downarrow}_\subseteq$. $\mathsf{lookback}_0$-consistent empirical models can therefore be seen to be contextuality empirical models on the cover $\Sigma{\downarrow}_\subseteq$.

Before proving proposition 4.2 we briefly recap the assumptions underlying macrorealism.

Appendix B.0.1. Macrorealism We now recall the assumptions which are termed macrorealism and often occur when studying temporal setups in the literature.

Leggett and Garg showed that if we take classicality to include the assumptions of macroscopic realism and non-invasiveness, then the a quantum system is able to produce correlations which cannot be produced by any such classical system [17]. We state these assumptions below for clarity:

(i) *Macroscopic Realism.* The system is always in one of several distinct macroscopic states at any given time.

(ii) *Non-invasiveness.* It is possible to measure the system to determine its state without changing the state of the system or affecting its subsequent dynamics.

Based on these assumptions, a macro-realistic hidden variable theory can be derived. In particular, it can be shown that for a macro-realistic system $\mathcal{S}$ undergoing a sequence of measurements by N parties each with a given measurement setting, the probability of obtaining the string of outcomes $(o_0, ..., o_n)$ given a string of input settings $(i_0, ..., i_n)$ factorises as follows:

$$p(o_0, ..., o_n | i_0, ..., i_n) = \sum_{\lambda \in \Lambda} p(\lambda) p(o_0|i_0) p(o_1|i_1) ... p(o_n|i_n). \tag{B.3}$$

¶ Importantly, this $\downarrow$ is with respect to the ordering induced by subset inclusion $\subseteq$ and not the prefix relation $\leq$.

Here, λ is in the literature called a *hidden variable*, and Λ is a set of hidden variables.

The validity of these assumptions placed on a classical system have been extensively debated. Here, we do not wish to comment further but rather show how the framework developed in the preceding sections produces, for zero depth lookback, correlations which factorise identically to Equation B.3. Thus we obtain the correlations of macrorealistic hidden variable systems using a sheaf-theoretic approach.

Proposition Appendix B.1. *Given a temporal measurement scenario $\mathcal{M}$ and an empirical $e \in \mathsf{EM}^{(0)}(\mathcal{M})$, if e is extendable then the correlations factorise as in Equation B.3.*

Proof. If e is classical then there exists a global section $h \in \mathcal{D}_{\mathbb{R}} \circ \mathcal{S}_0(\Sigma)$. From Theorem 4.1, we can equivalently consider whether the image empirical model $w \in \mathcal{C}_0(e)$ is classical, since classicality is preserved and reflected. We have seen that sections on $\mathcal{M}'$ are of the form $s : V \to O$ with V being some set of measurements. The proof of factorisation then follows directly from a proof contained in [3]. Given a section $s \in O^{X'}$ let $\delta_s(s')$ be the distribution which is 1 when $s = s'$ and zero otherwise. Moreover, given a subset $V \subseteq X'$ we can restrict the distribution $\delta_s|_V$. It can be shown that $\delta_s|V = \delta_{s|V}$. It then follows that

$$w_\sigma(s) = \sum_{s' \in O', s'|_\sigma = s} h(s') = \sum_{s' \in O^{X'}} \delta_{s'|\sigma}(s) \cdot h(s') = \sum_{s' \in O^{X'}} \delta_{s'}|\sigma(s) \cdot h(s'). \tag{B.4}$$

Moreover, it can be checked that the following factorisation holds

$$\delta_{s'}|\sigma(s) = \prod_{m \in \sigma} \delta_{s'}|\{m\}(s|\{m\}). \tag{B.5}$$

Altogether this says that

$$e_\sigma(s) = \sum_{s' \in O^{X'}} \prod_{m \in \sigma} \delta_{s'}|\{m\}(s|\{m\}) \cdot h(s'). \tag{B.6}$$

If we take $\sigma = m^0_{i_0}, \ldots, m^n_{i_n}$ (so that the choice of measurement at time step j is $\{m^j_0, m^j_1\}$) and $s(m_{i_k}) = o_k$, then this is exactly Equation B.3 where each $s' \in O^{X'}$ is identified with a hidden variable $\lambda \in \Lambda$ and $h(s')$ gives the probability of obtaining the hidden variable s'. $\square$

Appendix C. Proof of Proposition 4.3

Proposition 4.3 builds a correspondence between the framework of temporal scenarios as developed in this manuscript and that of definite causal scenarios as developed in [14].

In [14] they introduce a measurement scenario as a triple $\langle \Omega, \underline{I}, \underline{O} \rangle$, where

(i) Ω is a poset of agents.

(ii) $\underline{I} := (I_\omega)_{\omega \in \Omega}$ is an input set for each agent.

(iii) $\underline{O} = \{I_\omega\}_{\omega\in\Omega}$ is an output set for each agent.

We briefly recap their frameowrk by the full details can be found in [14].

They work with a locale $\mathcal{L}$, which is a kind of pointless topology, which they define as follows.

Let $\Lambda(\Omega)$ consist of the lowersets of the set of agents Ω, so that any $\lambda \in \Lambda(\Omega)$ is closed below with respect to the order relation.

They then define a locale of inputs, $\mathcal{L}_\Sigma$, which will form the domain of the event sheaf. This locale is defied as

$$\mathcal{L}_\Sigma := \sum_{\lambda\in\Lambda(\Omega)} \prod_{\omega\in\lambda} (\mathcal{P}(I_\omega)\backslash\{0\}). \tag{C.1}$$

An element $U \in \mathcal{L}_\Sigma$ can be written as

$$(U_\omega)_{\omega\in\lambda_U} \tag{C.2}$$

where λ_U is some element in $\Lambda(\Omega)$.

Now, the sheaf of events, which they denote $\mathcal{E}_\Sigma$, assigns to each $U \in \mathcal{L}_\Sigma$ the set of functions

$$\{f : \prod_{\omega\in\lambda_U} \left(\left(\prod_{\omega'\leq\omega} U_{\omega'} \right) \to U_\omega \right)\}. \tag{C.3}$$

Their manuscript should be consulted for an example.

