



Information-Theoretic Interpretation of Quantum Formalism

Michel Feldmann¹ 

Received: 8 December 2020 / Accepted: 21 April 2023 / Published online: 21 May 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

We present an information-theoretic interpretation of quantum formalism based on a Bayesian framework and devoid of any extra axiom or principle. Quantum information is construed as a technique for analyzing a logical system subject to classical constraints, based on a question-and-answer procedure. The problem is posed from a particular batch of queries while the constraints are represented by the truth table of a set of Boolean functions. The Bayesian inference technique consists in assigning a probability distribution within a real-valued probability space to the joint set of queries in order to satisfy the constraints. The initial query batch is not unique and alternative batches can be considered at will. They are enabled mechanically from the initial batch, quite simply by transcribing the probability space into an auxiliary Hilbert space. It turns out that this sole procedure leads to exactly rediscover the standard quantum information theory and thus provides an information-theoretic rationale to its technical rules. In this framework, the great challenges of quantum mechanics become simple platitudes: Why is the theory probabilistic? Why is the theory linear? Where does the Hilbert space come from? In addition, most of the paradoxes, such as uncertainty principle, entanglement, contextuality, nonsignaling correlation, measurement problem, etc., become straightforward features. In the end, our major conclusion is that quantum information is nothing but classical information processed by a mature form of Bayesian inference technique and, as such, substantial with Aristotelian logic.

Keywords Quantum theory · Quantum information · Bayesian inference · Information theory

✉ Michel Feldmann
michel.feldmann@polytechnique.org

¹ AX, Paris, France

1 Introduction

Basically, data are stored in a definite register, but in 1948 Shannon [1] construed a sequence of symbols as a stochastic process, giving rise to information theory. He thus joined the core concepts of thermodynamics, revealed by the pioneering work of Léo Szilard on Maxwell's demon dating back to 1929 [2, 3], opening a new horizon sometimes viewed as the ultimate explanatory principle in physics [4, 5]. Nowadays, classical information theory focuses essentially on uncertain discrete variables. In 1957, Jaynes incorporated the Shannon's concept of entropy in the Bayesian inference theory [6]. Later, contemplating quantum mechanical formalism, Jaynes noted in 1989 [7] that this formalism is strongly reminiscent of the Bayesian model. C. Rovelli considered in 1996 a reformulation of quantum mechanics in terms of information theory [8]. More explicitly, Caves et al. [9] proposed in 2002 in a seminal paper especially endorsed by Mermin [10], to understand quantum probability within a Bayesian framework. Fuchs coined the term "QBism" [11] for "Quantum Bayesianism" to describe this view.

1.1 Motivation

In this paper, we propose an explicit framework to implement the concept of "QBism". We start from a logical system subject to logical constraints and regard calculation as a Bayesian estimation [12] of the variables involved. This means applying probability theory as an alternative tool to solve the problem, although the uncertainty about the solution sought has nothing to do with that of a conventional random issue. This nevertheless works because standard probability laws are just the extension of Aristotelian logic rules to cases where the variables are uncertain, as pointed out by Cox [13] and Jaynes [6]. Technically, this implies *taking probabilities for the very unknowns* of the problem instead of the variables themselves and next equating the calculation to a problem of inference.

Let us start with an informal draft of the model.

1.2 Quantum Information in a Nutshell

We consider quantum information as a technique for analyzing a logical system under constraints, represented by the truth table of a set of Boolean functions. As a result, the system behaves like a memory containing a maximum of (say) N bits of information. The issue is to extract this information. The classic solution would be to consider once and for all the relevant Boolean variables and to use a discrete algorithm to find their truth values.

By contrast, we consider the memory as whole and address all possible binary queries, introducing a question-and-answer procedure. The problem remains nevertheless initially posed from a particular batch of N dichotomous queries and therefore represented by a particular batch of N Boolean variables. Technically, they

display a set, say Ω , of 2^N distinct classical states, ω . These classical states span a real-valued probability space, say \mathcal{P} , of dimension 2^N as

$$\mathcal{P} = \text{Span}_{\omega \in \Omega} (\omega)$$

The Bayesian inference technique consists in *estimating the probability* $\mathbb{P}(\omega)$ of the classical states to be valid, given the constraints regarded as *prior information*. This is achieved very simply, roughly by assigning a probability 1 to the Boolean constraints required to be valid and a probability 0 to the Boolean constraints required to be invalid. Then, we can calculate the probability $\mathbb{P}(\omega)$ of the classical states ω by solving a linear system in the probability space \mathcal{P} . As a result, the Bayesian expectation of any Boolean function and more generally of any linear combination of classical states, called *observable*, are immediately computed as a linear combination of classical state probabilities.

Even if we start from a particular batch of queries, this initial batch is by no means the only one and other batches can be addressed at will. They are necessary anyway to exhaust the set of all relevant observables because each particular observable calls for a specific batch of queries.

A major novelty of the model is that every new query batch can be enabled from the initial batch quite simply, by introducing an auxiliary Hilbert space, say \mathcal{H} , as

$$\mathcal{H} = \text{Span}_{\omega \in \Omega} (|\omega\rangle),$$

and then, transcribing the initial probability space \mathcal{P} into \mathcal{H} . It turns out that each new basis in \mathcal{H} defines by reverse transcription a new query batch in a new probability space \mathcal{P}' . Thus, it is easy to enable any query batch, merely by changing the basis in \mathcal{H} .

For convenience, the memory is called *Bayesian theater* while the set of constraints is called *Bayesian prior*. In addition, a batch of N dichotomous queries is called an *observation window* and especially, the initial batch is called the *source window*. When the queries are independent, the window is called “principal”. Otherwise, the window is said to be “twisted”.

This construction completely rediscovers standard quantum information theory while completely ignoring this theory (with the exception of terminology, kept for clarity).

1.3 Main New Results

Listed below are the main new insights provided by the model in both quantum information and Bayesian inference theories. Some of these are very surprising because they are at odds with current beliefs.

1.3.1 Nature of Quantum Information

The major point is that quantum information is nothing but classical information processed by an elaborate Bayesian inference technique. It is the relevant tool for managing the responses to a set of binary queries. Its scope is therefore universal, well beyond physics.

1.3.2 Major Feature of Bayesian Representation

The Bayesian representation of a single Boolean variable is very different from its deterministic representation. The main reason is that in this latest case, its expression is determined in advance. By contrast, in the Bayesian representation, the Boolean variable has no reason to coincide with a specific query of the current window. As a result, it is generally represented by a set of 2^N joint probabilities corresponding to the classical states generated by the N queries of the question-and-answer procedure.

1.3.3 Entanglement

Entanglement is in no way a characteristic of the system itself, but only expresses that the current binary queries are not mutually independent. In other words, entanglement is the aftermath of a *twisted* information window. Therefore, the concept of *entanglement* is not *per se* an intrinsic resource but a Bayesian artifact that expresses the non-independence of the current batch of variables. This seems surprising since it is often believed that entanglement is intrinsic and therefore cannot be changed by changing the observation window. But this is only true for local operations and classical communication (LOCC) and not in general. Among all observation windows, there is at least one optimal batch in which the queries are mutually independent. In this particular window, called “principal window” as opposed to “twisted window”, the problem is strictly classical.

1.3.4 Measurement

A measurement is defined naturally as the Bayesian estimate of an observable, which solves the so-called “measurement problem” as previously stated by Caves et al. [9]. Retrieving all the information stored in the memory usually requires several observation windows, but in return, this often generates some redundancy expressed by the uncertainty principle.

1.3.5 Uncertainty Principle

The iconic *uncertainty principle* expresses simply the obvious fact that it is impossible, by using two observation windows, to retrieve more information than is stored in the memory. Quantitatively, the uncertainty principle is expressed by standard

entropic bounds, namely the Maassen and Uffink [14] and the more precise Frank and Lieb [15] inequalities. Now, the present model provides a concrete and intuitive basis for these relationships. They are by no means a mysterious property of the quantum world.

1.3.6 Observables

Observable are computed as linear combinations of classical state probabilities. The existence of non commutative observables is a trivial consequence of the fact that each particular observable calls for a specific batch of queries and therefore is defined with respect to a specific observation window, called its “proper window”.

1.3.7 Gauge Group

The gauge group is the symmetry group of the Hilbert space \mathcal{H} , expressing non-uniqueness of the transcription of the Bayesian probability state into \mathcal{H} . It turns out that this group is just another expression of the Bayesian prior, up to contextual parameters. In other words, symmetry of \mathcal{H} is enough to describe the system. The model provides an *explicit derivation* of the gauge group as a combination of *unitary* and *antiunitary* operators. While antiunitary operators play an important but somewhat mysterious role in standard physics, they are now naturally introduced into the current model.

1.3.8 Window Contextuality

The *window contextuality* is the free choice of a particular batch of binary queries and gives rise to the famous “paradoxes”, like violation of Bell’s inequalities, perfectly rational in the present model. The model provides a concrete and intuitive basis for the contextually dependent aspects of quantum objects. Technically, the changes of binary queries, *a priori* complicated in the probability space \mathcal{P} , are simply expressed by unitary operators acting on the Hilbert space \mathcal{H} .

2 Background

This section introduces the memory accessible by a question-and-answer procedure and depicts its organization as a Kolmogorov probability space.

2.1 Classical Register

A classical register is a finite set X capable of storing classical information. We will only deal with binary degrees of freedom.

Definition 1 (*Discrete degree of freedom*) A discrete degree of freedom is one dichotomic choice.

2.2 Boolean Algebra

First, we must assign a query to each degree of freedom. We identify the classical register with a binary Boolean algebra, still denoted by X , with a batch of N Boolean variables X_i , for $i \in \llbracket 1, N \rrbracket$ taking value in $\{0, 1\}$. We adopt the symbol “1” for “valid” and “0” for “invalid”. We name *complete assignment*, x , a full assignment to the N variables and *partial assignment* an assignment to less than N variables. We note \bar{X}_i the negation of X_i . Finally, we call *literal* a variable or its negation. Obviously, this choice is a matter of gauge since we could rename $\bar{X}_i = Y_i$ and $X_i = \bar{Y}_i$. Let us term “discrete Boolean gauge” this choice. This initial allocation is done *once and for all* and its simultaneous inversion for all variables is simply a change of terminology.

Definition 2 (*Discrete Boolean gauge*) The discrete Boolean gauge is the initial allocation of a Boolean variable or its negation to all N degrees of freedom.

Given two Boolean formulas f_1 and f_2 , it is convenient to note $(f_1; f_2)$ (with a semicolon) the conjunction $f_1 \wedge f_2$ and (f_1, f_2) (with a comma) the disjunction $f_1 \vee f_2$. We name *partial requirement* a partial register of *literals*, that is a conjunction of literals, e.g. $(X_i; \bar{X}_j; X_k)$ and *complete requirement* (or *classical state*), ω , a conjunction of N literals, e.g. $\omega = (X_1; \bar{X}_2; \dots; X_N)$, which is satisfiable by a complete assignment x_ω , e.g. $x_\omega = (1; 0; \dots; 1)$. Clearly, there are 2^N different complete assignments and therefore 2^N complete requirements. In multivariate information analysis [16] these complete requirements are called *atoms* and the particular atom labelled $\varpi_0 = (0, 0, \dots, 0)$ is referred to as the *empty atom*. Clearly, the fact that a particular atom is the empty atom depends on the discrete Boolean gauge, Definition (2). Throughout this paper, we will use indifferently the terms “complete requirement”, “classical state” or “atom”. Let $\Omega \stackrel{\text{(def)}}{=} \{\omega\}$ denote the set of classical states.

On the other hand, with up to N variables, it is possible to construct 2^{2^N} different Boolean formulas, $f : \Omega \rightarrow \{0, 1\}$, described, e.g. as full disjunctive normal forms, i.e. reunion of complete requirements. Thus, any Boolean function can be described as a disjunction $(\omega_1, \omega_2, \dots, \omega_\ell)$ of $\ell \leq 2^N$ classical states ω_i . In particular, the *tautology* $I : \Omega \rightarrow \{1\}$ corresponds to the reunion of all 2^N classical states.

2.3 Bayesian Algebra

We propose to treat any *Boolean function* as a *random event* and account for the constraints by a set of equations between the *probabilities* of the relevant *requirements* (partial or complete), as explained below in Sect. (3.1). For this, we use the Bayesian theory of inferences [6] and regard henceforth the Boolean variables X_i as *random variables* taking values on the alphabet $\{0, 1\}$. We will name *Bayesian algebra*

such a mathematical object composed of a classical Boolean algebra endowed with a Bayesian probability structure.

In general, the hypotheses are specified by a set of constraints. We regard these constraints as a Bayesian *prior*, that is an *ensemble of definite conditions*, say (Λ) , e.g, a set of Boolean formulas compelled to be valid or invalid. We will see that it is convenient to consider more generally a linear combination of Boolean classical states in the probability space. Now, *the probability of any event will be conditional on (Λ)* .

2.3.1 Kolmogorov Probability Space

The basic sample set is the ensemble $\Omega = \{\omega\}$ of all mutually exclusive 2^N complete requirements, labelled by the 2^N complete assignments x_ω . Since the cardinality of Ω is finite, the power set $\mathfrak{P}(\Omega)$, of cardinality 2^{2^N} , is a sigma-algebra \mathcal{T} , pointwise identical to the ensemble of all Boolean functions. This means that any *event* is just a Boolean formula, that is a finite set of atoms. Next, we have to introduce an unknown probability measure \mathbb{P} on \mathcal{T} conditional on (Λ) . Finally, the Kolmogorov probability space associated with the prior (Λ) is $(\Omega, \mathcal{T}, \mathbb{P})$.

In general there is a number of probability distributions \mathbb{P} compatible with a prior (Λ) . We will define later the selection of a single solution as the “source contextuality”.

2.3.2 Notation

Throughout this paper, we will specifically name *unknowns* the conditional *probability* of complete or partial requirements, not to be confused with the variables or Boolean functions subject to randomness. Except when mentioned otherwise, we will use a shorthand to describe the unknowns, namely $\mathbb{P}(i)$ for $\mathbb{P}(X_i = 1|\Lambda)$, $\mathbb{P}(-i)$ for $\mathbb{P}(X_i = 0|\Lambda)$, $\mathbb{P}(i; -j)$ for $\mathbb{P}(X_i = 1; X_j = 0|\Lambda)$, etc. (for $i, j \dots \in [[1, N]]$). Similarly, we will use $\mathbb{P}(\omega)$ for $\mathbb{P}(\omega = 1|\Lambda)$. We will often call *partial probability* an unknown like $\mathbb{P}(i; -j)$ with less than N literals and *complete probability* an unknown $\mathbb{P}(\omega)$ with N literals. An unknown labeled k without further detail will be denoted by p_k , e.g. we may have $p_k = \mathbb{P}(i; -j)$. An array of unknowns will be denoted by $p = (p_k)$.

