Master of Operations Research, Combinatorics and Optimization

Master Informatique / Master Mathématiques & Applications

Contextuality as Automata: Open Generators and the Dynamics of Quantum Phenomena

Louis TRIOULEYRE--ROBERJOT

2026-06-24

This internship was done at the *Laboratory of Informatics of Grenoble* (LIG) in the *CAPP - Calculi, Algorithms, Programs and Proofs* team.

Is under the supervision of *Mehdi Mhalla*.

Defended before a jury composed of:

↪ Van-Dat Cung
↪ Nadia Brauner
↪ Hippolyte Dourdent

How an actual entity becomes constitutes what that actual entity is; so that the two descriptions of an actual entity are not independent. Its 'being' is constituted by its 'becoming.' This is the 'principle of process'.

— Alfred North Whitehead

This work is supported by the French National Research Agency in the framework of the "France 2030" program: ANR 15 IDEX 02 and ANR-11-LABX-0025-01 for the LabEx PERSYVAL.

Acknowledgements

I would like to thank Mehdi Mhalla for his constant presence, his trust, and the many discussions that shaped this work. This internship was an intense and deeply enriching experience. It allowed me to discover research in a very direct way: starting almost from a blank page, trying to understand an intuition, struggling with it, and slowly seeing a direction, a language, and a form of understanding emerge.

Abstract

This report develops a discrete and dynamic approach to contextuality. The starting point is operational: an experiment is first recorded as finite events before it is normalized into a probability table. Normalization is useful for prediction, but it erases the level of evidence, the arithmetic constraints imposed by finite counts, and the possibility of viewing an observation as a partial process that may still be completed. We therefore replace probability tables by integer countings. This is not a mere change of scale. Integer counts preserve the level of evidence, the arithmetic constraints of finite repetition, and the temporal structure of how observations are produced. Most importantly, they make it possible to define objects that have no analogue in the probabilistic framework: the *contextual completion deficit*, which measures how much work remains to close a partial explanation, and the *interruption profile*, which records how each coordinate of an observation has advanced in its past-to-future trajectory. After normalization, these dynamic quantities reduce to the standard contextual fraction. Before normalization, they carry strictly more information. The work is situated in the sheaf-theoretic approach to contextuality, where contextuality is an obstruction to gluing local data into a global section [AB11], and in the theory of the contextual fraction [ABM16]. Compatible observations form an affine semigroup, deterministic explanations become integer sums of global sections, and open generators record the residual future required to close a partial explanation. The guiding motivation comes from discussions with Mehdi MHALLA: understanding is not merely the possession of a final model, but the capacity to interrupt a process, distinguish what is already stabilized from what remains open, and reconstruct its admissible futures. In this sense, the report is a mathematical test of a broader idea: a scientific explanation should keep track of how evidence is produced, stabilized, interrupted, and completed. Contextuality provides a sharp laboratory for this idea, because it separates local observational consistency from the existence of a global classical explanation.

The report is organized around this passage from static records to dynamic explanations. In Chapter 1, we motivate the operational stance: start from observations, keep the finite record, and ask what hidden structure could have produced it. In Chapter 2, we recall the sheaf-theoretic formulation of contextuality and translate it into integer countings. The first contribution is this *translation itself*. By replacing probabilities with integer counts, contextuality becomes a question about integer decomposition in a semigroup. This shift is motivated by expressiveness: the probabilistic framework erases the level of evidence and the temporal order of production, while the integer framework keeps them. The contextual fraction is recovered as a normalized special case, but the integer setting also admits the *completion deficit* and the *interruption profile*, which have no probabilistic analogue. In Chapter 3, we introduce the second contribution: *dynamic contextual automata*. A generator is treated as an elementary process, while an interrupted experiment is represented by a dynamic state with closed copies and open copies. This turns partial observations into meaningful mathematical objects. In Chapter 4, we apply the framework to the $(2, 2, 2)$ Bell scenario. We define the finite-level Tsirelson bound $T(t)$ as the best quantum-realizable integer CHSH score at level t. This makes explicit an arithmetic point: every finite level gives a rational score, and the Tsirelson value $2\sqrt{2}$ appears only as an asymptotic limit. Using results from the NPA hierarchy, we approximate this bound and show how the integer framework connects to standard quantum relaxations. In Chapter 5, we compare

our new framework with the contextual fraction of [ABM16]. The closed deficit of an integer counting recovers the contextual fraction after normalization, so the dynamic framework agrees with the standard sheaf-theoretic measure on its static coordinate. In Chapter 6, we explore several dynamic constraints for characterizing quantum correlations. The investigation tests past/future product constraints, the interruption profile k_η that records the per-coordinate temporal structure of an explanation, and a constraint-programming formulation for computing admissible explanations. Each constraint reveals different aspects of the classical-quantum boundary, and the exploration shows how the dynamic framework enriches the standard contextual fraction with temporal structure.

The integer framework recovers the contextual fraction and the sheaf-theoretic obstructions in a direct, constructive, and finite way. It also introduces a genuinely dynamic layer: the interruption profile records the internal temporal structure of a process, not only its global outcome. This construction applies whenever a process can be decomposed into elementary steps and interrupted at intermediate stages, allowing the framework to distinguish observations with the same statistics but different underlying dynamics.

Contents

Context and Motivation

CHAPTER 1

1.1. Why start from observation?

Scientific theories do not begin with reality given all at once. They begin with phenomena: outcomes, traces, correlations, and reports produced through concrete experimental interfaces. A hidden variable, a state, a Hilbert space vector, or an automaton is therefore not direct access to reality itself. It is a model introduced to organize what is observed. We do not try to reveal a final hidden world. We ask which hidden structures can explain the visible data, how much structure they require, and what they still fail to explain. In that sense, the word **model** should be read operationally. A model is a disciplined way of connecting observed phenomena to a possible hidden organization. On the contrary, it makes the task more precise. Since the hidden structure is not directly given, it must be reconstructed under constraints. It must fit the observations, respect the operational limitations of the experiment, and use a controlled amount of structure. A model is then judged by its explanatory power, and by its coherence.[†]

1.2. Truth, understanding, and interruption

Truth is used in an operational and provisional sense. A model is accepted when it is the best available explanation of the observed phenomena. If the observations change, or if the class of admissible models changes, the accepted model may change as well. This does not make truth arbitrary. The choice of a model is constrained by data, by mathematical consistency, and by the cost of the hidden resources used in the explanation.

Understanding is slightly different. A model may give correct predictions while remaining too opaque to reuse. Here, a model is understood when it can be compressed, transmitted, interrupted, and reconstructed without losing its explanatory role.

A simple analogy is curve fitting. Suppose that an experiment produces a sequence of points (x_i, y_i). In a classical analysis, one may fix a class of models, for instance polynomials of degree at most 3, and choose the polynomial that best fits the first n points: One then says that the model is good if the final error is small enough. This report asks for a more dynamic requirement. The explanation procedure itself should be observable and interruptible. After each new point, we obtain a new model, a new error, and a transition from the model at step n to the model at step $n+1$. A good explanatory process should not only end with a small error. Its successive updates should also follow a coherent pattern. This is the sense in which we want a kind of meta-model. We do not only want a final explanation of the completed experiment. We also want to track the history of

[†]A model must be small and coherent. A model that replicates predictions without being understood is still a bad model.

partial explanations, their errors, and their possible continuations. If the experiment is interrupted at time n, the current state should still have a meaning.

This point will become central later. A probability table is a completed summary. It has already forgotten the order of events, the number of occurrences before normalization, and the possible unfinished mechanisms that produced the data. We need a language that keeps enough information to restart the process after an interruption.

The report therefore takes seriously a simple methodological principle: before normalizing, count; before completing, keep the prefix; before replacing a mechanism by a probability distribution, ask what hidden process could have produced it.

1.3. *Existing routes toward the quantum model*

The question of which model best explains quantum phenomena is not new. Bell's theorem showed that local hidden-variable models cannot reproduce all quantum predictions [Bel64]. The CHSH inequality gave a particularly clear experimental and mathematical form of this obstruction [Cla+69]. These results changed the status of hidden mechanisms: the difficulty is not merely that the hidden variables are unknown, but that some families of local observations cannot be glued into a single classical global assignment.

Several lines of work explore whether general operational principles can characterize quantum theory, for instance through generalized probabilistic theories [Bar07] or axiomatic reconstructions [CDP11, Har01, MM11]. These works illustrate a difficulty: if the axioms are too weak, one obtains theories larger than quantum theory; if they are too strong, the reconstruction may simply hide the quantum structure inside the assumptions.

The same difficulty appears from another direction in the resource-theoretic work of [Ans+26]. A *resource theory* does not study physical systems only as isolated objects. It asks which transformations are possible when one restricts the allowed operations. In the deterministic classical setting considered there, a model f can be converted into a model g by free operations exactly when the image of f is at least as large as the image of g. This gives a simple order on deterministic mechanisms. The quantum case is harder. As [Ans+26] point out, a CPTP map is the quantum analogue of a stochastic map, not of a probability distribution over functions. This means that quantum Shannon theory, where CPTP maps play the role of channels, is not a resource theory of causal influence for the same reasons that classical Shannon theory is not one. The quantum analogue of a probability distribution over functions remains an open challenge, and finding it is necessary for developing a quantum resource theory of causal influence.

The example of non-signaling correlations makes this tension concrete. Popescu and Rohrlich showed that there are non-signaling correlations stronger than quantum correlations [PR94]. Thus, no-signaling alone does not characterize quantum theory. The quantum set lies between the classical local polytope and the larger non-signaling polytope. One open question is why nature seems to occupy this intermediate region.

A second difficulty appears even inside the mathematical formalism of quantum correlations. Different completions of the model are not equivalent like you could see bellow. One may use finite-dimensional tensor-product models, infinite-dimensional limits, or commuting-operator models. The relation between these descriptions is connected to Tsirelson's problem and to Connes' embedding problem [Jun+11].

The distinction has a physical meaning. In a *finite tensor-product* model, Alice and Bob are described by two separate Hilbert spaces, and the joint system is represented by $\mathcal{H}_A \otimes \mathcal{H}_B$. This is close to the usual laboratory picture of two separated devices. In the *commuting-operator* model, one uses a single Hilbert space $\mathcal{H}$ and only requires Alice's and Bob's observables to commute, for example $A_x B_y = B_y A_x$. This algebraic condition says that the two measurements are compatible: changing their order does not change the predicted statistics. It is natural in algebraic quantum field theory, where observables localized in spacelike separated regions are required to commute. However, it is less directly tied to a finite experimental mechanism than a concrete tensor decomposition.

The issue can be stated in terms of correlation sets. Let C_{q} denote the set of correlations obtained from finite-dimensional tensor-product strategies. Let C_{qs} denote the spatial tensor-product model, where the Hilbert spaces may be infinite-dimensional. Let $C_{\mathrm{qa}} := \overline{C_{\mathrm{q}}}$ denote the approximate quantum set, obtained by closing the finite-dimensional set. Finally, let C_{qc} denote the commuting-operator set. These sets satisfy the inclusions

$$C_{\mathrm{q}} \subset C_{\mathrm{qs}} \subset C_{\mathrm{qa}} \subset C_{\mathrm{qc}}.$$

The inclusions are not only technical details. Slofstra showed that finite-dimensional quantum correlations need not form a closed set, so one can have $C_{\mathrm{q}} \neq C_{\mathrm{qa}}$ [Slo19]. Tsirelson's problem asks, in one standard formulation, whether the approximate model and the commuting-operator model coincide, that is, whether $C_{\mathrm{qa}} = C_{\mathrm{qc}}$. The theorem MIP* = RE implies a negative answer: there are non-local games whose commuting-operator value cannot be approximated by finite-dimensional tensor-product strategies, hence $C_{\mathrm{qa}} \subsetneq C_{\mathrm{qc}}$ [Ji+22].

A parallel phenomenon appears when the probabilistic gap is removed. In the zero-gap formulation of MIP*, denoted MIP$^*_0$, the probabilistic gap is removed: the verifier accepts with probability exactly 1 or strictly less than 1. One might expect that removing the continuous probabilistic layer simplifies the problem. On the contrary, Mousavi, Nezhadi, and Yuen show that the resulting class is exactly coRE [MNY20]. Related work on commuting-operator formulations yields a similar computability-theoretic behaviour [Lin25]. Thus even after removing the probabilistic layer, the underlying structure remains rich. This is one motivation for the present report: interesting structure can persist at the level of exact acceptance, beyond probabilities.

These results do not decide which mathematical completion is physically real. They show a more modest point: the word "quantum" does not determine a unique operational object unless the allowed idealization is specified. A finite experiment gives finite data, with finite precision. It can rule out restricted finite-dimensional models, or give lower bounds on dimension, but it cannot by itself certify an exact infinite-dimensional Hilbert space or a genuinely non-tensor commuting-operator realization. Thus, the separation $C_{\mathrm{qa}} \subsetneq C_{\mathrm{qc}}$ is mathematically decisive, while its direct physical interpretation remains more delicate.

This motivates the constructive stance of the present work. The aim is not to decide the ontology of Hilbert spaces. It is to work at the level where the experiment is actually recorded: visible events, integer counts, compatibility equations, and finite generators. A property proved at this level has a direct *operational meaning* before any passage to closures, limits, or operator-algebraic completions. In this sense, staying close to the experimental record is not only a simplification. *It is a way to separate physically accessible structure from structure introduced by the mathematical completion of the model.*

1.4. *From state spaces to constructive counts*

Most operational reconstructions start with a space of states, a set of effects, and rules for composing systems (for example the General Probability Theory). This is natural when one wants to reconstruct the full probabilistic formalism. In this report, we take a different starting point. We begin with finite observations and integer counts.

The reason is simple. A normalized probability distribution erases the level at which it was obtained. For example, consider two possible events A and B. The distribution $\mathbb{P}(A) = 1$ and $\mathbb{P}(B) = 0$ may come from the count $(1, 0)$, from the count $(1000, 0)$, or from any count $(t, 0)$ with $t > 0$. After normalization, all these situations become identical. Empirically, however, they are not the same: one observation and one thousand observations do not carry the same amount of evidence. Keeping integer counts preserves this scale.

The second reason is that errors in an experiment come from two fundamentally different sources. The first source is **observational error**: imprecisions in the measurement, finite resolution of detectors, or fluctuations in repeated runs. The second source is **model error**: the choice of the theoretical function f that maps raw data to a reported quantity. To see why this distinction matters, consider a standard situation in physics. One measures a set of raw data $x = (x_1, ..., x_n)$, each carry some observational uncertainty σ_{x_i}. A theoretical model $f : \mathbb{R}^n \to \mathbb{R}$ then maps these data to a derived quantity $y = f(x_1, ..., x_n)$. The combined uncertainty on y is classically estimated by propagation of errors: $\sigma_y^2 = \sum_{i=1}^{n} \left|\frac{\partial f}{\partial x_i}\right|^2 \sigma_{x_i}^2$. Here, the σ_{x_i} reflect the first type of error, while the choice of f itself carries the second type. The problem is the following. If the experimental stores only the final value $f(x)$ and its propagated uncertainty σ_y, then the day the model f is revised or replaced by a more accurate function g derived from a new theory, then the entire experiment must be repeated. The raw data x have been discarded. The information that would have allowed recomputing $g(x)$ is gone. This is one operational motivation for delaying normalization and for maintaining a record at the level of integer counts and unprocessed observations: it retains the maximal amount of information that a later, better model can exploit.

This is where contextuality enters. In the sheaf-theoretic approach, contextuality is an obstruction to gluing local data into a global section [AB11]. This language is well suited to our purpose because it separates local observation from global explanation. In the discrete version used here, local probability tables are replaced by integer count tables. Compatibility becomes equality of marginal counts on overlaps. Noncontextuality becomes the possibility of reconstructing the compatible count as an integer sum of deterministic global assignments.

The change from probabilities to counts is not only cosmetic. It turns convex geometry into semigroup geometry. Instead of asking only whether a point lies in a convex hull, we ask how an integer count decomposes into irreducible generators. The Hilbert basis of the relevant semigroup then gives candidate elementary bricks of explanation. In CHSH, these bricks include local deterministic generators and lifted PR-type generators. This does not mean that PR boxes are accepted as physical final states. It means that they appear as elementary algebraic pieces of the compatible counting structure, and therefore must be interpreted.[‡]

[‡]In this reading, the generators are not physical objects but elementary possibilities: they express the conditions under which an event can exist and be observed.

1.5. Quantum phenomena as unfinished stabilization

This perspective gives a useful way to read quantum phenomena. A quantum state will not be treated here as a primitive object whose meaning is already fixed. We instead ask whether some quantum features can be understood as effects of projection and incomplete stabilization. Superposition suggests that several hidden histories may give the same visible phenomenon. Entanglement suggests that local visible pieces may fail to factorize because they still come from one common hidden process. This is deliberately only a **guiding interpretation**. The automaton model is the concrete way in which we make this interpretation usable: it represents observation as a process, allows this process to be interrupted, and permits different hidden paths to lead to the same visible count. In this language, a full observation is one that has been completely produced by the automaton. An incomplete observation is still meaningful: it records what has already been seen, while keeping track of what could still happen next. This remaining part is the information needed to complete the process.

The aim is not to replace Hilbert spaces. It is to build a small discrete language in which such ideas can be tested. Contextuality is used as a laboratory for a broader question: how can visible phenomena be reconstructed from hidden processes without pretending that the hidden process is directly given?

Technical Preliminaries

CHAPTER 2

In this section we will introduce the technical concept from generic contextuality to our all new theory of discrete contextuality.

2.1. What is contextuality ?

Contextuality is the failure of a family of local observations to come from one global assignment. In a *measurement scenario* $\langle X, \mathcal{M}, O\rangle$, each context $C \in \mathcal{M}$ specifies a set of measurements that can be performed together, and a local outcome is a *section* $s : C \to O$. A classical deterministic explanation would require a single global section $g : X \to O$ whose restriction $g|_C$ explains the outcome seen in every context. Thus contextuality is not merely randomness: it is an obstruction to gluing compatible local descriptions into one global description. This report uses this standard idea in a discrete form

2.1.1. Classical sheaf theory

In the following, we will use the same notation that in [AB11]. We will **add** our theory on the top of this formalisze and we will discuss how consider discrete contextuality with this notation. The goal is to model the experiment itself, rather than only the final probability distribution. We consider one or several laboratories equipped with measurement devices. Let X denote the set of available *measurements*, let $\mathcal{M}$ denote the set of *measurement contexts*, i.e. the subsets of measurements that can be **performed jointly**, and let O be a finite set of *possible outcomes*. An experimental run consists in choosing a context $C \in \mathcal{M}$ and observing an *outcome assignment* $s : C \to O$. Thus the primitive observational events are pairs (C, s), where C is the performed context and s is the observed local outcome.

For example, in the classical $(2, 2, 2)$ Bell scenario, there are two parties, Alice and Bob, each with two possible measurements. We write

$$X = \{a_1, a_2, b_1, b_2\}$$

where a_1, a_2 are Alice's measurements and b_1, b_2 are Bob's measurements. The jointly performable contexts are

$$\mathcal{M} = \{\{a_1, b_1\}, \{a_1, b_2\}, \{a_2, b_1\}, \{a_2, b_2\}\},$$

and the outcome set is

$$O = \{0, 1\}.$$

This expresses the operational constraint that Alice and Bob can each choose one measurement per run, but Alice cannot perform both of her measurements simultaneously, so $\{a_1, a_2\} \notin \mathcal{M}$. Likewise, $\{b_1, b_2\} \notin \mathcal{M}$. In this sense, the scenario specifies not only which outcomes may be observed, but also **which observations can coexist in a single experimental run**.

The technical point of this section is the following. A context table is only local data. A genuine empirical model is obtained only when all context tables can be compared on overlaps and when they have a common level. Therefore we must distinguish three objects:

a local counting N_C, a family $N = (N_C)_{C\in\mathcal{M}}$, and the corresponding vector $N \in \mathbb{N}^V$ indexed by visible contextual events. The notation is intentionally the same, but the type of the object changes with the context.

DEFINITION 2.1. *A **measurement scenario** is a triple $\langle X, \mathcal{M}, O\rangle$ where X is a set of measurements, O is a set of outcomes, and $\mathcal{M} \subseteq \mathcal{P}(X)$ is a family of measurement contexts. The **event presheaf** associated with this scenario is the functor[†] $\mathcal{E} : \mathcal{P}(X)^{\mathrm{op}} \to \mathsf{Set}$ defined by $\mathcal{E}(U) := O^U$ for every subset $U \subseteq X$. If $U \subseteq V \subseteq X$, the restriction map $\mathcal{E}(V) \to \mathcal{E}(U)$ sends a joint assignment $s : V \to O$ to its restriction $s|_U : U \to O$. Thus, an element $s \in \mathcal{E}(U)$ is a local assignment of outcomes to all measurements in U. In particular, for a **context** $C \in \mathcal{M}$, an element $s \in \mathcal{E}(C)$ represents a possible joint outcome for the measurements in C. Let R be a semiring. For any finite set Y, we define the functor $\mathcal{D}_R : \mathsf{Set} \to \mathsf{Set}$ like*

$$\mathcal{D}_R(Y) := \{d : Y \to R \mid \mathrm{supp}(d) \text{ is finite}\}.$$

*Here Y is only the set on which the distribution is defined; it is not a level. The support is $\mathrm{supp}(d) = \{y \in Y \mid d(y) \neq 0\}$. The **total weight** of d is $|d| := \sum_{y\in Y} d(y)$. For a fixed level $t \in R$, we write*

$$\mathcal{D}_R^t(Y) := \{d \in \mathcal{D}_R(Y) \mid |d| = t\} \text{ and } \mathcal{D}_R^\bullet := \cup_{t\in R}\, \mathcal{D}_R^t$$

The argument Y is the underlying set of outcomes. In the probabilistic case one usually fixes $R = \mathbb{R}_{\geq 0}$ and $t = 1$. In the discrete setting one takes $R = \mathbb{N}$ and keeps $t \in \mathbb{N}$ as the number of counted trials.

In the standard probabilistic setting, one works directly with normalized distributions ($t = 1$) and the grading is invisible. Here, we keep the grading explicit: each level t defines a slice of the theory, and the passage from one level to the next carries empirical meaning.

2.1.2. Empirical models as $\mathcal{D}_R \circ \mathcal{E}$

The event presheaf $\mathcal{E}$ describes which local outcome assignments are available over each set of measurements. To pass from possible assignments to empirical data, one composes it with a distribution or counting functor. Thus the relevant object is not only $\mathcal{E}$, but the composite presheaf $\mathcal{D}_R \circ \mathcal{E}$, defined by

$$(\mathcal{D}_R \circ \mathcal{E})(U) = \mathcal{D}_R(\mathcal{E}(U)).$$

An element $d_U \in \mathcal{D}_R(\mathcal{E}(U))$ assigns a weight to each local section $s : U \to O$. In particular, a local counting over a context C is an element $N_C \in \mathcal{D}_\mathbb{N}^\bullet(\mathcal{E}(C))$. This means only that N_C has some integer level. If $N_C \in \mathcal{D}_\mathbb{N}^t(\mathcal{E}(C))$, then the context C has exact level t. In all the following we suppose that $\mathcal{M}$ is a **measurement cover**, meaning $\cup_{C\in\mathcal{M}} C = X$, and that it is an antichain, meaning $C \subseteq D$ with $C, D \in \mathcal{M}$ implies $C = D$.

2.1.3. Compatibility

The first notion is compatibility. It comes from the sheaf-theoretic idea that two local observations must agree on the part that they have in common.

DEFINITION 2.2 (Compatible family). *Let R be a semiring. A family $e = (e_C)_{C\in\mathcal{M}}$ of elements $e_C \in \mathcal{D}_R(\mathcal{E}(C))$ is **compatible** if, for all contexts $C, D \in \mathcal{M}$, $e_C|_{C\cap D} = e_D|_{C\cap D}$.*

[†] A functor is a mapping between categories. See [Lei16, Chapter 1] for a gentle introduction.

In the probabilistic case, this says that marginal distributions coincide on overlaps. In the counting case, it says that marginal count vectors coincide exactly. Equivalently, for a counting family $N = (N_C)_{C\in\mathcal{M}}$ and for $u \in \mathcal{E}(C \cap D)$, compatibility means

$$\sum_{s\in\mathcal{E}(C),s|_{C\cap D}=u} N_C(s) = \sum_{r\in\mathcal{E}(D),r|_{C\cap D}=u} N_D(r).$$

We sometimes denote this whole family of homogeneous linear equations by $\delta N = 0$.

In the standard probabilistic setting, where $R = \mathbb{R}_{\geq 0}$, one normalizes the integer counts to obtain probability distributions. To illustrate this passage, consider the following joint probability table $p(a, b|x, y)$ for a quantum strategy in the $(2, 2, 2)$ Bell scenario (Table 1). The rows correspond to the measurement settings (x, y) chosen by Alice and Bob, and the columns to the joint outcomes (a, b).

Table 1: Joint probabilities $p(a, b|x, y)$ for a quantum strategy achieving the Tsirelson bound $\cos^2\left(\frac{\pi}{8}\right) \approx 0.8536$.