We will now show that, in the case that Ω is a total order, we can always form a temporal measurement scenario $\mathcal{M}$. The opposite direction is not true, given a temporal measurement scenario we cannot always form a measurement scenario of GP type as temporal measurement scenarios abstract away the idea of agents (this generalisation is akin to that made in going from non-locality to contextuality).

We take $\Omega = \{1, 2, ..., N\}$ with $1 < 2 < ... < N$. A section on $U \in \mathcal{L}_\mathcal{M}$ is a function for each agent in λ_U from the inputs of all agents which could have occurred before to an output of that agent. Therefore sections will be tuples $(f_1, ...f_i)$ where

$$f_i : U_1 \times ... \times U_i \to O_i. \tag{C.4}$$

GP measurement scenarios are simply temporal Bell scenarios: given $\mathcal{M}$ we define a temporal measurement scenario $\mathcal{M}' = \langle X', \Sigma', O' \rangle$ where

(i) $X = \{(\omega, i_\omega) | \omega \in [1, N] \wedge i_\omega \in I_\omega\}$
(ii) $\Sigma = \{(1, i_1)...(k, i_k) | \forall\omega \in [1, k].i_\omega \in I_\omega\}$.
(iii) $O_{(\omega,i_\omega)} = O_\omega$

A GP empirical model consists of, for each tuple of inputs on a lowerset of agents $\underline{i} \in \sum_{\lambda\in\Lambda(\Omega)} \left(\prod_{\omega\in\lambda} I_\omega\right)$, a distribution $e_{\underline{i}} \in \mathcal{D}_\mathbb{R} \circ \mathcal{E}(\underline{i})$. We denote the set of such empirical models $\mathsf{EM}_{\mathrm{GP}}(\mathcal{M})$.

We would like to show that there is a bijective correspondence between the set $\mathsf{EM}_{\mathrm{GP}}(\mathcal{M})$ of GP empirical models and the set $\mathsf{EM}(\mathcal{M})$ of temporal empirical models. A temporal empirical model consists of a distribution $e'_C \in \mathcal{D}_{\mathbb{R}} \circ \mathcal{S}(C)$ for each causally consistent set of sequences. A causally consistent subset C of Σ consists of sequences which are all a subset of some sequence $(1, i_1)(2, i_2)...(k, i_k)$ for some $k \in [1, N]$. Given a section $f \in \mathcal{S}(C)$ recall that we define

$$f^*((1, i_1), (2, i_2), ..., (k, i_k)) = f((1, i_1))f((1, i_1)(2, i_2))...f((1, i_1), (2, i_2), ..., (k, i_k)). \tag{C.5}$$

So provided C is a subset of Σ which is down-closed, since Ω is a total order this is equivalent to a GP section on an input $\underline{i} \in \sum_{\lambda \in \Lambda(\Omega)} \left(\prod_{\omega \in \lambda} I_\omega\right)$. Specifically, in this case a GP section is a function $g^* : i_1 i_2 ... i_k \mapsto o_1 o_2 ... o_k$ which factors as the functions

$$g_1 : \{i_1\} \to O :: i_1 \mapsto o_1 \tag{C.6}$$
$$g_2 : \{i_1 i_2\} \to O :: i_1 i_2 \mapsto o_2 \tag{C.7}$$
$$... \tag{C.8}$$

so that $g_l(i_1 ... i_l) = f((1, i_1))(2, i_2)...(l, i_l))$.

Note that although the cover of the temporal empirical model is larger, because C need not be down closed, since the strategy presheaf acts on entire sequences (rather than just individual measurements), this has no consequences —i.e., the no-signalling constraints do not impose any additional considerations by taking this larger cover. In fact, it is requires that this we take the larger cover to get the correct form of the mapping in Theorem 4.1.

Thus, down-closed causally consistent subsets $C \subseteq \Sigma$ are in bijection with the cover $I_\Sigma = \sum_{\lambda \in \Lambda(\Omega)} \left(\prod_{\omega \in \lambda} I_\omega\right)$, and moreover the set of sections $\mathcal{E}((\underline{i}))$ is isomorphic to the set of strategies $\mathcal{S}(C)$. As such, the data specifying a GP empirical model is identical to that of a temporal empirical model in the case that the agent poset Ω is a linear order.

Appendix D. Proof of Proposition 4.5

We prove Proposition 4.5 below.

Proof. (i) **sufficiency.** For item 1, note that if $j - i > k$ then measurements performed by agent i are never contained in the k-lookback of agent j (and vice versa). Thus, we obtain a square as an induced subcomplex. It can be shown (see e.g. [7]) there exists an induced subcomplex K which is Graham irreducible then this implies that Σ is Graham irreducible. Take $\sigma_1, \sigma_2, \sigma_3, \sigma_4 \in \Sigma$ where

$$\begin{aligned} \sigma_1[-1] &= (i, 0) \\ \sigma_2[-1] &= (i, 1) \\ \sigma_3[-1] &= (j, 0) \\ \sigma_4[-1] &= (j, 1) \end{aligned}$$

and where for every pair of sequences $\sigma_m[r] = \sigma_n[r]$ for all other agents r (not including i and j). Then the image scenario $\mathcal{C}_k(\mathcal{M})$ will have as a subcomplex the simplicial complex with vertices

$$\{\mathsf{lookback}_k(\sigma_1), \mathsf{lookback}_k(\sigma_2), \mathsf{lookback}_k(\sigma_3), \mathsf{lookback}_k(\sigma_4)\}$$

and maximal faces

$$\begin{cases} \{\mathsf{lookback}_k(\sigma_1), \mathsf{lookback}_k(\sigma_3)\} \\ \{\mathsf{lookback}_k(\sigma_1), \mathsf{lookback}_k(\sigma_4)\} \\ \{\mathsf{lookback}_k(\sigma_2), \mathsf{lookback}_k(\sigma_3)\} \\ \{\mathsf{lookback}_k(\sigma_2), \mathsf{lookback}_k(\sigma_4)\} \end{cases}$$