For clarity, we use most of the time the term “classical” in its usual acceptance, as opposed to “quantum”, although this term remains vague at this stage.

2.3.3 Observation Window

Up to Sect. (4.1), we ignore communication channels and only consider a single viewpoint. This means that we are given a classical register and investigate what we can infer from the known assumptions. All parameters, either input data in the prior (Λ) or observable entries (q_ω) , rely to a single batch of binary variables, what we call a single *observation window*. We will discuss later the possibility of reformulating the same problem by using other batches of queries, that is, in our terminology, other

“observation windows”. This defines the concept of general system and requires the construction of transition mappings between successive windows: Eventually, the reunion of all windows within a global atlas, that we call a “Bayesian theater” will make use of a complex Hilbert space endowed with a density operator. We will refer to the initial static issue as the *source observation window*.

2.3.4 Universal Equations

Since the probability laws are just an extension of Aristotelian logic the following relations are universal:

$$\mathbb{P}(\pm i; \pm j; \pm k; \dots) \geq 0 \quad (1)$$

$$1 = \mathbb{P}(i) + \mathbb{P}(-i) \quad (2)$$

$$\mathbb{P}(i) = \mathbb{P}(i; j) + \mathbb{P}(i; -j) \quad (3)$$

$$\mathbb{P}(i; j) = \mathbb{P}(i; j; k) + \mathbb{P}(i; j; -k) \quad (4)$$

etc. where i, j, k, \dots are signed integers and $|i|, |j|, |k|, \dots \in [[1, N]]$ are distinct.

Note that accounting for Eqs. (2, 3, 4, etc.), Eq. (1) implies that

$$0 \leq \mathbb{P}(\pm i; \pm j; \pm k; \dots) \leq 1. \quad (5)$$

2.3.5 Bayesian LP System

To secure the unknowns to be indeed probabilities, the relevant universal equations must mandatory be added to the constraints. As a result, the LP problem considered in Sect. (3.1) below is specific, as opposed to a general LP system. Its solutions are indeed in the interval $[0, 1]$ and its specific polytope will prove to be a simplex. It will be called “Bayesian LP system”.

3 Source Observation Window

This section poses the logical issue at hand. We start with a particular batch of queries, referred to as the “source window”. The problem is defined by a set of hypotheses to be satisfied. In the present Bayesian model, they are viewed as a *prior*, say (Λ) . In principle, the prior is composed of Boolean formulas, that is *events* of the sigma-algebra \mathcal{T} , required to be valid or invalid. However, beyond Boolean formulas which can take only two values, it is convenient to use a more general concept, namely, “observables”.

Definition 3 (*Observable*) An observable Q is a real-valued function of the classical states on the register, defined as

$$Q : \Omega \rightarrow \mathbb{R} : \omega \mapsto Q(\omega) = q_\omega. \tag{6}$$

We will denote the array (q_ω) by q . Notice that while the set of Boolean functions \mathcal{T} is closed under complement, unions and intersections, the set of observables is closed under linear combinations and is in fact nothing else than the dual of the probability space. This will be detailed in Sect. 3.2 below.

3.1 Linear Programming Problem

Bayesian inference of the variables consists in calculating their probability from the prior. To ensure that the solutions are indeed probabilities, we must first include in the prior the relevant universal equations, Eqs. (2, 3, 4, etc.). Next, we propose that the prior be simply incorporated by assigning a probability of 1 to logical functions required to be valid and a probability of 0 to logical functions required to be invalid. For instance, a partial requirement $(X_i; \bar{X}_j; X_k)$, compelled to be valid or invalid in the Boolean algebra, is trivially encoded as $\mathbb{P}(i; -j; k) = 1$ or 0 respectively. A Boolean function defined as a disjunction of classical states $f = (\omega_1, \omega_2, \dots, \omega_\ell)$ and required to be valid or invalid in the Boolean algebra, is encoded as $\sum_i \mathbb{P}(\omega_i) = 1$ or 0, because the classical states, ω_i , are disjoint, etc. In this way, any logical constraint is translated into a *linear* expression, that is to say, an *observable*.

Subsequently, the full prior, comprising both the specific equations and the relevant universal constraints is formulated as a linear programming (LP) problem in stack variables [17] within a convenient real-valued vector space.

Proposition 1 *The Bayesian probability p_i are the solution of a LP system,*

$$\begin{aligned} Ap &= b \\ \text{subject to } p &\geq 0 \end{aligned} \tag{7}$$

where $p = (p_i)$ is a real-valued positive unknown vector, $A = (a_{i,j})$ a real matrix and $b = (b_j)$ a real vector, while $p \geq 0$ stands for $\forall i, p_i \geq 0$.

The number of unknowns p_i , say n , is based on the particular formulation, that is the partial and complete probabilities explicitly involved. In explicit computation, it is crucial to have a minimum set of unknowns. On the contrary, for a theoretical discussion, it is necessary to take the full set of complete probabilities as unknowns, even if the number $n = 2^N$ is exponential in N . We will adopt this choice from Sect. (3.2). Let $m > 0$ denote the number of rows of the matrix, so that A is a $m \times n$ matrix. We will assume that the non-independent rows have been eliminated and that m is also the rank of the system.

It remains to complete the computation by solving this LP problem, Eq. (7). A feasible solution is a numerical vector of unknowns, p , that satisfies the prior (Λ), that is Eq. (7), and therefore defines a probability distribution \mathbb{P} on the sample set Ω and thus a probability measure on the sigma-algebra \mathcal{T} .

If the problem is inconsistent, the system is unfeasible. *A priori*, if the problem is well posed and admits a solution, one might expect that the system will provide a deterministic solution. In reality, most feasible LP problems do not accept deterministic solutions. This simply means that the initial batch variables are not mutually independent.

Proposition 2 *When the LP problem accepts a deterministic solution, the binary variables X_i of the source window are mutually independent.*

Proof A deterministic solution is trivially a separable joint probability which implies that the variables X_i are mutually independent (see Sect. 3.5.1). \square

When the LP system is feasible but does not accept a deterministic solution, a deterministic solution exists nevertheless but in another window, namely a “principal window” defined in Sect. (5.2).

In general, the rank m of the matrix A is less than n and thus, there is a continuous set of solutions. This arises when for some reason the Bayesian prior (Λ) is not specific enough. For example, in quantum mechanics, a set of data may be fundamentally out of control of the experimenter. Thus, the particular probability distribution to be used has to be fixed by an exogenous choice. The system is said to be to context-dependent. Let us define precisely what we term “contextuality”.

Definition 4 (*Contextuality*) A system is context-dependent when the involved probability distribution depends on an exogenous choice.

Given that contextuality has also other causes in general systems (Sect. 4.1, below), we will refer to this property as the *source contextuality*.

Definition 5 (*Source contextuality*) Source contextuality expresses the possibility of choosing a particular feasible probability distribution among the solutions of the source LP problem.

A particular solution must be chosen by a selection rule. This rule will be said to fix a particular “context”. Thus, *source contextuality is a piece of intrinsic information specified at the outset* in addition to the Bayesian prior.

3.2 Real Probability Space \mathcal{P}

We now assume that the unknowns $p = (p_\omega)$ are specifically the 2^N complete probabilities of the classical states, i.e. $p_\omega = \mathbb{P}(\omega = 1 | \Lambda)$ with $\omega \in \Omega$. This can easily be achieved by eliminating the partial probabilities using the universal equations, Eqs. (3, 4, ...). Then

$$p \in \mathbb{R}^\Omega = \text{Span}_{\omega \in \Omega} (\omega).$$

We denote by \mathcal{P} this real-valued vector space and \mathcal{P}^* its dual space, both of dimension $n = d = 2^N$. Clearly, from Definition (3), the dual space \mathcal{P}^* is the space of the observables defined in the source window. As long as a single window is concerned, no metric is required. We will indifferently refer to \mathcal{P} as the “real probability space” or the “LP space”.

Notation When there is no risk of confusion, we will use the same symbols $\omega, \omega', \omega_i, \dots$ to designate either the classical states in Ω or the different labels in \mathcal{P} and \mathcal{P}^* .

- We note $\tilde{\omega} \in \mathcal{P}$, with $\omega \in \Omega$, the basis vectors in \mathcal{P} , i.e. $\tilde{\omega} = (p_{\omega'})$ with $p_{\omega'} = \delta_{\omega'\omega}$. They describe deterministic probability distributions. The basis is denoted by $\tilde{\Omega} \stackrel{(\text{def})}{=} \{\tilde{\omega}\}$ or simply Ω when no confusion can occur.
- A covector in the dual space \mathcal{P}^* is denoted $q = (q_\omega)$ with $\omega \in \Omega$. A covector defines an observable on the register, $Q(\omega) = q_\omega$.
- A dual form $(\mathcal{P}^*, \mathcal{P}) \rightarrow \mathbb{R}$ is denoted $\langle qp \rangle$, where $q \in \mathcal{P}^*$ and $p \in \mathcal{P}$.
- We will note $\tilde{\omega}^*$ the canonical basis covectors in \mathcal{P}^* defined by.
- An observable defined by a covector $q = (q_\omega)$ with $q_\omega \geq 0 (\forall \omega \in \Omega)$ is called *non-negative*.
- A Boolean function f defines an observable $F(\omega)$, that is a non-negative dual form whose associated covector $f = (f_\omega)$ is the indicator function of f in Ω . In particular, a basis covector $\tilde{\omega}^*$ defines an observable $F(\omega') = \langle \tilde{\omega}^* \tilde{\omega}' \rangle$ that we will simply denote $\tilde{\omega}^*$ when no confusion can occur.

Expectation The value $\langle Q \rangle$ of a dual form $\langle qp \rangle$ with respect to the probability distribution $\mathbb{P}(\omega) = p_\omega$, is trivially the *expectation value* of the observable $Q(\omega) = q_\omega$.

$$\langle Q \rangle = \sum_{\omega \in \Omega} Q(\omega) \mathbb{P}(\omega) = \sum_{\omega \in \Omega} q_\omega p_\omega = \langle qp \rangle$$

Proposition (1) can be rewritten as

Theorem 1 (Bayesian formulation) *Any LP system, Eq. (7), can be expressed as the following Bayesian problem,*

$$(\Lambda) : \text{ Given } m - 1 \text{ observables } A_\ell \text{ assign } \mathbb{P} \text{ on } \Omega \text{ subject to } \langle A_\ell \rangle = b_\ell, \quad (8)$$

where $\ell \in [[1, m - 1]]$. In addition, it is possible to assume that the expectation of the observables A_ℓ is zero, that is $b_\ell = 0$.

Proof In Eq. (7), without loss in generality, assume that one row is the normalization constraint that is the tautology. We reserve the index $\ell = 0$ to this normalization equation, namely, $A_0 = I, a_{0,\omega} = 1, \forall \omega \in \Omega$ and $b_0 = 1$. Clearly, each row, now labeled ℓ , defines a covector, $a_\ell = \sum_{\omega} a_{\ell,\omega} \tilde{\omega}^*, (\ell \in [[1, m - 1]])$. It is a partial

constraint expressed as the expectation of an observable $A_\ell(\omega) = a_{\ell,\omega}$. Indeed, $\sum_{\omega} a_{\ell,\omega} p_{\omega} = b_{\ell}$ means $\langle A_{\ell} \rangle = b_{\ell}$.

Now, Eq. (7) can be reformulated as follows: Assign a probability distribution \mathbb{P} on Ω , given that the expectation of m independent observables A_{ℓ} are subject to $\langle A_{\ell} \rangle = b_{\ell}$. Since normalization is implicit in probability theory, Eq. (7) can be expressed as Eq. (8). We can assume that $b_{\ell} = 0$ for $\ell > 0$ because otherwise, we can replace A_{ℓ} by $A_{\ell} - b_{\ell}A_0$. The converse is obvious. Now, the system, Eq. (8) depicts a standard Bayesian problem [6]. \square

It is convenient to first address the simplest problem, in which the prior is reduced to the normalization equation.

3.2.1 Tautology

Irrespective of the current prior (Λ), consider the following Bayesian LP system in the probability space \mathcal{P} ,

$$\sum_{\omega \in \Omega} p_{\omega} = 1$$

(9)

subject to $p_{\omega} \geq 0$

Any solution $p = (p_{\omega})$ of this system describes a potential probability distribution \mathbb{P} on Ω and conversely any probability distribution is a solution of Eq. (9). The d classical deterministic states $\omega \in \Omega$ label both the basis vector $\tilde{\omega} \in \mathcal{P}$ and the extreme points of a convex polytope, \mathcal{W}_i , of dimension $d - 1$ with d vertices, that is a $(d - 1)$ -simplex, known as “probability simplex” or “Choquet simplex” in convex geometry. In the present context, we will call this polytope, \mathcal{W}_i , the d -dimensional *tautological simplex*.

Definition 6 (*Tautological simplex \mathcal{W}_i*) The “tautological simplex” in the d -dimensional vector space \mathcal{P} is the $(d - 1)$ -simplex

$$\mathcal{W}_i = \operatorname{conv}_{\omega \in \Omega} (\tilde{\omega})$$

(10)

Proposition 3 *The entries p_{ω} in Eq. (9) represent both the d components of p in \mathcal{P} and the d barycentric coordinates of the point p on the tautological simplex \mathcal{W}_i . In other words, the distinction between barycentric and contravariant components vanishes on \mathcal{W}_i .*

Proof Since $\sum_{\omega \in \Omega} p_{\omega} = 1$, the two formulations mean $p = \sum_{\omega} p_{\omega} \tilde{\omega}$. \square

Since \mathcal{W}_i is a simplex, the barycentric coordinates are uniquely defined. The set of its extreme points $\tilde{\Omega} = \{\tilde{\omega}\}$ forms its *Choquet boundary* and describes the deterministic distributions.

Proposition 4 Any basis subspace of \mathcal{P} is specified by a Boolean function compelled to be valid.