A	B	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
a	b	$\frac{1}{2}$	0	0	$\frac{1}{2}$
a'	b	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
a	b'	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
a'	b'	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

The four contexts are $C_1 = \{a, b\}$, $C_2 = \{a, b'\}$, $C_3 = \{a', b\}$, and $C_4 = \{a', b'\}$. Two contexts that share a measurement must agree on its marginal.

- *Alice's marginal on a* : The contexts $C_1 = \{a, b\}$ and $C_2 = \{a, b'\}$ both contain a. Summing over Bob's outcomes in each context gives:

$$p(a = 0 \mid C_1) = p(0, 0|0, 0) + p(0, 1|0, 0) = \frac{1}{2} + 0 = \frac{1}{2},$$

$$p(a = 0 \mid C_2) = p(0, 0|0, 1) + p(0, 1|0, 1) = \frac{3}{8} + \frac{1}{8} = \frac{1}{2}.$$

  These two values coincide. The same holds for $a = 1$: both marginals equal $\frac{1}{2}$. Thus the two contexts agree on the distribution of a.
- *Bob's marginal on b* : The contexts $C_1 = \{a, b\}$ and $C_3 = \{a', b\}$ both contain b. Summing over Alice's outcomes:

$$p(b = 0 \mid C_1) = p(0, 0|0, 0) + p(1, 0|0, 0) = \frac{1}{2} + 0 = \frac{1}{2},$$

$$p(b = 0 \mid C_3) = p(0, 0|1, 0) + p(1, 0|1, 0) = \frac{3}{8} + \frac{1}{8} = \frac{1}{2}.$$

  Again the marginals coincide. The same check works for every shared measurement across all four contexts.

In summary, compatibility means that each measurement, when it appears in two different contexts, induces the same marginal distribution. The table is not free: the 16 entries are constrained by 8 independent marginal equations (two for each shared measurement), leaving only 8 degrees of freedom. This is precisely the condition $\delta N = 0$ applied to the CHSH cover. This can be called the **no-signaling condition**, because it says that the

choice of measurement on one side does not affect the marginal distribution on the other side, that is the precise definition of what signaling mean.

2.1.4. From presheaves to hypergraphs

The sheaf-theoretic and the hypergraph formulation describe the same local-to-global problem from two complementary viewpoints. In the sheaf-theoretic approach, the event presheaf $\mathcal{E}$ assigns to each set of measurements $U \subseteq X$ the set of local sections $\mathcal{E}(U) = O^U$. Contextuality is then an obstruction to gluing local data into a global section [AB11].

The hypergraph viewpoint keeps the same operational information but makes the linear constraints explicit. Its vertices are the visible events (C, s). Its hyperedges encode the tests that a table must satisfy: context normalization and marginal compatibility. This is close in spirit to the combinatorial approach to contextuality scenarios developed by Acín, Fritz, Leverrier, and Sainz [Ací+15]. It is also related to the graph-theoretic exclusivity viewpoint of Cabello, Severini, and Winter [CSW14], although here we keep the measurement-cover and marginal-compatibility structure visible. In this report, the sheaf language explains what must be glued, while the hypergraph and matrix language explains how the gluing constraints act on integer counts.

To make this concrete, consider the CHSH scenario. The visible event space is $V = \{(a, b|x, y) \mid a, b, x, y \in \{0, 1\}\}$ ($(a|x) \coloneqq$ (outcome|measurement/input)), which has 16 elements. If we adopt the ordering $0 \mapsto 0, 0 \mapsto 1, 1 \mapsto 0, 1 \mapsto 1$ for the question-to-answer map, or $0|0, 1|0, 0|1, 1|1$ for the answer-to-question map, then a counting vector $N \in \mathbb{N}^V$ can be represented as a 4×4 matrix. The rows represent Alice's outcome-measurement pairs $(a|x)$, and the columns represent Bob's outcome-measurement pairs $(b|y)$. Each cell (i, j) then corresponds to the joint event $(a, b|x, y)$, and records the count $N(a, b|x, y)$ for that combination. Figure 1 illustrates this representation.

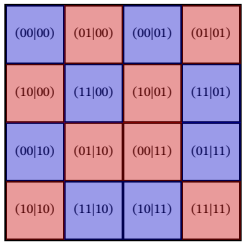

Figure 1: Matrix representation of a counting vector $N \in \mathbb{N}^V$ in the CHSH scenario. Rows correspond to Alice's pairs $(a|x)$, columns to Bob's pairs $(b|y)$, and each cell to the joint event $(a, b|x, y)$. The blue color is for the winning event of the CHSH game and red is for the loosing event.

In this matrix view, the context-normalization condition $A_{\mathcal{M}} N = t\,\mathbb{1}$ says that the sum over each context block equals t. The marginal-compatibility condition $\delta N = 0$ says that the marginals are consistent across contexts that share a measurement.

Concretely, all these constraints can be summarized by the **contextual hypergraph** of the Bell $(2, 2, 2)$ scenario, shown in Figure 2. The 16 visible events are the vertices. The hyperedges encode two types of constraints:

- **Context normalization** (blue): for each context (x, y), the four events $(a, b|x, y)$ with $a, b \in \{0, 1\}$ must have counts summing to t. This gives 4 hyperedges.
- **Marginal compatibility** (red for Alice, green for Bob): for each fixed measurement-outcome pair of one party, the marginal must be independent of the other party's measurement. For Alice, fixing $(a|x)$ requires that summing over Bob's outcomes gives

the same result for $y = 0$ and $y = 1$. For Bob, fixing $(b|y)$ requires the same for Alice's measurement x. This gives $4 + 4 = 8$ hyperedges.

In total, the hypergraph has 12 hyperedges. A vector $N \in \mathbb{N}^V$ represents a compatible counting if and only if it satisfies all 12 hyperedge constraints simultaneously.

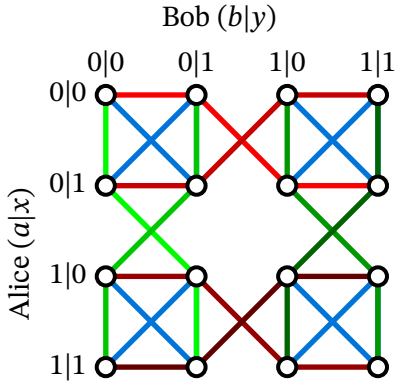

Figure 2: Contextual hypergraph of the Bell $(2, 2, 2)$ scenario. The hyperedge are represented by the same color.

Each hyperedge constrains its vertices: in the probabilistic setting, the sum of their probabilities must equal 1; in the counting setting, the sum of their counts must equal the same level t.

2.1.5. Global sections and contextuality

A global section is an assignment of an outcome to every measurement at once. Formally, it is an element $g \in \mathcal{E}(X)$, (can be views as a function $g : X \to O$). The intuitive meaning is simple: before any context is chosen, every measurement $x \in X$ already has a predetermined outcome value $g(x) \in O$.[‡] Thus, a global section represents a deterministic **hidden-variable** assignment. If the experimenter later chooses a context $C \in \mathcal{M}$, the observed local outcome is not chosen anew; it is just the restriction $g|_C : C \to O$ of the already fixed global assignment. In other words, determinism means that all counterfactual outcomes are jointly defined. Even if two measurements cannot be performed together, such as a and a' in the Bell scenario, a deterministic model still assigns values to both of them in advance.

2.2. From probabilities to integer counts: the discrete level ——

The classical sheaf-theoretic framework, as presented above, works with R-valued distributions for an arbitrary semiring R. In the probabilistic case ($R = \mathbb{R}_{\geq 0}$), one normalizes to $t = 1$ and obtains a single convex space. Our contribution begins with a different choice: we take $R = \mathbb{N}$ and keep the level t as an explicit parameter. This changes the normalization condition from $\sum_s e_C(s) = 1$ to $\sum_s N_C(s) = t$, and replaces the convex structure by a graded semigroup structure. The level t is not imposed by hand; it emerges from compatibility, as the next proposition shows.

[‡]The formalism is intentionally simplified for pedagogical clarity.

2.2.1. Compatibility creates the level

For a local counting N_C, define its local level by $t_C := |N_C| = \sum_{s\in\mathcal{E}(C)} N_C(s)$. At first, the numbers t_C may depend on the context C. The point is that compatibility can force them to become equal.

PROPOSITION 2.3 (Compatibility creates the level). *Suppose that the intersection graph of the contexts is connected, where C is adjacent to D when $C \cap D \neq \emptyset$. If a counting family $N = (N_C)_{C\in\mathcal{M}}$ is compatible, then all local levels $|N_C|$ are equal. Hence there is a unique integer t such that $\forall C \in \mathcal{M}, \quad |N_C| = t$.*

Proof.

If $C \cap D \neq \emptyset$ and N is compatible, then the marginals of N_C and N_D on $C \cap D$ are equal. Marginalization preserves total weight, so $|N_C| = |N_D|$. Since the intersection graph is connected, this equality propagates from one context to every other context. Thus all local levels are equal to a common integer t. Uniqueness is immediate, because $t = |N_C|$ for any context C.

Thus the level is not just an extra parameter added by hand. On a connected cover, it is the common normalization scale forced by compatible integer data.

DEFINITION 2.4 (Discrete empirical model of level t). *A **discrete empirical model of level** t is a compatible family $N = (N_C)_{C\in\mathcal{M}}$ such that $N_C \in \mathcal{D}^t_{\mathbb{N}}(\mathcal{E}(C))$ for every context C.*

2.2.2. Why the level matters

It may look as if the integer model contains no new information, because one can divide a level-t count by t and obtain a probability table. This would miss the main point. The level t is not only a denominator. It records the amount of observation that supports the table in each context. For instance, the count $(1, 0)$ and the count $(1000, 0)$ give the same normalized probability $(1, 0)$. They do not have the same empirical status. The second table is supported by many more repetitions. In a standard statistical interpretation, this usually means that the sampling uncertainty decreases at the scale $\frac{1}{\sqrt{t}}$. We do not need to assume such a statistical model in this report, but the intuition is useful: t is an **experimental resolution scale**. It is not a subjective probability of confidence. It is the amount of counted evidence that remains visible in the formal object.

Thus a discrete empirical model is not only a point in a polytope. It is an integer point at a definite level. The same probability may appear at several levels, but these levels correspond to different amounts of evidence and to different possible intermediate histories.

***Remark* (An arithmetic effect of the level).**

Consider a partial CHSH count at level 6. Suppose that one context already contains the counts $(3\ 1\ 1\ 1)$. After normalization, this gives probabilities $\left(\frac{1}{2}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}\right)$. Now suppose that a marginal compatibility equation involves two unknown probabilities a and b and has the form $\frac{1}{2} + \frac{1}{6} + a + b = 1$. Over probabilities, this gives the continuum of solutions $a + b = \frac{1}{3}$ with $a, b \geq 0$. At level 6, however, the unknowns must come from integer counts: $a = \frac{A}{6}$ and $b = \frac{B}{6}$ with $A, B \in \mathbb{N}$. The same equation becomes $3 + 1 + A + B = 6$, hence $A + B = 2$. The continuum has been cut down to the finite arithmetic set $(A, B) \in \{(0, 2), (1, 1), (2, 0)\}$.

> The point is not uniqueness. The point is that the level cuts the probabilistic continuum by an **arithmetic lattice**. This arithmetic rigidity disappears after normalization.

This is the reason why the transition to automata is natural. If the level records how much of a process has stabilized, then it is also natural to ask what exists before stabilization. A count of size 3 in CHSH is not a completed level-t model. But it is not meaningless. It may be the visible past of a generator that has not closed yet.

The conceptual chain is therefore the following. The level t describes closed stabilized observations. The Hilbert basis gives elementary closed bricks. Generator automata describe how these bricks can be produced step by step. Residuals record the future that is still needed for closure. The dynamic model below is designed to keep exactly this missing information: what has already been seen, and what still has to happen for the process to close.

2.3. The discrete framework: hypergraphs and semigroups —

2.3.1. Vector and hypergraph packaging

The vector notation is only a packaging of the sheaf-theoretic definition above. Define the visible event space

$$V \coloneqq \coprod_{C\in\mathcal{M}} \mathcal{E}(C) = \{(C,s) \mid C \in \mathcal{M} \text{ and } s \in \mathcal{E}(C)\}.$$

Giving a family $(N_C)_{C\in\mathcal{M}}$ is the same thing as giving a vector $N \in \mathbb{N}^V$, by the rule $N(C,s) \coloneqq N_C(s)$. The context hypergraph is $H_{\mathcal{M}} = (V, E_{\mathcal{M}})$, where $E_{\mathcal{M}} = \{e_C \mid C \in \mathcal{M}\}$ and $e_C \coloneqq \{(C,s) \mid s \in \mathcal{E}(C)\}$. Let $A_{\mathcal{M}}$ be its incidence matrix. Then $(A_{\mathcal{M}}N)_C = \sum_{s\in\mathcal{E}(C)} N(C,s) = |N_C|$. Therefore the equation $A_{\mathcal{M}}N = t\,\mathbb{1}$ expresses the **common-level condition** which is guaranteed to exist when the intersection graph is connected (Proposition 2.3). It does not express compatibility by itself.

The compatibility equations are homogeneous. If D denotes their matrix, the vector form of a *discrete empirical model of level* t is $A_{\mathcal{M}}N = t\,\mathbb{1}$ and $DN = 0$. Equivalently, using the notation $\delta N = 0$ for the second family of equations, the *compatible counting semigroup* is

$$\mathcal{S} \coloneqq \big\{(N,t) \in \mathbb{N}^V \times \mathbb{N} \mid A_{\mathcal{M}}N = t\,\mathbb{1} \text{ and } \delta N = 0\big\}.$$

For a fixed level t, we write

$$\mathcal{S}_t \coloneqq \big\{N \in \mathbb{N}^V \mid (N,t) \in \mathcal{S}\big\}.$$

We have that $\mathcal{S}$ is a semigroup by linearity of the constraints: if (N,t) and (N',t') belong to $\mathcal{S}$, then $(N+N', t+t')$ also belongs to $\mathcal{S}$. We can write also $\mathcal{S}_{\text{ns}}$ for non-signaling discrete model. If the two families of constraints are assembled into one augmented matrix, then $A_{\text{aug}}N = (t\,\mathbb{1}, 0)$. The zero part is important: compatibility equations are homogeneous, while normalization equations carry the level t.

2.3.2. From convex mixtures to Hilbert bases

This discrete framework differs from the standard probabilistic one. In the usual convex setting, if e and e' are empirical models, then every convex combination $re + (1-r)e'$ with $r \in [0,1]$ is again an empirical model. The geometry is therefore convex.

In our setting, this operation is not primitive. An arbitrary real coefficient r does not preserve integer counts. Instead, the natural operation is addition. Hence the appropriate

algebraic object is not first a convex set, but an **affine semigroup**. The analogue of a convex decomposition is therefore a *Hilbert basis* decomposition. The Hilbert basis $H(\mathcal{S})$ of $\mathcal{S}$ is the finite set of irreducible elements of the semigroup. Its elements are the elementary compatible counting models that cannot be decomposed into smaller compatible counting models. Thus, instead of asking whether a model is a convex mixture of deterministic models, we ask which Hilbert generators are required to build it as an integer sum.

2.3.3. Discrete noncontextuality and strong contextuality

Given a global section $g \in \mathcal{E}(X) = \mathbb{N}^V$ (or $g : X \to O$), we define its deterministic counting vector $d_g \in \mathbb{N}^V$ by $d_g(C,s) \coloneqq \begin{cases} 1 \text{ if } s = g|_C \\ 0 \text{ otherwise.} \end{cases}$ Using the matrix representation of Figure 1, d_g places one unit per context block. For $\mathcal{M} = \{\{a,b\},\{a',b\},\{a,b'\},\{a',b'\}\}$ (or $\{(0,0),(1,0),(0,1),(1,1)\}$ in the hypergraph representation[§]) and $g = (a \mapsto 0, a' \mapsto 1, b \mapsto 0, b' \mapsto 0)$:

$$d_g = \left(\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right)$$

The 1's are not placed independently in each block. Because g assigns a fixed outcome to each measurement, the same outcome appears in every context that contains that measurement. This creates a characteristic "square" pattern: the 1's form a rectangle across the four context blocks. This pattern is typical of all global sections (Table 1) and reflects the compatibility condition: a global section automatically satisfies no-signaling. In incidence form, $A_{\mathcal{M}} d_g = \mathbb{1}$ and $\delta d_g = 0$; with the augmented convention, this is $A_{\text{aug}} d_g = (\mathbb{1}, 0)$.

More generally, an arbitrary *integer counting of global hidden assignments* is an element $c \in \mathcal{D}^{\bullet}_{\mathbb{N}}\mathcal{E}(\boldsymbol{X})$ (not to be confused with $N_C \in \mathcal{D}^t_{\mathbb{N}}\mathcal{E}(C)$). Here $c(g)$ counts how many times the global assignment $g : X \to O$ is used. Such a global counting produces a contextual-event counting by the linear map $\Phi(c) \coloneqq \sum_{g \in \mathcal{E}(X)} c(g) d_g \in \mathbb{N}^V$. If $|c| = t$, then $\Phi(c)$ has common level t and is compatible. Thus Φ maps global hidden countings into the empirical semigroup $\mathcal{S}$.

Indeed, fix a context C. Since each deterministic vector d_g puts exactly one unit on the event $\left(C, g|_C\right)$ and zero on the other events of the same context, we have

$$\sum_{s \in \mathcal{E}(C)} \Phi(c)(C,s) = \sum_{s \in \mathcal{E}(C)} \sum_{g \in \mathcal{E}(X)} c(g) d_g(C,s) = \sum_{g \in \mathcal{E}(X)} c(g) = |c| = t.$$

So every context has the same level t. Compatibility is also automatic. If $C, D \in \mathcal{M}$ and $u \in \mathcal{E}(C \cap D)$, then

$$\sum_{s|_{C\cap D} = u} \Phi(c)(C,s) = \sum_{g : g|_{C\cap D} = u} c(g) = \sum_{r|_{C\cap D} = u} \Phi(c)(D,r).$$

Both sides count exactly the same global assignments: those whose restriction to the overlap $C \cap D$ is u. Hence $\delta\Phi(c) = 0$.

A discrete empirical model $N \in \mathbb{N}^V$ is *noncontextual* if it can be written as an integer sum of deterministic global sections. That is, there exist coefficients $\{c_g\}_{g \in \mathcal{E}(X)} \in \mathbb{N}$ (or $c \in \mathcal{D}^{\bullet}_{\mathbb{N}}\mathcal{E}(X)$) such that $N = \sum_{g \in \mathcal{E}(X)} c_g d_g$ (or $N = \Phi(c)$). The coefficient c_g counts how many times the deterministic assignment g is used in the construction of N. If such a decomposition exists, then the apparent contextual data can be explained by a classical

[§]We identify a, b with question 0 and a', b' with question 1.

hidden-variable mechanism: each run secretly chooses a global assignment g, and the context merely reveals the part $g|_C$ corresponding to the measurements actually performed.

If no such decomposition exists, the model is *contextual* in the discrete sense. It is compatible as a counting model, because it satisfies the common-level constraints, but it cannot be reconstructed as a sum of predetermined global assignments.

In the notation above, this says simply that the noncontextual integer models are the image $\mathcal{S}_{\mathbf{nc}} := \Phi(\mathcal{D}^{\bullet}_{\mathbb{N}}\mathcal{E}(X)) \subseteq \mathcal{S}_{\mathrm{ns}}$. This notation is useful because it separates two questions. The semigroup $\mathcal{S}_{\mathrm{ns}}$ contains all compatible integer empirical models. The subsemigroup $\mathcal{S}_{\mathrm{nc}}$ contains only those compatible models that can be generated from global deterministic assignments. Thus $\mathcal{S}_{\mathrm{nc}}$ is the integer analogue of the classical or local polytope (in this sense, we can also write $\mathcal{S}_{\mathrm{loc}}$ instead of $\mathcal{S}_{\mathrm{nc}}$), while $\mathcal{S}_{\mathrm{ns}}$ is the integer analogue of the compatible, or no-signaling, object.

PROPOSITION 2.5. *The semigroup $\mathcal{S}_{\mathrm{nc}}$ is generated by the deterministic vectors $(d_g, 1)$ with $g \in \mathcal{E}(X)$, and $\mathcal{S}_{\mathrm{nc}} \subseteq \mathcal{S}_{\mathrm{ns}}$.*

Proof.

By definition, every element of $\mathcal{D}^{\bullet}_{\mathbb{N}}(\mathcal{E}(X))$ is a finite counting $c : \mathcal{E}(X) \to \mathbb{N}$. Hence $\Phi(c) = \sum_{g \in \mathcal{E}(X)} c(g) d_g$. Therefore every element of $\mathcal{S}_{\mathrm{nc}}$ is an integer sum of deterministic vectors. Conversely, every integer sum of deterministic vectors is obtained by taking $c(g)$ equal to the multiplicity of d_g in the sum. Thus these vectors generate $\mathcal{S}_{\mathrm{nc}}$.

In the generic case, the Hilbert basis of $\mathcal{S}_{\mathrm{nc}}$ is obtained from these deterministic generators after removing duplicates, because two distinct global assignments may induce the same contextual vector if the measurement cover does not separate them. In the usual separated scenarios, and in particular in CHSH, the deterministic vectors are distinct and irreducible; then the Hilbert basis of $\mathcal{S}_{\mathrm{nc}}$ is exactly $\mathcal{H}_{\mathrm{nc}} = \{(d_g, 1) \mid g \in \mathcal{E}(X)\}$.

Therefore contextuality is a property of a compatible counting $N \in \mathcal{S}_{\mathrm{ns}}$: it holds when N cannot be written as an integer sum of deterministic global sections, that is, when $N \notin \mathcal{S}_{\mathrm{nc}}$.

We shall only use the support version briefly. For each context C, define $\mathrm{supp}(N_C) := \{s \in \mathcal{E}(C) \mid N_C(s) > 0\}$. A compatible model N is *strongly contextual* when no global section remains compatible with these supports:

$$\nexists g \in \mathcal{E}(X) \quad \text{such that} \quad \forall C \in \mathcal{M},\ g|_C \in \mathrm{supp}(N_C).$$

In this report, this notion is mainly used as orientation. The main constructions below concern compatible countings, noncontextual decompositions, and dynamic generator explanations.

We therefore keep three levels distinct:

- **compatibility** or **non-signaling**: the local count tables have a common level and agree on overlaps, equivalently $A_{\mathcal{M}} N = t\,\mathbb{1}$ and $\delta N = 0$;
- **noncontextuality** or **classical**: the compatible model decomposes as an integer sum of deterministic global sections;
- **strong contextuality**: not even one deterministic global section is compatible with the support of the model.

2.4. First results: Contextuality viewed with integer polynomials

These first results show that Ehrhart polynomials can be used to characterise exactly the integer models of a semigroup of compatible countings.

To be clear, the $(2,2,2)$ scenario refers to the case with two parties, each having two measurement settings, and each measurement having two possible outcomes. The CHSH setting refers to a game following the $(2,2,2)$ scenario where the winning condition is $a \oplus b = x \wedge y$.

We now record the elementary consequences of the previous definitions for the $(2,2,2)$ scenario. For a context $C \in \mathcal{M}$ and a local section $s \in \mathcal{E}(C)$, the coordinate $N(C,s)$ may also be written $N(s \mid C)$. In the Bell notation this becomes $N(ab \mid xy)$: conditionally on the chosen questions (x,y), one counts the observed answers (a,b). In CHSH the vertex set is $V = \{(ab \mid xy) \mid a,b,x,y \in \{0,1\}\}$, hence $|V| = 16$. The contextual hypergraph has 12 hyperedges: four context-normalization edges and eight non-signaling marginal edges. Each edge has size 4, and each vertex belongs to exactly 3 edges. If A_{aug} is the corresponding incidence matrix (encoding both context normalization and marginal compatibility), then the discrete non-signaling equation is $A_{\text{aug}}N = t\,\mathbb{1}$. This is the same object as the non-signaling polytope, but seen before normalization. Indeed, if $G_{\text{NS}} := \{p \in \mathbb{R}^V_{\geq 0} \mid A_{\text{aug}}p = \mathbb{1}\}$, then the equation $A_{\text{aug}}N = t\,\mathbb{1}$ is equivalent, for $t > 0$, to $p := N/t \in G_{\text{NS}}$. Thus

$$\{N \in \mathbb{N}^V \mid A_{\text{aug}}N = t\,\mathbb{1}\} = tG_{\text{NS}} \cap \mathbb{Z}^V.$$

The semigroup $\mathcal{S}_{\text{ns}}$ is therefore the semigroup of integer points in the cone over G_{NS}:

$$\mathcal{S}_{\text{ns}} := \{(N,t) \in \mathbb{N}^V \times \mathbb{N} \mid A_{\text{aug}}N = t\,\mathbb{1}\} = \{(N,t) \mid N \in tG_{\text{NS}} \cap \mathbb{Z}^V\}.$$

The polytope G_{NS} is the normalized slice $t = 1$ of this cone, while the discrete models at level t are the integer points in the dilated slice tG_{NS}. The local part is obtained by replacing G_{NS} by the local polytope $L_{\text{loc}} := \text{conv}\{d_\lambda \mid \lambda \text{ deterministic}\}$. The first admissible nonzero level is therefore $t = 1$, thus $|N|_1 = 4$. At this level, $A_{\text{aug}}N = \mathbb{1}$ forces N to select exactly one event on every hyperedge. These exact **transversals** (of the hypergraph $H_{\mathcal{M}}$) are precisely the local deterministic strategies, i.e. assignments $a = f(x)$ and $b = g(y)$. Thus the first level contains only classical generators. The first intrinsically non-local (or contextual) CHSH generators occur at $t = 2$, thus $|N|_1 = 8$. Because a PR box has probabilities in $\left\{0, \frac{1}{2}\right\}$ and therefore is not an integer point at level $t = 1$. Multiplying it by 2 gives an integer vector $r \in \{0,1\}^{16}$ with $A_{\text{aug}}r = 2\,\mathbb{1}$. Hence its first discrete lift is at level $t = 2$. Let $d_\lambda \in \{0,1\}^{16}$ denote the 16 local deterministic generators, with $A_{\text{aug}}d_\lambda = \mathbb{1}$, and let $r_\mu = 2p^{\text{PR}}_\mu \in \{0,1\}^{16}$ denote the 8 lifted PR generators, with $A_{\text{aug}}r_\mu = 2\,\mathbb{1}$. The natural candidate Hilbert basis, i.e. the set of points that generate any other points in $\mathcal{S}_{\text{ns}}$ is

$$\mathcal{H}_{\text{ns}} := \{(d_\lambda, 1)\}_{\lambda=1}^{16} \cup \{(r_\mu, 2)\}_{\mu=1}^{8}.$$

To visualize what $\mathcal{H}_{\text{ns}}$ looks like, see Table 1 and Table 2.