which is a square and therefore irreducible via a sequence of Graham reduction steps. For item 2, if $l - i > k$ some some party l then the measurements performed by agent i are never in the k-lookback of agent l. On the contrary, if $l - j \leq k$ then l will contain measurements performed by j in its $k-$lookback. Again are again able to find an induced subcomplex of the image scenario $\mathcal{C}_k(\mathcal{M})$ which is not Graham reducible. This time take $\sigma_1, \sigma_2, \sigma_3, \sigma_4 \in \Sigma$ where

$$\begin{array}{rcllrcl} \sigma_1[-1] & = & (i,0) & & & & \\ \sigma_2[-1] & = & (i,1) & & & & \\ \sigma_3[-1] & = & (l,0) & \& & \sigma_3[j] & = & (j,0) \\ \sigma_4[-1] & = & (l,0) & \& & \sigma_4[j] & = & (j,1) \end{array}$$

and where for every pair of sequences $\sigma_m[r] = \sigma_n[r]$ for all other agents r (not including i, j and l). Then the image scenario $\mathcal{C}_k(\mathcal{M})$ will again have a square as the induced subcomplex with vertices

$$\{\mathsf{lookback}_k(\sigma_1), \mathsf{lookback}_k(\sigma_2), \mathsf{lookback}_k(\sigma_3), \mathsf{lookback}_k(\sigma_4)\}$$

and maximal faces

$$\begin{cases} \{\mathsf{lookback}_k(\sigma_1), \mathsf{lookback}_k(\sigma_3)\} \\ \{\mathsf{lookback}_k(\sigma_1), \mathsf{lookback}_k(\sigma_4)\} \\ \{\mathsf{lookback}_k(\sigma_2), \mathsf{lookback}_k(\sigma_3)\} \\ \{\mathsf{lookback}_k(\sigma_2), \mathsf{lookback}_k(\sigma_4)\} \end{cases}$$

(ii) **necessity.** In order to prove necessity we prove the converse statement—that is, if the condition on the input set sizes and the separation between agents i and j does not hold then this implies that Σ'_k is acyclic, and therefore reducible. Therefore suppose first that we do not have at least two agents i and j with both $|I_i| \geq 2$ and $|I_j| \geq 2$. We denote measurements for i as (i, s) with $s \in I_i$ and same for j. We therefore have either only one agent in $\mathcal{A}$ with $I_i \geq 2$. Then the number of maximal sequences in Σ is given by $|I_i|$, and the image scenario has $|I_i|$ vertices: $\mathcal{C}_k(\sigma)$ for each $\sigma \in \Sigma$. This image is always acyclic. This is because vertices $v \in \Sigma'_k$, which we recall are elements of X^* and therefore string of measurements, will either (a) contain a single measurement of the form (j, s) or (b) not contain any (j, s)'s.

All vertices for which (b) holds will fall into a single context $C \in \Sigma'_k$. Vertices for which (a) holds will always lie in a single maximal context $C_{b_i} \in \Sigma'_k$, and therefore all vertices for which (a) holds can be removed in a Graham reduction step. This leaves us only with vertices of type (b). But since these lie in the single context C, and after removal of vertices (a) this is the only context removing, the complex is acyclic. Now suppose we do have $|I_i| \geq 2$ and $|I_j| \geq 2$ for some agents i and j, but that both conditions 1 and 2 do not hold. If conditions 1 and 2 do not hold then the vertices containing (j, s) always occur in a single maximal context, and can be removed. Likewise, vertices containing neither (j, s) nor (i, s') vertices occur in a single maximal context and can be removed in a Graham reduction step. This now leaves vertices containing (i, s) measurements. These cannot be in the same context, since they never occur in the same sequence. The result is an acyclic complex.

$\square$

Appendix E. Proof of Theorem 7.1

Proof. Suppose that $\mathcal{C}_k(\mathcal{M})$ is not acyclic, so that we prove the contrapositive. By Proposition 4.5 this implies that there are at least two agents q and p, $q \leq p$ with either (i) or (ii). Consider case (i), as case (ii) follows analogously. If (i) holds then we restrict $\mathcal{C}_k(\mathcal{M})$ to the vertices

$$\{\mathsf{lookback}_k((1,i_1)(2,i_2)...(p,i_p)), \mathsf{lookback}_k((1,i_1)(2,i_2)...(p,i'_p)),$$
$$\mathsf{lookback}_k((1,i_1)(2,i_2)...(q,i_q)), \mathsf{lookback}_k((1,i_1)(2,i_2)...(q,i'_q))\}$$

W.l.o.g. we take $i_p = i_q = 0$ and $i'_p = i'_q = 1$. The restriction of Σ'_k to this set of vertices, which we call V, will contain four faces of the form

$$\{\mathsf{lookback}_k((1,i_1)(2,i_2)...(p,l_p)), \mathsf{lookback}_k((1,i_1)(2,i_2)...(q,l_q))\}$$

for $l_p, l_q \in \{0,1\}$. In other words, the restricted complex is a square. We define an empirical model e on the restricted scenario to be a PR box. We now wish to define an empirical model d on $\mathcal{C}_k(\mathcal{M})$ which simulates e and which pullsback to give a $\mathsf{lookback}^{(e)}_k$-nonclassical empirical model on $\mathcal{M}$. Note that since $p - q > k$ we have that $\mathsf{lookback}_k((1,i_1)(2,i_2)...(p,l_p))$ and $\mathsf{lookback}_k((1,i_1)(2,i_2)...(q,l_q))$ do not overlap on any agents. Define the empirical model d such that in any context C the nodes not in V are deterministically sent to 0. We would like to show that the pullback $C^*_k(d)$ is $\mathsf{E}_{\mathsf{lookback}^{(e)}_k}-$nonclassical. Since it is guaranteed to be $\mathsf{E}_{\mathsf{lookback}_k}-$nonclassical (from Corollary 4.1), the set of $\mathsf{lookback}_k$-consistent global sections consistent with the support of $\mathcal{C}^*_k(d)$ is empty. However, since if a global strategy is $\mathsf{lookback}^{(e)}_k$-consistent this implies that it is $\mathsf{lookback}_k$-consistent (because if there is agreement on the last k measurement events then there is agreement on the last k measurements), this implies that the set of global $\mathsf{lookback}^{(e)}_k$-consistent strategies is empty, and thus $\mathcal{C}^*_k(d)$ is $\mathsf{E}_{\mathsf{lookback}^{(e)}_k}$-nonclassical.