Proof Any basis subspace is the direct sum of one-dimensional subspaces \mathcal{P}_i , each spanned by a basis vector $\tilde{\omega}_i$ so that the direct sum $\mathcal{P}_1 \oplus \mathcal{P}_2 \cdots \oplus \mathcal{P}_\ell$ is specified by $f = (\omega_1, \omega_2, \dots, \omega_\ell) = 1$. □

3.2.2 Current Bayesian LP System

Return now to the current Bayesian LP system, Eq. (8) associated with the prior (Λ) . Suppose that the system is feasible and consider the set of solutions. It is convenient to single out two subspaces containing these solutions, respectively the *affine subspace* and the *effective probability space*.

Definition 7 (*Affine subspace P_Λ*) The affine subspace P_Λ is the affine set of the solutions.

Definition 8 (*Effective probability space \mathbb{W}_{d-m+1}*) The effective probability space \mathbb{W}_{d-m+1} is the linear span of the solutions.

Specific polytope \mathcal{W}_Λ In fact, from standard LP theory, the locus of the solutions is a specific polytope \mathcal{W}_Λ . Using Definition (6), we have from a simple inspection

$$\mathcal{W}_\Lambda = \text{conv}(w_k) = P_\Lambda \cap \mathcal{W}_l = \mathbb{W}_{d-m+1} \cap \mathcal{W}_l. \tag{11}$$

In addition, for Bayesian LP systems, this polytope is compact and convex and will prove to be a simplex in \mathbb{W}_{d-m+1} with $r = d - m + 1$ vertices. Let $\{w_k\}$ with $k \in [[1, r]]$ denote its extreme points.

Definition 9 (*Simplicial system*) A simplicial system is a LP problem whose specific polytope is either an isolated point or a simplex.

Proposition 5 Any Bayesian LP system is simplicial.

Proof Express the LP system in the effective probability space \mathbb{W}_{d-m+1} . In this space, using the vertices $\{w_k\}$ as basis vectors, the specific polytope is the tautological simplex. □

Definition 10 (*Specific simplex \mathcal{W}_Λ*) The solutions of the LP system, Eq. (7) are located on a simplex \mathcal{W}_Λ , called “specific simplex”, with $r = d - m + 1$ vertices.

3.2.3 Source Contextuality

In general, there are a number of solutions to the current Bayesian system Eq. (8) located on the specific simplex \mathcal{W}_Λ . The choice of a single solution, say $w_\Lambda \in \mathcal{W}_\Lambda$, specifies the “source context”.

3.2.3.1 Default Context Suppose first that there is no extra constraint, which we call the “default context”. The standard Bayesian solution is then the *most likely distribution*, determined by the maximum entropy principle [18], that is a generalization of the Laplace’s principle of indifference. This requires to consider a *uniform probability density* of dimension $d - m$ over the affine subspace P_Λ , normalized to unity on the convex hull of the specific polytope. The center of mass \tilde{c} is the mean point with respect to this uniform hull density. The default context corresponds to $w_\Lambda = \tilde{c}$. Equivalently, it is the barycenter of the vertices.

3.2.3.2 Other Contexts Other context could be specified by assigning a non-uniform hull density over the specific polytope, but more simply, we will specify the context by means of a discrete probability distribution over the vertices of the specific simplex, which we will call the *contextual probability distribution*.

3.3 Representation of the Bayesian States

A Bayesian state is represented by a specific simplex \mathcal{W}_Λ and a contextual distribution.

3.3.1 Working Distribution, Contextual Distribution

Technically, we need only to specify the mean point w_Λ on the simplex with respect to the contextual distribution because the details can be derived from the framework. Let us name this mean point the “working distribution”.

Definition 11 (*Working distribution*) The working distribution $w_\Lambda \in \mathcal{W}_\Lambda$ is the mean point with respect to an auxiliary probability distribution Σ_μ on the vertices of the specific simplex.

Let $r = d - m + 1$, let w_i be the r vertices of the simplex \mathcal{W}_Λ and let $\Sigma_\mu = \{\mu_i\}$ be the set of barycentric coordinates of $w_\Lambda \in \mathcal{W}_\Lambda$ on the simplex, that is,

$$w_\Lambda = \sum_{i=1}^r \mu_i w_i \quad \text{where} \quad \mu_i \geq 0 \quad \text{and} \quad \sum_{i=1}^r \mu_i = 1$$

Therefore, w_Λ is well the center of mass of the vertices $\{w_i\}$ weighted by $\{\mu_i\}$.

Definition 12 (*Contextual probability distribution*) The contextual probability distribution Σ_μ is the set of barycentric coordinates $\{\mu_i\}$ specifying the working distribution in the simplex.

Let us compute its entropy.

Definition 13 (*Working entropy*) The working entropy $\mathbb{H}(w_\Lambda)$ is the Shannon entropy S_w of the working distribution w_Λ .

$$S_w = \mathbb{H}(w_\Lambda) \stackrel{\text{(def)}}{=} \sum_{\omega \in \Omega} -w_{\Lambda,\omega} \log_2 w_{\Lambda,\omega}. \tag{12}$$

The working entropy is rather a Bayesian parameter and has little to do with a real uncertainty. It will be identified later with the so-called “window entropy”, Definition (42). Indeed, we will show that the quantity $N - \mathbb{H}(w_\Lambda)$ represents the amount of information that can be extracted from the current observation window. By contrast, the “simplicial entropy” defined in the next section will directly represent a form of uncertainty.

3.3.2 Simplicial Quantum States

It turns out that a compact formulation of the Bayesian state, comprising the specific simplex and a contextual probability distribution, is the exact equivalent of the standard quantum state restricted to the source window. We propose to call it “a simplicial quantum state”.

Definition 14 (*Simplicial quantum state*) A simplicial quantum state is the pair $(\Sigma_\mu, \mathcal{W}_\Lambda)$ of a contextual probability distribution $\Sigma_\mu = \{\mu_i\}$ and a specific simplex \mathcal{W}_Λ . The working distribution is the mean point $w_\Lambda = \sum_{i=1}^r \mu_i w_i$ where $r = d - m + 1$. The simplicial quantum state will be designated indifferently by one of the pairs $(\Sigma_\mu, \mathcal{W}_\Lambda)$ or $(w_\Lambda, \mathcal{W}_\Lambda)$.

Let us compute the entropy of the contextual distribution.

Definition 15 (*Simplicial entropy S_μ*) The simplicial entropy of a simplicial quantum state $(\Sigma_\mu, \mathcal{W}_\Lambda)$ is the Shannon entropy of the contextual distribution

$$S_\mu \stackrel{\text{(def)}}{=} \mathbb{H}(\Sigma_\mu) = \sum_{i=1}^r -\mu_i \log_2 \mu_i. \tag{13}$$

We will use indifferently the terms “simplicial entropy” or “contextual entropy”.

To sum up, we encountered two forms of entropy, the working entropy $\mathbb{H}(w_\Lambda)$ on the sample set and the simplicial entropy $\mathbb{H}(\Sigma_\mu)$ on the simplex. The two forms of entropy obviously differ in the source window, for instance the simplicial entropy of a pure state (defined just below) is zero, which is not the case in general for the working entropy.

However, they will merge in a “principal window” (Proposition 28 below). At last, they are both bounded above by the storage capacity of the register, i.e. N bits.

The simplicial entropy is closely related to the von Neumann entropy of standard quantum information. It turns out that the von Neumann entropy is actually the lower bound of all simplicial entropies over all windows, defined in general systems, Sect. (5). This will lead to a more substantial interpretation of the von Neumann entropy in terms of information theory in Theorem (5) below.

3.3.3 Pure States

When the simplex \mathcal{W}_Λ is reduced to an isolated point, we have a *pure state*. This means that the rank m of the LP-system, Eq. (7) is equal to the dimension of the space, $m = d$ and thus $r = d - m + 1 = 1$. There is a single feasible solution, $w_\Lambda = (w_{\Lambda,\omega})$ and the polytope $\mathcal{W}_\Lambda = \{w_\Lambda\} \subset \mathcal{W}_i$ is trivially identical to the working distribution. At last there is a single feasible probability distribution,

$$\mathbb{P}(\omega = 1|\Lambda) \stackrel{\text{(def)}}{=} w_{\Lambda,\omega}$$

The simplicial entropy is zero. Finally, the expectation of any observable $Q(\omega) = q_\omega$ reads trivially

$$\langle Q \rangle = \langle qw_\Lambda \rangle = \sum_{\omega \in \Omega} q_\omega w_{\Lambda,\omega}. \tag{14}$$

Definition 16 (*Pure and mixed simplicial quantum states*) A simplicial quantum state is pure when the specific simplex is reduced to a single point. Otherwise, the state is mixed.

3.3.4 Mixed States

When the rank $m > 0$ is less than d the prior does not uniquely determine the solution of the system and therefore the working probability w_Λ is defined by the contextual distribution Σ_μ . In this case, from Definition (16) the simplicial state is termed “mixed”.

Let μ_i be the simplicial coordinates of w_Λ in \mathcal{W}_Λ . We have, using $r = d - m + 1$

$$\mathbb{P}(\omega = 1|\Lambda) \stackrel{\text{(def)}}{=} w_{\Lambda,\omega} = \sum_{i=1}^r \mu_i w_{i,\omega} \quad \text{with} \quad \sum_{i=1}^r \mu_i = 1 \tag{15}$$

As a result, for any observable $Q(\omega) = q_\omega$, we have

$$\langle Q \rangle = \langle qw_\Lambda \rangle = \sum_{i=1}^r \mu_i \langle qw_i \rangle = \sum_{\omega \in \Omega} \sum_{i=1}^r \mu_i q_\omega w_{i,\omega} \tag{16}$$

This equation is also valid for pure states, with $m = d$, $\mu_1 = 1$ and $w_1 = w_\Lambda$. Furthermore, when the contextual distribution is deterministic, that is when all μ_i

coefficients except one are zero, e.g. $\mu_1 = 1$, the mixed state reduces to a pure state, e.g. $w_\Lambda = w_1$.

3.4 Measurement with Respect to a Simplicial Quantum State

Let us now turn to the measurement of an observable of the source observation window with respect to a simplicial quantum state $(w_\Lambda, \mathcal{W}_\Lambda)$. We simply identify the Bayesian measurement with the expectation value of the observable with respect to the working distribution.

3.4.1 Expectation of an Observable

Let $q = (q_\omega)$ be a covector, corresponding to an observable Q .

Definition 17 (*Quantum expectation $\langle Q \rangle$*) The quantum expectation of an observable $Q(\omega) = q_\omega$ is the expectation $\langle Q \rangle = \langle qw_\Lambda \rangle$ with respect to the working distribution w_Λ .

The expectation was previously computed in Eqs. (14, 16).

3.4.2 Projective Measurement

Let $\Gamma = \{\gamma\}$ denote a finite set. Define an ensemble of mutually disjoint Boolean functions $\{f_\gamma, \gamma \in \Gamma\}$ such that the reunion of all f_γ is the tautology. Equivalently, let $\{f_\gamma = (f_{\gamma,\omega}), \gamma \in \Gamma\}$ be the indicators F_γ of f_γ in \mathcal{P}^* , such that $\sum_\gamma f_{\gamma,\omega} = 1$ for all $\omega \in \Omega$, i.e. $\sum_\gamma F_\gamma = I$.

A projective measurement is defined as

$$\gamma \in \Gamma \mapsto p(\gamma) = \mathbb{P}(f_\gamma = 1 | \Lambda_\mu) = \langle f_\gamma w_\Lambda \rangle = \langle F_\gamma \rangle \geq 0.$$

From Proposition (4), a projective measurement means expanding the working distribution w_Λ with respect to the set of subspaces defined by the Boolean functions f_γ . This is similar to a projective measurement in standard quantum information, but restricted to the source window. In particular, when $\Gamma = \Omega$, $\{f_\omega = \tilde{\omega}, \omega \in \Omega\}$, $p(\omega) = \mathbb{P}(\omega)$.

3.4.3 General Measurement

Let $\Gamma = \{\gamma\}$ denote a finite set. Define an abstract resolution of the tautology in the source window, that is a set of non-negative forms in \mathcal{P}^* , $\{q_\gamma = (q_{\gamma,\omega})\}$ (with $\gamma \in \Gamma$), such that $\sum_\gamma q_{\gamma,\omega} = 1$ for all $\omega \in \Omega$, i.e. $\sum_\gamma q_\gamma = I$. Since $q_{\gamma,\omega}$ is not necessarily 0 or 1, q_γ is not necessarily associated with a Boolean function, but corresponds to a

positive observable Q_γ and $\sum_\gamma Q_\gamma = I$. A general measurement in the source window is defined by

$$\gamma \in \Gamma \mapsto p(\gamma) = \langle q_\gamma w_\lambda \rangle = \langle Q_\gamma \rangle.$$

This is similar to a particular positive-operator valued measure (POVM) in quantum information, *when the involved observables commute*.

3.5 Pair of Registers

The combination of two registers brings together most of the peculiarities of standard quantum information. This is the purpose of this section. Some developments are rather lengthy, but the results will be used later.

Consider a global classical register X_c composed of two distinct subregisters X_a and X_b . Let (Λ_c) denote a global Bayesian prior. Let N_a, N_b and $N_c = N_a + N_b$ be the numbers of binary variables in X_a, X_b and X_c respectively. Let $\mathcal{P}_a, \mathcal{P}_b$ and \mathcal{P}_c denote the probability spaces corresponding to X_a, X_b and X_c of dimension $d_a = 2^{N_a}, d_b = 2^{N_b}$ and $d_c = 2^{N_c}$ respectively. We have $\mathcal{P}_a \otimes \mathcal{P}_b = \mathcal{P}_c, N_a + N_b = N_c$ and $d_a \times d_b = d_c$.

Notation The classical states, e.g. in Ω_c , are noted $\omega_{c,i}, i \in [1, d_c]$.

3.5.1 Separability and Entanglement of a Single Probability Distribution

Definition 18 (*Separability, entanglement*) A probability distribution, $\mathbb{P}_c(\omega_a; \omega_b)$ on a global register, $X_c = (X_a, X_b)$, is separable with respect to a partition into the two distinct subregisters X_a and X_b , if

$$\begin{aligned} &\mathbb{P}_c(\omega_a; \omega_b) = \mathbb{P}_a(\omega_a) \times \mathbb{P}_b(\omega_b), \\ \text{subject to } &\sum_{\omega_a \in \Omega_a} \mathbb{P}_a(\omega_a) = \sum_{\omega_b \in \Omega_b} \mathbb{P}_b(\omega_b) = \sum_{\omega_c \in \Omega_c} \mathbb{P}_c(\omega_c) = 1. \end{aligned} \tag{17}$$

Otherwise, the joint distribution is *entangled*.