PROPOSITION 2.6. *The set of integer models $\mathcal{H}_{\text{ns}}$ is a Hilbert basis of the semigroup $\mathcal{S}_{\text{ns}}$.*

Proof.

By construction, all elements of $\mathcal{H}_{\text{ns}}$ belong to $\mathcal{S}_{\text{ns}}$. The deterministic generators $(d_\lambda, 1)$ are irreducible. The PR generators $(r_\mu, 2)$ are also irreducible. Indeed, any non-trivial decomposition $(r_\mu, 2) = (N_1, t_1) + (N_2, t_2)$ would force $t_1 = t_2 = 1$. Hence we would have

$AN_i = \mathbb{1}$ for $i = 1, 2$, so N_1 and N_2 would be two local deterministic models. It would follow that $p_\mu^{\mathrm{PR}} = \frac{1}{2}N_1 + \frac{1}{2}N_2$, which would express a PR box as a convex combination of deterministic boxes, a contradiction.

It remains to prove that $\mathcal{H}_{\mathrm{ns}}$ generates the whole semigroup. We proceed by induction on t. If $t = 0$, then $AN = 0$, and since every vertex belongs to at least one edge, we necessarily get $N = 0$.

We first need the following support lemma.

LEMMA 2.7 (Support lemma). *For every nonzero $(N, t) \in \mathcal{S}_{\mathrm{ns}}$, there exists $(U, s) \in \mathcal{H}_{\mathrm{ns}}$ such that $U(v) \le N(v)$ for every vertex $v \in V$.*

Proof.

The proof is computer-assisted. The finite step is an exhaustive test of supports. Let $S \subseteq V$ be a nonempty subset of the 16 vertices. We say that S is **realizable** if there is an element $(m, t) \in \mathcal{S}_{\mathrm{ns}}$ such that $\mathrm{supp}(m) = S$. Since $|V| = 16$, there are $2^{16} - 1 = 65535$ nonempty supports to test. For each nonempty subset $S \subseteq V$, the computation solves the following feasibility problem. Find a vector $q \in \mathbb{R}^V_{\ge 0}$ and a scalar $\tau \ge 1$ such that $Aq = \tau\mathbf{1}$, with the support constraints $q_v \ge 1$ for $v \in S$, $q_v = 0$ for $v \notin S$.

If this system has no solution, then no element of $\mathcal{S}_{\mathrm{ns}}$ can have support S. Indeed, any integer element (m, t) with support S would also be a real solution by taking $q = m$ and $\tau = t$.

If the system has a solution, then S is realizable. The matrix A and all support constraints have rational coefficients. Hence a feasible rational point exists whenever the rational polyhedron is nonempty. Multiplying by a common denominator gives an integer vector $m \in \mathbb{N}^{16}$ and an integer level $t \in \mathbb{N}$ such that $Am = t\mathbf{1}$ and $\mathrm{supp}(m) = S$. Thus $(m, t) \in \mathcal{S}_{\mathrm{ns}}$.

The exhaustive computation then applies a second test to every realizable support S. It checks whether there exists one of the proposed generators $(u, s) \in \mathcal{H}_{\mathrm{ns}}$ such that $\mathrm{supp}(u) \subseteq S$. The computation verifies that this condition holds for every realizable nonempty support $S \subseteq V$.

We now use this finite verification to prove the lemma. Let $(m, t) \in \mathcal{S}_{\mathrm{ns}}$ be nonzero, and set $S := \mathrm{supp}(m)$. Then S is a realizable nonempty support. By the exhaustive verification, there is a generator $(u, s) \in \mathcal{H}_{\mathrm{ns}}$ with $\mathrm{supp}(u) \subseteq S$.

Every deterministic generator and every lifted PR generator has entries in $\{0, 1\}$. Therefore, for each vertex $v \in V$, there are two cases. If $u_v = 0$, then $u_v \le m_v$ is immediate. If $u_v = 1$, then $v \in \mathrm{supp}(u) \subseteq S = \mathrm{supp}(m)$, so $m_v \ge 1$ and again $u_v \le m_v$. Hence

$$u_v \le m_v \quad \text{for every} \quad v \in V.$$

Thus every nonzero element of the semigroup contains one of the 24 proposed generators as a sub-counting. This proves the lemma.

We now prove generation by induction on the level t. The case $t = 0$ was treated above. Assume that every element of level strictly smaller than t is generated by $\mathcal{H}_{\mathrm{ns}}$, and let $(N, t) \in \mathcal{S}_{\mathrm{ns}}$ with $t \ge 1$.

By the support lemma, there exists a generator $(u, s) \in \mathcal{H}_{\mathrm{ns}}$ such that $u_v \le N_v$ for every $v \in V$. Hence the difference

$$M := N - u$$

is again an element of $\mathbb{N}^{16}$. Since $(u, s) \in \mathcal{S}_{\mathrm{ns}}$, we have

$$AM = A(N-u) = (t-s)\mathbb{1}.$$

Thus $(M, t-s) \in \mathcal{S}_{\text{ns}}$. Moreover, $s \in \{1,2\}$, so $t-s < t$. By the induction hypothesis, $(M, t-s)$ is a sum of elements of $\mathcal{H}_{\text{ns}}$. Adding (u,s) to this sum gives a decomposition of (N,t).

Therefore, $\mathcal{H}_{\text{ns}}$ generates $\mathcal{S}_{\text{ns}}$. Finally, no element of $\mathcal{H}_{\text{ns}}$ belongs to the subsemigroup generated by the others, since this would give a non-trivial decomposition of an irreducible element. Hence $\mathcal{H}_{\text{ns}}$ is minimal, with respect to inclusion, among generating families. In other words, it is a Hilbert basis.

This means that, at the discrete level, non-signaling models are built from two types of **irreducible bricks**: local deterministic bricks of level 1 and PR bricks of level 2 like you can see in the Figure 3.

Finally, we could ask how many integer points are in this semigroup. We define $Q_{\text{NS}}(t) \coloneqq \#\left(tG_{\text{NS}} \cap \mathbb{Z}^V\right)$ and $Q_{\text{loc}}(t) \coloneqq \#\left(tL_{\text{loc}} \cap \mathbb{Z}^V\right)$, then these functions count the number of discrete models at level t in the non-signaling and local cases. Since the affine dimension is 8, these functions have degree 8; Q_{loc} is an Ehrhart polynomial [Ehr05], while Q_{NS} is a quasi-polynomial of period 2 because PR vertices have denominator 2. A direct interpolation from exact integer-point counts gives

$$Q_{\text{loc}}(t) = 1 + \frac{71}{21}t + \frac{6053}{1260}t^2 + \frac{697}{180}t^3 + \frac{91}{45}t^4 + \frac{13}{18}t^5 + \frac{31}{180}t^6 + \frac{31}{1260}t^7 + \frac{1}{630}t^8$$

and

$$Q_{\text{NS}}(t) = \frac{17}{10080}t^8 + \frac{17}{630}t^7 + \frac{7}{36}t^6 + \frac{37}{45}t^5 + \frac{1607}{720}t^4 + \frac{359}{90}t^3 + \frac{290}{63}t^2 + \frac{332}{105}t + \frac{63}{64} + \frac{(-1)^t}{64}.$$

Thus the number of discrete CHSH models with size $k \coloneqq |N|_1$ is 0 if 4 does not divide k, and is $Q_{\text{NS}}\left(\frac{k}{4}\right)$ if $k = 4t$. The local polynomial has normalized volume $8! \cdot \frac{1}{630} = 64$, whereas the non-signaling quasi-polynomial has normalized volume[¶] $8! \cdot \frac{17}{10080} = 68$. Consequently, the asymptotic non-local part $Q_{\text{NS}}(t) - Q_{\text{loc}}(t)$ has leading term $\frac{1}{10080}t^8$, and the asymptotic proportion of non-local integer models inside the non-signaling ones is 1/17.

[¶]The normalized volume gives a geometric measure of the polytope. It estimates the number of integer points at each level t, helps evaluate the difficulty of integer programming problems over the polytope, and reveals structural properties such as the presence of non-integer vertices.

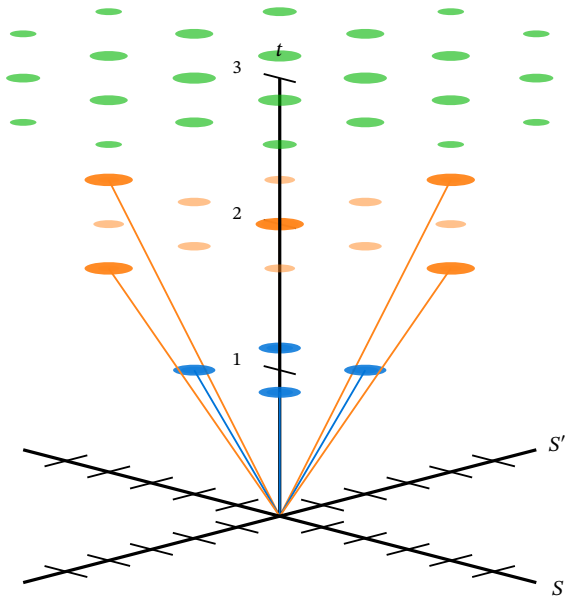

Figure 3: Cone representation of $\mathcal{S}_{\text{ns}}$. Each point represents a compatible integer model $N \in \mathcal{S}_{\text{ns}}$. The blue points are local deterministic models ($t = 1$), the orange points are the PR models ($t = 2$), and the semi-transparent points are integer combinations of these generators. The axes (S, S') project each model via its CHSH scores: $S(N) = E(0,0) + E(0,1) + E(1,0) - E(1,1)$ and $S'(N) = E(0,0) + E(0,1) - E(1,0) + E(1,1)$, where $E(x,y) = N(0,0|x,y) + N(1,1|x,y) - N(0,1|x,y) - N(1,0|x,y)$ is the integer correlator for context (x,y). Point sizes reflect projection multiplicities.

Viewing contextuality as an automaton process

CHAPTER 3

The previous section described compatible counts as elements of an affine semigroup. This is a static description. It tells us which completed integer tables are possible at each level t. The level is important because it records the scale at which the observation has stabilized.

In CHSH, a completed compatible count has total size $k := |N|_1 = 4t$. Hence, completed objects appear only at sizes divisible by 4. The dynamic question is therefore unavoidable: what can be said after only one, two, or three visible events have occurred? Such a prefix is not a completed empirical model, but it may still be the beginning of a process that will later close into one.

The automaton model introduced here is a way to make this dynamic reading concrete. It does not redefine contextuality. It gives a constructive description of how a compatible count can be produced, interrupted, and later completed. A completed count is built from elementary generators. An interrupted count keeps track of the generators that have already closed and of those that are still open. The open part contains information about what remains to be emitted in order to close the process.

The following sections develop this automaton model step by step. We first introduce generators as elementary complete processes. We then attach residual automata to them, define hidden dynamic states, and finally add constraints that select which trajectories are admissible. The goal is to keep the constructive information that is lost when one records only the final probability table.

3.1. Basic automaton model

3.1.1. Generators as complete processes

Let V be the visible event space introduced above. A *count* is an element of $\mathbb{N}^V$. A *generator* is a finite counting vector $g \in \mathbb{N}^V$ that is regarded as one **elementary complete process**. Generators are not introduced as arbitrary formal objects. They are meant to be elementary bricks of stabilization. A closed empirical table is not explained by a probability point alone, but by a finite production of such bricks. In $(2, 2, 2)$ bell scenarios, the previous semigroup computation gives the intended examples. The local deterministic generators are the minimal classical bricks. The lifted PR generators are the minimal non-local bricks of the compatible semigroup. A local deterministic generator has total size 4. A lifted PR generator has total size 8. The automata below do not change these bricks. They only describe their progressive production.

At this stage, we do not need to assume that these are the only possible generators in every scenario. The set of generators is a modelling choice, but it is a constrained choice: it must be rich enough to reconstruct the completed countings that we want to explain, and structured enough to make intermediate states meaningful.

DEFINITION 3.1 (Generator family). *A **generator family** is a finite set $\mathcal{G} \subseteq \mathbb{N}^V$ whose elements are called generators. Each $g \in \mathcal{G}$ has a level $\ell(g) \in \mathbb{N}$ such that $(g, \ell(g)) \in \mathcal{S}$. The*

family is ***closed-complete*** *for $\mathcal{S}$ if every model (element of $\mathcal{S}$) can be written as a finite sum of generators:*

$$\forall (N,t) \in \mathcal{S}, \quad \exists c : \mathcal{G} \to \mathbb{N}, \quad N = \sum_{g \in \mathcal{G}} c(g)g \text{ and } t = \sum_{g \in \mathcal{G}} c(g)\ell(g).$$

This condition is the closed, semigroup-level requirement. It says that the chosen generators can explain every completed compatible counting. When $\mathcal{G}$ is the Hilbert basis of $\mathcal{S}$, this condition is automatic. In practice, the choice of $\mathcal{G}$ must be made carefully: if it is too small, some counts cannot be explained; if it is too large, the explanation loses structure. **This point is crucial for understanding the framework:** the model, or theory, chosen to account for a phenomenon can be more or less fine-grained. The choice of generators is therefore part of the modelling process. A more precise theory requires families of generators with higher levels $l(g)$, whereas a less precise theory may only require generators of lower levels.

3.1.2. The automaton of one generator

A generator is *complete*, but it can be produced in several orders. To model this, we attach a small automaton to each $g \in \mathcal{G}$. For $v \in V$, let $\varepsilon_v \in \mathbb{N}^V$ be the unit vector at v. Thus $\varepsilon_v(v) = 1$ and $\varepsilon_v(w) = 0$ for $w \neq v$.

DEFINITION 3.2 (Residual automaton of a generator). *Let $g \in \mathcal{G}$. The residual automaton $\mathcal{A}_g$ has states $\{R \in \mathbb{N}^V \mid 0 \leq R \leq g\}$. The initial state is $R = g$, and the final state is $R = 0$. If $R(v) > 0$, there is a transition $R \to R - \varepsilon_v$ labelled by v. This transition emits the visible event v and decreases the residual part by one unit at v.*

A ***copy*** *of the generator g is a running instance of the automaton $\mathcal{A}_g$. When the state reaches $R = 0$, the copy is* ***closed****: all events have been emitted. If the state is $R > 0$, the copy is* ***open****: the residual R records the events that remain to be emitted before the copy can close.*

The state R is the part of g that remains to be emitted. The emitted part is $g - R$. Thus an interruption of the automaton at residual state R gives a past/future decomposition $g = (g - R) + R$. The first term is already visible. The second term is the future required to *close* this copy of the generator.

Remark **(A four-event generator).**

> Suppose that $g = \varepsilon_{v_1} + \varepsilon_{v_2} + \varepsilon_{v_3} + \varepsilon_{v_4}$. The automaton $\mathcal{A}_g$ starts at residual state g. If the first three emitted events are v_1, v_2, v_3, then the residual state is $R = \varepsilon_{v_4}$ and the visible count is $g - R = \varepsilon_{v_1} + \varepsilon_{v_2} + \varepsilon_{v_3}$. This prefix has size 3. It is not a closed generator. It is nevertheless meaningful, because it knows that the missing future is ε_{v_4}.

3.1.3. Global states and open generators

An experiment can contain several copies of generators at the same time. Some copies are already closed. Other copies are open and have only been partially emitted. The correct hidden state must remember this distinction. To define this notion properly, we first introduce the concept of a **multiset**, that is, a set whose elements may appear with multiplicity. Traditionally, a multiset is represented by a set together with a multiplicity function. In our setting, instead of writing expressions such as $\sum_{x \in M} x \cdot m(x)$ where m is the multiplicity function, we use the notation $\sum_{x \text{ in } M} x$, which can be read as a sum over an array in the

sense of a programming language. The $x \in M$ represents the condition that x is an element of the multiset M and $x \texttt{ in } M$ enumerates with multiplicity the elements of M.

DEFINITION 3.3 (Dynamic state). *A **dynamic state** is a pair $\eta = (c_\eta, O_\eta)$. Here $c_\eta : \mathcal{G} \to \mathbb{N}$ counts the closed copies of each generator. The object O_η is a finite multiset of open copies. An open copy is a pair (g, R), where $g \in \mathcal{G}$ is the complete generator and $R \in \mathbb{N}^V$ is its residual part, with $0 < R \leq g$. The **visible counting** associated with η is $N_\eta := \sum_{g\in\mathcal{G}} c_\eta(g)g + \sum_{(g,R) \text{ in } O_\eta}(g - R)$.*

We may equivalently define a dynamic state using only a multiset $\mathbb{O}_\eta$, where elements with null residuals, such as $(g, 0)$, are also allowed. In this formulation, $N_\eta = \sum_{(g,R) \text{ in } \mathbb{O}_\eta}(g - R)$, so closed copies correspond precisely to open copies with null residual.

We will note $\mathbb{m}(g, R)$ the multiplicity of the open copy (g, R) in $\mathbb{O}_\eta$.

The first sum is the contribution of closed copies. The second sum is the contribution of open copies: each open copy has already emitted $g - R$ and still owes the residual R. The future *residual* of the state is $R_\eta := \sum_{(g,R) \text{ in } O_\eta} R$. If no new generator is opened, then closing all current open copies would produce the completed count

$$F_\eta := N_\eta + R_\eta = \sum_{(g,R) \text{ in } \mathbb{O}_\eta} g.$$

Thus the natural *signature* of an interrupted state is the pair $\Phi(\eta) := (N_\eta, F_\eta)$. It records both the stabilized visible count (**the past**) and the completed count (**the futur**) that would be obtained if all open generators were closed

Remark.

> There is no reason for the map $\eta \to N_\eta$ to be injective. Two different hidden states may have the same visible count but different open generators and different residual futures. This is why the dynamic theory cannot be reduced to the visible counting vector alone.

In the following, we extend the notation N_η to $c_\eta : \mathcal{G} \to \mathbb{N}$ that is the number of closed generators, but also to $o_\eta : \mathcal{G} \to \mathbb{N}$ that is the number of open generators. Then by consequence $\forall g \in \mathcal{G}, \sum_{0\leq R\leq g} \mathbb{m}_\eta(g, R) = \mathbb{m}_\eta(g, 0) + \sum_{0<R\leq g} \mathbb{m}_\eta(g, R) = c_\eta + o_\eta$.

3.1.4. Petri-net reading and concurrency

The previous definitions can be read as a Petri net. A mark is a multi set of tokens. In our case, a token is either a closed copy of a generator or an open copy with a residual state. The marking is therefore the hidden interface of the experiment. There are three elementary kinds of transitions.

- **Opening**: choose $g \in \mathcal{G}$ and create a new open copy (g, g).
- **Emission**: if an open copy is (g, R) and $R(v) > 0$, emit the event v and replace (g, R) by $(g, R - \varepsilon_v)$.
- **Closing**: if an open copy has residual 0, replace it by one closed copy of g.

The emission transitions on two different open copies are independent. They can be performed in either order and lead to the same marking. This is the basic Petri-net notion of concurrency: the model does not force a single total order between independent local emissions. The observer may later apply the Parikh[†] compression and keep only the count

[†]A reference to Parikh vector define by counting the letter use in a word.

N_η, but the hidden marking still remembers which copies are open and what remains to be produced.

PROPOSITION 3.4 (Closed states recover the semigroup). *Suppose that $\mathcal{G}$ is closed-complete for $\mathcal{S}$. If a dynamic state η has no open copy, then $(N_\eta, t_\eta) \in \mathcal{S}$, where $t_\eta := \sum_{g\in\mathcal{G}} c_\eta(g)\ell(g)$. Conversely, every $(N,t) \in \mathcal{S}$ is represented by at least one dynamic state with no open copy.*

Proof.

If O_η is empty, then $N_\eta = \sum_{g\in\mathcal{G}} c_\eta(g)g$. Since each $(g, \ell(g))$ belongs to $\mathcal{S}$ and $\mathcal{S}$ is closed under addition, (N_η, t_η) belongs to $\mathcal{S}$. Conversely, closed completeness gives a decomposition of every $(N,t) \in \mathcal{S}$ as a finite sum of generators. This decomposition defines a state with the corresponding closed multiplicities and with no open copy.

3.2. From observations to theory: the three-layer architecture -

We now describe the full architecture of a dynamic contextual theory. The framework separates three layers: the raw observations, the dynamic explanations, and the constraint that selects admissible dynamics.

3.2.1. Layer 1: The sequence of observations

An experiment that can be interrupted produces not a single completed counting, but a sequence of partial observations. At each interruption time i, the experimenter records an integer counting $N_i \in \mathbb{N}^V$. The full experimental record is the sequence $\boldsymbol{N_{1:k}} := (N_1, ..., N_k)$. Each N_i is a snapshot of the process at time i. It records what has been seen so far, but not how it was produced. The sequence carries more information than the final count alone: it preserves the order of events, the intermediate states, and the possibility of interruption.

3.2.2. Layer 2: The dynamic states and their transitions

For each observation N_i, the theory assigns a set of possible *hidden explanations* $\mathcal{H}_\mathcal{G}(\boldsymbol{N_i}) := \{\eta_i \mid N_{\eta_i} = N_i\}$. Each η_i remembers which generators are closed (stabilized) and which are open (still in progress). The *transition condition* $\eta_i \rightsquigarrow \eta_{i+1}$ says that the explanation at time $i+1$ is compatible with the explanation at time i. No copy that already exists may disappear. Moreover, a copy cannot change its generator, and its residual future can only shrink. Then the dynamics preserves the identity of the elementary process while allowing part of its remaining future to be emitted.

To define this condition formally, we use the extended multiset $\mathbb{O}_\eta$ rather than O_η. Recall that $\mathbb{O}_\eta$ also contains closed copies, represented as pairs $(g, 0)$. Since $\mathbb{O}_\eta$ is a multiset, a function between two multisets must distinguish individual occurrences. We therefore introduce the **occurrence set** of a multiset $\mathbb{O}$ as

$$\hat{\mathbb{O}} := \{(g, R, i) \mid (g, R) \texttt{ in } \mathbb{O}, \quad 1 \le i \le \mathbb{m}(g, R)\}.$$

Each triple (g, R, i) represents one specific copy: g is the generator, R is the residual, and i distinguishes copies that share the same pair (g, R).

DEFINITION 3.5 (Transition condition). *Let η and η' be two dynamic states. We define the condition $\eta \rightsquigarrow \eta'$ by the existence of a function $\varphi : \hat{\mathbb{O}}_\eta \to \hat{\mathbb{O}}_{\eta'}$ such that:*

(1) φ is injective: for all $x, y \in \hat{\mathbb{O}}_\eta$, if $x \neq y$ then $\varphi(x) \neq \varphi(y)$;
(2) φ preserves the generator and reduces the residual: if we write for every $(g, R, i) \in \hat{\mathbb{O}}_\eta$, $\varphi(g, R, i) = (g', R', i')$, then $g' = g$ and $0 \leq R' \leq R$.

We can also say: for each generator $g \in \mathcal{G}$ and each residual $0 \leq R \leq g$, there exists an injection $\sigma_{g,R} : \{1, \ldots, \mathbb{m}_\eta(g, R)\} \to \{1, \ldots, \mathbb{m}_{\eta'}(g, R')\}$ such that $R'_{\sigma_{g,R}(i)} \leq R_i$ for every i. The map $\sigma_{g,R}$ is only required to be injective, not bijective. This reflects the fact that a transition may open new copies of a generator, but it cannot close or destroy existing ones: every copy present in η must still be present in η', possibly with a smaller residual.