$\square$

Appendix F. A Constructive Proof of Proposition 4.4 for $N = 2$ Agents

We give a constructive proof of Proposition 4.4 in the main text, showing that, when full lookback is allowed, every empirical model for two parties for produces correlations which are classically realisable. We therefore consider the unrestricted sheaf of strategies $\mathcal{S}$, noting that for $k \geq N-1$ $\mathcal{S}(U) \cong \mathcal{S}_k(U)$.

We specifically consider temporal Bell scenarios, and since for this case our framework is equivalent to that in [14], we use theirs for clarity of presentation.

We begin with $N = 2$ parties in order to gain some intuition in the simplest setup, and then show how this can be generalised. First we explicitly write down what a global section is. It is a choice of distribution $d \in \mathcal{D}\left((I_A \to O_A) \times (I_A \times I_B \to O_B)\right)$. An empirical model is realised from a global section as follows

(i) Two functions $f : I_A \to O_A$ and $g : I_A \times I_B \to O_A \times O_B$ are sampled from the global distribution d. (f, g) can be thought of as specifying the hidden variable for a particular run of the experiment.

(ii) Given an input $(i, j) \in I_A \times I_B$, the outputs R and S are evaluated as $R = f(i)$ and $S = g(i, j)$

(iii) (R, S) is returned.

As a result, the empirical model, if arising from a global section d, will have probabilities

$$e_d(i,j)[R,S] = \sum_{f,g.R=f(i),S=g(i,j)} d[f,g] \tag{F.1}$$

On the other hand, empirical data is obtained from observation as follows (this can be done whether a global section exists or not). Actually observing the system gives two distributions:

$$p : I_A \to \mathcal{D}(O_A) \tag{F.2}$$

$$q : I_A \times I_B \times O_A \to \mathcal{D}(O_B) \tag{F.3}$$

From this we get an empirical model $e_{p,q} : I_A \times I_B \to \mathcal{D}(O_A \times O_B)$ as follows

(i) Sample R from the distribution $p(i)$ and S from the distribution $q(i, j, R)$.

(ii) Return (R, S)

Then the empirical model is defined $e_{p,q}(i,j)[R,S] = p(i)[R]q(i,j,R)[S]$.

Given (p, q), how might we define d? Take it to be

$$d[f,g] = \prod_{i \in I_A, j \in I_B} p(i)\,[f(i)]\,q(i,j,f(i))\,[g(i,j)]\,. \tag{F.4}$$

We would then like to check if

(i) The distribution d is normalised.

(ii) Defining d in this way results in $e_d(i,j)[R,S] = e_{p,q}(i,j)[R,S]$. i.e. the observed distributions are consistent with marginalisation of the global section.

Function	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
g_0	0	0	0	0
g_1	0	0	0	1
g_2	0	0	1	0
...	0	...	...	...
g_8	0	1	1	1
g_9	1	0	0	0
...	1	...	...	...
g_{16}	1	1	1	1

Table F1: Function definitions for two agents. A function $g_i : I_A \times I_B \to O_A \times O_B$ is a deterministic global section specifying, for every pair of inputs (i, j) for the parties A and B, an output $g_i(i, j)$. There are 16 such distinct functions.

Appendix F.1. Normalisation

In order to check that d is a valid distribution we should first check that

$$\sum_{f,g} \prod_{i \in I_A} p(i)\,[f(i)] \prod_{j \in I_B} q(i, j, f(i))[g(i, j)] = 1. \tag{F.5}$$

First we note that since p and q are by assumption valid distributions, we have that for any $i \in I_A$,

$$\sum_{o \in O_A} p(i)[o] = 1 \tag{F.6}$$

and for any (i, j, R),

$$\sum_{o \in O_B} q(i, j, R)[o] = 1. \tag{F.7}$$

This will be used in simplifying expressions in what follows. We begin by considering, for fixed f, the sum over g and products over outputs. First consider terms with fixed f– so we take $f := f_0$ (see table F1 on definitions of functions). We then obtain

$$\begin{aligned}\sum_{z=0}^{15} &(p(0)[f_0(0)]q(0, 0, f_0(0))[g_z(0, 0)]q(0, 1, f_0(0))[g_z(0, 1)] \\ &\cdot p(1)[f_0(1)]q(1, 0, f_0(1))[g_z(1, 0)]q(1, 1, f_0(1))[g_z(1, 1)])\end{aligned} \tag{F.8}$$

In order to simplify this, note that g_0 and g_8 agree on all inputs except $(0, 0)$, where one is equal to 1 and the other to 0. The same applies to g_1 and g_9, g_2 and g_{10}, and so on.