On the other hand, the concept of marginal distribution is related to the joint distribution $\mathbb{P}_c(\omega_a; \omega_b)$ by the *conditional probability* $\mathbb{P}_c(\omega_a | \omega_b)$ thanks to Bayes' law,

$$\mathbb{P}_c(\omega_a; \omega_b) = \mathbb{P}_b(\omega_b) \times \mathbb{P}_c(\omega_a | \omega_b).$$

Define a particular separable probability distribution $\mathbb{P}'_c(\omega_c)$ still on Ω_c as the product of the two marginal distributions $\mathbb{P}_a(\omega_a)$ and $\mathbb{P}_b(\omega_b)$, namely,

$$\mathbb{P}'_c(\omega_a; \omega_b) \stackrel{\text{(def)}}{=} \mathbb{P}_a(\omega_a) \times \mathbb{P}_b(\omega_b). \tag{18}$$

Define finally

$$S(\mathbb{P}_c \parallel \mathbb{P}'_c) = \sum_{\omega_c \in \Omega_c} \mathbb{P}_c(\omega_c) \log_2 \frac{\mathbb{P}_c(\omega_c)}{\mathbb{P}'_c(\omega_c)} \geq 0. \tag{19}$$

Proposition 6 *The global probability \mathbb{P}_c is separable with respect to the partition (X_a, X_b) if and only if its relative entropy with respect to the product $\mathbb{P}'_c(\omega_c) = \mathbb{P}_a(\omega_a) \times \mathbb{P}_b(\omega_b)$ of the marginal distribution in \mathcal{P}_a and \mathcal{P}_b is zero, that is, $S(\mathbb{P}_c \parallel \mathbb{P}'_c) = 0$.*

Proof We have $S(\mathbb{P}_c \parallel \mathbb{P}'_c) \geq 0$ because a relative entropy is non-negative. In addition, $S(\mathbb{P}_c \parallel \mathbb{P}'_c)$ is the minimum value over all possible relative entropies $S(\mathbb{P}_c \parallel \mathbb{P}''_c)$ for all separable distributions $\mathbb{P}''_c(\omega_a; \omega_b) = \mathbb{P}''_a(\omega_a) \times \mathbb{P}''_b(\omega_b)$, since we have by expanding Eq. (19),

$$S(\mathbb{P}_c \parallel \mathbb{P}_a \times \mathbb{P}_b) - S(\mathbb{P}_c \parallel \mathbb{P}''_a \times \mathbb{P}''_b) = -S(\mathbb{P}_a \parallel \mathbb{P}''_a) - S(\mathbb{P}_b \parallel \mathbb{P}''_b) \leq 0.$$

Therefore, $0 \leq S(\mathbb{P}_c \parallel \mathbb{P}_a \times \mathbb{P}_b) \leq S(\mathbb{P}_c \parallel \mathbb{P}''_a \times \mathbb{P}''_b)$. The minimum of $S(\mathbb{P}_c \parallel \mathbb{P}''_c)$ is zero if $\mathbb{P}'_a = \mathbb{P}_a$, $\mathbb{P}'_b = \mathbb{P}_b$ and $\mathbb{P}_c = \mathbb{P}_a \times \mathbb{P}_b$. □

To sum up, we have the following result:

Proposition 7 *A global probability distribution w_c governing a pair of distinct classical registers subject to a global prior is generally entangled with respect to the pair of registers. The amount of entanglement is characterized by the relative entropy between the global distribution and the product of its marginal distributions, Eq. (19). When the relative entropy is zero, the distribution w_c is separable and equal to the product of its marginals.*

3.5.2 Partial Simplicial Quantum State

The restriction of a global LP system to a subregister will be termed “partial LP system”. In essence, the problem is to reconstruct the effective probability subspace in the subregister. Technically, the reduction is implemented with respect to the current working distribution at work in the global system, that is on the simplicial quantum state, but the reduced specific simplex is actually independent of the working distribution. We will use indifferently the terms “partial”, “reduced” and “marginal” when no confusion can occur.

While the concept of *separable probability distribution* is not ambiguous, the situation is more subtle for simplicial quantum state. For convenience, set the following definitions, where every vertex of the specific simplex is viewed as a single probability distribution.

Definition 19 (*Separable simplex*) A simplex is separable with respect to a partition between two subregisters if all of its vertices are separable. Otherwise, the simplex is twisted.

Definition 20 (*Separable LP system*) A LP system is separable with respect to a partition between two subregisters if its specific simplex is separable. Otherwise, the LP system is twisted.

Definition 21 (*Separable simplicial quantum state*) A simplicial quantum state (w_c, \mathcal{W}_c) is separable with respect to a partition between two subregisters if its specific simplex \mathcal{W}_c is separable, irrespective of the working distribution w_c . Otherwise, the simplicial quantum state is twisted.

Definition 22 (*Product state*) A simplicial quantum state (w_c, \mathcal{W}_c) is a product state with respect to a partition between two subregisters if it results merely from the simple concatenation of the two registers X_a and X_b , meaning that the registers are defined independently, each subjected to its own constraint set.

Definition 23 (*Completely divisible state*) A simplicial quantum state (w_c, \mathcal{W}_c) is completely divisible if it results from the concatenation of N independent 1-bit registers X_i , each subjected to its own constraint set.

3.5.2.1 Reduction of a Pure State Assume first that the Bayesian system (Λ_c) in \mathcal{P}_c accepts a unique solution, i.e. depicts a pure state $w_c = (w_{c,(\omega_a;\omega_b)})$. The rank of the LP system is $m_c = d_c$. As a simplicial quantum state, its simplex is $\{w_c\}$ and the state is noted $(w_c, \{w_c\})$ or just w_c for simplicity. The rank of the state is $r_c = d_c - m_c + 1 = 1$ and the effective probability space $\mathbb{W}_c = \text{Span}(w_c)$ is of dimension 1.

Proposition 8 (*Reduction of a pure simplicial quantum state*) The restriction to \mathcal{P}_a of a global pure state, $w_c(\omega_c) = \mathbb{P}_c(\omega_a;\omega_b) \in \mathcal{P}_c$, is a partial simplicial quantum state (w_a, \mathcal{W}_a) whose specific simplex \mathcal{W}_a is the convex hull of the points $\tilde{v}_{\omega_b} \in \mathcal{P}_a$

$$\mathcal{W}_a = \text{conv}(\tilde{v}_{\omega_b}) \quad ; \quad \tilde{v}_{\omega_b} \stackrel{(\text{def})}{=} \sum_{\omega_a \in \Omega_a} \mathbb{P}_c(\omega_a|\omega_b) \tilde{\omega}_a \quad (20)$$

Its rank r_a is thus the rank of the set of vectors $\{\tilde{v}_{\omega_b}\}$ and the rank m_a of the associated LP system is $m_a = d_a - r_a + 1$. The working distribution w_a is the marginal in \mathcal{P}_a of the probability distribution w_c in \mathcal{P}_c .

When the global pure state w_c is separable, $r_a = 1$ and the partial simplicial quantum state is also a pure state $(w_a, \{w_a\})$.

Proof The restriction of the pure state $w_c \in \mathcal{P}_c$ to \mathcal{P}_a comprises by definition its marginal, $w_a = (w_{a,\omega_a})$, as

$$w_a = \sum_{\omega_a \in \Omega_a} \sum_{\omega_b \in \Omega_b} \mathbb{P}_c(\omega_a; \omega_b) \tilde{\omega}_a = \sum_{\omega_b \in \Omega_b} \mathbb{P}_c(\omega_b) \sum_{\omega_a \in \Omega_a} \mathbb{P}_c(\omega_a | \omega_b) \tilde{\omega}_a \tag{21}$$

where $\mathbb{P}_c(\omega_b) \stackrel{\text{(def)}}{=} \sum_{\omega_a \in \Omega_a} w_{c,(\omega_a; \omega_b)} = w_{b, \omega_b} = \mathbb{P}_b(\omega_b)$. Let $v_{\omega_b, \omega_a} \stackrel{\text{(def)}}{=} \mathbb{P}_c(\omega_a | \omega_b)$, that is

$$v_{\omega_b, \omega_a} = \begin{cases} w_{c,(\omega_a; \omega_b)} / w_{b, \omega_b} & \text{if } w_{b, \omega_b} \neq 0 \\ 0 & \text{if } w_{b, \omega_b} = 0. \end{cases} \tag{22}$$

Construct the vector set $\{\tilde{v}_{\omega_b} | \omega_b \in \Omega_b\} = \{(v_{\omega_b, \omega_a})\}$ in \mathcal{P}_a . Then, each vector $\tilde{v}_{\omega_b} \neq 0$ is a probability distribution in \mathcal{P}_a . Define $v_{\omega_b} = \mathbb{P}_c(\omega_b)$ and let r_a denote the rank of $\{\tilde{v}_{\omega_b}\}$. As a result, from Eq. (21), we have

$$w_a = \sum_{\omega_b \in \Omega_b} v_{\omega_b} \tilde{v}_{\omega_b} \in \mathcal{P}_a \tag{23}$$

In other words, the working distribution in \mathcal{P}_a is determined by the barycentric coefficients $v_{\omega_b} = \mathbb{P}_c(\omega_b)$. Since by hypothesis the outcomes ω_b are no more involved in the partial states, the coefficients v_{ω_b} are regarded henceforth as exogenous. As a result, the set of feasible solutions in \mathcal{P}_a is the full polytope $\text{conv}(\tilde{v}_{\omega_b})$ and its extreme points $\{w_{ai}\}$ are a subset of $\{\tilde{v}_{\omega_b}\}$. This polytope is actually the tautological simplex \mathcal{W}_a in the effective probability space $\mathbb{W}_a = \text{Span}(\tilde{v}_{\omega_b})$ with basis $\{w_{ai}\}$ in \mathcal{P}_a . Thus, the pair of this simplex \mathcal{W}_a and the initial marginal distribution w_a , Eq. (21), defines a simplicial quantum state (w_a, \mathcal{W}_a) in the probability space \mathcal{P}_a . Since the global simplex \mathcal{W}_c is reduced to a single point in isolation, there is only one choice for w_c and therefore there is a unique partial LP system.

When w_c is separable, $\mathbb{P}_c(\omega_a | \omega_b) = \mathbb{P}_c(\omega_a)$ irrespective of ω_b and

$$\tilde{v}_{\omega_b} = \sum_{\omega_a \in \Omega_a} \mathbb{P}_c(\omega_a | \omega_b) \tilde{\omega}_a = \sum_{\omega_a \in \Omega_a} \mathbb{P}_c(\omega_a) \tilde{\omega}_a = w_a$$

so that the simplex \mathcal{W}_a is reduced to the marginal distribution in isolation $\{w_a\}$. □

Proposition 9 *A pure separable simplicial quantum state is a product state.*

Proof The two independent LP systems are trivially e.g. in \mathcal{P}_a , $\langle \tilde{\omega}_a \rangle = \mathbb{P}_a(\omega_a)$ and in \mathcal{P}_b , $\langle \tilde{\omega}_b \rangle = \mathbb{P}_b(\omega_b)$. The concatenation leads in \mathcal{P}_c to $\langle \tilde{\omega}_c \rangle = \mathbb{P}_c(\omega_c)$ with $\omega_c = (\omega_a; \omega_b)$ so that $\mathbb{P}_c(\omega_c) = \mathbb{P}_a(\omega_a) \times \mathbb{P}_b(\omega_b)$. □

3.5.2.2 Construction of a global simplicial quantum state from a pair of reduced states Given two arbitrary simplicial quantum states in \mathcal{P}_a and \mathcal{P}_b , it is always possible to construct a compatible global state in \mathcal{P}_c .

Proposition 10 *There is always a non-empty set of global simplicial quantum states compatible with an arbitrary pair of partial simplicial quantum states.*

Proof The set of compatible global simplicial quantum state contains the product state and is thus non-empty. □

As a result, the restriction of a global simplicial quantum state to a subregister is always possible. Even if the global state (w_c, \mathcal{W}_c) is pure, the partial states (w_a, \mathcal{W}_a) and (w_b, \mathcal{W}_b) are generally mixed, with the exception of separable pure states $(w_c, \{w_c\})$. The simplicial entropy of the subsystem can thus be greater than the entropy of the full system and therefore *the simplicial entropy is not extensive*. Again, this property is a simple consequence of the Bayesian method and corresponds to the partial trace in standard quantum information theory.

3.5.3 Local Consistency and Non-signaling Correlations

Consider two correlated subregisters X_a, X_b and the partial sample sets Ω_a, Ω_b . The joint distribution $\mathbb{P}_c(\omega_c)$ is defined on the Cartesian product $\Omega_c = (\Omega_a, \Omega_b)$. From the definition of a partial subsystem, a local observer has only access to the variables of one subsystem and can only take into account the corresponding marginal probabilities. In other words, *each subsystem endowed with its marginal probability distribution is self-consistent* and can be considered in isolation.

Proposition 11 *The correlations between two partial subsystems subject to a global Bayesian prior are non-signaling.*

Proof From Proposition (10), whatever the second subsystem, the two partial subsystems are compatible. Therefore, any measurement in a subsystem is unable to provide information on the other subsystem. \square

The non-signaling property is less trivial when some input variables are implicit and considered as parameters. Then, for clarity, the actual variable set can be complemented so that the implicit variables become genuine variables as opposed to only parameters. Similarly, this is also an important feature of the partial trace in quantum information.

We first proved this conclusion in the context of the EPR paradox [19]. The expression “non-signaling correlations” was coined by Barrett et al [20] after a proposal by Popescu and Rohrlich to regard “nonlocality” as an axiom of quantum physics [21].

3.5.4 “Purification” of (w_a, \mathcal{W}_a) into \mathcal{P}_c

We saw that computing a partial LP system in a single observation window is similar to calculating the partial trace in quantum formalism. This suggests to consider the equivalent of a purification of the simplicial quantum state (w_a, \mathcal{W}_a) in \mathcal{P}_a with $r_a > 1$ vertices into a pure state w_c in \mathcal{P}_c . For convenience, we note “purification” (with quotes) this operation on the source probability space.