The full fibre of *dynamic explanations* for the sequence is

$$\mathcal{H}_{\mathcal{G}}(\boldsymbol{N_{1:k}}) := \{(\eta_1, \ldots, \eta_k) \mid \forall i, \eta_i \in \mathcal{H}_{\mathcal{G}}(N_i) \text{ and } \eta_i \rightsquigarrow \eta_{i+1}\}.$$

This fibre may contain many trajectories. Two different trajectories can explain the same observation sequence with different distributions of open and closed copies and different residual futures.

3.2.3. Layer 3: The dynamic constraint ξ

The fibre $\mathcal{H}_{\mathcal{G}}(N_{1:k})$ is typically too large to be useful by itself. The third layer introduces a constraint ξ that filters the fibre to select only those trajectories that satisfy a global stability condition.

The dynamic fit under constraint ξ (**predicate**) is

$$\mathcal{H}_{\mathcal{G},\xi}(\boldsymbol{N_{1:k}}) := \{(\eta_1, \ldots, \eta_k) \in \mathcal{H}_{\mathcal{G}}(N_{1:k}) \mid (\eta_1, \ldots, \eta_k) \vDash \xi\}.$$

The observation sequence is *explained* by the theory $(\mathcal{G}, \xi)$ if this set is non-empty.

3.2.4. The theory $\mathbb{T} = (\mathcal{G}, \xi)$

A dynamic contextual theory is a pair $\mathbb{T} = (\mathcal{G}, \xi)$, where $\mathcal{G}$ is the **structure** and ξ is the **dynamics**.

The **structure** $\mathcal{G}$ specifies the elementary processes: which generators are available, what their levels are, and how they decompose compatible countings. The choice of $\mathcal{G}$ is a modelling decision. A coarse theory uses only low-level generators. A fine-grained theory uses higher-level generators that can explain more complex observations.

The **dynamics** ξ specifies how the generators can evolve over time. It does not determine a unique trajectory. It constrains the set of admissible trajectories, removing those that violate a stability condition. The dynamics is what makes the theory **predictive**: it says not only what can be observed, but how the hidden process must behave between observations.

Remark.

> If $\mathcal{G}$ is not complete (see Definition 9.1), then the choice of $\mathcal{G}$ already imposes a structural restriction: some completed observations cannot be generated. If $\mathcal{G}$ is complete, however, the structure alone does not select a proper subclass of completed observations. It only re-expresses them as sums of elementary bricks. In that case, a non-trivial theory requires an additional constraint on admissible trajectories. This is the role of ξ.

3.2.5. The transition condition is a modelling choice

The transition condition $\eta \rightsquigarrow \eta'$ is part of the modelling assumptions of the meta-theory. It is not forced by the static semigroup description. Other choices could be imagined, for example allowing a copy to split, two copies to merge, or a copy to change its generator. We do not include such transitions here.

The choice made in Definition 3.5 is more restrictive. It assumes that a copy keeps its identity through time, keeps the same generator, and can only decrease its residual part. In other words, an elementary process may continue to emit events, but it cannot become another process and it cannot erase what has already been emitted. This gives a simple class of theories in which interruptions have a stable meaning. This class seams natural for the automaton interpretation developed here.

PROPOSITION 3.6 (Monotonicity of visible counts). *Let $\mathcal{G}$ a generator family, if $(\eta_1, \dots, \eta_k) \in \mathcal{H}_{\mathcal{G}}(N_{1:k})$, then the observed countings are monotone: $N_1 \leq N_2 \leq \dots \leq N_k$ componentwise in $\mathbb{N}^V$.*

Proof.

It is enough to prove the claim for one transition $\eta \rightsquigarrow \eta'$. By definition, there exists an injective map $\varphi : \hat{\mathbb{O}}_\eta \to \hat{\mathbb{O}}_{\eta'}$ such that every $(g, R, i) \in \hat{\mathbb{O}}_\eta$ is sent to some (g, S, j) with $S \leq R$. For each $x = (g, R, i) \in \hat{\mathbb{O}}_\eta$ write $\varphi(x) =: (g, S_x, j_x)$ with $S_x \leq R$. The visible contribution of x in η is $g - R$, whereas the visible contribution of $\varphi(x)$ in η' is $g - S_x$. Since $S_x \leq R$, we have $g - R \leq g - S_x$ componentwise. We now sum over all occurrences. The total visible contribution of $\mathbb{O}_\eta$ is $N_\eta = \sum_{x \in \hat{\mathbb{O}}_\eta} (g_x - R_x)$. We claim that $N_\eta \leq N_{\eta'}$. The argument proceeds in three steps:

$$N_\eta = \sum_{x \in \hat{\mathbb{O}}_\eta} (g_x - R_x) \underset{1}{\leq} \sum_{x \in \hat{\mathbb{O}}_\eta} (g_x - S_x) \underset{2}{=} \sum_{y \in \varphi(\hat{\mathbb{O}}_\eta)} (g_y - S_y) \underset{3}{\leq} \sum_{y \in \hat{\mathbb{O}}_{\eta'}} (g_y - S_y) = N_{\eta'}.$$

(1) For each $x \in \hat{\mathbb{O}}_\eta$, the term $g_x - R_x \leq g_x - S_x$ componentwise, because $S_x \leq R_x$.
(2) Since φ is injective, the map $x \mapsto \varphi(x)$ is a bijection from $\hat{\mathbb{O}}_\eta$ onto $\varphi(\hat{\mathbb{O}}_\eta)$. Reindexing the sum over the image gives $\sum_{x \in \hat{\mathbb{O}}_\eta} (g_x - S_x) = \sum_{y \in \varphi(\hat{\mathbb{O}}_\eta)} (g_y - S_y)$.
(3) Since $\varphi(\hat{\mathbb{O}}_\eta) \subseteq \hat{\mathbb{O}}_{\eta'}$, we decompose:

$$\sum_{y \in \hat{\mathbb{O}}_{\eta'}} (g_y - S_y) = \sum_{y \in \varphi(\hat{\mathbb{O}}_\eta)} (g_y - S_y) + \underbrace{\sum_{y \in \hat{\mathbb{O}}_{\eta'} \setminus \varphi(\hat{\mathbb{O}}_\eta)} (g_y - S_y)}_{\geq 0 \text{ since } S_y \leq g_y}.$$

Dropping the non-negative remainder gives the inequality. Now let $(\eta_1, \dots, \eta_k) \in \mathcal{H}_{\mathcal{G}}(N_{1:k})$. For every $i < k$, the definition of the fibre gives $\eta_i \rightsquigarrow \eta_{i+1}$ and $N_{\eta_i} = N_i$. Applying the previous paragraph gives $N_i = N_{\eta_i} \leq N_{\eta_{i+1}} = N_{i+1}$. Thus $N_1 \leq N_2 \leq \dots \leq N_k$.

***Remark* (Monotonicity is independent of $\mathcal{G}$).**

The implication proved in Proposition 3.6 does not require any completeness assumption on $\mathcal{G}$. Once a trajectory of dynamic states exists, monotonicity follows only from the transition condition: old copies are mapped injectively to new copies, and their residuals can only decrease.

The converse question is different. A monotone sequence $N_1 \leq \dots \leq N_k$ admits a dynamic explanation for every choice of visible countings if and only if $\mathcal{G}$ is complete.

This notion is defined in the appendix Definition 9.1. There, Proposition 9.2 shows that dynamic completeness is equivalent to the condition that the generators cover all visible events, and Proposition 9.3 proves the converse of monotonicity under this assumption.

3.2.6. A first example of ξ: bounding open generators

The symbol ξ denotes a constraint on trajectories. It is not fixed by the definition of a dynamic contextual automaton. It is an additional modelling choice. A first possible choice, and the one that initially motivated this layer, is to bound the number of open copies of each generator.

For a state η, write $o_\eta(g) \coloneqq \sum_{0<R\leq g} \mathbb{m}_\eta(g,R)$ for the number of open copies of g, and write $c_\eta(g) \coloneqq \mathbb{m}_\eta(g,0)$ for the number of closed copies of g. Fix a slope $\lambda \geq 0$ and a tolerance $B \in \mathbb{N}$. The open-generator constraint is

$$\xi_{\lambda,B}: \quad o_{\eta_i}(g) \leq \lambda c_{\eta_i}(g) + B \quad \forall i \in [k], \quad \forall (\eta_1, \ldots, \eta_k) \in \mathcal{H}_{\mathcal{G}}(N_{1:k}), \quad \forall g \in \mathcal{G}.$$

This condition says that a trajectory is admissible only if, at each interruption time, each generator has a controlled number of open copies relative to its closed copies. The tolerance B allows a bounded amount of unfinished work. The slope λ allows the system to carry more unfinished copies once it has already stabilized many closed copies of the same generator.

This simple constraint already has a dynamic effect. Suppose that an observation N forces many open copies of a generator g. This can be measured by the minimal open pressure

$$\omega_g(N) \coloneqq \min_{\eta \in \mathcal{H}_{\mathcal{G}}(N)} \big(o_\eta(g) - \lambda c_\eta(g)\big).$$

If $\omega_g(N) > B$, then every hidden explanation of N violates ξ_B. Hence no trajectory satisfying ξ_B can pass through this observation. If $\omega_g(N)$ is close to B, then the trajectory is not yet impossible, but its future is strongly constrained: it cannot keep opening new copies of g without also emitting residual events that help existing copies close.

***Remark* (Magnet analogy).**

The constraint $\xi_{\lambda,B}$ behaves like a magnet guiding a trajectory. It does not determine the path: the dynamics of opening, emitting, and closing generators is not fixed by $\xi_{\lambda,B}$ alone. But if the trajectory strays too far, meaning that the number of open copies grows too large relative to the closed ones, the constraint creates a wall that pushes the process back toward stabilization. The magnet does not choose the destination. It prevents the trajectory from escaping into unbounded instability. More precisely, $\xi_{\lambda,B}$ does not select a unique trajectory $\eta_1 \rightsquigarrow \eta_2 \rightsquigarrow \ldots \rightsquigarrow \eta_k$. It filters the set of admissible trajectories by removing those that violate the stability bound at some step.

What ξ is not. It is important to clarify the status of this constraint.

- ξ does not determine a unique hidden state. The same observation sequence may have several admissible trajectories.
- ξ does not choose which event will be emitted next. It only removes continuations that would make the hidden state unstable.
- ξ is not a hidden-variable assignment in the usual sense. It does not assign predetermined values to all measurements. It constrains the evolution of the explanatory process.
- ξ is not part of the visible data. It is a theoretical filter placed on the fibre $\mathcal{H}_{\mathcal{G}}(N_{1:k})$.

In this sense, ξ is closer to a stability law than to a hidden cause. It says which dynamic explanations are admissible. It does not say that one specific explanation is the real one.

Quantum viewed by dynamic contextual automata

CHAPTER 4

The previous sections established the static and dynamic tools: generators, dynamic explanations, and the admissibility constraint ξ. This chapter applies the framework to the $(2, 2, 2)$ Bell scenario. The goal is to define a finite-level Tsirelson bound $T(t)$: the best quantum-realizable integer CHSH score at level t. This makes explicit an arithmetic point that is invisible in the probabilistic setting: every finite level gives a rational score, and the Tsirelson value $2\sqrt{2}$ appears only as an asymptotic limit. Crucially, this operational bound can be approximated using the same tools as the standard approach, in particular the NPA hierarchy. The integer framework therefore does not replace the standard quantum relaxations; it reuses them while keeping the finite-level structure that normalization erases. This chapter does not aim to prove new theoretical results. It illustrates how known techniques from quantum information theory, in particular the NPA hierarchy, can be applied to the specific setting required by our formalization: recovering finite-level integer bounds from an irrational asymptotic value.

4.1. The Tsirelson bound at level t

We have seen that a discrete empirical model of level t is a compatible integer counting $N \in \mathbb{N}^V$ with $A_{\mathcal{M}} N = t\,\mathbb{1}$ and $\delta N = 0$. In the CHSH scenario, the *integer correlator* for context (x, y) is $E(x, y) \coloneqq N(0, 0|x, y) + N(1, 1|x, y) - N(0, 1|x, y) - N(1, 0|x, y)$. The *CHSH score* of a counts N is

$$S(N) \coloneqq E(0, 0) + E(0, 1) + E(1, 0) - E(1, 1).$$

Each correlator $E(x, y)$ is an integer bounded by t, so $S(N)$ is an integer bounded by $4t$. The normalized score $S(N)/t$ recovers the usual probabilistic CHSH expression.

A counts N lies on the *isotropic line*[†] if all four contexts share the same winning count: there exists $w \in \mathbb{N}$ such that $N(0, 0|x, y) + N(1, 1|x, y) = w$ for $(x, y) \in \{(0, 0), (0, 1), (1, 0)\}$, and $N(0, 1|1, 1) + N(1, 0|1, 1) = w$. Equivalently, the correlators satisfy $E(0, 0) = E(0, 1) = E(1, 0) = -E(1, 1) =: c$. The CHSH score is then $S(N)/t = \frac{4c}{t}$. At the quantum optimum, $w = \lfloor t p^* \rfloor$ with $p^* = \frac{2+\sqrt{2}}{4}$, giving

$$T_{\text{iso}}(t) \coloneqq \frac{4(2\lfloor t p^* \rfloor - t)}{t}.$$

Let Q denote the *quantum set*: the set of correlations achievable by some finite-dimensional quantum strategy.

DEFINITION 4.1 (Tsirelson bound at level t). *Let $\mathcal{S}_t$ denote the set of compatible integer countings at level t (those satisfying $A_{\mathcal{M}} N = t\,\mathbb{1}$ and $\delta N = 0$). The* ***quantum-realizable***

[†]One could also say "uniform", since the winning count is the same across all contexts. We prefer "isotropic" to emphasize that the construction is not tied to a specific number of contexts or to the CHSH scenario, and can be extended to more general settings.

countings at level t are $\mathcal{Q}_t := \mathcal{S}_t \cap t \cdot Q$. That is, $\mathcal{Q}_t$ consists of the integer countings at level t whose normalization $\frac{N}{t}$ lies in the quantum set Q. The ***Tsirelson bound*** *at level t is*

$$T(t) := \max_{N \in \mathcal{Q}_t} \frac{S(N)}{t}.$$

Since $\mathcal{Q}_t$ is a finite set, the maximum is attained and $T(t)$ is a rational number.

The Tsirelson bound $T(t)$ is the best normalized CHSH score achievable by any quantum strategy using exactly t repetitions of the experiment.

4.2. The Tsirelson bound is an asymptotic idealization

The celebrated Tsirelson bound states that quantum mechanics cannot produce a CHSH score exceeding $2\sqrt{2}$. In the standard probabilistic framework, this bound is derived from the algebraic structure of Hilbert-space observables. However, from the discrete perspective, the situation is fundamentally different. All the calculation of this part can be found in Section 9.2.4 .

PROPOSITION 4.2 ($\sqrt{2}$ requires infinite experiments). *For every finite level $t \in \mathbb{N}$, the Tsirelson bound satisfies $T(t) < 2\sqrt{2}$. The Tsirelson value is recovered only in the limit:*

$$\lim_{t \to \infty} T(t) = 2\sqrt{2}.$$

Proof.

The proof proceeds in two steps.

(1) **Rationality.** For any $N \in \mathcal{Q}_t$, each correlator $E(x, y)$ is an integer and $S(N)$ is an integer. Hence $\frac{S(N)}{t} \in \mathbb{Q}$. Because $2\sqrt{2} \notin \mathbb{Q}$ then $\frac{S(N)}{t} \neq 2\sqrt{2}$ for any finite t.

(2) **Convergence.** On the isotropic line, one can construct explicit quantum strategies that approach $2\sqrt{2}$ as t grows. The optimal winning probability at level t is $w^*(t) = \lfloor tp^* \rfloor$ where $p^* = \frac{2+\sqrt{2}}{4}$. The resulting score is $T_{\text{iso}}(t) = \frac{4(2w^*(t)-t)}{t}$, and the gap satisfies $2\sqrt{2} - T_{\text{iso}}(t) = \frac{8 \cdot \text{frac}(tp^*)}{t}$, which tends to 0 as $t \to \infty$. Since $T(t) \geq T_{\text{iso}}(t)$, the result follows.

This theorem has a striking interpretation: the $\sqrt{2}$ of Tsirelson is not a parameter of the discrete model. It is an emergent, asymptotic, purely abstract quantity. No finite experiment can ever reach it. The irrationality of $\sqrt{2}$ creates an arithmetic obstruction: at every finite level t, the quantum score is rational, and the gap to $2\sqrt{2}$ is controlled by the Diophantine properties of $\sqrt{2}$ through its continued fraction expansion $\sqrt{2} = 1 + \cfrac{1}{2+\cfrac{1}{2+\cfrac{1}{\cdots}}}$.

Remark.

The gap formula admits a natural decomposition:

$$2\sqrt{2} - T_{\text{iso}}(t) = 2\Big(\sqrt{2} - r_t\Big),$$

where $r_t := \frac{4\lfloor tp^* \rfloor - 2t}{t}$ is the rational approximation to $\sqrt{2}$ at level t. This decomposition reflects the structure of the CHSH game. The score splits into two complementary pairs of contexts:

$$S(N) = \underbrace{(E(0,0) + E(1,0))}_{\text{pair 1}} + \underbrace{(E(0,1) - E(1,1))}_{\text{pair 2}}.$$

Each pair ideally contributes $t\sqrt{2}$, but at level t each contributes only tr_t. The per-pair deficit is $t\left(\sqrt{2} - r_t\right)$, and the total deficit is twice this value.

4.3. *Approximation via the NPA hierarchy*

Computing $T(t)$ exactly is a hard combinatorial problem. One would need to check, for each integer counting N, whether its normalization N/t can be realized by a genuine quantum strategy. This requires finding a Hilbert space, a quantum state, and measurement operators satisfying all the rules of quantum mechanics, which is not directly tractable.

The NPA hierarchy [NPA08] provides a systematic way to approximate the quantum set from the outside. The idea is as follows. Instead of searching explicitly for the quantum state and operators, one constructs a matrix Γ whose entries represent correlations between operators. This matrix must satisfy a key constraint: it must be positive semidefinite, $\Gamma \succeq 0$. This condition forces the correlations to be consistent with a genuine quantum system. The matrix Γ must also satisfy linear constraints that encode the algebraic rules of quantum measurements: projectors, exclusivity of outcomes, normalization, and commutativity between distant parties.

By allowing products of measurement operators up to degree k, one obtains a relaxation $\mathcal{Q}_k$ of the true quantum set $\mathcal{Q}$. The larger k is, the tighter the relaxation. Each $\mathcal{Q}_k$ is described by a semidefinite program (SDP): one maximizes the CHSH score over all matrices Γ that satisfy the positive-semidefiniteness and algebraic constraints. The NPA bound at level k is $\omega_k := \max_{p \in \mathcal{Q}_k} S(p)$. This is a semidefinite program, solvable by standard SDP solvers. Its value satisfies $\omega_k \geq 2\sqrt{2}$ for every k, and ω_k decreases to $2\sqrt{2}$ as $k \to \infty$.

The important point for our purposes is that the NPA hierarchy gives a finite description of an outer approximation to the quantum set. By solving the dual SDP for several objective functions, one obtains a finite system of M linear inequalities that outer-approximates $\mathcal{Q}_k$ [NPA08]. These inequalities can then be used as cuts in an integer program.

4.3.1. The MILP for $T_k(t)$

To compute $T_k(t)$ over integer countings, one replaces $p(a,b|x,y)$ by $\frac{N(a,b|x,y)}{t}$ and optimizes:

$$\begin{aligned} T_k(t) = \max \quad & \frac{1}{t} \sum_{a,b,x,y} (-1)^{a \oplus b \oplus (x \wedge y)} N(a,b,x,y) \\ \text{subject to} \quad & A_{\mathcal{M}} N = t\,\mathbb{1} \quad \text{(normalization)}, \\ & \delta N = 0 \quad \text{(no-signaling)}, \\ & \sum_{a,b,x,y} \alpha^{(m)}_{a,b,x,y} N(a,b,x,y) \leq t\beta_m \quad \forall m \quad \text{(NPA cuts)}, \\ & N(a,b,x,y) \in \mathbb{N} \quad \forall a,b,x,y. \end{aligned}$$

This MILP is solvable by standard solvers (Gurobi, SCIP, CPLEX). The NPA cuts prune the integer search space: they eliminate countings whose normalization lies outside $\mathcal{Q}_k$, without requiring the solver to handle semidefinite constraints.

4.3.2. Sandwich

PROPOSITION 4.3 (NPA sandwich). *For every finite t and every NPA level k:*

$$T_{\text{iso}}(t) \quad \leq \quad T(t) \quad \leq \quad 2\sqrt{2} \quad \leq \quad T_k(t).$$

Proof.

(1) $T_{\text{iso}(t)} \leq T(t)$**.** The isotropic strategy is a particular quantum strategy.
(2) $T(t) \leq 2\sqrt{2}$**.** The Tsirelson bound holds for all quantum correlations.
(3) $2\sqrt{2} \leq T_{k(t)}$**.** Since $\mathcal{Q} \subseteq \mathcal{Q}_k$, the Tsirelson-optimal correlation lies in $\mathcal{Q}_k$. At level t, one can approximate it by an integer counting, so the maximum over $\mathcal{Q}_k$ is at least $2\sqrt{2}$.

The both direct and NPA approch squeeze the Tsirelson bound: $T(t) \leq T_k(t)$, and both converge to $2\sqrt{2}$ as t and k grow.

4.3.3. The see-saw method for lower bounds

The NPA hierarchy approximates the quantum set $\mathcal{Q}$ from the outside. It therefore gives upper bounds on $T(t)$. Lower bounds require explicit quantum strategies whose normalized countings have high CHSH scores at level t. Since such strategies are hard to find by direct enumeration, one can use the *see-saw method*, also called alternating optimization [AMP24].

The idea is to relax the optimization over quantum strategies as a sequence of semidefinite programs. Fix a dimension d for the Hilbert space. Starting from a randomly chosen quantum state $\rho \in \mathcal{H}_A \otimes \mathcal{H}_B$ and measurement operators $\{A_x^a\}, \{B_y^b\}$, the algorithm alternates between three steps:

1. **Fix the state, optimize the measurements.** For each party separately, solve an SDP to find the measurement operators that maximize the CHSH score while keeping ρ and the other party's measurements fixed.
2. **Fix the measurements, optimize the state.** Solve an SDP to find the state ρ that maximizes the CHSH score given the current measurements.
3. **Repeat** until convergence (or for a fixed number of iterations).

The best CHSH score found across multiple random initializations provides a lower bound $T_{\text{seesaw}}(t)$ on $T(t)$.

***Remark* (See-saw: no convergence guarantee).**

The see-saw method is a heuristic. The optimization landscape is biconvex (convex in the state when the measurements are fixed, and vice versa), but **not jointly convex**. Therefore:

- There is no guarantee of convergence to the global optimum.
- The algorithm may get stuck in local optima.
- Different random initializations may yield different results.

4.4. The three levels at each t

At each finite level t, the discrete framework separates three sets of integer countings:

- **Local:** $\mathcal{S}_{\text{loc}}(t)$ contains the noncontextual integer models. These satisfy $\frac{S(N)}{t} \leq 2$.

- **Quantum:** $\mathcal{S}_{\text{quant}}(t) = \mathcal{Q}_t$ contains the quantum-realizable integer models. These satisfy $\frac{S(N)}{t} \leq T(t) < 2\sqrt{2}$.
- **No-signaling:** $\mathcal{S}_{\text{ns}}(t)$ contains all compatible integer models. These satisfy $\frac{S(N)}{t} \leq 4$.

The inclusions $\mathcal{S}_{\text{loc}}(t) \subseteq \mathcal{S}_{\text{quant}}(t) \subseteq \mathcal{S}_{\text{ns}}(t)$ hold at every level. The Tsirelson bound $T(t)$ is the frontier of the quantum set inside the no-signaling set. It is a rational number that depends on the arithmetic of t and on the Diophantine properties of $\sqrt{2}$.

***Remark* (Non-monotonicity of $T(t)$).**

> The function $t \mapsto T(t)$ is not monotone. For example, $T(7) = 2.571 < T(6) = 2.667$. This happens because the arithmetic structure of the integer lattice changes with t: some levels admit particularly good rational approximations to the quantum set, while others do not.

Fractional contextuality is equivalent to half distance

CHAPTER 5

The previous chapters introduced the integer counting framework and its dynamic extension. We now show that this construction recovers the contextual fraction of Abramsky, Barbosa, and Mansfield [ABM16]. The goal is to translate their argument into the integer counting setting and to show that the two frameworks coincide at the level of the non-contextual part.

This result was not expected. The framework was built independently, guided by the idea that integer counts preserve information lost by normalization. The fact that it recovers an existing measure of contextuality gives it additional value: it can be compared directly with the literature. Moreover, the dynamic framework adds a second layer: the completion deficit, which measures how far the non-classical residue is from being closed (it can be views as the dual of contextual fraction).