Focusing only on g_0 and g_8 for now, we obtain

$$
\begin{aligned}
&p(0)[f_0(0)]q(0,0,f_0(0))[g_0(0,0)]q(0,1,f_0(0))[g_0(0,1)]\\
&\cdot p(1)[f_0(1)]q(1,0,f_0(1))[g_0(1,0)]q(1,1,f_0(1))[g_0(1,1)]+\\
&p(0)[f_0(0)]q(0,0,f_0(0))[g_8(0,0)]q(0,1,f_0(0))[g_8(0,1)]\\
&\cdot p(1)[f_0(1)]q(1,0,f_0(1))[g_8(1,0)]q(1,1,f_0(1))[g_8(1,1)]\\
=&p(0)[f_0(0)]p(1)[f_0(1)]\\
&\quad \cdot q(0,1,f_0(0))[g_{0,8}(0,1)]q(1,0,f_0(1))[g_{0,8}(1,0)]q(1,1,f_0(1))[g_{0,8}(1,1)]\\
&\quad (q(0,0,f_0(0))[g_0(0,0)]+q(0,0,f_0(0))[g_8(0,0)])\\
=&p(0)[f_0(0)]p(1)[f_0(1)]\\
&\quad \cdot q(0,1,f_0(0))[g_{0,8}(0,1)]q(1,0,f_0(1))[g_{0,8}(1,0)]q(1,1,f_0(1))[g_{0,8}(1,1)]\\
&\quad (q(0,0,f_0(0))[0]+q(0,0,f_0(0))[1])\\
=&p(0)[f_0(0)]p(1)[f_0(1)]\\
&\quad \cdot q(0,1,f_0(0))[g_{0,8}(0,1)]q(1,0,f_0(1))[g_{0,8}(1,0)]q(1,1,f_0(1))[g_{0,8}(1,1)]\\
&\quad (1)
\end{aligned}
$$

Here, we use $g_{0,8}$ to indicate that either g_0 or g_8 could be used to evaluate the input and both would give the same answer (note thtat this can only be done on inputs on which they agree). The first equality then follows from factorisation based on this observation, the second by noting from the table that $g_0(0,0)=0$ and $g_8(0,0)=1$, and the last from normalisation of the distribution $q(1,1,f_0(0))$. The same reasoning can be applied when grouping together terms $z=1$ and $z=9$, $z=2$ and $z=10$, ..., and $z=7$ and $z=18$ in Equation (23). Note that we now, for example, have an expression involving g_0 (or g_8) evaluated only on 01, 10 and 11. In order to get rid of another one of these terms, note that $g_{1,9}$ agrees with $g_{0,8}$ on $(0,1)$ and $(1,0)$ but not $(1,1)$. Therefore we find

$$
\begin{aligned}
&p(0)[f_0(0)]p(1)[f_0(1)]\\
&\quad \cdot q(0,1,f_0(0))[g_{0,8,1,9}(0,1)]q(1,0,f_0(1))[g_{0,8,1,9}(1,0)]\\
&\quad \cdot (q(1,1,f_0(1))[g_{0,8}(1,1)]+q(1,1,f_0(1))[g_{1,9}(1,1)])\\
=&p(0)[f_0(0)]p(1)[f_0(1)]\\
&\quad \cdot q(0,1,f_0(0))[g_{0,8,1,9}(0,1)]q(1,0,f_0(1))[g_{0,8,1,9}(1,0)]\\
&\quad \cdot (q(1,1,f_0(1))[0]+q(1,1,f_0(1))[1])\\
=&p(0)[f_0(0)]p(1)[f_0(1)]\\
&\quad \cdot q(0,1,f_0(0))[g_{0,8,1,9}(0,1)]q(1,0,f_0(1))[g_{0,8,1,9}(1,0)]\,(1)
\end{aligned}
$$

Again using normalisation of $q(0,1,f_0(1))$. We get the same type of terms for the groupings $(2,3,10,11)$, $(4,5,12,13)$ and $(6,7,14,15)$. Now, however, we are only evaluating g_z terms on the inputs $(0,1)$ and $(1,0)$. Looking at the groupings $(0,8,1,9)$ and $(2,3,10,11)$ (i.e. looking at g_z with z taken to be equal to one representative in the group, such as taking g_0 and g_2 for example) we find that they agree on $(0,1)$ but not on $(1,0)$. The same is true for the grouping $(4,5,12,13)$ and $(6,7,14,15)$. Taking the first grouping, we find that

$$\begin{aligned}
&p(0)[f_0(0)]p(1)[f_0(1)]\\
&\cdot q(0,1,f_0(0))[g_{0,8,1,9,2,3,10,11}(0,1)]\\
&(q(1,0,f_0(1))[g_{0,8,1,9}(1,0)] + (q(1,0,f_0(1))[g_{2,3,10,11}(1,0)])\\
&= p(0)[f_0(0)]p(1)[f_0(1)]\\
&\cdot q(0,1,f_0(0))[g_{0,8,1,9,2,3,10,11}(0,1)]\,(q(1,0,f_0(1))[0] + (q(1,0,f_0(1))[1])\\
&= p(0)[f_0(0)]p(1)[f_0(1)]q(0,1,f_0(0))[g_{0,8,1,9,2,3,10,11}(0,1)]
\end{aligned}$$

Similarly taking the second grouping we get

$$\begin{aligned}
&p(0)[f_0(0)]p(1)[f_0(1)]\\
&\cdot q(0,1,f_0(0))[g_{4,5,12,13,6,7,14,15}(0,1)]\\
&(q(1,0,f_0(1))[g_{4,5,12,13}(1,0)] + (q(1,0,f_0(1))[g_{6,7,14,15}(1,0)])\\
&= p(0)[f_0(0)]p(1)[f_0(1)]\\
&\cdot q(0,1,f_0(0))[g_{4,5,12,13,6,7,14,15}(0,1)]\,(q(1,0,f_0(1))[0] + (q(1,0,f_0(1))[1])\\
&= p(0)[f_0(0)]p(1)[f_0(1)]q(0,1,f_0(0))[g_{4,5,12,13,6,7,14,15}(0,1)]
\end{aligned}$$