Consider the LP system of rank m_a in \mathcal{P}_a with $m_a = d_a - r_a + 1$ extreme points, w_i . It is possible to construct a “purification” of (w_a, \mathcal{W}_a) in \mathcal{P}_c .

Proposition 12 (“Purification”) *A simplicial quantum state (w_a, \mathcal{W}_a) in a probability space \mathcal{P}_a can be considered as the partial system of a pure state w_c in a probability space $\mathcal{P}_c = \mathcal{P}_a \otimes \mathcal{P}_b$.*

Proof Start from

$$w_a = \sum_{i=1}^{r_a} \mu_i w_i \in \mathcal{W}_a \subset \mathcal{P}_a. \tag{24}$$

where μ_i are the simplicial coordinates of w_a . Define an auxiliary space \mathcal{P}_b and suppose that $d_b \geq r_a$. Construct an arbitrary set of r_a independent vectors v_i in the tautological simplex \mathcal{W}_b in \mathcal{P}_b , i.e. $v_i \in \mathcal{W}_b \subset \mathcal{P}_b$ for $i \in [[1, r_a]]$. Construct a probability distribution $w_c = (w_{c,\omega_c}) = (w_{c,(\omega_a;\omega_b)}) \in \mathcal{P}_c = \mathcal{P}_a \otimes \mathcal{P}_b$ as

$$w_c = \sum_{i=1}^{r_a} \mu_i w_i \otimes v_i \quad \text{i.e.} \quad w_{c,(\omega_a;\omega_b)} = \sum_{i=1}^{r_a} \mu_i w_{i,\omega_a} v_{i,\omega_b}$$

We have clearly,

$$\sum_{\omega_c \in \Omega_c} w_{c,\omega_c} = \sum_{\omega_a \in \Omega_a} \sum_{\omega_b \in \Omega_b} w_{c,(\omega_a;\omega_b)} = \sum_{i=1}^{r_a} \mu_i \sum_{\omega_a \in \Omega_a} w_{i,\omega_a} \sum_{\omega_b \in \Omega_b} v_{i,\omega_b} = 1$$

so that w_c is indeed a probability distribution in \mathcal{P}_c and from Eq. (24)

$$\sum_{\omega_b \in \Omega_b} w_{c,(\omega_a;\omega_b)} = \sum_{i=1}^{r_a} \mu_i w_{i,\omega_a} \sum_{\omega_b \in \Omega_b} v_{i,\omega_b} = \sum_{i=1}^{r_a} \mu_i w_{i,\omega_a} = w_{a,\omega_a}.$$

Then, $w_a \in \mathcal{P}_a$ is effectively the marginal of $w_c \in \mathcal{P}_c$. The “purification” is completed. □

Depending upon the particular set of distributions $\{v_i\}$ in \mathcal{P}_b there is a number of possible solutions. For simplicity, it is possible to select v_i specifically among the basis vectors in \mathcal{P}_b . Label $\omega_b \in [[1, d_b]]$ the basis vectors $\tilde{\omega}_b$ in \mathcal{P}_b . Consider the set of r_a basis vectors $\tilde{\omega}_b \in \mathcal{P}_b$ for $\omega_b \in [[1, r_a]]$. For ease of exposition, rename ω_b the dummy subscript i in Eq. (24). Rewrite $w_a = \sum_{\omega_b=1}^{r_a} \mu_{\omega_b} w_{\omega_b}$ and set $v_{\omega_b} = \tilde{\omega}_b \in \mathcal{P}_b$ for $\omega_b \in [[1, r_a]]$. Construct the specific probability distribution $w_c = (w_{c,(\omega_a;\omega_b)}) \in \mathcal{P}_c = \mathcal{P}_a \otimes \mathcal{P}_b$ as

$$w_c = \sum_{\omega_b=1}^{r_a} \mu_{\omega_b} w_{\omega_b} \otimes \tilde{\omega}_b \quad \text{then} \quad w_{c,(\omega_a;\omega_b)} = \begin{cases} \mu_{\omega_b} w_{\omega_b,\omega_a} & \text{if } \omega_b \in [[1, r_a]] \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

Partial systems and “purifications” in real probability spaces are formally equivalent to partial traces and purifications in Hilbert spaces.

4 Introducing an Auxiliary Hilbert Space

We constructed a Bayesian probability space, \mathcal{P} , based on an initial batch of N binary queries, constituting the source observation window. In this section, we address the issue of changing the observation window. This leads to the construction of an auxiliary Hilbert space.

4.1 Changing the Current Observation Window

Changing the observation window constitutes a form of contextuality that we will call “window contextuality”.

Definition 24 (*Window contextuality*) Window contextuality corresponds to the free choice of a particular batch of binary queries.

Obviously, there is a close connection between the current batch of queries and the current *sample set* Ω of the Kolmogorov probability space.

Proposition 13 *There is a one-to-one correspondence between the sample set Ω defined in the current window and the current batch of binary queries.*

Proof By definition, the basic sample set Ω is the ensemble $\{\omega\}$ of the 2^N mutually exclusive classical states describing the joint probability distribution of all queries in the current window. \square

For simplicity, when no confusion can occur, we will name Ω both the probability sample set and the corresponding query batch.

4.2 Introducing a Hermitian Metric

By hypothesis, all batches of queries concern the same logical system of total probability 1. Therefore, each observation window Ω depicts a particular resolution of this total probability. Thus, irrespective of the current state, for any probability distribution $p = (p_\omega)$, we have

$$\forall \Omega : \sum_{\omega \in \Omega} p_\omega = 1 \quad (26)$$

Proposition 14 *Any resolution of the tautology defines a particular observation window. Therefore, to change the observation window, it suffices to change the sample set Ω .*

Proof Any resolution of the tautology defines a sample set Ω and thus, from Proposition (13), an observation window. Conversely, the sample set Ω defines a unique set of queries, that is a unique observation window. \square

For convenience, let us use an equivalent formulation to Eq. (26), as

$$\forall \Omega : \sum_{\omega \in \Omega} \left| \sqrt{p_\omega} e^{i\theta} \right|^2 = 1 \tag{27}$$

where $\theta(\omega)$ is an arbitrary gauge parameter introduced for reasons which will appear later. The real valued probability space on the source window was defined as

$$\mathcal{P} = \text{Span}_{\omega \in \Omega} (\omega).$$

Now, Eq. (27) suggests to introduce a Hermitian metric on a complex valued auxiliary space, namely, a finite dimensional Hilbert space \mathcal{H} associated to the probability space \mathcal{P} and defined as

$$\mathcal{H} = \text{Span}_{\omega \in \Omega} (|\omega\rangle),$$

where $|\omega\rangle$ represent the 2^N basis vectors.

We now propose to transcribe the probability space \mathcal{P} into \mathcal{H} , so as to keep in \mathcal{H} the dual forms of \mathcal{P} . It turns out that this corresponds to the standard ‘‘Born rule’’. Afterwards, we will specialize this transcription when \mathcal{P} is endowed with a simplicial quantum state.

4.3 Born Rule

In the source window, the covectors, $q = (q_\omega) \in \mathcal{P}^*$ are dual of the probability distributions $p = (p_\omega) \in \mathcal{P}$. Irrespective of the current simplicial state, using a relevant transcription rule, it is possible to obtain a similar correspondence in the Hilbert space \mathcal{H} .

To this end, we propose to transcribe any specific probability vector p in \mathcal{P} into a rank 1 projection operator Π acting on \mathcal{H} as

$$p \stackrel{\text{(def)}}{=} (p_\omega) \in \mathcal{P} \mapsto \Pi \stackrel{\text{(def)}}{=} |\sqrt{p}\rangle \langle \sqrt{p}| \quad \text{where } |\sqrt{p}\rangle = \sum_{\omega \in \Omega} \sqrt{p_\omega} e^{i\theta} |\omega\rangle \in \mathcal{H} \tag{28}$$

Again θ is a gauge parameter. From Eq. (26), $\text{Tr}(\Pi) = 1$. We adopt the standard terminology of *density operator* to denote a Hermitian operator of unit trace. Let $\mathcal{D}(\mathcal{H})$ denote the set of density operators. Clearly, $\Pi \in \mathcal{D}(\mathcal{H})$.

Furthermore, we propose to transcribe any covector $q \in \mathcal{P}^*$ into a Hermitian diagonal operators Q , as

$$q \stackrel{\text{(def)}}{=} (q_\omega) \in \mathcal{P}^* \mapsto Q \stackrel{\text{(def)}}{=} \text{Diag}_{\omega \in \Omega}(q_\omega) \in \text{Herm}(\mathcal{H}) \tag{29}$$

where $\text{Herm}(\mathcal{H})$ is the set of Hermitian operators acting on \mathcal{H} .

Now, $\text{D}(\mathcal{H}) \subset \text{Herm}(\mathcal{H})$ while $\text{Herm}(\mathcal{H})$ is a vector space over \mathbb{R} . Therefore, $\text{Tr}(\text{Q}\Pi)$ can be viewed as a dual form in $\text{Herm}(\mathcal{H})$. For ease of exposition, we will say that $\text{Tr}(\text{Q}\Pi)$ is a dual form in \mathcal{H} while $\langle \text{q}p \rangle$ is a dual form in \mathcal{P} .

Proposition 15 *When using the transcription rules, Eqs. (28, 29) for vectors and covectors respectively, the dual forms in \mathcal{P} are conserved in \mathcal{H} .*

Proof We have from a simple calculation,

$$\langle \text{q}p \rangle = \text{Tr}(\text{Q}\Pi) = \sum_{\omega \in \Omega} q_{\omega} p_{\omega}. \quad (30)$$

□

We propose to identify this result with the standard “Born rule”.

Definition 25 (*Born rule*) The Born rule expresses the conservation of the dual forms in the transcription $\mathcal{P} \rightarrow \mathcal{H}$, Eq. (30).

4.4 Transcription of the Current Probability Space Into the Hilbert Space

The above results, Sect. (4.2), are valid for any probability vector $p \in \mathcal{P}$. We now specifically consider the current probability space endowed with a simplicial quantum states $(w_{\lambda}, \mathcal{W}_{\lambda})$. For clarity, we slightly change the terminology, i.e, p is now the working distribution w_{λ} and the projector Π is a special case of density operator, say ρ_{λ} .

4.4.1 Transcription of Observables

The transcription of observables does not depend on the current state. Therefore we just have to repeat Eq. (29).

Definition 26 (*Observables in the source window*) In the source window the observables $Q = (q_{\omega}) \in \mathcal{P}^*$ are transcribed into diagonal Hermitian operators acting on \mathcal{H} as

$$\mathcal{P}^* \rightarrow \text{Herm}(\mathcal{H}) \quad Q = (q_{\omega}) \mapsto Q = \text{Diag}(q_{\omega}) \in \text{Herm}(\mathcal{H}) \quad (31)$$

Conversely, a diagonal Hermitian operator acting on \mathcal{H} represents an observable. We will later change the basis in \mathcal{H} . As a result, irrespective of the basis, any Hermitian operator acting on \mathcal{H} represents an observable. The operator Q will remain Hermitian but obviously will no more remain diagonal.

Proposition 16 *Observables are represented by Hermitian operators on the Hilbert space.*

Proof Any Hermitian operator is diagonal in some basis. □

The fact that Q is diagonal in the source window is expressed by saying that the source window is a “proper window” of the observable Q .

Definition 27 (*Proper window of an observable*) The proper window of an observable Q on a Hilbert space \mathcal{H} is an observation window in which the Hermitian operator Q is diagonal.

Any observable is thus calculated in its proper window. So that all the observables can be calculated, all the observation windows must therefore be taken into account.

4.4.2 Transcription of a Pure State

Consider a pure simplicial state $(w_\Lambda, \{w_\Lambda\})$.

Proposition 17 *It is possible to transcribe a pure simplicial quantum state $(w_\Lambda, \{w_\Lambda\})$ as a density matrix $\rho_\Lambda = |a\rangle\langle a|$, with $|a\rangle = \sum_{\omega \in \Omega} a_\omega |\omega\rangle$*

$$a_\omega = e^{i\theta} \sqrt{w_{\Lambda,\omega}} \tag{32}$$

where $\theta(\omega)$ is a gauge parameter (possibly 0).

Proof This corresponds to Eq. (28). □

In order to address mixed states, we find convenient to call “Gleason’s vector” the vector $|a\rangle \in \mathcal{H}$.

Definition 28 (*Gleason’s vector*) A Gleason’s vector is any unit vector $|a\rangle \in \mathcal{H}$ obtained by transcription of a pure state.

4.4.3 Transcription of a Mixed State

A mixed simplicial state, $(w_\Lambda, \mathcal{W}_\Lambda)$ or equivalently $(\Sigma_\mu, \mathcal{W}_\Lambda)$, is defined by a simplex \mathcal{W}_Λ composed of $r_\Lambda > 1$ extreme points w_i in \mathcal{P}_Λ and a set $\Sigma_\mu = \{\mu_i\}$ of simplicial coordinates.

Proposition 18 *A mixed simplicial quantum state $(\Sigma_\mu, \mathcal{W}_\Lambda)$ can be transcribed as a density operator ρ_Λ . Each extreme point w_i of the simplex is transcribed independently as a projector $|a_i\rangle\langle a_i|$, where the vector $|a_i\rangle$ is the Gleason’s vector of the pure state w_i , while the simplicial coordinates μ_i are conserved. Then*

$$w_\Lambda = \sum_{i=1}^{r_\Lambda} \mu_i w_i \rightarrow \rho_\Lambda = \sum_{i=1}^{r_\Lambda} \mu_i |a_i\rangle\langle a_i|. \quad (33)$$

In general, the unit vectors $|a_i\rangle$ are not orthogonal. An orthonormal array $|e_i\rangle$ is easily obtained by diagonalizing the density operator ρ_Λ as,

$$\rho_\Lambda = \sum_{i=1}^{r_\Lambda} \lambda_i |e_i\rangle\langle e_i|. \quad (34)$$

Proof The working distribution w_Λ can be viewed as a weighted combination of $r_\Lambda > 1$ auxiliary pure states of working distributions w_i in \mathcal{P}_Λ for $i \in [1, r_\Lambda]$. Since the weighting coefficients μ_i are independent of the simplex itself, the mixed state must be transcribed for consistency as the same weighted combination of the r transcribed projectors $|a_i\rangle\langle a_i|$ of the auxiliary pure states w_i . From Eq. (32), we obtain Eq. (33) and next Eq. (34) conserving the same rank r_Λ . \square

Gauge selection can be derived for the gauge selection of a pure state. To this end, we can use the *purification* procedure defined in Sect. (3.5.4).