In the first part, we recall the classical sheaf-theoretic setting with $R = \mathbb{R}_{\geq 0}$.

5.1. Overview of [ABM16]

In the standard sheaf-theoretic approach, an empirical model is a compatible family of probability distributions

$$e = (e_C)_{C \in \mathcal{M}}, \quad e_C \in \mathcal{D}_{\mathbb{R}_{\geq 0}}(\mathcal{E}(C)), \quad \sum_{s \in \mathcal{E}(C)} e_C(s) = 1.$$

Compatibility means $e_C|_{C \cap D} = e_D|_{C \cap D}$ for all contexts C, D. A global section $g : X \to O$ is **consistent** with e if $e_C\big(g|_C\big) > 0$ for every context C. The set of consistent global sections is

$$S_e(X) = \big\{ g \in \mathcal{E}(X) \mid \forall C \in \mathcal{M}, e_C\big(g|_C\big) > 0 \big\}.$$

When $S_e(X) = \varnothing$, the model is strongly contextual: no single global assignment is compatible with the support of every context. If no global section explains the entire model, one can ask whether a global section explains a **fraction** of it. For $r \in (0, 1]$, define

$$S_e^{\geq r}(X) = \big\{ g \in \mathcal{E}(X) \mid \forall C \in \mathcal{M}, e_C\big(g|_C\big) \geq r \big\}.$$

A global section $g \in S_e^{\geq r}(X)$ is one whose restriction to each context has probability at least r. The sets form an anti-monotone family: if $r \leq r'$, then $S_e^{\geq r}(X) \supseteq S_e^{\geq r'}(X)$. Their union recovers the possibilistic notion: $\bigcup_{r \in (0,1]} S_e^{\geq r}(X) = S_e(X)$. The key observation is that a global section in $S_e^{\geq r}(X)$ can explain away a fraction r of the model.

PROPOSITION 5.1 (See [ABM16, Proposition 3.2]). *If $g \in S_e^{\geq r}(X)$, then e admits a convex decomposition $e = r\delta_g + (1 - r)e'$ where $\delta_g \in \mathcal{D}_{\mathbb{R}_{\geq 0}}\mathcal{E}(X)$ is the deterministic model at g, i.e. $\delta_g(g') = \begin{cases} 1 \text{ if } g = g' \\ 0 \text{ otherwise} \end{cases}$ and e' is another empirical model.*

This says: a fraction r of the model is explained by the classical history g. The remainder $(1-r)$ is another model e'.

5.1.1. The overlap problem

Now suppose we find **two** global sections:

- $g_1 \in S_e^{\geq r_1}(X)$, explaining a fraction r_1;
- $g_2 \in S_e^{\geq r_2}(X)$, explaining a fraction r_2.

Can we combine them to explain a fraction $r_1 + r_2$ such that $e \stackrel{?}{=} r_1\delta_{g_1} + r_2\delta_{g_2} + (1 - r_1 - r_2)e''$.

The answer is **no**, in general. The problem is that g_1 and g_2 may explain the **same part** of the model. If both assign the same local outcome s in a context C, and $e_C(s) = 0.6$, then $r_1 = 0.4$ and $r_2 = 0.4$ would require $r_1 + r_2 = 0.8 > 0.6 = e_C(s)$, which is impossible. The explanations **overlap**. In the possibilistic case, this is harmless: $1 \vee 1 = 1$. In the probabilistic case, $0.4 + 0.4 = 0.8 > 0.6$: double counting is real.

***Example 5.1* (Why fractional sections cannot be naively added).**

Consider three measurements $X = \{a, b, c\}$ with contexts $C_1 = \{a, b\}$, $C_2 = \{b, c\}$, $C_3 = \{a, c\}$ and outcomes $O = \{0, 1\}$. The following is a compatible (no-signaling) probability model:

C	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$	total
$\{a,b\}$	0.35	0.15	0.25	0.25	1
$\{b,c\}$	0.30	0.20	0.10	0.40	1
$\{a,c\}$	0.30	0.10	0.20	0.40	1

Pick two global sections:

- $g_1 = (a \mapsto 0, b \mapsto 0, c \mapsto 0)$: its restrictions are $(0,0)$ in every context, with probabilities 0.35, 0.30, 0.30. So $g_1 \in S_e^{\geq 0.30}(X)$.
- $g_2 = (a \mapsto 0, b \mapsto 0, c \mapsto 1)$: its restrictions are $(0,0)$ in C_1, $(0,1)$ in C_2, $(0,1)$ in C_3, with probabilities 0.35, 0.10, 0.20. So $g_2 \in S_e^{\geq 0.10}(X)$.

Both sections are in $S_e(X)$: every restriction has positive weight. Each one, taken alone, yields a valid decomposition: $e = 0.30\delta_{g_1} + 0.70e'$ and $e = 0.10\delta_{g_2} + 0.90e''$. Now try to **combine** them: $e \stackrel{?}{=} 0.30\delta_{g_1} + 0.10\delta_{g_2} + 0.60e'''$. At context C_1, both g_1 and g_2 restrict to the **same** local assignment $(a \mapsto 0, b \mapsto 0)$. The proposed mixture places mass $0.30 + 0.10 = 0.40$ on this event, but $e_{C_1}(a \mapsto 0, b \mapsto 0) = 0.35 < 0.40$. This forces $e'''_{C_1}(a \mapsto 0, b \mapsto 0) < 0$: **impossible**.

The root cause is that g_1 and g_2 draw from the same local pool of probability. Each one individually can claim up to 0.35 from the event $(a \mapsto 0, b \mapsto 0)$ in C_1, but not both at once. The fractional weights r_1 and r_2 are *not independent*: they compete at shared local assignments.

This is the overlap problem: fractional explanations cannot simply be added because they may draw from the same pool of probability at some context. The next construction, subdistributions, is designed to prevent exactly this double counting.

5.1.2. Subdistributions

To handle overlap, [ABM16] introduce *subdistributions*: functions whose weights sum to **at most** 1, instead of exactly 1. An $\mathbb{R}$-*subdistribution* on a set S is a function $c : S \to \mathbb{R}_{\geq 0}$

with finite support and total weight $\omega(c) = \sum_{s \in S} c(s) \leq 1$. The weight $\omega(c)$ is the fraction of the model explained by c. The subdistribution framework allows multiple global sections to be combined without exceeding the probability of any local event. The set of *consistent subdistributions* is

$$C_e(X) \coloneqq \left\{ c \in \mathcal{SD}_{\mathbb{R}_{\geq 0}}(\mathcal{E}(X)) \mid \forall C \in \mathcal{M}, \forall s \in \mathcal{E}(C), e_C(s) \geq c|_C(s) \right\}.$$

where $\mathcal{SD}_{\mathbb{R}_{\geq 0}} : \mathsf{Set} \to \mathsf{Set}$ is the *subdistributions functor*[†]. The condition $e_C(s) \geq c|_C(s)$ means: the mass that c places on global assignments restricting to s must not exceed the probability that e gives to s. This condition prevents double counting: if two global sections g_1 and g_2 both restrict to s in context C, then $c(g_1) + c(g_2) \leq e_C(s)$.

PROPOSITION 5.2 (See [ABM16, Proposition 3.5]). *If $c \in C_e(X)$, then*

$$e = \left(\sum_{g \in \mathcal{E}(X)} c(g)\delta_g \right) + (1 - \omega(c))e'.$$

This decomposes e into a non-contextual part of weight $\omega(c)$ and a remainder e'.

5.1.3. The non-contextual fraction

The *non-contextual fraction* of e is the maximum weight achievable by a consistent subdistribution:

$$\mathrm{NCF}(e) = \max\{\omega(c) \mid c \in C_e(X)\}.$$

The *contextual fraction* is

$$\mathrm{CF}(e) = 1 - \mathrm{NCF}(e).$$

This maximum exists because $C_e(X)$ is closed and bounded in $\mathbb{R}^{|\mathcal{E}(X)|}$, hence compact by Heine–Borel, and ω is continuous. The dual linear program produces a Bell inequality whose normalised violation by e equals $\mathrm{CF}(e)$. Thus the contextual fraction is both a measure of non-classicality and a witness of Bell violation.

5.2. Translation to integer counts

We now translate the entire construction into the integer counting setting. Let $N \in \mathbb{N}^V$ be a compatible integer counting of level t. The normalized model is $e_N = N/t$. Because we have the same problem of overlap problem, see Section 9.2.3, we want to define an integer analogue of consistent subdistributions.

DEFINITION 5.3 (Integer consistent subcountings). *The set of **consistent subcountings** is*

$$C_N(X) = \left\{ z \in \mathbb{N}^{\mathcal{E}(X)} \mid \forall C \in \mathcal{M}, \forall s \in \mathcal{E}(C), \sum_{g|_C = s} z(g) \leq N_C(s) \right\}.$$

Equivalently, in matrix form:

$$C_N(X) = \left\{ z \in \mathbb{N}^{\mathcal{E}(X)} \mid Mz \leq N \right\},$$

where M is the incidence matrix with $M_{(C,s),g} = 1$ if $g|_C = s$ and 0 otherwise.

The weight of a subcounting z is $\omega(z) = \sum_g z(g) = \mathbf{1}^\perp z$.

[†]This is categorial theory notation, see [ABM16] to more details

***Remark* (Subcountings as closed generators).**

In the integer setting, $C_N(X)$ has a direct interpretation in terms of the dynamic framework. A vector $z \in \mathbb{N}^{\mathcal{E}(X)}$ assigns to each global section g a multiplicity $z(g) \in \mathbb{N}$: the number of times g is used as a closed (deterministic) generator. The constraint $Mz \leq N$, or equivalently $\sum_{g:g|_C=s} z(g) \leq N_C(s)$ for every local event (C, s), says exactly that the closed generators must *fit inside* N: no local event receives more mass from the deterministic explanation than was actually observed. In the notation of the dynamic framework, if η is a dynamic state with closed part $P_\eta = Mz$ (the closed part is $\sum_{g\in\mathcal{G}} c_\eta(g) \cdot g$), then $z(g) = c_\eta(g)$ counts how many times g appears as a closed generator. The condition $z \in C_N(X)$ is then simply $P_\eta \leq N$, i.e. the closed part does not exceed the observation.

PROPOSITION 5.4 (Integer decomposition). *If $z \in C_N(X)$, then $N = Mz + R$ where $R = N - Mz \geq 0$. Here Mz is the* ***closed part*** *(a combination of deterministic global sections) and R is the* ***open part*** *(the residue not yet explained by classical histories).*

Proof.

By definition, $Mz \leq N$ componentwise, so $R = N - Mz \geq 0$. The matrix Mz is compatible because each column of M is a deterministic model, and compatibility is preserved under non-negative integer combinations.

This decomposition is the integer analogue of Proposition 5.2. It says exactly what the dynamic framework says: an observation N splits into a closed part Mz (stabilized, non-contextual) and an open part R (still in progress, potentially contextual).

5.3. The past distance recovers the contextual fraction

We now state the main theorem linking the two frameworks.

PROPOSITION 5.5 (The closed deficit equals the contextual fraction). *Let $N \in \mathbb{N}^V$ be a compatible integer counting of level t, and let $e_N = N/t$. Define the* closed deficit *as the minimal mass not explained by deterministic global sections (or maximal mass explained by global section):*

$$D^-(N) := \min_{z\in C_N(X)} |N - Mz|_1.$$

Then

$$D^-(N) = |N|_1 \cdot \mathrm{CF}(e_N).$$

Equivalently, $\dfrac{D^-(N)}{|N|_1} = \mathrm{CF}(e_N)$.

Proof.

Since $Mz \leq N$, we have $|N - Mz|_1 = |N|_1 - |Mz|_1$. Each column of M has exactly $|\mathcal{M}|$ entries equal to 1 (one per context), so $|Mz|_1 = |\mathcal{M}| \cdot \mathbf{1}^\perp z$. Also $|N|_1 = |\mathcal{M}|\, t$. Therefore:

$$D^-(N) = |\mathcal{M}|\, t - |\mathcal{M}| \max_{z\in C_{N(X)}} \mathbf{1}^\perp z = |\mathcal{M}| \left(t - \max_{z\in C_{N(X)}} \mathbf{1}^\perp z \right).$$

Now set $c = \dfrac{z}{t}$. The constraint $Mz \leq N$ becomes $Mc \leq \dfrac{N}{t} = e_N$. Hence

$$\max_{z\in C_{N(X)}} \mathbf{1}^\perp z = t \max_{c\geq 0, Mc\leq e_N} \mathbf{1}^\perp c = t \cdot \mathrm{NCF}(e_N).$$

Substituting:

$$D^{-}(N) = |\mathcal{M}|\, t(1 - \mathrm{NCF}(e_N)) = |N|_1 \cdot \mathrm{CF}(e_N).$$

This theorem says: the minimal mass of a compatible counting that cannot be explained by deterministic global sections is **exactly** the contextual fraction of the normalized model, multiplied by the total count.

5.4. *The dynamic enrichment*

The sheaf-theoretic framework stops at the contextual fraction: it measures **how much** of the model is non-classical. The dynamic framework adds a second layer: it measures **how far** the non-classical residue is from being closed. For this reason, the distances are first defined for a dynamic state itself, and only then minimized over the dynamic explanations of a fixed observation.

This also provides a natural visualization: each dynamic state lies on a trajectory between its closed past and its completed future, and the distances measure how far the state has progressed along this trajectory.

DEFINITION 5.6 (Past and future distances). *Let $\mathcal{G}$ be a family of generators, and let $\eta = (c_\eta, O_\eta)$ be a dynamic state. Its closed part is $P_\eta := \sum_{g\in\mathcal{G}} c_\eta(g)g$. Its visible emitted counting is $N_\eta := P_\eta + \sum_{(g,R)\ \mathrm{in}\ O_\eta}(g - R)$. Its completed counting is $F_\eta := P_\eta + \sum_{(g,R)\ \mathrm{in}\ O_\eta} g$, obtained by closing every open generator. The **past and future distances** of η are*

$$d^-_\eta := |N_\eta - P_\eta|_1, \quad d^+_\eta := |F_\eta - N_\eta|_1.$$

Thus d^-_η is the emitted mass carried by open generators, while d^+_η is the residual mass still needed to close them.

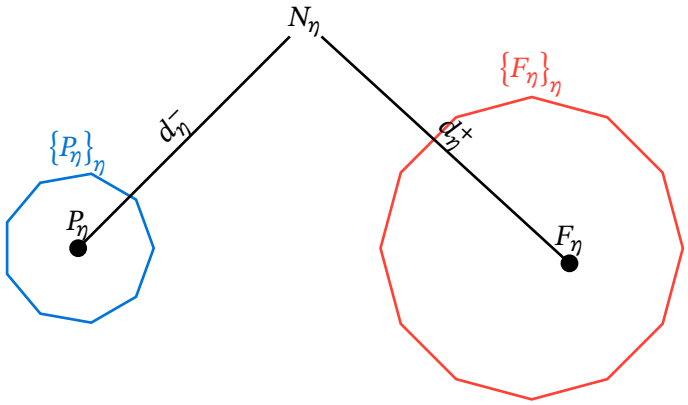

Figure 4: This represent the past and future distance for a fixed dynamic states η.

We can see in Figure 4 that we have one past and future distance by dynamic states $\eta \in \mathcal{H}_\mathcal{G}(N)$.

If η explains the observed counting N, this means $N_\eta = N$. In that case the same quantities may be written as $d^-_\eta = |N - P_\eta|_1$ and $d^+_\eta = |F_\eta - N|_1$, but N is not part of the intrinsic definition of the distances.

For a fixed observed counting N, let $\mathcal{H}_{\mathcal{G}}(N) \coloneqq \{\eta \mid N_\eta = N\}$ be the set of dynamic explanations of N over $\mathcal{G}$. We define $D^-(N) \coloneqq \min_{\eta \in \mathcal{H}_{\mathcal{G}}(N)} d^-_\eta$. We can see that this is the same that $\min_{z \in C_N(X)} |N - Mz|_1$.

For the future distance we have two possible choice, either we fixed η to me the one realize the minimum for $D^-(N)$ and we define the $D^+(N) \coloneqq d^+_\eta$, or we minimize overall explanations of N without constraint. The second choice would be more natural if we were only interested in the future distance as a measure of how far N is from being closed. But our goal is to enrich the contextual fraction, which is already captured by the past distance. We want to measure how far the **residue** left by the optimal past choice is from being closed. The reason is that $D^-(N)$ already selects the largest closed, non-contextual part of N. The relevant future cost is therefore the cost of completing the residue left by such an optimal past choice. We define the residual distance

$$\boldsymbol{D^{\curvearrowright}(N)} \coloneqq \min\{d^+_\eta \mid \eta \in \mathcal{H}_{\mathcal{G}}(N) \wedge d^-_\eta = D^-(N)\}.$$

DEFINITION 5.7 (Tight distances). *Let $\mathcal{G}$ be a complete family of generators. Thus, for every visible counting $N \in \mathbb{N}^V$, the fibre $\mathcal{H}_{\mathcal{G}}(N)$ is non-empty. For a dynamic state $\eta \in \mathcal{H}_{\mathcal{G}}(N)$, let P_η denote its closed part and let F_η denote its completed part, obtained by closing all open generators. We define $d^-_\eta \coloneqq |N - P_\eta|_1$ and $d^+_\eta \coloneqq |F_\eta - N|_1$.*
*The **integer contextual deficit** of N, i.e. the non-normalised contextual fraction of the empirical model associated with N, is*

$$D^-(N) \coloneqq \min_{\eta \in \mathcal{H}_{\mathcal{G}}(N)} d^-_\eta.$$

*The **contextual completion deficit** of N is the minimal future deficit among the states that already minimize the past deficit:*

$$D^{\curvearrowright}(N) \coloneqq \min\{d^+_\eta \mid \eta \in \mathcal{H}_{\mathcal{G}}(N) \text{ and } d^-_\eta = D^-(N)\}.$$

The pair

$$(D^-(N), D^{\curvearrowright}(N)) = \left(\underbrace{|N|_1 \cdot \mathrm{CF}(e_N)}_{\text{contextual fraction (sheaf theory)}} , \underbrace{D^{\curvearrowright}(N)}_{\text{completion cost after optimal past extraction}} \right)$$

*is called the **tight contextual profile** of N. We define the dual of this contextual profile, or **dual tight contextual profile** as the following pair :*

$$(D^{\curvearrowleft}(N), D^+(N)) \coloneqq \left(\min\{d^-_\eta \mid \eta \in \mathcal{H}_{\mathcal{G}}(N) \text{ and } d^+_\eta = D^+(N)\}, \min\{d^+_\eta \mid \eta \in \mathcal{H}_{\mathcal{G}}(N)\}\right)$$

The contextual completion deficit can be visualized as shown in Figure 5.

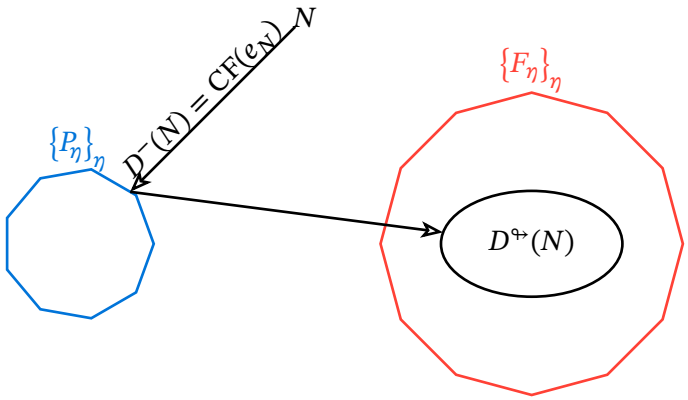

Figure 5: Representation of the integer contextual deficit and the contextual completion deficit. The polygon $\{P_\eta\}$ and $\{F_\eta\}$ represent the set of all possible past P_η and completion F_η such that $\eta \in \mathcal{H}_{\mathcal{G}}(N)$.

5.5. *Summary of the correspondence*

The following table summarizes the parallel between the probabilistic and integer frameworks.

Table 1: Complete correspondence between the probabilistic sheaf-theoretic framework and the integer counting framework.

Sheaf theory (probabilistic)	Our framework (integer)	Relationship
$e_C \in \mathcal{D}_{\mathbb{R}}(\mathcal{E}(C))$	$N_C \in \mathbb{N}^{\mathcal{E}(C)}$	$e_C = \frac{N_C}{t}$
$\sum_s e_{C(s)} = 1$	$\sum_s N_{C(s)} = t$	normalization
$S_e^{\geq r}(X)$	$S_N^{\geq r}(X)$	same definition
$e = r\delta_g + (1-r)e'$	$N = r\delta_g + R$	fractional decomposition
$C_e(X) = \{c : Mc \leq e\}$	$C_N(X) = \{z : Mz \leq N\}$	$c = \frac{z}{t}$
$e = \sum c(g)\delta_g + (1-\omega(e))e'$	$N = Mz + R$	subdistribution decomposition
$\mathrm{NCF}(e) = \max \omega(c)$	$\mathrm{NCF}(e_N) = \max \mathbf{1}^\perp \frac{z}{t}$	same LP
$\mathrm{CF}(e) = 1 - \mathrm{NCF}(e)$	$\frac{D^-(N)}{\lvert N\rvert_1} = \mathrm{CF}(e_N)$	**main theorem**
χ	$D^{\leftrightarrow}(N)$	**dynamic enrichment**

The last row is the main result: the normalized closed deficit in the integer framework equals the contextual fraction in the probabilistic framework. The two measures are the same quantity, seen from two sides of the normalization barrier.

Exploring dynamic constraints for quantum non-locality

CHAPTER 6

This chapter explores how the dynamic framework developed in the previous chapters can be used to investigate quantum correlations. We then explain how this objective can be approached through the guiding ideas of interruption, discrete reconstruction, and dynamic explanation. The investigation starts from simple scalar constraints, tests them on concrete examples, and shows how their limitations lead to finer objects like the interruption profile. The results of this section are still a work in progress.

6.1. Objective: a dynamic theory for the finite quantum set

The previous chapters built the static and dynamic language. We now state the target of the programme more explicitly. At each finite level t, let $\mathcal{Q}_t$ be the set of compatible integer countings whose normalization belongs to the Hilbert-space quantum set:

$$\mathcal{Q}_t := \left\{ N \in \mathcal{S}_t \mid \frac{N}{t} \in Q \right\}.$$

Here $\mathcal{S}_t$ is the set of compatible integer countings at level t, and Q is the usual quantum set of correlations. The ideal goal is to find a dynamic contextual theory $\mathbb{T}_q = (\mathcal{G}_q, \xi_q)$ such that, for every finite level t and every compatible counting $N \in \mathcal{S}_t$,

$$N \in \mathbb{T}_q \quad \Leftrightarrow \quad N \in \mathcal{Q}_t.$$

The notation $N \in \mathbb{T}_q$ means that N admits at least one admissible dynamic explanation under the theory:

$$N \in \mathbb{T}_q \quad :\Leftrightarrow \quad \mathcal{H}_{\mathcal{G}_q, \xi_q}(N) \neq \varnothing.$$

This chapter does not claim to have found such a theory. It formulates the target clearly and presents the first explorations toward it.

6.2. Warm-up: open-generator stability

The first constraint we tried was to control the number of open generators. For a dynamic state $\{\eta_i\} \in \mathcal{H}_{\mathcal{G}}(N_{1:k})$, let o_i be the number of open copies and let c_i be the number of closed copies of the state η_i. The basic stability constraint is

$$\xi_{\lambda,B} : \quad o_i \leq \lambda c_i + B \quad \forall i.$$

The parameter λ is a slope and B is a buffer. This is the formal version of the magnet analogy from Remark 3.6: the constraint does not determine the next event, but it prevents a trajectory from accumulating too many unfinished copies compared with the number of stabilized copies.

This first constraint was useful as a sanity check, because it showed that open-generator debt increases when one moves from local points toward PR-like points. It counts how many generators are open, but it does not measure how much of the observed past is still not closed, nor how much future remains to be emitted. For the quantum problem, the

more informative object is therefore not only o_i, but the past/future distance discuss in Chapter 5.

6.3. *First candidate: the tight contextual product constraint* —

The *tight contextual product constraint* is the condition $\xi_A : \sqrt{D^-(N) \cdot D^{\not\to}(N)} \leq A$. This quantity is the geometric mean of the past distance $D^-(N)$ and the future distance $D^{\not\to}(N)$. The constraint therefore keeps both distances small at the same time. We choose it as a first candidate because it expresses the working hypothesis that quantum countings lie between two successive stabilized steps: they are not fully closed, but their remaining future cost is still controlled by the already observed past. This interpretation is speculative. The role of the meta-theory is precisely to make such hypotheses testable: we can compare different choices of ξ and ask which of them, if any, captures quantum behavior. We can ask also that our observation live exactly between our past and future with $|D^-(N) - D^+(N)| \leq \varepsilon$ for example.