Finally, adding these two terms together we get

$$\begin{aligned}
&p(0)[f_0(0)]p(1)[f_0(1)]q(0,1,f_0(0))[g_{0,8,1,9,2,3,10,11}(0,1)]\\
&+ p(0)[f_0(0)]p(1)[f_0(1)]q(0,1,f_0(0))[g_{4,5,12,13,6,7,14,15}(0,1)]\\
&= p(0)[f_0(0)]p(1)[f_0(1)]\,(q(0,1,f_0(0))[g_{0,8,1,9,2,3,10,11}(0,1)]\\
&+q(0,1,f_0(0))[g_{4,5,12,13,6,7,14,15}(0,1)])\\
&= p(0)[f_0(0)]p(1)[f_0(1)]\,(1)\\
&= p(0)[f_0(0)]p(1)[f_0(1)]
\end{aligned}$$

This will be true for every term in the sum over f, therefore we obtain

$$\begin{aligned}
&p(0)[f_0(0)]p(1)[f_0(1)] + p(0)[f_1(0)]p(1)[f_1(1)]\\
&p(0)[f_2(0)]p(1)[f_2(1)] + p(0)[f_3(0)]p(1)[f_3(1)]\\
&= p(0)[f_{0,1}(0)](p(1)[f_0(1)] + p(1)[f_1(1)])\\
&+ p(0)[f_{2,3}(0)](p(1)[f_2(1)] + p(1)[f_3(1)])\\
&= p(0)[f_{0,1}(0)] + p(0)[f_{2,3}(0)]\\
&= 1
\end{aligned}$$

where we have used normalisation of $p(0)$ and $p(1)$ to simplify the expression. It therefore follows that d is normalised.

Appendix F.2. Empirical Agreement

Next we need to check that e_d and $e_{p,q}$ empirically agree. In order for this to happen we must have that

$$\begin{aligned} & e_d(i,j)[R,S] = e_{p,q}(i,j)[R,S] \\ \implies & p(i)[R]q(i,j,R)[S] = \sum_{f,g:R=f(i),S=g(i,j)} \prod_{i' \in I_A} p(i')[f(i')] \prod_{j \in I_B} q(i',j',f(i'))[g(i',j')] \end{aligned}$$

Without loss of generality take $i = j = 0$, $R = S = 0$. Thus we are only summing over functions f_k with $0 = f_k(0)$ and g_r with $0 = g_r(0,0)$. We therefore have $k \in \{0,1\}$ and $r \in \{0,1,2,3,4,5,6,7\}$. We then proceed to group together terms as before, but noting that we can only group together terms in the sum if they do occur. In particular, again fixing a choice of f (say $f = f_0$), we must proceed differently in summing over g. Begin by grouping single terms into the pairs $(0,1),(2,3),(4,5),(6,7)$, which will get rid of the $g_z(1,1)$ terms (since for example g_0 and g_1 agree on all inputs except $(1,1)$).

This leaves us with an expression

$$\begin{aligned} & p(0)[f_0(0)]p(1)[f_0(1)] \\ & (q(0,0,f_0(0))[g_{0,1}(0,0)]q(0,1,f_0(0))[g_{0,1}(0,1)]q(1,0,f_0(1))[g_{0,1}(1,0)] + ...) \end{aligned}$$

where ... indicates similar expressions for all other groupings (just replacing the subindex).

Next we can group $(0,1)$ and $(2,3)$, and $(4,5)$ and $(6,7)$. This gets rid of the $(i,j) = (1,0)$ terms as for example g_0 and g_2 are different on the inputs $(1,0)$. This gives the expression

$$\begin{aligned} & p(0)[f_0(0)]p(1)[f_0(1)] \\ & (q(0,0,f_0(0))[g_{0,1,2,3}(0,0)]q(0,1,f_0(0))[g_{0,1,2,3}(0,1)] + ...) \end{aligned}$$

Finally, we can group together the $(0,1,2,3)$ and $(4,5,6,7)$ functions getting rid of the $(i,j) = (0,1)$ input terms:

$$\begin{aligned} & p(0)[f_0(0)]p(1)[f_0(1)] \\ & (q(0,0,f_0(0))[g_{0,1,2,3,4,5,6,7}(0,0)]) \end{aligned}$$

This is true for both f_0 and f_1 (since we do not consider other f's in the sum). This gives

$$\begin{aligned} & p(0)[f_0(0)]p(1)[f_0(1)]q(0,0,f_0(0))[g_{0,1,2,3,4,5,6,7}(0,0)] \\ & + p(0)[f_1(0)]p(1)[f_1(1)]q(0,0,f_1(0))[g_{0,1,2,3,4,5,6,7}(0,0)] \\ & = p(0)[f_0(0)]p(1)[f_0(1)]q(0,0,f_0(0))[0] \\ & + p(0)[f_1(0)]p(1)[f_1(1)]q(0,0,f_1(0))[0] \\ & = p(0)[0]q(0,0,0)[0](p(1)[0] + p(1)[1]) \\ & = p(0)[0]q(0,0,0)[0] \end{aligned}$$

Therefore showing empirical agreement for $i = j = R = S = 0$. Other choices for these variables lead to equality of the statement in a similar fashion.

Function	$(0,0,...,0)$	$(0,0,...,0,1)$	...	$(1,1,...,1,0)$	$(1,1,...,1,1)$
$f_0^{(k)}$	0	0	...	0	0
$f_1^{(k)}$	0	0	...	0	1
$f_2^{(k)}$	0	0	...	1	0
...	0	...	...	...	...
$f_{2^k/2}^{(k)}$	0	1	...	1	1
$f_{2^k/2+1}^{(k)}$	1	0	...	0	0
...	1	...	...	...	...
$f_{2^k}^{(k)}$	1	1	...	1	1

Table G1: Function definitions for N agents. A function $f_i^{(k)} : I_A \times I_B \to O_A \times O_B$ is a deterministic global section specifying, for every pair of inputs (i, j) for the parties A and B, an output $g_i(i, j)$. There are 16 such distinct functions.

Appendix G. A Constructive Proof of Proposition 4.4 for N Agents

We now generalise the proof in the previous Appendix to N parties. We restate the Proposition from the main text and then show how it follows by induction.