Proposition 19 *The transcription of a mixed simplicial state can be implemented by (1) “purifying” this mixed state, (2) transcribing the simplicial pure state to obtained a standard quantum pure state and (3) tracing out this pure state.*

Proof We proceed in three steps. (1) “Purify” the simplicial quantum state $\{w_\Lambda, \mathcal{W}_\Lambda\}$ of rank r_Λ defined in the real probability space \mathcal{P}_Λ into a pure state w_c living in an auxiliary space $\mathcal{P}_c = \mathcal{P}_\Lambda \otimes \mathcal{P}_b$, as described in Sect. (3.5.4). (2) Transcribe the pure state w_c into a projection operator $|c\rangle\langle c|$ defined in a Hilbert space $\mathcal{H}_c = \mathcal{H}_\Lambda \otimes \mathcal{H}_b$. (3) Compute the partial trace over \mathcal{H}_b of the projection operator $|c\rangle\langle c|$ to obtain the relevant density operator ρ_Λ in \mathcal{H}_Λ . Step (1) has been defined in Sect. (3.5.2). Consider a real probability space \mathcal{P}_b of dimension $d_b \geq r_\Lambda$. Assume that $d_b = r_\Lambda$ and select the set of r_Λ basis vectors in \mathcal{P}_b , as described by Eq. (25),

$$w_c \stackrel{\text{(def)}}{=} \sum_{\omega_b=1}^{r_\Lambda} \mu_{\omega_b} w_{\omega_b} \otimes \tilde{\omega}_b \quad \text{then} \quad w_{c,(\omega_\Lambda;\omega_b)} = \mu_{\omega_b} w_{\omega_b,\omega_\Lambda}$$

where we changed the dummy subscripts “ i ” into “ ω_b ” for clarity.

Step (2) has been defined just above (Proposition 17). Let us denote $|c\rangle$ the Gleason’s vector and $c_{(\omega_\Lambda;\omega_b)}$ its entries.

Step (3) is a standard operation in quantum information with a unique solution. Resuming the subscripts “ ω_b ” into “ i ”, we obtain

$$\rho_\Lambda = \text{Tr}_b(|c\rangle\langle c|) = \sum_{i=1}^{r_\Lambda} \mu_i |a_i\rangle\langle a_i|$$

We have recovered Eq. (33) as required. \square

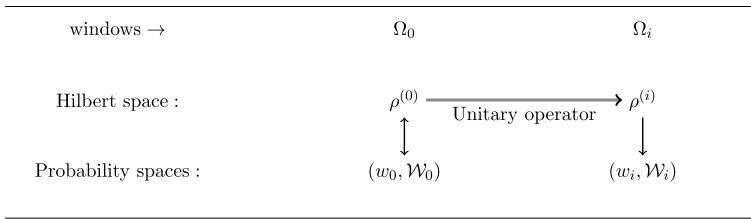


Fig. 1 General system. The source window Ω_0 is associated with a probability space, \mathcal{P}_0 endowed with a simplicial quantum state, (w_0, \mathcal{W}_0) . The space \mathcal{P}_0 is transcribed into an auxiliary Hilbert space \mathcal{H} with a basis also called Ω_0 and the simplicial quantum state, (w_0, \mathcal{W}_0) is transcribed into a density operator $\rho^{(0)}$. When changing the basis in \mathcal{H} from $\Omega_0 \rightarrow \Omega_i$, the window changes accordingly, the density operator expression changes from $\rho^{(0)}$ to $\rho^{(i)}$ and the new simplicial quantum state (w_i, \mathcal{W}_i) is computed by reverse transcription $\mathcal{H} \rightarrow \mathcal{P}_i$

Gauge Selection Any particular feasible Gleason’s vector $|c\rangle$ constructed in Step (2) corresponds to a gauge selection, as described for pure states.

4.4.4 Transcription of Dual Forms

By construction, dual forms are conserved.

Theorem 2 *Irrespective of the gauge, the expectation of an observable is*

$$Q(w_\Lambda) = \text{Tr}(Q\rho_\Lambda) \tag{35}$$

Proof $Q(w_\Lambda) = \sum_{\omega \in \Omega} q_\omega w_{\Lambda, \omega}$. Born’s rule applies. □

5 General Systems

In this section and the followings, we will treat the auxiliary Hilbert space \mathcal{H} as the standard Hilbert space of conventional quantum information theory. The difference is that every new basis of \mathcal{H} defines a particular observation window Ω_i and therefore a new probability space \mathcal{P}_i (Fig. 1).

In each observation window, the “factual” entity is the probability space while the Hilbert space is an auxiliary tool. Therefore, the ultimate justification for any rule in \mathcal{H} must be found in the probability spaces \mathcal{P}_i , obtained by reverse transcription. In particular, identical reverse transcriptions express an exact symmetry of \mathcal{H} .

Let us address the reverse transcription technique.

5.1 Reverse Transcription into a Source System

Reverse transcription is always possible, so any window can be considered a source window with the exception of a few exceptional windows which we will call “blind”.

5.1.1 Reverse Transcription of a Pure State

Let $\rho_\Lambda = |e\rangle\langle e|$ denote a pure density matrix in \mathcal{H} . Construct a real probability space \mathcal{P} of dimension $d = 2^N$. From Eq. (32), the working distribution is $w_\Lambda = |e|^2 \in \mathcal{P}$, i.e. $w_{\Lambda,\omega} = |e_\omega|^2$. The simplex \mathcal{W}_Λ is reduced to the isolated vertex $\{w_\Lambda\}$.

Proposition 20 *A density operator $\rho_\Lambda = |e\rangle\langle e|$ of rank 1 is reverse-transcribed as a simplex $\mathcal{W}_\Lambda = \{w_\Lambda\}$ composed of an isolated vertex $w_\Lambda = (w_{\Lambda,\omega})$ with $w_{\Lambda,\omega} = |e_\omega|^2$.*

LP System The vector $w_\Lambda = \mathbb{P}(\omega)$ is trivially the solution of the Bayesian linear system $p = |e|^2$ of rank $m = d$ as

$$\text{Assign } \mathbb{P} \text{ subject to } \langle \tilde{\omega}^* \rangle = |e_\omega|^2 \quad (\forall \omega \in \Omega)$$

where $\tilde{\omega}^*$ is the indicator function corresponding to the classical state ω . The normalization of the probability distribution arises from the normalization of e .

5.1.2 Reverse Transcription of a Mixed State

We are going to propose two different techniques to reverse backwards mixed states: One using purification is rather opaque but always valid. The second generalizes the technique used for pure states but is not valid for “blind windows” defined just below.

Start from a density operator ρ_Λ of rank r acting on a standard Hilbert space \mathcal{H} (Eq. 34) as

$$\rho_\Lambda = \sum_{i=1}^r \lambda_i |e_i\rangle\langle e_i|,$$

where the r vectors $|e_i\rangle$ form an orthonormal array in \mathcal{H} . Construct a real probability space \mathcal{P} of dimension $d = 2^N$ and let \mathcal{W}_i be the tautological simplex in \mathcal{P} , Definition (6). Construct the vectors $v_i = |e_i|^2 = (v_{i,\omega}) \in \mathcal{W}_i$ as $v_{i,\omega} = |e_{i,\omega}|^2$ and $w_\Lambda = \sum_{i=1}^r \lambda_i v_i$. Clearly, $w_\Lambda \in \mathcal{P}$ is a probability distribution. The rank of the vectors $\{v_i\}$ is crucial.

5.1.2.1 Regular Versus Blind Windows A “regular” window is a window in which transcription and reverse transcription are reversible. Thus, the rank is conserved by reverse transcription.

Definition 29 (*Regular window, blind window*) A window of rank r is “regular” when the r extreme orthonormal vectors $|e_i\rangle$ in the Hilbert space are reverse transcribed as a system $v_i = |e_i|^2$ of same rank r in the probability space. Otherwise, the window is called “blind”.

In particular, a pure window is trivially regular. It carries N information bits. By contrast, we will see that a blind window carries no information (see Sect. 7.5 below). As a result, a blind window is unable to serve as a “source window”.

5.1.2.2 Reverse Transcription By Purification of the Window Let \mathcal{H}_b be an auxiliary Hilbert space of dimension r . It is always possible to purify the mixed state into a Hilbert space $\mathcal{H}_c = \mathcal{H} \otimes \mathcal{H}_b$ of dimension $d \times r$, and next to reverse transcribe the pure state into a probability space $\mathcal{P}_c = \mathcal{P} \otimes \mathcal{P}_b$ as in Sect. (5.1.1). The quantum state $(w_\Lambda, \mathcal{W}_\Lambda)$ is then computed by applying Proposition (8).

Alternatively, it is possible to reverse transcribe a regular window by extending the method used in pure windows as follows.

5.1.2.3 Reverse Transcription of a Regular Window Construct the r -dimensional subspace $\mathbb{W}_r = \text{Span}_i(v_i) \subseteq \mathcal{P}$ and the tautological simplex \mathcal{W}_r in \mathcal{P} . Identify \mathbb{W}_r with an effective probability space and define the polytope

$$\mathcal{W}_\Lambda = \mathcal{W}_r \cap \mathbb{W}_r$$

The simplex \mathcal{W}_Λ admits r equivalent vertices, say w_j . Since w_Λ is a probability distribution and $w_\Lambda \in \mathbb{W}_r$, then $w_\Lambda \in \mathcal{W}_\Lambda$ so that

$$w_\Lambda = \sum_{j=1}^r \mu_j w_j,$$

for a specific set of simplicial coefficients μ_j . Finally, the reverse transcribed simplicial quantum state is $(w_\Lambda, \mathcal{W}_\Lambda)$. From the demonstration in Sect. (4.4.3), this explicit method is consistent with the purification procedure and provides the same result.

Finally, we reach the final result,

Theorem 3 (Quantum state) *A quantum state can be represented either by a standard density operator ρ_Λ in a Hilbert space \mathcal{H} or by a simplicial quantum state, i.e. a working distribution w_Λ within a simplex \mathcal{W}_Λ in a real probability space \mathcal{P} . For a definite simplicial state $(w_\Lambda, \mathcal{W}_\Lambda)$ in \mathcal{P} , the corresponding density operator ρ_Λ in \mathcal{H} is defined up to a gauge selection.*

5.1.3 Reverse Transcription of an Observable

We are given an observable Q , i.e. an Hermitian operator acting on a Hilbert space. Recall from Definition (3) that an observable is a real-valued function on a sample set Ω . Therefore, the observable can only be reverse transcribed in its proper window.

Theorem 4 *Any observable represented by a Hermitian operator Q acting on the Hilbert space \mathcal{H} can be reversed transcribed in its proper window as a covector q_ω in the probability space of \mathcal{P} .*

Proof Construct the probability space \mathcal{P} of the Q proper window. The entries of $q \in \mathcal{P}^*$ are the eigenvalues of Q . \square

5.2 Principal Window

It is always possible to *diagonalize* the density matrix ρ_Λ in \mathcal{H} by means of a unitary channel. This particular window is called “principal window” because it contains on its own all the Shannon information of the Bayesian theater, although in fact the principal basis is not unique when the eigenvalues are not all distinct.

Definition 30 (*Principal window, twisted window*) A principal window is an observation window in which the density operator is diagonal. Otherwise, the window is twisted.

Let $|\omega_i\rangle$ be the d basis vectors of the Hilbert space \mathcal{H} in a principal observation window. Let $|e_i\rangle$ denote the eigenvectors normalized to unity and λ_i the eigenvalues of the density operator ρ_Λ . Since ρ_Λ is diagonal, we have $|e_i\rangle = |\omega_i\rangle$ up to arbitrary gauge phase factors. After reordering the basis vectors if necessary, we can assume that the eigenvalues λ_i are sorted in descending order. The density operator reads

$$\rho_\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0). \quad (36)$$

Then, $\sum_i \lambda_i = 1$ and irrespective of the gauge factors

$$\rho_\Lambda = \sum_{i=1}^r \lambda_i |\omega_i\rangle\langle\omega_i|,$$

Proposition 21 *In a principal window, the expression of the density operator ρ_Λ is independent of the gauge.*

Proof Gauge transformations just change the phases of the Gleason’s vector. Diagonal matrices are not affected. \square

The Hilbert space \mathcal{H} is the direct sum of the eigensubspaces h_k of the density operator ρ_Λ as $\mathcal{H} = \bigoplus_k h_k$. Let A_k denote the orthogonal projector of \mathcal{H} on h_k and let n_e be the number of distinct values of multiplicity d_k . Let α_k be the common eigenvalues λ_i in h_k , ending with $\alpha_{n_e} = 0$, so that $\sum_k n_k \alpha_k = 1$. Then, irrespective of the gauge,

$$\rho_\Lambda = \sum_{k=1}^{n_e} \alpha_k A_k. \quad (37)$$

5.2.1 Reverse Transcription of a Principal Window

The reverse transcription of a principal window leads to a strictly conventional joint probability problem on the principal sample set Ω , with the distribution $\mathbb{P}(\omega_i) = \lambda_i$. As a result, the principal window can immediately be interpreted in terms of standard probability distribution on the Boolean classical states.