Following the guiding idea of Chapter 1 (illustrated by the curve fitting example), the model allows for feedback during the experiment. At each step, the observer may decide that the current closed explanation is too rigid, temporarily lower confidence in the extracted past, admit more candidate explanations, and inspect the future cost of these alternative explanations. In this sense, the model does not only predict outcomes. It also revises its own explanatory criteria while the experiment is running. In curve fitting, one usually optimizes a fixed criterion on the data collected so far. Here, the criterion itself can move: we ask what happens to the future cost if the past criterion is relaxed by a controlled amount. If relaxing the past suddenly makes many previously rejected explanations viable, and if their future cost remains moderate, then the model has learned from its own uncertainty. If the future cost explodes, then the current past criterion was already close to the best available explanation.

To formalize this, let $\alpha \geq 1$ be a parameter that measures how often the observer expects to be wrong. We interpret α as the frequency of revision, or equivalently as the rate at which the observer is ready to mistrust the current extraction of the past. The natural relaxed set is then

$$\boldsymbol{\mathcal{D}}_{\alpha}^{\not\to}(\boldsymbol{N}) \coloneqq \min\left\{ d_\eta^+ \mid \eta \in \mathcal{H}_{\mathcal{G}}(N) \text{ and } d_\eta^- \leq D^-(N) + \left\lceil \frac{|N|_1}{\alpha} \right\rceil \right\}.$$

The case $\alpha = 1$ corresponds to the tight setting: only the explanations whose past deficit realizes the contextual fraction are kept. As α increases, the admissible past deficit grows, but only by a controlled amount compared with the total size of the observation. This is a way to make the theory explore alternative explanations without abandoning the static analysis entirely like we can see in Figure 6.

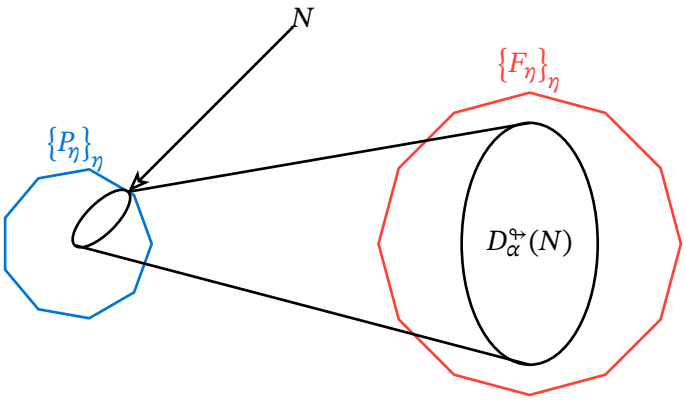

Figure 6: Same as Figure 5, but with a relaxed past distance. The set of possible completions is larger.

6.4. *Testing variations of the product constraint*

Several variations were tried, and this section reports what they revealed.

For a single observation N, a direct idea is to control the product between a past distance and a future distance. This gives several possible constraints. The strongest version asks every explanation to satisfy the bound $d_\eta^- d_\eta^+ \leq A \quad \forall \eta \in \mathcal{H}_\mathcal{G}(N)$. A weaker version only asks that at least one explanation realizes a small product: $\min_{\eta \in \mathcal{H}_\mathcal{G}(N)} d_\eta^- d_\eta^+ \leq A$. Finally, one can use the global distances extracted from the observation: $D^-(N)D^{\leftrightsquigarrow}(N) \leq A$. The relaxed variants replace $D^{\leftrightsquigarrow}(N)$ by $D_\alpha^{\leftrightsquigarrow}(N)$. The dual variants exchange the roles of past and future. For a sequence of observations $N_{1:k}$, the same idea can be lifted to the dynamic setting. Instead of testing one counting, we test the worst step of the trajectory. A typical dynamic constraint has the form $\max_i D^-(N_i)D_\alpha^{\leftrightsquigarrow}(N_i) \leq A$. This condition says that the trajectory is accepted only if no observed step accumulates too much past/future tension. More refined versions can optimize over the whole sequence of explanations $(\eta_1, \ldots, \eta_k)$, but the principle is the same: the dynamic constraint controls the evolution of the explanatory debt along the experiment.

Preliminary computations suggest that all these constraints are equivalent and behave like fractional contextuality: they detect non-classicality, but do not isolate the quantum region. This motivates the search for finer objects, such as the interruption profile, that can capture more structure than a single scalar distance.

6.5. *Interruption profiles: from scalar distances to trajectory*

The scalar constraints tested above compress a dynamic explanation η into one or two numbers, such as d_η^- and d_η^+. This is useful, but it also loses information: if we represent η only as a point between its closed past P_η and its completed future F_η, then we forget how the interruption is distributed across coordinates. The exploration of scalar constraints showed that they detect non-classicality, but do not isolate the quantum region. This observation suggested looking for a finer object.

The dynamic point of view suggests a finer object. An interrupted state is not only located between P_η and F_η. It also tells us which coordinates have already advanced and

which coordinates remain delayed. Thus the line from P_η to F_η serves only as a reference trajectory, not as a replacement for the state itself.

Let η be a dynamic explanation, with closed part P_η, visible part N_η, and completed part F_η. For every coordinate $v \in V$, define

$$a_\eta(v) \coloneqq N_\eta(v) - P_\eta(v), \quad b_\eta(v) \coloneqq F_\eta(v) - N_\eta(v), \quad h_\eta(v) \coloneqq a_\eta(v) + b_\eta(v) = F_\eta(v) - P_\eta(v).$$

Here $a_\eta(v)$ is the amount already emitted after the closed past, $b_\eta(v)$ is the amount still missing before closure, and $h_\eta(v)$ is the total height of the interruption at coordinate v. Since $P_\eta \le N_\eta \le F_\eta$ these numbers are non-negative integers. The scalar distances are recovered by summing them:

$$d_\eta^- = \sum_{v \in V} a_\eta(v), \quad d_\eta^+ = \sum_{v \in V} b_\eta(v).$$

Thus d_η^- and d_η^+ forget how the emitted and residual masses are distributed among the coordinates.

The *active coordinates* are those for which something really happens between past and future: $V_\eta^{\mathrm{act}} \coloneqq \{v \in V \mid h_\eta(v) > 0\}$. Instead of introducing rational local times, we keep the construction integer. If V_η^{act} is non-empty, set $\boldsymbol{\Delta_\eta} \coloneqq \operatorname{lcm}\{h_\eta(v) \mid v \in V_\eta^{\mathrm{act}}\}$. Then each $h_\eta(v)$ divides Δ_η. We define the *local resolution factor* $q_\eta(v) \coloneqq \Delta_\eta / h_\eta(v)$ and the *discrete interruption time* $k_\eta(v) \coloneqq a_\eta(v) q_\eta(v)$. The number $k_\eta(v)$ is an integer between 0 and Δ_η. It is the position of the coordinate v on the common grid of resolution $1/\Delta_\eta$. If $k_\eta(v) = 0$, the coordinate has not yet advanced beyond the closed past. If $k_\eta(v) = \Delta_\eta$, it is already closed. Intermediate values describe partial advancement.

DEFINITION 6.1 (Interruption profile). *Let $\eta \in \mathcal{H}_{\mathcal{G}}(N)$ be a dynamic explanation. If V_η^{act} is non-empty, let Δ_η, q_η, and k_η be defined as above. The **interruption profile** of η is the map*

$$k_\eta : \begin{cases} V_\eta^{\mathrm{act}} \to \{0, \dots, \Delta_\eta\} \\ \quad v \mapsto k_\eta(v) \end{cases}$$

If V_η^{act} is empty, the state is already closed and the profile is empty.

A small two-coordinate calculation is given in Example 9.1.

The interruption profile k_η is the raw per-coordinate information. It tells us, for each active coordinate, at which discrete time the interruption has occurred. Crucially, it does not compress this information into a scalar. Two explanations η and η' can have the same values of d_η^- and d_η^+ while having very different profiles k_η and $k_{\eta'}$. The profile is therefore a genuinely dynamic object: it records *how* the interruption has occurred, not only *how much* has been emitted and how much remains.

A first way to visualize this information is to count how many coordinates have the same interruption time. We define the histogram $\mathrm{Hist}_\eta(k) \coloneqq \#\{v \in V_\eta^{\mathrm{act}} \mid k_\eta(v) = k\}$. It gives a coarse picture of the profile: instead of remembering which coordinate has which time, it records how the active coordinates are distributed along the internal trajectory from the closed past P_η to the completed future F_η.

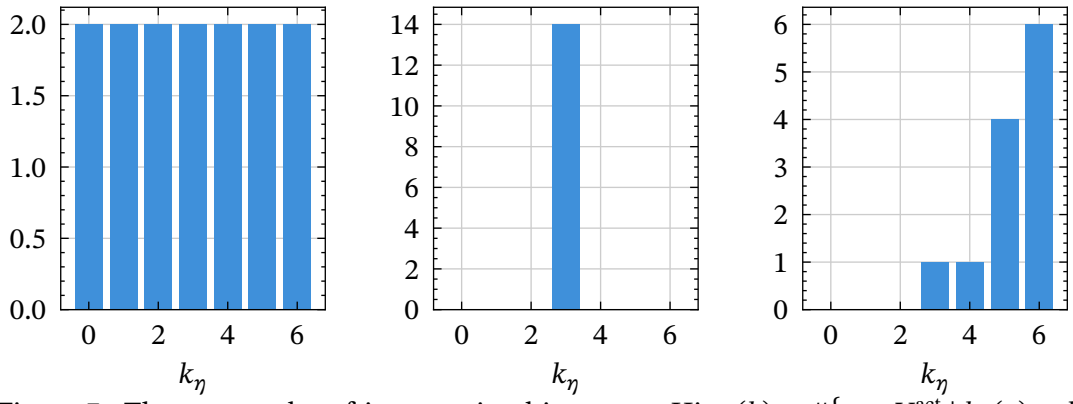

Figure 7: Three examples of interruption histograms $\mathrm{Hist}_\eta(k) = \#\{v \in V_\eta^{\mathrm{act}} \mid k_\eta(v) = k\}$ for an observation vector of dimension $\dim(N) = 14$, assuming all coordinates are active. Left: the interruption times are uniformly distributed between the closed past P_η and the completed future F_η. Middle: all coordinates have the same interruption time, so the process is synchronized at one intermediate position. Right: most coordinates have large values of k_η, so the observation is closer to F_η than to P_η.

The histogram suggests an analogy with the way a statistical profile appears in quantum mechanics. In a quantum experiment, one detection is point-like. Only after many repeated detections does a spatial distribution appear, usually described by $|\psi(x,t)|^2$. Here, each active coordinate gives one discrete interruption time $k_\eta(v)$, and the histogram reveals the global distribution of these times.

The analogy is only partial. The histogram $\mathrm{Hist}_\eta(k)$ is not a wave function: it has no phase, no complex amplitude, and no interference structure. It is closer to a discrete density of interruption times. The difference is also important: in ordinary wave mechanics, the horizontal axis is physical space, while here the horizontal axis is the internal time between P_η and F_η. In this sense, Hist_η gives a density-like view of the interruption, but in the temporal direction of the dynamic explanation rather than in physical space.

The full profile k_η still contains more information than the histogram. The histogram only counts how many coordinates occur at each interruption time, whereas the profile remembers *which* coordinate has *which* time. This coordinate-wise information can be used to impose constraints with a physical interpretation. For example, one may require two coupled CHSH coordinates to be synchronized, require one coordinate to be emitted before another, or bound the delay between two coordinates by a condition such as $|k_\eta(v) - k_\eta(v')| \le B$. More generally, a causal relation $v \to v'$ could be represented by an ordering constraint $k_\eta(v) \le k_\eta(v')$.

These constraints do not only measure how much mass has been emitted or remains to be completed. They constrain the internal organization of the process. This is why they cannot be expressed by a single scalar such as d_η^-, d_η^+, or their product. At this exploratory stage, the interruption profile therefore appears as a more expressive object: it provides a place where synchronization, ordering, bounded delay, and causal-type constraints can be tested directly.

6.6. Experimental observations

The constraints were tested on a family of observations parametrized by a visibility parameter. Fix a level t, and let $\{N_v^t\}_v$ be a family of integer countings whose normalization approximates a probability model of the form $P_v = vP_{\text{PR}} + (1-v)P_{\text{white}}$. Here, v is the visibility parameter, P_{PR} is a PR-type box, and P_{white} is the uniform noise distribution.

For each value of v, we computed a score associated with the constraint ξ. The simple product constraint did not show a specific signature near the quantum boundary. Its behaviour was close to fractional contextuality: it detected non-classicality, but did not isolate the quantum region. This observation motivated the search for finer objects, such as the interruption profile.

To compute these scores, we needed to describe the admissible explanations

$$\eta \in \mathcal{H}_{\mathcal{G},\xi}(N).$$

Once these explanations are available, we can compute d_η^-, d_η^+, the profile k_η, and the discrete moments for each of them, and then evaluate the candidate metric. The constraint-programming formulation below makes this computation explicit.

6.7. Algorithmic implementation

To compute the explanations used by the previous metrics, we encoded the search problem as a Constraint Programming (CP) instance. A solution of the CP instance is an explanation η: it tells which generators are used, which residual parts remain open, and how these residual parts are transported from one observation to the next.

In practice, we use CP-SAT, which integrates an integer solver and supports both feasibility checking and solution enumeration. If this constraint program has a solution m^i, then m^i defines the multiplicity function of the dynamic state $\eta_i \in \mathcal{H}_{\mathcal{G},\xi}(N_i)$.

$$\begin{aligned}
&\textit{Variables}: && \boldsymbol{m_{g,R}^i} \in \mathbb{N} \quad g \in \mathcal{G}, 0 \le R \le g, i \in [k] && \\
& && \boldsymbol{f_{R\to S}^{g,i}} \in \mathbb{N} \quad g \in \mathcal{G}, 0 \le S \le R \le g, i \in [k] && \\
&\textit{Constraints}: && \sum_{g\in\mathcal{G}} \sum_{0\le R\le g} (g-R)\cdot m_{g,R}^i = N_i && \forall i \in [k] \\
& && \sum_{S:S\le R} f_{R\to S}^{g,i} = m_{g,R}^i && g \in \mathcal{G}, 0 \le R \le g, i \in [k] \\
& && \sum_{S:S\ge R} f_{R\to S}^{g,i} \le m_{g,R}^{i+1} && g \in \mathcal{G}, 0 \le R \le g, i \in [k-1] \\
& && \xi &&
\end{aligned}$$

The variable $m_{g,R}^i$ counts how many copies of the generator g are used at step i with residual part R. The visible contribution of such a copy is $g - R$. The first constraint therefore says that the sum of all visible contributions must reconstruct the observation N_i. The variable $f_{R\to S}^{g,i}$ describes how many copies of g move from residual state R at step i to residual state S at the next step. The second constraint conserves the number of copies leaving a residual state. The third constraint ensures that the flow entering a residual state at step $i+1$ does not exceed the number of copies available there. The final line adds the chosen admissibility constraint ξ.

The choice of CP is also justified by the nature of the constraints collected in ξ. These include product constraints of the form $D^-(N) \cdot D^+(N)$. Such bilinear constraints are naturally expressed in a CP framework, whereas a pure integer linear solver would require additional reformulation.

Remark.

This formulation can be seen as a **multi-commodity flow** problem. The commodities are the k counting vectors $N_1, \ldots, N_k$ to be decomposed. Each commodity i must be transported through the network of generators: the source is the target vector N_i, and the admissible paths are sequences of generators $g \in \mathcal{G}$ selected with residual R at each step. The variables $m^i_{g,R}$ record how many units of commodity i are routed through the node (g, R), while the flow variables $f^{g,i}_{R \to S}$ track how residual capacity evolves from R to S between consecutive levels. The conservation constraints ensure that the total inflow into each node matches the outflow, and the coupling constraint $\sum_{S:S \geq R} f^{g,i}_{R \to S} \leq m^{i+1}_{g,R}$ prevents a commodity from consuming more residual capacity than is available at the next level.

Dynamic theories as categories of explanatory paths

CHAPTER 7

The preceding chapters introduced dynamic states as explanations of partial observations. A state records which generator copies have already closed and which copies remain open. This is more informative than a final probability point, but it is still only a snapshot. The present chapter develops the next step: a dynamic theory should evaluate not only states or final observations, but also the paths by which explanations are built.

This objective follows the methodological requirement of Chapter 1. An explanation should remain meaningful after an interruption. It should also be possible to compare two local explanations, to construct a common description when they can be synchronized, and to revise the evaluation of an earlier part of a process when a later comparison supplies a finer interpretation. The proposed language has three levels:

- the hidden level of automaton paths;
- the compressed level of diagrammatic paths; and
- the evaluative level of additive path costs.

The constructions below define a framework for investigating dynamic constraints. They do not yet derive quantum theory, a Bell inequality, or the Tsirelson bound. Their purpose is to state precisely what a path-based theory would have to constrain and how local descriptions could be reconciled.

7.1. First level: the automaton path category

The transition condition in Definition 3.5 defines when one dynamic state can evolve into another. To study paths, we must retain not only the existence of a transition but also the transition that was chosen.

We define the *transition graph* $\mathrm{Tr}_{\mathcal{G}}$ associated with $\mathcal{G}$ by the graph where node are the dynamic states $\eta \in \mathcal{H}_{\mathcal{G}}$ and where each directed edge represents one elementary transition $\eta \rightsquigarrow \eta'$ where $|N_{\eta'}|_1 - |N_\eta|_1 = 1$. It can be views as the product of the automaton of each generator.

DEFINITION 7.1 (Automaton path category).

Let $\mathrm{Tr}_{\mathcal{G}}$ *be the transition graph of the dynamic automaton associated with* $\mathcal{G}$*. Its vertices are the dynamic states* $\eta \in \mathcal{H}_{\mathcal{G}}$*. Each directed edge represents one* ***elementary transition*** *of the automaton. The category* $\mathbf{Aut}_{\mathcal{G}} \coloneqq \mathrm{Path}\big(\mathbf{Tr}_{\mathcal{G}}\big)$ *is therefore the path category of this graph. Its morphisms are finite paths* $\eta_0 \xrightarrow{\tau_1} \eta_1 \xrightarrow{\tau_2} \cdots \xrightarrow{\tau_m} \eta_m$*, where every* τ_j *is an elementary transition.*

More generally, the morphisms $\tau \coloneqq \eta \xrightarrow{\tau} \eta'$ *that is call* ***Trait*** *can be a single elementary transition, any kind of transition such that* $\eta \rightsquigarrow \eta'$ *or the identity morphism* $\eta \xrightarrow{\mathrm{id}} \eta$*.*

The trait τ should not be identified with a formal difference $\eta' - \eta$. Dynamic states are multisets of copies and are not, by themselves, elements of a vector space. One can extract additive shadows of a trait, $\Delta N_\tau \coloneqq N_{\eta'} - N_\eta, \quad \Delta R_\tau \coloneqq R_{\eta'} - R_\eta$, but these quantities do not

determine the transition. Two different copies can emit the same visible event and produce the same pair $(\Delta N_\tau, \Delta R_\tau)$ while remaining distinct hidden operations. The morphism τ retains precisely the information that the additive difference forgets.

The original relation $\eta \rightsquigarrow \eta'$ says that at least one compatible evolution exists. The category $\mathbf{Aut}_{\mathcal{G}}$ is finer: it can contain several morphisms between the same two states, and it records their compositions. This distinction is necessary if a later constraint is meant to evaluate a trajectory rather than only its endpoints.

7.2. Second level: diagrammatic compression

The automaton path category is structurally rich. It retains the full marking of open copies and may therefore contain many paths that are irrelevant for a coarse comparison. We now compress it into a category of diagrammatic paths. We compare the residual size $|R_\eta|_1$, rather than the past size $|P_\eta|_1$, because the residual directly records the visible events still required to close the copies that are currently open. Along the curve $t \mapsto \big(|N_{\eta_t}|_1, |P_{\eta_t}|_1\big)$, the second coordinate may undergo a downward jump of arbitrary size. The past size alone does not determine how much future production is required by the current hidden state.

In contrast, the curve $t \mapsto \big(|N_{\eta_t}|_1, |R_{\eta_t}|_1\big)$ has elementary moves with a direct automaton interpretation. Emitting one event from an already open copy changes the diagram by $(1, -1)$: the visible count increases by one and the residual decreases by one. If a generator has size 4, then opening a new copy and emitting its first event produces the increment $(1, 3)$, since the copy has emitted one event and still has three events to emit.

The residual coordinate therefore restricts the possible local moves. Instead of allowing downward jumps of arbitrary size, it describes the evolution of a dynamic state through the elementary production and closure of its generator copies.

DEFINITION 7.2 (Trajectory diagram functor).

Let E a generator family. Let Aut_E *the automaton path category. Its* ***diagram category*** $\mathbf{Diag}_E$ *is the image of the functor* $\mathcal{D}_E(\mathrm{Aut}_E)$, Diag_E *is the image category of* $\mathcal{D}_E : \mathrm{Aut}_E \to \mathrm{Diag}_E$. *By specifying a functor* $\mathcal{D}_E$ *we define the associed category* Diag_E.

The typical example will be the $1-$diagram *functor, let* $\mathcal{D}^1$ *the functor define by* $\mathcal{D}^1(\eta) \coloneqq \big(|N_\eta|_1, |R_\eta|_1\big)$ *and* $\mathcal{D}^1(\tau \coloneqq \eta \rightsquigarrow \eta') = \Delta\big(|N_{\eta'}|_1 - |N_\eta|_1, |R_{\eta'}|_1 - |R_\eta|_1\big)$.

You can views the $\mathcal{D}^1(\mathrm{Aut}_E)$ category as a bounded lattice. This define point that are accessible and some other that are not. For example, you cannot have a state with $|R_\eta| = 4$ if your observation is of size 1 given a generator family with generator of size four. Opening a copy of generator g gives the projected displacement $\big(0, |g|_1\big) \circ (1, -1) = \big(1, |g|_1 - 1\big)$. Thus a generator of size 4 produces the familiar initial direction $(1, 3)$, whereas each later emission decreases the residual coordinate by one unit.

The functor $\mathcal{D}_E$ forgets the identities of generator copies, the distribution of the residual among coordinates, and the exact transition witness. It retains the visible mass and total residual debt along the path. The first coordinate is visible. The second coordinate is inferred from the explanatory state. Hence the diagram is not an observation alone; it is a compressed interface between observation and explanation.

The compression is deliberately non-injective. Two automaton paths may have the same image in $\mathbf{Diag}_E$. The fibre above a diagram path is the collection of hidden explanations that share the same compressed history.

7.3. Third level: cost functors

The third level deliberately has no environment-specific geometry (no specify generator family). It is a fixed category of additive costs. The environment-dependent information is not a chosen unit or measure: it is the functor that evaluates diagrammatic paths in this category.

DEFINITION 7.3 (Cost category). *The **cost category** **Cost** has one object, denoted by ★:*

$$\mathrm{Ob}(\mathbf{Cost}) = \{\star\}.$$

Its morphisms are the natural numbers:

$$\mathrm{Hom}_{\mathbf{Cost}}(\star, \star) = \mathbb{N}.$$

For $a, b \in \mathbb{N}$, composition is addition:

$$b \circ a := a + b.$$

The identity morphism is 0*.*

DEFINITION 7.4 (Weight functor). *Let E be a generator family with diagram category $\mathbf{Diag}_E$. A **weight functor** is a functor $\mathcal{W}_E : \mathbf{Diag}_E \to \mathbf{Cost}$. It sends every diagram point P to ★. It is determined by the natural-number cost $\mathcal{W}_E(e)$ assigned to each elementary diagram move e (generating morphisms). For a path $\delta = e_1; \ldots; e_m$, functoriality gives $\mathcal{W}_E(\delta) = \mathcal{W}_E(e_1) + \ldots + \mathcal{W}_E(e_m)$.*

The functor $\mathcal{W}_E$ is part of the theory. It may assign the residual size to an elementary transition, for example

$$\mathcal{W}_E(e) = y \quad \text{for } e : P \to P', (x, y) := \mathcal{D}^1(P).$$

In this case, a diagram path δ has cost

$$\mathcal{W}_E(\delta) = \sum_{i=1}^{m} Y_{i-1}.$$

Multiplying this assignment by 2 or 7 gives another valid cost functor. These factors are not introduced as intrinsic units of a single environment. Their meaning arises only when cost functors from distinct environments are compared through a common refinement.

7.4. Environments as three-level structures

The word *environment* now has a precise categorical meaning. It does not refer to a new primitive trait or to a privileged observer. It is a local explanatory apparatus: an automaton category, a diagrammatic representation, and a cost functor to the fixed category **Cost**.