Proposition Appendix G.1. *Given a temporal Bell scenario $\mathcal{M}$, every $\mathsf{lookback}_k$-consistent empirical model $e \in \mathsf{EM}^{(k)}(\mathcal{M})$ is E_{lookback_k} classical for $k \geq N-1$.*

Proof. The proof will proceed by induction. Assume therefore that the statement is true for $z-1$ parties, so that we must show that it holds for z parties. Label the input of party $i \in [1, z]$ as I_i, and let the distribution $p_i : I_1 \times ... \times I_i \times .. \times O_{i-1} \to \mathcal{D}(O_i)$. Moreover, write $f^{(i)} : I_1 \times ... \times I_i \to O_i$ for the function sampled from the global distribution d.

We then would like to show that given any empirical model

$$e_{p_1,...,p_z}(i_1, i_2, ..., i_z)[R_1, ..., R_N] = p_1(i_1)[R_1]p(i_1, i_2, R_1)[R_2]...p_k(i_1, ..., i_z.R_1, ..., R_{z-1})[R_l], \tag{G.1}$$

it could always have arisen from marginalisation of a global section d^z which we define as follows

$$d^l[f^{(1)}, f^{(2)}, ..., f^{(z)}] = \prod_{i_1 \in I_1, ..., i_z \in I_z} p_1(i_1)[f^{(1)}(i_1)], ..., p_z(i_1, ..., i_z, f^{(1)}(i_1), ..., f^{(z-1)}(i_{z-1}))[f^z(i_1, ..., i_z)] \tag{G.2}$$

We again need to check

(i) Normalisation
(ii) Empirical Agreement

Normalisation. We must check that $\sum_{f^{(1)},f^{(2)},...,f^{(z)}} d^z[f^{(1)}, f^{(2)}, ..., f^{(z)}] = 1$. We begin by considering only certain values in the sum for all i except i_z. Therefore fix $f^{(i)} = f_0^{(i)}$, where $f_0^{(i)}$ sends all arguments to zero. We are then left with the expression

$$\prod_{i_1\in I_1,...} p_1(i_1)[f^{(1)}(i_1)],...,\times \sum_{f^{(z)}} \prod_{i_1\in I_1,...i_z\in I_z} p_l(i_1,...,i_z,f^{(1)}(i_1),...,f^{(z-1)}(i_{z-1}))[f^z(i_1,...,i_z)] \tag{G.3}$$

By the way that we define $f^{(z)}$ (see Table G1), the functions $f_0^{(z)}$ and $f_1^{(z)}$ agree on all inputs strings except for the one which contains all ones, in which case $f_0^{(z)}$ returns 0 and $f_1^{(z)}$ returns 1.

We therefore have that those two terms in the sum become

$$\prod_{i_1\in I_1,...i_z\in I_z\backslash(1,1,...,1)} p_z(i_1,...,i_z,f_0^{(1)}(i_1),...,f_0^{(z-1)}(i_{z-1}))[f^z_{0,2^z/2+1}(i_1,...,i_z)] \tag{G.4}$$

$$\times\Big(p_z(1,...,1,f_0^{(1)}(1),...,f_0^{(z-1)}(1))[f_0^z(1,...,1)] + p_z(1,...,1,f_0^{(1)}(1),...,f_0^{(z-1)}(1))[f_1^z(1,...,1)]\Big) \tag{G.5}$$

$$= \prod_{i_1\in I_1,...i_z\in I_z\backslash(1,1,...,1)} p_z(i_1,...,i_z,f_0^{(1)}(i_1),...,f_0^{(z-1)}(i_{z-1}))[f^z_{0,1}(i_1,...,i_z)]\times 1 \tag{G.6}$$

where the equality follows from normalisation of the probability distribution p_z and the fact that $f_0^{(z)}$ and $f_1^{(z)}$ differ on their mappings of $(1,1,...,1)$.

The same is true for the pair of functions $f_2^{(z)}$ and $f_3^{(z)}$, and more generally $f_i^{(z)}$ and $f_{i+1}^{(z)}$. In this manner, we eliminate the parts of the terms containing the string with all ones as input. We next note that ignoring the 1111..1 input, as that has now been eliminated, functions $f_0^{(z)}/f_1^{(z)}$ and $f_2^{(z)}/f_3^{(z)}$ agree on all inputs strings except for $(1,1,1,...,1,0)$, the one which contains all ones except for the last entry, which is a zero, in which case $f_{0,1}^{(z)}$ returns 0 and $f_{2,3}^{(z)}$ returns 1. In this manner we can eliminate the terms with $(1,1,1,...,1,0)$. This grouping process continues, so that we end up with groups of $4(=2^2)$ and then $8(=2^3)$, and so on, until we arrive at 2^{z-1}, in which the functions will be split into two groups. We will have $f_i^{(z)}$ with $1\in[0,2^z/2]$ in the first and $f_i^{(z)}$ with $i\in[2^z/2+1,2^z]$ in the second. Disregarding all inputs strings (which have by now been eliminated in previous groupings) except for the string of all zeros $(0,0,...,0)$, we see that the first group of functions maps this string to 0 while the second maps it to 1. This then eliminates all input string and we are left with

$$\prod_{i_1\in I_1,...} p_1(i_1)[f^{(1)}(i_1)],..., \tag{G.7}$$

Recall however that the original expression included a sum over $f^{(1)},...,f^{(z-1)}$. Including this gives

$$\sum_{f^{(1)},f^{(2)},...,f^{(z-1)}} \prod_{i_1\in I_1,...} p_1(i_1)[f^{(1)}(i_1)],..., \tag{G.8}$$

This, however, is equal to 1 because it is the sum over all arguments of d^{z-1} which is by assumption a valid probability distribution. Thus d^z is a valid probability distribution.