Proposition 22 (Principal probability distribution) *A principal window is always regular. Its reverse transcription into a probability space \mathcal{P} leads to a simplicial quantum state $(w_\Lambda, \mathcal{W}_\Lambda)$ describing a strictly classical distribution. The vertices w_i of the simplex \mathcal{W}_Λ are basic vectors in \mathcal{P} , i.e. deterministic states, $w_i = \tilde{\omega}_i, \forall i \in [[1, r]]$ and the principal probability distribution is $\mathbb{P}(\omega_i) = \lambda_i, \forall i \in [[1, d]]$.*

Proof Construct a real-valued d -dimensional probability space \mathcal{P} with basis $\{\tilde{\omega}_i\}$. Let $w_i = \tilde{\omega}_i \in \mathcal{P}$ for $i \in [[1, r]]$, so that the rank of the vector set $\{w_i\}$ is r . Define $\mathcal{W}_\Lambda = \text{conv}(w_i)$ and $w_{\Lambda, \omega_i} = \lambda_i$, so that the working distribution is

$$w_\Lambda = (w_{\Lambda, \omega_i}) = \sum_{i=1}^r \lambda_i \tilde{\omega}_i$$

By inspection, from Eq. (33), the direct transcription of the quantum state $(w_\Lambda, \mathcal{W}_\Lambda)$ is indeed the diagonal operator ρ_Λ . In addition, the rank r of the density operator ρ_Λ is equal to the number of vertices of the simplex \mathcal{W}_Λ , which proves that the system is regular. □

Proposition 23 (Independent binary variables) *A principal window corresponds to a batch of N mutually independent Boolean variables.*

Proof In a principal window, all basic vectors of the probability space are deterministic solutions of the LP problem. Therefore, from Proposition (2) and Sect. 3.5.1, the binary variables are mutually independent. □

From Eq. (37) any projector A_k is a diagonal Hermitian operator with entries 0 or 1 and $\text{Tr}(A_k) = d_k$. It also represents an observable whose reverse transcription is the indicator function of some Boolean function, say A_k . For simplicity, we note A_k both the Boolean function itself in the Boolean algebra and its indicator function in \mathcal{P}^* .

The Principal Bayesian LP Problem Now, we aim to recover the Bayesian system, that is the relevant LP problem in \mathcal{P} .

Proposition 24 *When $r < d$, the principal LP problem can be formulated as*

$$(\Lambda) : \text{ Given } d - r \text{ classical states } \omega_{\mathcal{P}} \text{ assign } \mathbb{P} \text{ subject to } \langle \tilde{\omega}_{\mathcal{P}}^* \rangle = 0. \quad (38)$$

When $r = d$, the prior (Λ) is reduced to the implicit normalization equation.

Proof The r basis vectors $\tilde{\omega}_i$ span the effective probability space $\mathbb{W}_r \subseteq \mathcal{P}$ and the specific simplex \mathcal{W}_λ is the tautological simplex \mathcal{W}_r in \mathbb{W}_r . The r vertices of the simplex in the probability space \mathcal{P} are therefore the basic vectors $\tilde{\omega}_i$ for $i \in [[1, r]]$. Complement the r basis vectors $\tilde{\omega}_i$ by $d - r$ other basis vectors $\tilde{\omega}_i$ in \mathcal{P} . In Eq. (38), $\tilde{\omega}_i^*$ denote the $d - r$ indicator functions corresponding to the classical states ω_i . \square

As a result, the core of any Bayesian system is simply limited to its order r . Consequently, the main actual input is the contextual distribution.

The indicator function $A_{n_e} \in \mathcal{P}^*$ depicts the $d_{n_e} = d - r$ vertices $\tilde{\omega}_i'$ of zero probability, $\lambda_{n_e} = 0$. Taking into account the other contextual multiplicities, let $A_k \in \mathcal{P}^*$ denote the indicator function of the union of all d_k vertices $\tilde{\omega}_i$ corresponding to the same probability α_k . Since the eigenvalues are sorted in descending order, A_k is the indicator function of a set of basic vectors with contiguous indexes, say k_1 to k_2 , with d_k non zero entries, for instance, A_k may be the covector $(0, 0, 0, 1, 1, 1, 1, 0, 0, 0) \in \mathcal{P}^*$.

Now, for all $k \in [[1, n_e]]$, the dual form $\langle A_k p \rangle$ with $p \in \mathcal{P}$ is $\langle A_k p \rangle = \sum_{i=k_1}^{k_2} p_i$ while the expectation $\langle A_k w_\lambda \rangle$ is $\langle A_k \rangle = d_k \alpha_k$. Clearly, the system is invariant under arbitrary permutation of the d_k indexes of same mixed probability α_k . This defines a *contextual symmetry*.

Definition 31 (*Contextual symmetry*) A contextual symmetry is a transformation of the sample set Ω in a principal window, leaving invariant the mixed probability distribution.

Proposition 25 *The contextual symmetry group is the direct product $S_{d_1} \times S_{d_2} \times \dots \times S_{d_{n_e}}$ of the permutation symmetric groups of degree d_k .*

Proof Any product of vertex permutations of same mixed probability α_k is a contextual symmetry by definition. \square

Note that from Proposition (24), strictly speaking, the symmetry group does not depend on the context but on the core LP problem.

5.2.2 Fundamental Theorem

A principal window depicts a conventional probability problem, composed of d *deterministic outcomes* mutually exclusive, namely $\omega_i \in \Omega$ with $i \in [[1, d]]$, and a standard probability distribution, $\Sigma_\lambda = \{\lambda_i\}$, on the sample set Ω . Only $r \leq d$ probability masses λ_i are non-zero.

Theorem 5 (Fundamental theorem) *Any density operator ρ_λ of spectrum $\Sigma_\lambda = \{\lambda_i\}$ in a Hilbert space \mathcal{H} is the image by a unitary channel of a strictly conventional probability problem consisting in drawing one object among d deterministic classical states $\omega_i \in \Omega$ with respect to the contextual probability distribution Σ_λ . The*

Shannon entropy of this conventional problem is the von Neumann entropy of the density operator on \mathcal{H} .

Proof Just diagonalize the density operator. □

In fact, much of this has been known since von Neumann [22] and the only novelty lies in the interpretation: Now, from Theorem (5), *a Bayesian theater represents a quite classical logical system*. In particular, entanglement is a property of the variable batch and not of the problem itself.

Proposition 26 *A pure quantum state depicts a deterministic distribution expressed in the principal sample set.*

Proof By definition, a pure state is of rank 1 and thus deterministic in a principal window. □

This can be expressed in striking form: With a relevant discrete Boolean gauge, a pure state represents just a reset register.

Proposition 27 *It is always possible to choose a discrete Boolean gauge so that the deterministic distribution of a pure state coincides with the empty atom ϖ_0 , that is a reset register composed of N zeros, $(0, 0, \dots, 0)$.*

Proof This is a consequence of the discrete Boolean gauge definition, Definition (2). For each variable required to be valid, just change its gauge so that the variable becomes required to be invalid. □

5.2.3 Information Expressions

In a principal window, three probability distributions are identical: (1) the working distribution w_λ in the sample set Ω , (2) the simplicial distribution μ_i of the contextual distribution in Σ_μ and (3) the distribution λ_i in the spectrum Σ_λ of the density operator ρ_λ .

Entropy Let us recall the definition of the entropy of these different distributions in general.

Definition 32 (*Forms of entropy*)

- The entropy of the working distribution w_λ in a particular window is the *working entropy* $S_w = \mathbb{H}(w)$.
- The entropy of the contextual distribution in a particular window is the *simplicial entropy* $S_\mu = \mathbb{H}(\Sigma_\mu)$. We will use interchangeably the terms “simplicial entropy” and “contextual entropy”.

- The entropy of the Bayesian theater is the *von Neumann entropy* $S_\Lambda = S(\rho_\Lambda) = \mathbb{H}(\Sigma_\Lambda)$. We will use interchangeably the terms “von Neumann entropy” and “mixed entropy”.

The von Neumann entropy $S(\rho_\Lambda)$ is invariant under a unitary channel and can be regarded as the global “theater entropy” while the window entropy S_w and the simplicial entropy S_μ are window-dependent by definition.

Proposition 28 *In a principal window, we have*

$$S_\Lambda = S_\mu = S_w.$$

Proof By simple inspection, in a principal window, the three distributions are identical and therefore the entropies are identical as well. \square

By contrast, in a general window the three entropies are distinct.

Proposition 29 (Jaynes’ inequality) *The von Neumann entropy $S(\rho_\Lambda) = \mathbb{H}(\Sigma_\Lambda)$ is bounded above by the simplicial entropy in any window $S_\mu = \mathbb{H}(\Sigma_\mu)$.*

$$\mathbb{H}(\Sigma_\Lambda) \leq \mathbb{H}(\Sigma_\mu) \quad (39)$$

Proof This inequality is due to Jaynes (Ref. [23], Appendix A). \square

Proposition 30 *The von Neumann entropy is the lower bound of the simplicial entropy over all possible windows.*

$$S_\Lambda = \min_{\text{windows}} (S_\mu).$$

Proof From Jaynes’ inequality, Proposition (29), $S_\Lambda \leq S_\mu$. From Proposition (28), the inequality is saturated in a principal window. \square

The upper bound of the simplicial entropy is trivially $\log r$ when the working distribution coincides with the center of mass of the specific simplex.

At last, it is convenient to define also the overall information, or von Neumann negentropy, as $\mathbb{I}(\rho_\Lambda) = N - S(\rho_\Lambda)$.

Definition 33 (*von Neumann information*) The von Neumann information, or von Neumann negentropy of a density operator ρ_Λ acting on a d -dimensional Hilbert space is $\mathbb{I}(\rho_\Lambda) = N - S(\rho_\Lambda)$, where $d = 2^N$ and $S(\rho_\Lambda) = -\text{Tr}(\rho_\Lambda \log_2 \rho_\Lambda)$.

Theorem 6 *The information content of a N bit Bayesian theater endowed with density operator ρ_Λ is the von Neumann information $N - S(\rho_\Lambda)$*

Proof By reverse transcription in the principal window, this corresponds to the probability distribution $\{\lambda_i\}$. \square

6 Gauge Transcription Group

The gauge group stems from the transcription of a simplicial quantum state (w, \mathcal{W}) from the information domain \mathcal{P} to a geometric domain, \mathcal{H} . This transcription is not uniquely defined, which generates the group, say \mathcal{G} . In the present model, it is thus intrinsic.

By construction, it expresses the exact symmetry of the Hilbert space. Remarkably, it represents the Bayesian prior, except for its contextual distribution.

Since the probability distribution is conserved, from Wigner theorem, gauge operators are either *unitary* or *antiunitary*.

Choosing another transcription is called “changing the gauge”. This can be done either by using a new Hilbert space while keeping the same window (what is called “global gauge”), or by keeping the same Hilbert space but changing the window (which is called “local gauge”).

For simplicity, we will only detail the global gauge.

6.1 Global Gauge

Start from a given simplicial quantum state and consider its transcription into a Hilbert space. The transcription is performed from the source window which determines a particular gauge, say g , and requires a particular Hilbert space, \mathcal{H}_g . Actually, the particular source window itself is widely indifferent because it is straightforward to perform the transcription from any other regular window.

Proposition 31 *For any gauge g , it is possible to construct a unique Hilbert space \mathcal{H}_g irrespective of the regular source window used for the transcription.*

Proof Transcribe the simplicial quantum from a source window. This defines a gauge g and determines both a particular density operator and, by reverse transcription, a particular simplicial quantum state in every window (Fig. 1). Now just decide that some particular regular window is the new source window and that the corresponding density operator is precisely the result of the direct transcription with the same gauge referred to as g . □

Definition 34 (*Global gauge*) A global gauge representation g is the specific transcription of the logical system into a specific Hilbert space \mathcal{H}_g .

6.2 Changing the Global Gauge

Consider a second gauge g' and therefore a new Hilbert space $\mathcal{H}_{g'}$. Let ρ_g and $\rho_{g'}$ denote the density operators acting on \mathcal{H}_g and $\mathcal{H}_{g'}$, respectively. First, make sure that as far as $g' \neq g$, \mathcal{H}_g and $\mathcal{H}_{g'}$ are distinct.

Proposition 32 *When the gauge is global, distinct gauges require distinct Hilbert spaces.*

Proof From Proposition (21), irrespective of the gauge, the density operators are identical in a principal window. If the Hilbert spaces were the same for every gauge, the density operators would be also identical in every windows and the gauges would not be distinct. \square

Proposition 33 *Any change from a gauge g to a gauge g' maps the eigensubspaces of ρ_g onto the eigensubspaces of $\rho_{g'}$.*

Proof Since the expressions of the density operators are identical in the two principal windows, the eigensubspaces are globally invariant. \square

Let $\Theta : \mathcal{H}_g \rightarrow \mathcal{H}_{g'}$ denote a gauge operator. For every window, the bases in the two Hilbert spaces are identical. As a result, when changing the source window, the expression of Θ changes according to the transition matrix.

Proposition 34 *Using another source window Ω' obtained from the initial source window Ω by a unitary transition matrix $U \in U(d)$, the gauge operator $\Theta \in \mathfrak{G}$, whether unitary or antiunitary is expressed as*

$$\Theta' = U\Theta U^{-1} \tag{40}$$

Proof Since the gauge is global, the two bases Ω and Ω' are by hypothesis identical in the two distinct Hilbert spaces \mathcal{H}_g and $\mathcal{H}_{g'}$. As a result, the transition unitary matrices U are also identical. Let $|\psi\rangle_g$ denote a current vector of the Hilbert space \mathcal{H}_g .

$$\begin{array}{ccc} \mathcal{H}_g : & |\psi_g\rangle & \xrightarrow{U} & |\psi'_g\rangle \\ & \Theta \downarrow & & \downarrow \Theta' \\ \mathcal{H}_{g'} : & |\psi_{g'}\rangle & \xrightarrow{U} & |\psi'_{g'}\rangle \end{array}$$

From simple inspection of the commutative diagram we have $|\psi'_g\rangle = U|\psi_g\rangle$ and $|\psi'_{g'}\rangle = U|\psi_{g'}\rangle$ so that, irrespective of $|\psi_g\rangle$, $|\psi'_{g'}\rangle = \Theta'U|\psi_g\rangle = U\Theta|\psi_g\rangle$ and thus $\Theta' = U\Theta U^{-1}$. \square

Since from Proposition (31) the source window is indifferent, it is convenient to select henceforth the source window as a *principal window* corresponding to a batch of mutually independent Boolean variables. The gauge operators Θ can be unitary or antiunitary. Let us start by investigating the set of unitary operators, that is the unitary gauge group.

6.3 The Unitary Gauge Group \mathcal{G}

Obviously, the unitary transformations of a global gauge into another global gauge form a unitary group.

Definition 35 (*Unitary gauge group \mathcal{G}*) The unitary gauge group \mathcal{G} is the unitary transformation group of the global gauges.