DEFINITION 7.5 (Three-level environment). *A **three-level environment** is a tuple*

$$\mathbf{Env}_E := (\mathbf{Aut}_E, \mathbf{Diag}_E, \mathbf{Cost}; \mathcal{D}_E, \mathcal{W}_E)$$

consisting of an automaton path category, a diagram category, the fixed cost category, and functors

$$\mathbf{Aut}_E \xrightarrow{\mathcal{D}_E} \mathbf{Diag}_E \xrightarrow{\mathcal{W}_E} \mathbf{Cost}.$$

DEFINITION 7.6 (Environment morphism). *Let $\mathbf{Env}_A$ and $\mathbf{Env}_K$ be three-level environments. A morphism*

$$(L_A, r_A) : \mathbf{Env}_A \to \mathbf{Env}_K$$

consists of functors

$$L_A : \mathbf{Aut}_A \to \mathbf{Aut}_K, \quad r_A : \mathbf{Diag}_A \to \mathbf{Diag}_K,$$

such that

$$\mathcal{D}_K \circ L_A = r_A \circ \mathcal{D}_A$$

and

$$\mathcal{W}_K \circ r_A = \mathcal{W}_A.$$

The first equality says that refining an automaton path and then drawing it gives the same diagrammatic path as drawing it first and then refining its representation. The second equality says that diagrammatic refinement and cost evaluation are also compatible. Combining them gives

$$\mathcal{W}_K \circ \mathcal{D}_K \circ L_A = \mathcal{W}_A \circ \mathcal{D}_A.$$

Thus the evaluation of a path cannot be changed independently of the automaton and diagram maps. The comparison does not presuppose an intrinsic unit in A. Rather, it asks whether the local assignment $\mathcal{W}_A$ can be realized as the restriction of the common assignment $\mathcal{W}_K$. If no such functor $\mathcal{W}_K$ exists, then the apparent local scales cannot be glued in the proposed refinement.

For a trait $\tau : \eta \to \eta'$, the functor L_A may send one local transition to a more detailed path in $\mathbf{Aut}_K$. Hence a refinement preserves the cuts of a trajectory. Refining a complete local history is the same as refining its elementary transitions in their original order.

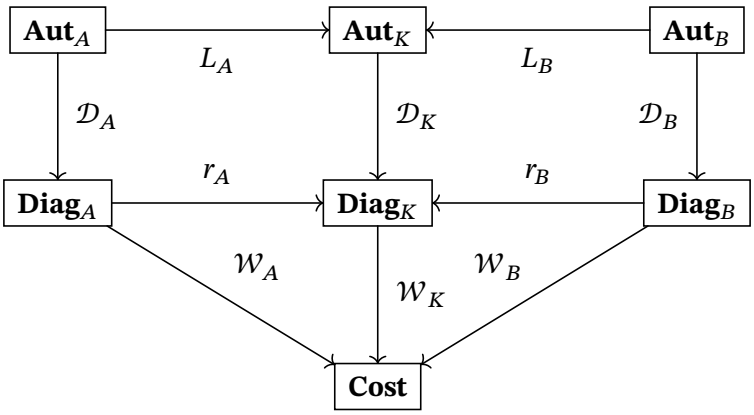

Figure 8: The automaton and diagram descriptions of A and B are each refined in the common environment K. The three diagram descriptions are then evaluated in the same cost category **Cost**.

7.5. Two notions of synchronization

Two environments can be synchronized at two different strengths. The distinction is essential.

7.5.1. Universal common refinement

Let $\mathbf{Env}_A$ and $\mathbf{Env}_B$ be two environments. A **common refinement** is an environment $\mathbf{Env}_K$ equipped with morphisms

$$i_A : \mathbf{Env}_A \to \mathbf{Env}_K, \quad i_B : \mathbf{Env}_B \to \mathbf{Env}_K.$$

If no common refinement exists in the chosen class of environments, then A and B cannot be synchronized in that class. If common refinements exist, a **universal common refinement** is a common refinement $\mathbf{Env}_K$ such that, for every other common refinement $\mathbf{Env}_X$ with morphisms

$$f_A : \mathbf{Env}_A \to \mathbf{Env}_X, \quad f_B : \mathbf{Env}_B \to \mathbf{Env}_X,$$

there exists a unique morphism

$$u : \mathbf{Env}_K \to \mathbf{Env}_X$$

satisfying

$$u \circ i_A = f_A, \quad u \circ i_B = f_B.$$

Thus K is least only with respect to the refinement order: every other common refinement receives a unique morphism from K. This does not mean that K has fewer states or fewer generators.

This construction generalizes the familiar least common multiple. In the category of integer resolutions with arrows

$$d \to D \quad \text{whenever } d \text{ divides } D,$$
$$2 \to 6, \quad 3 \to 6, \quad 6 = \mathrm{lcm}(2, 3).$$

The numerical PPCM is therefore a simple instance of a universal common refinement.

7.5.2. Observable reconciliation and common costs

An external environment does not normally have direct access to the full automaton categories of two other environments. It receives only their reported diagrammatic paths and costs. The operationally accessible construction is therefore weaker. An **observable reconciliation** is a diagram category $\mathbf{Diag}_K$ equipped with refinement functors

$$r_A : \mathbf{Diag}_A \to \mathbf{Diag}_K, \quad r_B : \mathbf{Diag}_B \to \mathbf{Diag}_K.$$

If no such diagram category exists, then the two reports cannot be reconciled at the observable level. If several such categories exist, one may seek a least observable reconciliation: a category $\mathbf{Diag}_K$ such that every other observable reconciliation $\mathbf{Diag}_X$ receives a unique functor $\mathbf{Diag}_K \to \mathbf{Diag}_X$ compatible with the two refinement functors. At this level, K synchronizes only the representations that are available for comparison. It need not know whether these representations arise from compatible hidden copies, residuals, or generator families.

A diagrammatic reconciliation admits a **common cost evaluation** when there is a functor $\mathcal{W}_K : \mathbf{Diag}_K \to \mathbf{Cost}$ such that $\mathcal{W}_K \circ r_A = \mathcal{W}_A, \quad \mathcal{W}_K \circ r_B = \mathcal{W}_B$. The two equations ask whether the local numerical assignments can be extended to the reconciled diagram.

This is where a common unit can emerge. For example, one environment may assign twice the residual weight to every elementary edge, while another assigns seven times that weight. These factors have no privileged meaning in isolation. They become comparable only when a common functor $\mathcal{W}_K$ is required to agree with both assignments after refinement. An observable reconciliation may exist while no such $\mathcal{W}_K$ exists; in that case, the diagrams reconcile but their costs do not.

7.5.3. The hidden lifting problem

The crucial question is whether an observable reconciliation can be lifted to a universal common refinement of environments. Write $g_A \coloneqq r_A \circ \mathcal{D}_A, \quad g_B \coloneqq r_B \circ \mathcal{D}_B$. If $\mathbf{Aut}_K$ and $\mathcal{D}_K : \mathbf{Aut}_K \to \mathbf{Diag}_K$ are fixed, finding automaton functors L_A and L_B such that

$\mathcal{D}_K \circ L_A = g_A$, and $\mathcal{D}_K \circ L_B = g_B$ is a standard **functor lifting problem**: the diagrammatic functors g_A and g_B must be lifted through $\mathcal{D}_K$.

If $\mathbf{Aut}_K$ or $\mathcal{D}_K$ is itself unknown, the problem is broader: one seeks a joint factorization of g_A and g_B through a common automaton category. The present chapter does not assume a general existence theorem for this factorization. The diagrammatic reconciliation may exist while no such hidden lift exists. For example, two diagram paths may have compatible visible mass and residual totals while requiring incompatible copy identities or residual allocations in any common automaton.

This gives a sheaf-like local-to-global problem. The commom diagram $\mathbf{Diag}_K$ act like a global section that reconciles the local diagrams. Local diagram paths may possess local lifts. Those lifts may agree on every reported overlap. Nevertheless, there may be no global automaton path that realizes all of them simultaneously. In that case, the observable data glue while the hidden explanatory structure does not admit a global section.

7.5.4. A simple change of scale

Sometimes two environments describe the same system using generator units of different sizes. Consider generators $g_A, g_B \in \mathbb{N}^V$ such that

$$g_A = 2g_B$$

where the equality holds in every coordinate. One complete copy of g_A then has the same visible counting as two complete copies of g_B.

For a simple example, let $A = \{g_A\}$, $B = \{g_B\}$, and $K = \{g_B\}$. The maps to K satisfy $L_A(g_A) = 2g_B$ and $L_B(g_B) = g_B$. Thus, when A produces one copy of g_A, the environment K describes this event as two separate copies of g_B. If the process stops after the first copy, K can keep only one copy of g_B. From the point of view of A, this copy can be read as one half of g_A. The environment A does not itself show this smaller scale.

The purpose of this example is only to compare two descriptions of the same event. Passing from A to K gives a more detailed description of a past event: one copy in A becomes two smaller copies in K. This difference may be useful when studying how an environment affects the description available to an observer. Any physical interpretation, such as a connection with superposition, requires an additional model.

7.6. Interpretation and programme

The path viewpoint changes the location of the theoretical question. Instead of asking only whether a final counting N belongs to a prescribed set, we ask which histories can explain it and whether those histories can be consistently compressed, synchronized, lifted, and evaluated.

The proposed hierarchy is

$$\mathbf{Aut} \to \mathbf{Diag} \to \mathbf{Cost}.$$

At each level, information is discarded. The final cost is less detailed than a diagram path, and a diagram path is less detailed than an automaton path. Conversely, a global property may become visible only after compression. A path inequality can constrain a whole family of hidden explanations without being reducible to a property of one final point.

This suggests one possible meaning of emergence in the present framework. A higher-level phenomenon is not introduced as an additional primitive object. It is a stable constraint on compressed paths that is inherited by the hidden automaton paths that

realize them. Whether quantum correlations admit such a characterization remains an open question. The framework only specifies a route by which this question can be made finite, constructive, and testable.

The immediate programme is therefore the following:

- construct small automaton path categories for CHSH and GHZ examples;
- compute diagrammatic reconciliations of selected local environments;
- test whether these reconciliations admit hidden automaton lifts;
- compare cost values of trajectories with identical final countings; and
- search for path constraints Ξ that distinguish local, quantum, and post-quantum families.

No claim is made here that the proposed cost functors already recover Bell inequalities, quantum mechanics, or the Tsirelson bound. The contribution of this chapter is a categorical language in which these questions can be posed as questions about paths, refinements, synchronization, and global lifts rather than only about isolated correlation points.

Conclusion

CHAPTER 8

This report started from a simple methodological choice: before normalizing a probability table, keep the integer counts. This choice is not a rejection of the probabilistic formalism. It is a way to preserve information that normalization erases: the level at which an observation was obtained, the arithmetic constraints imposed by finite repetition, and the possibility of viewing an observation as a partial process that may still be completed.

The first contribution is the translation itself. By replacing probabilities with integer counts, contextuality becomes a question about integer decomposition in an affine semigroup. The Hilbert basis of the compatible semigroup gives the elementary bricks of explanation. In the $(2, 2, 2)$ Bell scenario, these bricks are the local deterministic generators and the lifted PR generators. The Ehrhart and quasi-Ehrhart polynomials count how many integer models exist at each level. This is a change of perspective: instead of asking whether a point lies in a convex hull, we ask how an integer vector decomposes into irreducible generators.

The second contribution is the automaton model. A generator is treated as an elementary process that can be produced step by step. An interrupted experiment is represented by a dynamic state: a collection of closed copies and open copies, each with a residual part that records what remains to be emitted. This turns partial observations into meaningful mathematical objects. A prefix is not an error; it is the visible past of a process that still has a future. The transition condition between states preserves the identity of each generator and allows its residual to decrease. This is a modelling choice, but a natural one: it captures the idea that an elementary process may continue, but it cannot become another process.

We then illustrate how the standard Tsirelson bound appears in our framework as the asymptotic limit of the finite-level bound $T(t)$. The gap is controlled by the Diophantine properties of $\sqrt{2}$. Crucially, this operational bound can be approximated using the same tools as the standard approach: the NPA hierarchy gives outer approximations of the quantum set, and these approximations translate directly into integer linear programs. The integer framework does not replace the standard quantum relaxations; it reuses them while keeping the finite-level structure.

The third contribution is the recovery of the contextual fraction. The past distance $D^{-}(N)$ measures how much of a compatible counting can be explained by a non-contextual mechanism. After normalization, this distance recovers exactly the contextual fraction of [ABM16]. The dynamic framework therefore agrees with the standard sheaf-theoretic measure on its static coordinate. But it also carries more information: the completion deficit $D^{\looparrowright}(N)$ measures how much work remains to close a partial explanation.

We then explore dynamic constraints by testing several candidates on concrete examples. Scalar constraints, such as the product $d_\eta^- d_\eta^+$, detect non-classicality but do not isolate the quantum region. This observation points toward finer objects. The *interruption profile* breaks the compression imposed by scalar distances: it remembers how the interruption is distributed across coordinates, not only how much has been emitted and how much remains. Constraints on the profile can express synchronization conditions, emission

orderings, and dispersion bounds that are invisible to any scalar measure. The constraint-programming formulation makes these computations explicit.

The work presented here is exploratory. It does not claim to have found a dynamic theory that characterizes quantum phenomena. What it has done is build a language in which such a question can be posed precisely. The framework is constructive: it starts from finite observations, keeps the integer structure, and defines dynamic objects that record how a process is produced, interrupted, and completed. It is compatible with existing tools: the NPA hierarchy, the contextual fraction, and the sheaf-theoretic formulation all have natural translations into this language.

This opens a dimension of contextuality that has not yet been explored. By passing from a static description, where an empirical model is a point in a polytope, to a dynamic description, where it is the visible trace of an interruptible process, a whole field of experimentation becomes accessible. Which constraints on the interruption profile isolate the quantum region? How does the profile behave in scenarios more complex than CHSH? Can a theory be built that revises its own explanatory criteria during the experiment? These are concrete questions that the framework makes testable. The door is open for passing from a contextuality classified by static properties to a contextuality understood through its dynamics of production, interruption, and completion.

Bibliography

CHAPTER 8

[Ehr05] **"Ehrhart polynomials,"** *Combinatorial Commutative Algebra*, pp. 229–246, 2005. [Online]. Available: https://doi.org/10.1007/0-387-27103-1_12

[AMP24] Alastair A. Abbott, Mehdi Mhalla, and Pierre Pocreau, **"Improving social welfare in non-cooperative games with different types of quantum resources,"** *Quantum*, vol. 8, pp. 1376, 2024. [Online]. Available: https://arxiv.org/abs/2211.01687

[ABM16] Samson Abramsky, Rui Soares Barbosa, and Shane Mansfield, **"Quantifying contextuality via linear programming,"** *Proceedings of the 13th International Conference on Quantum Physics and Logic (QPL 2016)*, 2016. [Online]. Available: https://arxiv.org/abs/1604.07299

[AB11] Samson Abramsky and Adam Brandenburger, **"The Sheaf-Theoretic Structure Of Non-Locality and Contextuality,"** *New Journal of Physics*, vol. 13 no. 11, pp. 113036, 2011.

[Ací+15] Antonio Acín, Tobias Fritz, Anthony Leverrier, and Ana Belén Sainz, **"A combinatorial approach to nonlocality and contextuality,"** *Communications in Mathematical Physics*, vol. 334 no. 2, pp. 533–628, 2015. [Online]. Available: https://arxiv.org/abs/1212.4084

[Ans+26] Marina Maciel Ansanelli, Beata Zjawin, David Schmid, Yìlè Yīng, John H. Selby, Ciarán M. Gilligan-Lee, Ana Belén Sainz, and Robert W. Spekkens, **"The resource theory of causal influence and knowledge of causal influence,"** 2026. [Online]. Available: https://arxiv.org/abs/2512.11209

[Bal+24] Alkida Balliu, Sebastian Brandt, Xavier Coiteux-Roy, Francesco d'Amore, Massimo Equi, François Le Gall, Henrik Lievonen, Augusto Modanese, Dennis Olivetti, Marc-Olivier Renou, Jukka Suomela, Lucas Tendick, and Isadora Veeren, **"Distributed Quantum Advantage for Local Problems,"** 2024. [Online]. Available: https://arxiv.org/abs/2411.03240

[Bar07] Jonathan Barrett, **"Information processing in generalized probabilistic theories,"** *Physical Review A*, vol. 75 no. 3, pp. 32304, 2007. [Online]. Available: https://arxiv.org/abs/quant-ph/0508211

[Bel64] John S. Bell, **"On the Einstein Podolsky Rosen Paradox,"** *Physics Physique Fizika*, vol. 1, pp. 195–200, 1964.

[BW24] Pierre Botteron and Moritz Weber, **"Communication Complexity of Graph Isomorphism, Coloring, and Distance Games,"** 2024. [Online]. Available: https://arxiv.org/abs/2406.02199

[CSW14] Adán Cabello, Simone Severini, and Andreas Winter, **"Graph-Theoretic Approach to Quantum Correlations,"** *Physical Review Letters*, vol. 112 no. 4, pp. 40401, 2014. [Online]. Available: https://arxiv.org/abs/1010.2163

[CDP11] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti, **"Informational derivation of quantum theory,"** *Physical Review A*, vol. 84 no. 1, pp. 12311, 2011. [Online]. Available: https://arxiv.org/abs/1011.6451

[Cla+69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt, **"Proposed Experiment to Test Local Hidden-Variable Theories,"** *Physical Review Letters*, vol. 23 no. 15, pp. 880–884, 1969.

[Cui+25] David Cui, Laura Mančinska, Seyed Sajjad Nezhadi, and David E. Roberson, **"Quantum Perfect Matchings,"** 2025. [Online]. Available: https://arxiv.org/abs/2502.05136

[Har01] Lucien Hardy, **"Quantum Theory From Five Reasonable Axioms,"** 2001. [Online]. Available: https://arxiv.org/abs/quant-ph/0101012

[Ji+22] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen, **"MIP*=RE,"** 2022. [Online]. Available: https://arxiv.org/abs/2001.04383

[Jun+11] Marius Junge, Miguel Navascués, Carlos Palazuelos, David Pérez-García, Volkher B. Scholz, and Reinhard F. Werner, **"Connes' embedding problem and Tsirelson's problem,"** *Journal of Mathematical Physics*, vol. 52 no. 1, pp. 12102, 2011. [Online]. Available: https://arxiv.org/abs/1008.1142

[Lei16] Tom Leinster, **"Basic Category Theory,"** 2016. [Online]. Available: https://arxiv.org/abs/1612.09375

[Lin25] Junqiao Lin, **"MIPco=coRE,"** 2025. [Online]. Available: https://arxiv.org/abs/2510.07162

[MR20] Laura Mančinska and David E. Roberson, **"Quantum isomorphism is equivalent to equality of homomorphism counts from planar graphs,"** *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 661–672, 2020. [Online]. Available: https://arxiv.org/abs/1910.06958

[Man+25] Laura Mančinska, Pieter Spaas, Taro Spirig, and Matthijs Vernooij, **"Gap-preserving reductions and RE-completeness of independent set games,"** 2025. [Online]. Available: http://arxiv.org/abs/2505.05253

[MM11] Lluis Masanes and Markus P. Müller, **"A derivation of quantum theory from physical requirements,"** *New Journal of Physics*, vol. 13, pp. 63001, 2011. [Online]. Available: https://arxiv.org/abs/1004.1483

[MNY20] Hamoon Mousavi, Seyed Sajjad Nezhadi, and Henry Yuen, **"On the complexity of zero gap MIP*,"** 2020. [Online]. Available: https://arxiv.org/abs/2002.10490

[NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín, **"A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations,"** *New Journal of Physics*, vol. 10, pp. 73013, 2008. [Online]. Available: https://arxiv.org/abs/0803.4290

[PR94] Sandu Popescu and Daniel Rohrlich, **"Quantum nonlocality as an axiom,"** *Foundations of Physics*, vol. 24 no. 3, pp. 379–385, 1994.

[Slo19] William Slofstra, **"The set of quantum correlations is not closed,"** *Forum of Mathematics, Pi*, vol. 7, pp. e1, 2019. [Online]. Available: https://arxiv.org/abs/1703.08618

l

Appendix

CHAPTER 9

9.1. Nature of the internship

This internship was unusual because the object studied in this report was not the initial subject. The first plan was to explore links between quantum information, combinatorial structures, and localized complexity classes. In particular, the idea was to ask whether one could define a quantum cost of using entanglement when access to the entangled resource is constrained by locality.

9.1.1. The initial subject and the state of the art

This section is technical and can be skipped (go to Section 9.1.2) on a first reading. It describes the initial investigations done during the internship

The starting point was a central observation in quantum complexity theory: entangled multi-prover interactive proofs are extremely powerful, as captured by the equality $\mathrm{MIP}^* = \mathrm{RE}$. In the standard MIP^* model, a verifier interacts with spatially separated provers that may share arbitrary quantum correlations. The model does not impose any explicit structural constraint on how entanglement is distributed, nor on what information is operationally accessible to subsets of provers. This abstraction is theoretically useful, because it isolates the informational consequences of quantum non-locality. However, it also incorporates a strong implicit assumption: the availability of unstructured, globally coordinated correlations at no explicit cost.

In physically motivated settings, information propagation is constrained by locality. Interactions are limited by distance and propagation time, and causal structure bounds what can be coordinated within a given time window. Even when large-scale entangled resources are assumed to be pre-distributed, the ability to access and coordinate arbitrary distant subsystems is not free, since synchronization and correlation across long distances become resources in their own right.

The initial idea was to study models that retain quantum non-locality while enforcing explicit locality constraints, whether geometric, combinatorial, or causal. The goal was not to weaken MIP^* a priori, but to identify which aspects of its power rely on unconstrained global correlations, and which persist under structured, bounded access to entanglement.

To this end, we considered introducing a localized variant of entangled multi-prover interactive proofs, denoted $\mathrm{MIP}_0^*(k, n)$. The definition separates two levels. First, a global resource specifies how entanglement is distributed, for instance via a large entanglement graph state. Second, each effective interaction is constrained to exploit only local correlations, represented by induced subgraphs of size at most k. Informally, at any point in the protocol, only a "local group" of at most k provers may access and exploit the correlations supported on a k-vertex window of the global resource. To preserve the flexibility of multi-prover protocols, we allowed n independent copies of the resource state, enabling repeated use of the same local structure across many interactions, while preventing the protocol from implicitly assuming a single globally accessible instance with unbounded coordinated correlations.

This model aimed to capture a simple principle: quantum non-locality is not removed, but organized. The parameter k acts as a correlation granularity, or an effective locality

radius, bounding the entanglement that can be operationally mobilized within one interaction. The parameter n functions as a parallelism budget, modeling the ability to distribute multiple independent instances of the same resource without obtaining unconstrained global correlations inside a single instance. Consequently, any global power must arise from repetition, composition, and consistency constraints across local windows, rather than from immediate access to arbitrary global entanglement.

To explore this direction, we read four papers: [BW24], [Bal+24], [Man+25] and [Cui+25]. The goal was to find connections that could motivate the approach.

9.1.1a. Locality and quantum advantage in distributed computing

The paper [Bal+24] exhibits a separation between classical and quantum distributed computation for a problem specified purely by local constraints. The authors define an LCL problem in a bipartite black-white formalism, where white nodes represent vertices and black nodes represent edges, allowing uniform constraints to be expressed via multisets of input-output pairs. They introduce the *iterated GHZ* problem, which can be viewed as a local composition of GHZ-type nonlocal games whose consistency is locally checkable. Their main result shows that on graphs of maximum degree Δ, any classical strategy requires $\Omega(\Delta)$ rounds, whereas a quantum strategy achieves a constant number of rounds by locally distributing GHZ states and then performing measurements without further communication.

The key point for our purposes was how bounded radius induces structured uncertainty. The round elimination technique transforms a problem Π_i into a derived problem Π_{i+1} whose labels are sets of labels of Π_i. This "set lifting" encodes exactly what a node can no longer distinguish when one communication round is removed. Conceptually, this provided a template: instead of relying on implicit global coordination, one explicitly models the information available within a bounded neighborhood, replaces the inaccessible part by a structured description of possibilities, and studies how local constraints compose to enforce global correctness.

9.1.1b. Combinatorial characterizations of quantum games

A recurrent observation is that operational properties of quantum protocols can be equivalently expressed as structural properties of purely combinatorial objects. Several results in the literature illustrate this principle with remarkable precision.

In [BW24], the authors study the isomorphism game: given two graphs $\mathcal{G}$ and $\mathcal{H}$, if the graphs are isomorphic, then the game admits a perfect classical strategy. By extension, they define the notion of quantum isomorphism $\mathcal{G} \cong_q \mathcal{H}$ and non-signaling isomorphism $\mathcal{G} \cong_{\text{ns}} \mathcal{H}$, which hold whenever there exists a perfect quantum or non-signaling strategy, respectively. A key structural result is that non-signaling isomorphism is equivalent to fractional isomorphism, that is, $\mathcal{G} \cong_{\text{ns}} \mathcal{H}$ if and only if there exists a bistochastic matrix u such that $A_{\mathcal{G}} u = u A_{\mathcal{H}}$. This is a clean relaxation of the classical notion of isomorphism, which requires u to be a permutation matrix.