Empirical Agreement Next we would like to check empirical agreemnt i.e. we would like to check that

$$e_{d^z}(i_1, ..., i_z)[R_1, ..., R_z] = e_{p_1,...,p_z}(i_1, ..., i_z)[R_1, ..., R_z]. \tag{G.9}$$

If this is true, then we must have (by definition of the above) that

$$p_1(i_1)[R_1]p_2(i_1, i_2, R_1)[R_2] = \tag{G.10}$$

$$\sum_{\substack{f^{(1)},f^{(2)},...,:\\ R_1=f^{(1)}(i_1)\\ R_2=f^{(2)}=f^{(2)}(i_1,i_2)\\ ...}} \prod_{i'_1\in I_1} p_1(i'_1)[f^{(1)}(i'_1)] \prod_{i'_2\in I_2} p_2(i'_1, i'_2, f^{(1)}(i'_1))[f^{(2)}(i'_1, i'_2)]... \tag{G.11}$$

$$\prod_{i'_z\in I_z} p_z(i'_1, i'_2, ..., f^{(z-1)}(i'_1, ..., f^{(z-2)}(i'_1, ..., i'_{z-2})))[f^{(z)}(i'_1, ..., i'_z)] \tag{G.12}$$

We will again prove equality of Equation G.9 for arbitrary z using proof by induction. Therefore, assume it is true for $z-1$. Also for now consider at first the special case $i_1 = i_2 = ... = i_z = 0$ and $R_1 = R_2 = ... = R_z = 0$. Therefore, due to the condition on the sum, we only include in that sum functions with $f_l^{(z)}(0, 0, ..., 0) = 0$.

We begin by fixing the choice of $f^{(z-1)}, ..., f^{(1)}$. We choose $f_1^{(z-1)}, ..., f_1^{(1)}$ for example. We then, for $f^{(z)}$, group together its realisations with subindex $(1, 2)$, $(3, 4)$, ..., and $(\frac{2^z}{2} - 1, \frac{2^z}{2})$ (since any other choices do not map $(0, 0, ..., 0)$ to 0).

In an analogous fashion to the proof for normalisation, such a grouping together with conditions of normalisation of the p_i distributions gets rid of $f_l^{(z)}(1, 1, ..., 1)$ terms (since for example $f_0^{(z)}$ and $f_1^{(z)}$ agree on all inputs except $(1, 1, ..., 1)$). It is useful to describe an equivalence relation on the set of functions $f^{(z)} := \{f_i^{(z)} | i \in [1, 2^z]\}$ in order to formally describe these groupings. To this end, we let R_S, with $S \subseteq \prod_{i\in[1,z]} I_i$, be a relation on $f^{(z)}$ defined as follows

$$f_j^{(z)} R_S f_l^{(z)} \iff f_j^{(z)}(m) = f_l^{(z)}(m) \text{ when } m \notin S \text{ and } f_j^{(z)}(m) \neq f_l^{(z)}(m) \text{ when } m \in S. \tag{G.13}$$

Therefore the grouping of $(1, 2)$ etc. terms in $f^{(z)}$ results in summation instead over the quotient set $f^{(z)} \backslash R_{\{(1,1,...,1)\}}$. We then have the expression

$$\left(\prod_{i'_1\in I_1} p_1(i'_1)[f_1^{(1)}(i'_1)] \prod_{i'_2\in I_2} p_2(i'_1, i'_2, f_1^{(1)}(i'_1))[f_1^{(2)}(i'_1, i'_2)]... \right) \tag{G.14}$$

$$\sum_{\substack{f^{(z)}\backslash R_{(1,1,...,1)}:\\ 0=f^{(z)}((0,0,...,0))}} \prod_{\substack{(i_1,...,i_z)\in I_1\times...\times I_z\\ \backslash\{(1,1,...,1)\}}} p_z(i'_1, i'_2, ..., f_1^{(z-1)}(i'_1, ..., f_1^{(z-2)}(i'_1, ..., i'_{z-2})))[f^{(z)}(i'_1, ..., i'_z)] \tag{G.15}$$

The next grouping will be of functions which agree on all inputs except $(1, 1, ..., 1)$ and $(1, 1, ..., 1, 0)$. More generally, given a set S resulting in the grouping $f^{(z)} \backslash R_S$, we get an expression of the form

$$\left(\prod_{i_1' \in I_1} p_1(i_1')[f_1^{(1)}(i_1')] \prod_{i_2' \in I_2} p_2(i_1', i_2', f_1^{(1)}(i_1'))[f_1^{(2)}(i_1', i_2')] ... \right) \tag{G.16}$$

$$\sum_{\substack{f^{(z)} \backslash R_S: \\ 0 = f^{(z)}((0,0,...,0))}} \prod_{\substack{(i_1,...,i_z) \in I_1 \times ... \times I_z \\ \backslash S}} p_z(i_1', i_2', ..., f_1^{(z-1)}(i_1', ..., f_1^{(z-2)}(i_1', ..., i_{z-2}')))[f^{(z)}(i_1', ..., i_z')] \tag{G.17}$$

Specifically, when $S = \{(i_1, ..., i_z) \in I_1 \times ... \times I_z | i_1 \neq 1\}$, we get the following

$$\left(\prod_{i_1' \in I_1} p_1(i_1')[f_1^{(1)}(i_1')] \prod_{i_2' \in I_2} p_2(i_1', i_2', f_1^{(1)}(i_1'))[f_1^{(2)}(i_1', i_2')] ... \right) \tag{G.18}$$

$$p_z(0, 0, ..., 0, ..., 0))[0] \tag{G.19}$$

Note that the term in parentheses is $e_{d^{z-1}}(i_1, ..., i_{z-1})[R_1, ..., R_{z-1}]$, and since we have empirical agreement for $z - 1$ parties by assumption, then the expression becomes

$$p_1(0)[0]p_2(0, 0, 0)[0]...p_z(0, 0, ..., 0, ..., 0))[0] \tag{G.20}$$

The same holds for arbitrary choice of $i_1, ..., i_z$ and $R_1, ..., R_S$, as can be checked. We therefore have empirical agreement.

Now, since we know that the proposition holds for $z = 2$, by induction it holds for all N.

$\square$