A striking result, from [MR20], establishes that $\mathcal{G} \cong_{\text{ns}} \mathcal{H}$ is equivalent to $\forall$ tree $\mathcal{K}$, $\#\text{Hom}(\mathcal{K}, \mathcal{G}) = \#\text{Hom}(\mathcal{K}, \mathcal{H})$. This is a beautiful correspondence: on one side stands a game, which is a protocol with interactions, rules, and strategies, and on the other side stands a simple combinatorial property involving homomorphism counts. An analogous equivalence holds for quantum isomorphism: $\mathcal{G} \cong_q \mathcal{H}$ if and only if $\forall$ planar $\mathcal{K}$, $\#\text{Hom}(\mathcal{K}, \mathcal{G}) = \#\text{Hom}(\mathcal{K}, \mathcal{H})$. The fact that the laws of quantum mechanics can be captured by a condition on planar homomorphism counts is a striking illustration of the power of combinatorial characterizations.

A second important result is the construction of D-fractional isomorphism. The authors introduce a D-distance game that imposes that the output node of both players must be at the same distance from the input node, but only when the distance is below D. This forces a kind of locally consistent isomorphism. The key observation is that this local condition $\mathcal{G} \cong_q^D \mathcal{H}$ enforces a global property $\mathcal{G} \cong_{\text{frac}} \mathcal{H}$. This demonstrates that a local condition can enforce a global property.

9.1.1c. Quantum perfect matchings

The paper [Cui+25] studies a family of nonlocal games built around the notion of perfect matching. In the bipartite L-perfect matching game BPM_G, the verifier asks questions corresponding to vertices on one side of a bipartite graph, and the players answer with edges. The winning conditions are local: each answer must be incident to the queried vertex, and two answers must be compatible with being part of the same matching. Classically, having a perfect strategy is equivalent to the existence of an actual perfect matching.

What interested us was precisely this passage from local tests to a global combinatorial object. The verifier never sees the whole matching directly; it only checks local consistency conditions. Nevertheless, in the classical case these local checks force the existence of a global perfect matching. This was close to the kind of mechanism we were looking for: global structure should emerge from local compatibility constraints.

The quantum case gives a richer picture. The paper introduces a genuine notion of quantum perfect matching. The main characterization states that a graph G has a quantum perfect matching if and only if its line graph $L(G)$ admits a projective packing of the appropriate value. In this formulation, edges of the original graph are replaced by projectors, and the condition that two incident edges cannot both be selected becomes an orthogonality condition. This showed that a nonlocal game can sometimes be translated into a clean algebraic-combinatorial condition on a derived graph. The paper also contains an important warning: when the same ideas are extended to hypergraphs, deciding the existence of quantum perfect matchings becomes undecidable.

9.1.1d. Gap-preserving reductions for entangled-prover games

The paper [Man+25] studies gap-preserving reductions between entangled-prover games. Its main question is how to transform one nonlocal game into another while preserving the distinction between perfect strategies and strategies that are bounded away from perfect. The authors show that general synchronous games can be reduced to independent set games with only polynomial loss in the gap. The technically difficult part is the approximate case: a near-perfect quantum strategy only gives relations that are approximately true. The paper proves a stability theorem showing that such approximate measurement families can be rounded to genuine projective measurements without losing too much.

This paper was relevant because it gives a precise example of how local combinatorial constraints can retain the hardness of a much more general quantum game. The independent set game has a very simple rule: equal questions require equal answers, and different questions require non-adjacent answers. Yet, through the game graph construction and the stability theorem, this simple game family can encode the complexity of general entangled-prover games.

9.1.2. What these readings brought

The initial motivation was not only complexity-theoretic. My supervisor had developed a broader theoretical viewpoint which seemed to become more stable and coherent over

time. After many discussions, I became interested in building a toy model that could justify or at least motivate this way of seeing information, structure, and quantum behaviour.

The ideas from these papers were really the very beginning of the reflection. We tried for a while to work in this direction, but it quickly became clear that it would not lead to something interesting enough on its own. At the same time, my supervisor's intuition about stabilization, interruption, and reconstruction was becoming more precise. Seeing these two directions side by side made us change course: the localized complexity approach was abandoned in favour of the dynamic framework developed in this report.

9.1.3. How the project evolved

The first month was mainly devoted to the state of the art described above. The next two months were more exploratory. I tried several directions that did not immediately become the final report. I worked on the distributed-computing framework of [Bal+24] and tried to understand how it could be applied to examples such as C_5[†]. I also read about tensor networks, graph-state methods, the zero-gap problem, nonlocal games, and graph-theoretic encodings of quantum constraints. At the same time, I was thinking about the philosophical part of the project: what it means for an observation to be only a trace of a hidden process, and how a process can be constrained without being deterministic.

The turning point was the decision to describe observations as events emitted by elementary generators. A generator can be closed, open, or partially emitted. This led to the language of traces, count, automata, Petri nets, and dynamic states. From this point on, the project became more stable. The first concrete results were the integer and Ehrhart-type viewpoint on contextuality in Chapter 2, followed by the construction of dynamic contextual automata in Chapter 3.

Only at this late stage did we realize that some of the quantities introduced in this report were related to measures of contextuality, especially the contextual fraction [ABM16]. This was an important point: it showed that the formalism was not merely a language built for a specific intuition, but could be compared with existing mathematical structures.

[†]The game has five players arranged on a cycle. Player i receives a question x_i and answers a_i. The winning conditions are: $(a_{i-1} + a_i + a_{i+1})(a_{j-1} + a_j + a_{j+1}) = 1$ whenever $(x_{i-1}, x_i, x_{i+1}) = (0, 1, 0)$, $(x_{j-1}, x_j, x_{j+1}) = (0, 1, 0)$, and $|i - j \bmod 5| \geq 2$. If $x_i = 1$ for all i, then the condition is $\sum_i a_i = 1$.

9.2. Technical complements

9.2.1. What look like $\mathcal{H}_{\text{ns}}$?

Table 1: The local generator $\{(d_\lambda, 1)\}_{\lambda=1}^{16}$ can be views like Figure 1, with a matrix representation of a counting vector $N \in \mathbb{N}^V$ in the CHSH scenario. Rows correspond to Alice's pairs $(a|x)$, columns to Bob's pairs $(b|y)$, and each cell to the joint event $(a, b|x, y)$. In represent the matrix with the following order $0|0; 1|0; 0|1; 1|1$ for the two axis Alice and Bob.

$$\begin{pmatrix} 1&0&1&0\\0&0&0&0\\1&0&1&0\\0&0&0&0 \end{pmatrix} \quad \begin{pmatrix} 1&0&0&1\\0&0&0&0\\1&0&0&1\\0&0&0&0 \end{pmatrix} \quad \begin{pmatrix} 0&1&1&0\\0&0&0&0\\0&1&1&0\\0&0&0&0 \end{pmatrix} \quad \begin{pmatrix} 0&1&0&1\\0&0&0&0\\0&1&0&1\\0&0&0&0 \end{pmatrix}$$
$$\begin{pmatrix} 1&0&1&0\\0&0&0&0\\0&0&0&0\\1&0&1&0 \end{pmatrix} \quad \begin{pmatrix} 1&0&0&1\\0&0&0&0\\0&0&0&0\\1&0&0&1 \end{pmatrix} \quad \begin{pmatrix} 0&1&1&0\\0&0&0&0\\0&0&0&0\\0&1&1&0 \end{pmatrix} \quad \begin{pmatrix} 0&1&0&1\\0&0&0&0\\0&0&0&0\\0&1&0&1 \end{pmatrix}$$
$$\begin{pmatrix} 0&0&0&0\\1&0&1&0\\1&0&1&0\\0&0&0&0 \end{pmatrix} \quad \begin{pmatrix} 0&0&0&0\\1&0&0&1\\1&0&0&1\\0&0&0&0 \end{pmatrix} \quad \begin{pmatrix} 0&0&0&0\\0&1&1&0\\0&1&1&0\\0&0&0&0 \end{pmatrix} \quad \begin{pmatrix} 0&0&0&0\\0&1&0&1\\0&1&0&1\\0&0&0&0 \end{pmatrix}$$
$$\begin{pmatrix} 0&0&0&0\\1&0&1&0\\0&0&0&0\\1&0&1&0 \end{pmatrix} \quad \begin{pmatrix} 0&0&0&0\\1&0&0&1\\0&0&0&0\\1&0&0&1 \end{pmatrix} \quad \begin{pmatrix} 0&0&0&0\\0&1&1&0\\0&0&0&0\\0&1&1&0 \end{pmatrix} \quad \begin{pmatrix} 0&0&0&0\\0&1&0&1\\0&0&0&0\\0&1&0&1 \end{pmatrix}$$

Table 2: The PR generator (or lifted) $\{(r_\mu, 2)\}_{\mu=1}^{8}$ can be views like Table 1

$$\begin{pmatrix} 1&0&1&0\\0&1&0&1\\1&0&0&1\\0&1&1&0 \end{pmatrix} \quad \begin{pmatrix} 0&1&0&1\\1&0&1&0\\0&1&1&0\\1&0&0&1 \end{pmatrix} \quad \begin{pmatrix} 1&0&0&1\\0&1&1&0\\1&0&1&0\\0&1&0&1 \end{pmatrix} \quad \begin{pmatrix} 0&1&1&0\\1&0&0&1\\0&1&0&1\\1&0&1&0 \end{pmatrix}$$
$$\begin{pmatrix} 1&0&1&0\\0&1&0&1\\0&1&1&0\\1&0&0&1 \end{pmatrix} \quad \begin{pmatrix} 0&1&0&1\\1&0&1&0\\1&0&0&1\\0&1&1&0 \end{pmatrix} \quad \begin{pmatrix} 1&0&0&1\\0&1&1&0\\0&1&0&1\\1&0&1&0 \end{pmatrix} \quad \begin{pmatrix} 0&1&1&0\\1&0&0&1\\1&0&1&0\\0&1&0&1 \end{pmatrix}$$

9.2.2. Dynamic completeness

The main text uses only the monotonicity direction: if a dynamic trajectory exists, then the visible counts are monotone. The converse is elementary but not central to the construction, so we record it here.

DEFINITION 9.1 (Completeness). *A generator family $\mathcal{G} \subseteq \mathbb{N}^V$ is **complete** if every visible counting admits at least one dynamic explanation:*

$$\forall N \in \mathbb{N}^V, \quad \exists \eta, \quad N_\eta = N.$$

Equivalently, $\mathcal{H}_{\mathcal{G}}(N) \neq \varnothing$ for every $N \in \mathbb{N}^V$.

PROPOSITION 9.2 (Dynamic completeness is support coverage). *A generator family $\mathcal{G}$ is complete if and only if its generators cover all visible events:*

$$\forall v \in V, \quad \exists g \in \mathcal{G}, \quad g(v) > 0.$$

Proof.

Suppose first that the generators cover V. For each $v \in V$, choose a generator $g_v \in \mathcal{G}$ such that $g_v(v) > 0$. Let $\varepsilon_v \in \mathbb{N}^V$ be the unit vector at v, and define $R_v := g_v - \varepsilon_v$. Since $g_v(v) > 0$, we have $0 \leq R_v \leq g_v$. Thus (g_v, R_v) is an admissible copy in the extended

multiset formulation. Its visible contribution is $g_v - R_v = \varepsilon_v$. Now let $N \in \mathbb{N}^V$. For each $v \in V$, take $N(v)$ copies of (g_v, R_v). The resulting dynamic state η satisfies

$$N_\eta = \sum_{v \in V} N(v)(g_v - R_v) = \sum_{v \in V} N(v)\varepsilon_v = N.$$

Hence $\mathcal{G}$ is complete.

Conversely, suppose that the generators do not cover V. Then there exists $v \in V$ such that $g(v) = 0$ for every $g \in \mathcal{G}$. If (g, R) is any open or closed copy, with $0 \le R \le g$, then $(g - R)(v) = 0$. Therefore every dynamic state η satisfies $N_\eta(v) = 0$. In particular, the unit count ε_v cannot be explained. Thus $\mathcal{G}$ is not complete.

PROPOSITION 9.3 (Converse monotonicity under dynamic completeness). *Suppose that $\mathcal{G}$ is complete. For every finite sequence $N_{1:k} = (N_1, \dots, N_k)$ of visible countings in $\mathbb{N}^V$,*

$$\mathcal{H}_{\mathcal{G}}(N_{1:k}) \neq \varnothing \quad \Leftrightarrow \quad N_1 \le N_2 \le \dots \le N_k$$

componentwise.

Proof.

The implication from left to right is exactly Proposition 3.6.

Conversely, suppose that $N_1 \le \dots \le N_k$ componentwise. Since $\mathcal{G}$ is complete, Proposition 9.2 gives, for each $v \in V$, a generator $g_v \in \mathcal{G}$ with $g_v(v) > 0$. As in the previous proof, set $R_v := g_v - \varepsilon_v$, so that $g_v - R_v = \varepsilon_v$. For each time i, define a dynamic state η_i by taking $N_i(v)$ copies of (g_v, R_v) for every $v \in V$. Then

$$N_{\eta_i} = \sum_{v \in V} N_i(v)(g_v - R_v) = \sum_{v \in V} N_i(v)\varepsilon_v = N_i.$$

It remains to check the transition condition. For each $v \in V$, set $p_v := (g_v, R_v)$. Several visible events may give the same pair p_v, so we argue at the level of multiplicities. Let $\mathbb{m}_i(g, R)$ be the multiplicity of the pair (g, R) in $\mathbb{O}_{\eta_i}$. By construction, $\mathbb{m}_{i(g,R)} = \sum_{v \in V, p_v = (g,R)} N_i(v)$. Since $N_i(v) \le N_{i+1}(v)$ for every v, we get $\mathbb{m}_i(g, R) \le \mathbb{m}_{i+1}(g, R)$ for every pair (g, R). We now define the transition map explicitly. An occurrence of $\mathbb{O}_{\eta_i}$ is a triple (g, R, r) with $1 \le r \le \mathbb{m}_i(g, R)$. Define $\varphi_i(g, R, r) := (g, R, r)$. This is well-defined because $r \le \mathbb{m}_i(g, R) \le \mathbb{m}_{i+1}(g, R)$, so (g, R, r) is also an occurrence of $\mathbb{O}_{\eta_{i+1}}$. The map φ_i is injective, it preserves the generator, and it leaves the residual unchanged. Hence the residual condition is $R \le R$, which holds. Therefore $\eta_i \rightsquigarrow \eta_{i+1}$ for every $i < k$. Thus $(\eta_1, \dots, \eta_k) \in \mathcal{H}_{\mathcal{G}}(N_{1:k})$, so the fibre is non-empty.

9.2.3. The overlap problem in the integer setting

The overlap problem of Example 5.1 has an exact integer counterpart. Recall the scenario: three measurements $X = \{a, b, c\}$, contexts $C_1 = \{a, b\}$, $C_2 = \{b, c\}$, $C_3 = \{a, c\}$, and binary outcomes. The probabilistic model of Example 5.1 lifts to an integer counting at level $t = 20$:

Table 3: Integer counting at level $t = 20$, obtained by multiplying the probabilities of Example 5.1 by 20.

C	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$	total
$\{a,b\}$	7	3	5	5	20
$\{b,c\}$	6	4	2	8	20
$\{a,c\}$	6	2	4	8	20

The same two global sections appear, now with integer multiplicities:

- $g_1 = (a \mapsto 0, b \mapsto 0, c \mapsto 0)$: restrictions $(0,0)$ in every context, with counts $7, 6, 6$. The maximum multiplicity is $z(g_1) = \min(7,6,6) = 6$.
- $g_2 = (a \mapsto 0, b \mapsto 0, c \mapsto 1)$: restrictions $(0,0)$ in C_1, $(0,1)$ in C_2, $(0,1)$ in C_3, with counts $7, 2, 2$. The maximum multiplicity is $z(g_2) = \min(7,2,2) = 2$.

Individually, each gives a valid subcounting: $Mz_1 \leq N$ with $z_1 = (6, 0, \ldots, 0)$, and $Mz_2 \leq N$ with $z_2 = (0, 0, \ldots, 2, 0, \ldots)$. But their *sum* fails. At context C_1, both g_1 and g_2 restrict to the same local event $(a \mapsto 0, b \mapsto 0)$. The combined multiplicity is $6 + 2 = 8$, while $N_{C_1}(a \mapsto 0, b \mapsto 0) = 7$. The constraint

$$\sum_{g|_{C_1} = (a \mapsto 0, b \mapsto 0)} z(g) \leq 7$$

is violated. In the language of generators: we are trying to place 8 copies of a deterministic explanation on an event that was only observed 7 times. The constraint $Mz \leq N$ prevents this.

The integer and probabilistic overlap problems are the *same linear inequality*, only seen at different scales. Dividing by $t = 20$ recovers the probabilistic version: $\frac{6}{20} + \frac{2}{20} = 0.40 > 0.35 = \frac{7}{20}$. The constraint $Mz \leq N$ is the integer form of $Mc \leq e_N$ with $c = \frac{z}{t}$. In both cases, the root cause is identical: two global sections compete for the same local event, and their combined mass exceeds what the model allows.

This is why the consistent subcounting set $C_N(X)$ is defined with a $\leq$ constraint rather than equality. The $\leq$ allows multiple global sections to share a local event, up to its observed count. Without it, any attempt to combine generators would either overcount or force a rigid assignment that may not exist.

9.2.4. Detailed calculations for the gap formulas

This appendix records the explicit algebraic steps behind two identities used in the main text. Both concern the gap between the Tsirelson bound $2\sqrt{2}$ and the isotropic Tsirelson score $T_{\text{iso}}(t)$ at finite level t.

9.2.4a. Notation and starting point

We recall the definitions used throughout.

- $p^* := \left(2 + \sqrt{2}\right)/4$ is the optimal winning probability on the isotropic line.
- $w^*(t) := \lfloor t p^* \rfloor$ is the largest integer winning count compatible with the quantum set at level t.
- $T_{\text{iso}}(t) := (4(2w^*(t) - t))/t$ is the isotropic Tsirelson score at level t.
- $\text{frac}\,(x) := x - \lfloor x \rfloor$ denotes the fractional part of a real number x.

The goal is to prove two equalities:

1. $2\sqrt{2} - T_{\text{iso}}(t) = (8 \cdot \text{frac}(tp^*))/t$,
2. $2\sqrt{2} - T_{\text{iso}}(t) = 2\left(\sqrt{2} - r_t\right)$, where $r_t \coloneqq (4w^*(t) - 2t)/t$.

***9.2.4b*. Derivation 1: The fractional-part formula**

We want to show:

$$2\sqrt{2} - T_{\text{iso}}(t) = \frac{8 \cdot \text{frac}(tp^*)}{t}.$$

Since $p^* = \frac{2+\sqrt{2}}{4}$, we have $4p^* = 2 + \sqrt{2}$, hence $\sqrt{2} = 4p^* - 2$, and therefore $2\sqrt{2} = 8p^* - 4$. By definition,

$$T_{\text{iso}}(t) = \frac{4(2w^*(t) - t)}{t} = \frac{8w^*(t) - 4t}{t} = \frac{8w^*(t)}{t} - 4.$$

We compute the gap:

$$\begin{aligned} 2\sqrt{2} - T_{\text{iso}}(t) &= (8p^* - 4) - \left(\frac{8w^*(t)}{t} - 4\right) \\ &= 8p^* - 4 - \frac{8w^*(t)}{t} + 4 \\ &= 8p^* - \frac{8w^*(t)}{t} \\ &= \frac{8tp^* - 8w^*(t)}{t} \\ &= \frac{8(tp^* - w^*(t))}{t}. \end{aligned}$$

Then we identify the fractional part. By definition, $w^*(t) = \lfloor tp^* \rfloor$. Therefore

$$tp^* - w^*(t) = tp^* - \lfloor tp^* \rfloor = \text{frac}(tp^*).$$

Substituting:

$$2\sqrt{2} - T_{\text{iso}}(t) = \frac{8 \cdot \text{frac}(tp^*)}{t}.$$

***9.2.4c*. Derivation 2: The rational-approximation formula**

We want to show:

$$2\sqrt{2} - T_{\text{iso}}(t) = 2\left(\sqrt{2} - r_t\right),$$

where $r_t \coloneqq (4w^*(t) - 2t)/t$ is the rational approximation to $\sqrt{2}$ at level t.

First, verify that r_t approximates $\sqrt{2}$. From the definition of $T_{\text{iso}}(t)$:

$$T_{\text{iso}}(t) = \frac{8w^*(t) - 4t}{t} = \frac{4(2w^*(t) - t)}{t}.$$

Define $c_t \coloneqq (2w^*(t) - t)/t$. This is the integer correlator at level t, normalized by t. Then $T_{\text{iso}}(t) = 4c_t$, and

$$r_t = \frac{4w^*(t) - 2t}{t} = 2 \cdot \frac{2w^*(t) - t}{t} = 2c_t.$$

At the quantum optimum, each correlator equals $\sqrt{2}/2$, so $c^* = \sqrt{2}/2$ and $r^* = 2c^* = \sqrt{2}$. Thus r_t is the rational approximation to $\sqrt{2}$ obtained by replacing the ideal correlator c^* with the finite-level correlator c_t.

Then, we express the gap in terms of r_t. From Derivation 1, we already know:

$$2\sqrt{2} - T_{\text{iso}}(t) = 8p^* - \frac{8w^*(t)}{t}.$$

Now substitute $p^* = \frac{2+\sqrt{2}}{4}$:

$$8p^* = 8 \cdot \frac{2+\sqrt{2}}{4} = 2\Big(2+\sqrt{2}\Big) = 4 + 2\sqrt{2}.$$

Also,

$$\frac{8w^*(t)}{t} = 2 \cdot \frac{4w^*(t)}{t} = 2 \cdot \frac{4w^*(t) - 2t + 2t}{t} = 2(r_t + 2) = 2r_t + 4.$$

Therefore:

$$2\sqrt{2} - T_{\text{iso}}(t) = \Big(4 + 2\sqrt{2}\Big) - (2r_t + 4) = 2\sqrt{2} - 2r_t = 2\Big(\sqrt{2} - r_t\Big).$$

***9.2.4d*. Summary: The two equivalent forms**

The gap admits three equivalent expressions:

$$2\sqrt{2} - T_{\text{iso}}(t) = \frac{8 \cdot \text{frac}(t\,p^*)}{t} = 2\Big(\sqrt{2} - r_t\Big) = 2 \cdot \Big(\sqrt{2} - 2c_t\Big),$$

where $c_t = (2w^*(t) - t)/t$ is the normalized integer correlator.

The first form makes the Diophantine nature explicit: the gap is controlled by how close $t\,p^*$ is to an integer. The second form shows that the gap is twice the error of the rational approximation r_t to $\sqrt{2}$. The third form decomposes this further: the per-correlator gap is $\frac{\sqrt{2}}{2} - c_t$, and the total CHSH gap is four times this value (since there are four correlators), which simplifies to $2\Big(\sqrt{2} - 2c_t\Big)$.

***9.2.4e*. Connection to the CHSH game structure**

The factor of 2 in the formula $2\Big(\sqrt{2} - r_t\Big)$ has a structural origin. The CHSH score decomposes into two complementary pairs of contexts:

$$S = \underbrace{(E(0,0) + E(1,0))}_{\text{pair 1}} + \underbrace{(E(0,1) - E(1,1))}_{\text{pair 2}}.$$

On the isotropic line, each pair has the same value $2c_t$. At the quantum optimum, each pair equals $\sqrt{2}$. The per-pair deficit is therefore $\sqrt{2} - 2c_t = \sqrt{2} - r_t$, and the total deficit is twice this value. Equivalently, the CHSH game has three positive contexts and one negative context. The gap from the three positive contexts is $3\Big(\sqrt{2} - r_t\Big)$ (they are below ideal), and the gap from the negative context is $-\Big(\sqrt{2} - r_t\Big)$ (it is above ideal, i.e., less anti-correlated than optimal). The net gap is $3\Big(\sqrt{2} - r_t\Big) - \Big(\sqrt{2} - r_t\Big) = 2\Big(\sqrt{2} - r_t\Big)$. This factor of 2 is therefore a consequence of the CHSH game structure (3 positive minus 1 negative gives a net coefficient of 2), not an artifact of the approximation.

9.2.5. Example of an interruption profile

***Example 9.1* (Two-coordinate interruption profile).**

This example illustrates the integer construction of the interruption profile and its moments. Consider a dynamic explanation with two visible coordinates. Let

$$P_\eta = (4, 2), \quad F_\eta = (6, 10).$$

The total heights between past and future are

$$h_\eta = F_\eta - P_\eta = (2, 8).$$

Hence the common resolution is

$$\Delta_\eta = \text{lcm}(2, 8) = 8.$$

The local resolution factors are therefore

$$q_\eta(1) = \frac{8}{2} = 4, \quad q_\eta(2) = \frac{8}{8} = 1.$$

Now choose the interrupted visible state

$$N_\eta = (5, 4).$$

Then the emitted part after the closed past is

$$a_\eta = N_\eta - P_\eta = (1, 2).$$

The discrete interruption times are

$$k_\eta(1) = a_\eta(1)q_\eta(1) = 1 \cdot 4 = 4, \quad k_\eta(2) = a_\eta(2)q_\eta(2) = 2 \cdot 1 = 2.$$

Thus the two coordinates are not synchronized. The first coordinate is at time $4/8$, while the second coordinate is at time $2/8$. The interruption profile is therefore $k_\eta = (4, 2)$.