Proposition 35 *The unitary group \mathcal{G} is the group of unitary operators leaving invariant the eigensubspaces of the density operator expressed in any particular gauge.*

Proof From Proposition (21), the eigensubspaces of the density operator are invariant under every gauge transformation and conversely, any unitary transformation leaving invariant these eigensubspaces leaves invariant the density operator in any principal window and thus defines a gauge change. □

Constructing the Unitary Gauge Group \mathcal{G} We will hereafter regard the unitary gauge group \mathcal{G} as realized by unitary matrices acting on the d -dimensional Hilbert space \mathcal{H}_{g_0} for an arbitrary but fixed gauge g_0 and expressed in a common principal basis, so that the group is isomorphic to a subgroup of the standard unitary matrix group $U(d)$.

In the principal window, after reordering the basis vectors if necessary, suppose that the eigenvalues λ_i of the density operator ρ_{g_0} are sorted in descending order. Let $|\omega_i\rangle \in \mathcal{H}_{g_0}$ for $i \in [[1, d]]$ denote the basis vectors. The Hilbert space \mathcal{H}_{g_0} is the direct sum of the eigensubspaces \mathfrak{h}_k of the density operator ρ_{g_0} as $\mathcal{H}_{g_0} = \bigoplus_k \mathfrak{h}_k$. Let A_k denote the orthogonal projectors $\mathcal{H}_{g_0} \rightarrow \mathfrak{h}_k \subseteq \mathcal{H}_{g_0}$ and let n_e be the number of distinct eigenvalues α_k of multiplicity d_k , including possibly zero. Then, from Eq. (37),

$$\rho_{g_0} = \sum_{k=1}^{n_e} \alpha_k A_k$$

Proposition 36 *The unitary gauge group \mathcal{G} is a Lie group of dimension $\sum_k d_k^2$ isomorphic to the direct product $U(d_1) \times U(d_2) \times U(d_3) \cdots \times U(d_{n_e})$, where $U(d_k)$ are respectively the unitary groups acting on the d_k -dimensional eigensubspaces \mathfrak{h}_k of the density operator.*

Proof By construction, the Hilbert space \mathcal{H}_{g_0} is a linear representation of dimension d of the gauge group \mathcal{G} . On each subspace \mathfrak{h}_k of dimension d_k , ($k \in [[1, n_e]]$), \mathcal{G} acts as the full unitary group $U(d_k)$ so that any subspace \mathfrak{h}_k is a linear representation of dimension d_k . Finally \mathcal{H}_g is a completely decomposable representation of \mathcal{G} . As a result, each subgroup $U(d_k)$ is normal in \mathcal{G} and \mathcal{G} is the direct product $U(d_1) \times U(d_2) \times U(d_3) \cdots \times U(d_{n_e})$. The dimension of a unitary Lie group $U(d_k)$ is d_k^2 , so that the dimension of the n_e -tuple is $\sum_{k=1}^{n_e} d_k^2$. □

Conversely, the set of eigensubspaces $\{h_k\}$ determines the density operator, up to a possible rescaling of the mixed distribution $\{\alpha_k\}$ leaving the multiplicities unchanged, allowing just a modification of the source contextuality. By contrast, a complete rescaling of the mixed distribution $\{\lambda_i\}$ can e.g. increase the number of eigensubspaces, which would express a break of symmetry.

Proposition 37 *There is a one-to-one correspondence between the unitary gauge subgroups $U(d_k)$ and the intrinsic symmetry subgroups S_k , Definition (31). Moreover, the intrinsic symmetry group is a discrete subgroup of the Lie gauge group.*

Proof The unitary gauge group and the intrinsic symmetry group are both determined by the same set (d_k) of the n_e multiplicities. Moreover, from Proposition (25), the gauge group contains any permutation of the basis vectors in a principal window, leaving invariant the eigensubspaces, that is the intrinsic symmetry group. \square

Reversing the logic, the unitary gauge group \mathcal{G} determines to some extent the density operator. In fact, the group of gauges does not specify the contextual distribution and is equivalent to simply giving the specific simplex, that is the LP system.

Theorem 7 (Correspondence between the unitary gauge group and the quantum state) *The unitary gauge group \mathcal{G} determines the quantum state up to a rescaling of the mixed distribution $\{\alpha_k\}$. Conversely, the quantum state is specified by the set $\{d_k, \alpha_k\}$ with $\sum d_k = d$ and $\sum d_k \alpha_k = 1$ for $k \in [1, n_e]$.*

Proof The gauge groups are direct products of subgroups $U(d_k)$. Therefore the set $\{d_k\}$ is completely determined by \mathcal{G} . The eigenvalues $\{\alpha_k\}$ of the density operator can be arbitrary chosen provided they be positive, distinct and sum to 1 when accounting for the multiplicity. Therefore the quantum state is determined by the set $\{d_k, \alpha_k\}, k \in [1, n_e]$. \square

6.4 Invariant Observables and Noether Constants

By definition, the eigenprojectors A_k are invariant under the gauge group action. Consequently, their corresponding eigenvalues are the Noether constants of the gauge group.

Proposition 38 *The eigenprojectors A_k are invariant under the gauge group and commute with any group operator. They form a commutative POVM of mutually orthogonal observables. By reverse-transcription into any principal window, they are depicted by n_e indicator functions A_k corresponding to the union of the d_k classical states of same mixed probability α_k so that $\langle A_k \rangle = \langle A_k \rangle = d_k \alpha_k$.*

Proof By construction, the group operators leaves invariant the subspaces h_k . The projectors A_k on h_k commute with any group operator and therefore are invariant under the gauge group. They commute and have a common proper window, namely,

any principal window. They sum to the identity, $\sum_{k=1}^{n_c} A_k = \mathbb{1}_d$. Therefore, they form a commutative POVM of orthogonal observables. In a principal window, they are reverse-transcribed as indicator functions A_k . Finally $\langle A_k \rangle = \langle \mathbf{A}_k \rangle = d_k \alpha_k$. \square

Definition 36 (*Invariant observables and Noether constants*) The eigenprojectors A_k constitute a set of invariant observables. The Noether constants $\langle A_k \rangle = d_k \alpha_k$ are the expectation values of these observables.

Now it is possible to reformulate the correspondence between the gauge group and the quantum state, Theorem (7), in terms of these entities.

Theorem 8 *The Bayesian theater is completely determined by the n_e invariant observables A_k and the corresponding Noether constants, namely, the n_e expectations $\langle A_k \rangle = d_k \alpha_k$.*

The unitary gauge group \mathcal{G} does not exhaust all gauge transformations because the antiunitary operators have been omitted. Let us now investigate these antiunitary gauge changes, obtained by complex conjugation $\mathcal{H}_g \rightarrow \mathcal{H}_{g^*}$

6.5 The Conjugation Gauge Group \mathcal{C}

The conjugation group arises from the transcription of a real-valued space into a complex-valued space. Then any transcription and its complex conjugate are equivalent. In other words, conjugation is a gauge transformation.

Let $K : z \mapsto z^*$ denote the standard complex conjugation in \mathbb{C} . Consider the global conjugation gauge $\mathcal{H}_g \rightarrow \mathcal{H}_{g^*}$, obtained by changing each current vector, say $|\psi_g\rangle$, into its complex conjugate $|\psi_{g^*}\rangle = |\psi_g\rangle^*$ in the source window. Let $\mathbb{1}_d \times K$, or simply K when no confusion can occur, denote the diagonal matrix $\text{Diag}(K, K, \dots, K)$. Now from a theorem by E. Wigner [24], any antiunitary operator is of the form UK where U is unitary.

Proposition 39 *In a principal window, any antiunitary gauge operator Θ is the product GK of a unitary gauge operator $G \in \mathcal{G}$ by the matrix $\mathbb{1}_d \times K$.*

Proof Let C denote a conjugation gauge operator. Then $C = GK$ where G is unitary [24]. In a principal window the density operator ρ is real and invariant by any gauge operator. Therefore G is a unitary gauge operator. \square

It is possible to select for definiteness the initial conjugation operator in the principal source window as $C = \mathbb{1}_d \times K$. Let us term this matrix “conjugation gauge operator”.

Definition 37 (*Conjugation gauge operator C*) The conjugation operator C is expressed in a principal source window Ω by the matrix $\mathbb{1}_d \times K$ so that in this window

$$\mathbb{C} : \mathcal{H}_g \rightarrow \mathcal{H}_{g^*} \quad : \quad |\psi_g\rangle \mapsto |\psi_{g^*}\rangle = K|\psi_g\rangle = |\psi_g\rangle^*$$

Definition 38 (*Conjugation gauge group* \mathcal{C}) The conjugation gauge group is the involutive group $\mathcal{C} = \{\mathbb{1}_d, \mathbb{C}\}$

Finally, the full gauge group \mathfrak{G} is the semi-direct product of the discrete conjugation group \mathcal{C} and the continuous unitary group \mathcal{G} .

$$\mathfrak{G} = \mathcal{C} \rtimes \mathcal{G}$$

Conjugation expresses exotic symmetry of the Hilbert space.

7 Measurement

In this section, we recover the concept of measurement of standard quantum information, notably by using observables or positive operator value measurements (POVM). From Theorem (6), a Bayesian theater in a state ρ contains $N - S(\rho)$ information bits. The issue is to extract all or part of this information.

7.1 Observable Measurement

Let ρ denote the density operator and Q an observable, that is a Hermitian operator acting on the Hilbert space. From the Born rule, regardless of the density matrix and whatever the observable, its expectation is

$$\langle Q \rangle = \text{Tr}(\rho Q).$$

As such, this result does not directly provide strict information (i.e. in bits). To interpret it in terms of extracting information when possible, it is necessary to include the measurement in a positive operator value measurement (POVM), as discussed in the following sections.

7.2 POVM Measurement

The general measurement described in the source observation window, Sect. (3.4.3) can be equivalently reproduced in full generality in the Hilbert space. For commutative observables the measurement estimates the probability of outcomes collected from a unique viewpoint on the register. By contrast, for non commutative observables, the measurement estimates the probability of outcomes collected from different viewpoints.

Theorem 9 (General measurement) *General POVMs can be performed regardless of the density matrix and whatever the positive observables.*

Proof Just use the Born's rule. □

Let Γ be a finite set. Consider a resolution of the identity in \mathcal{H} described by a set of positive Hermitian operators $\{Q_\gamma\}_{\gamma \in \Gamma}$, not necessarily commutative nor diagonal in the current window, such that

$$Q_\gamma \geq 0; \quad \sum_{\gamma \in \Gamma} Q_\gamma = \mathbb{1}_d$$

The expectation of Q_γ is $\langle Q_\gamma \rangle = \text{Tr}(\rho Q_\gamma) \leq 1$.

By linearity, we have

$$\sum_{\gamma \in \Gamma} \langle Q_\gamma \rangle = \text{Tr} \left(\rho \times \sum_{\gamma \in \Gamma} Q_\gamma \right) = \text{Tr}(\rho) = 1.$$

Therefore, the array $\langle Q_\gamma \rangle$ is a probability distribution.

Definition 39 (*POVM probability distribution*) A POVM $\{Q_\gamma\}_{\gamma \in \Gamma}$ defines a probability distribution $p(\gamma) \stackrel{\text{(def)}}{=} \langle Q_\gamma \rangle$.

Let us address an information interpretation.

7.3 POVM Information Gain

From Theorem (6), a Bayesian theater in a state ρ contains $N - S(\rho)$ information bits. The POVM $\{Q_\gamma\}_{\gamma \in \Gamma}$ can be used to extract a fraction of this information. Practically, we will use the conventional *relative entropy* which depicts the gain of information between two states. The first state is the *completely random* distribution, $\rho_{\text{void}} = (1/d) \times \mathbb{1}_d$, with $d = 2^N$, representing an absence of information. The corresponding POVM probability distribution p_{void} reads

$$p_{\text{void}}(\gamma) = \text{Tr}(\rho_{\text{void}} Q_\gamma) = \frac{1}{d} \times \text{Tr}(Q_\gamma) = \frac{q_\gamma}{d} \leq 1 \tag{41}$$

where $q_\gamma = \text{Tr}(Q_\gamma)$. The second state is the current density operator ρ . Now, the ‘‘POVM information gain’’ $\mathbb{I}(\rho \parallel \Gamma)$ provided by the POVM (Γ) is the relative entropy $\mathbb{H}(p \parallel p_{\text{void}})$.

Definition 40 (*POVM information gain*) The POVM information gain $\mathbb{I}(\rho \parallel \Gamma)$ is the relative entropy $\mathbb{H}(p \parallel p_{\text{void}})$,

$$\mathbb{I}(\rho \parallel \Gamma) \stackrel{\text{(def)}}{=} \mathbb{H}(p \parallel p_{\text{void}}) = \sum_{\gamma \in \Gamma} p(\gamma) \log_2 \frac{p(\gamma)}{p_{\text{void}}(\gamma)} \tag{42}$$

If $\rho = \rho_{\text{void}}$ then $p = p_{\text{void}}$ and $\mathbb{I}(\rho \parallel \Gamma) = \mathbb{H}(p_{\text{void}} \parallel p_{\text{void}}) = 0$. The information gain of a completely random state is well zero.

Proposition 40 *The information gain $\mathbb{I}(\rho||\Gamma)$ is non-negative and less than the storage capacity N of the register and even of the total information $N - S(\rho)$ currently stored in the Bayesian theater.*

Proof A relative entropy is non-negative, so that so is the information gain $\mathbb{I}(\rho||\Gamma)$. It is trivially less than the storage capacity N of the register and even of the total information $N - S(\rho)$ currently stored in the Bayesian theater. Using Eq. (41) and $d = 2^N$

$$0 \leq \mathbb{I}(\rho||\Gamma) = N + \sum_{\gamma \in \Gamma} p(\gamma) \log_2 \frac{p(\gamma)}{q_\gamma} \leq N - S(\rho) \leq N \quad (43)$$

□

Information systems are generally characterized by their *entropy*, \mathbb{H} , which corresponds to a lack of information. Therefore, the concept of *information*, \mathbb{I} , defined in a Bayesian theater of N bits corresponds to an entropy of $\mathbb{H} \stackrel{\text{(def)}}{=} N - \mathbb{I}$. With this convention, $\mathbb{I}(\rho||\Gamma)$ corresponds to the entropy $\mathbb{H}(\Gamma)$. From Eq. (43) we obtain

$$\mathbb{H}(\Gamma) \stackrel{\text{(def)}}{=} N - \mathbb{I}(\rho||\Gamma) = \sum_{\gamma \in \Gamma} -p(\gamma) \log_2 \frac{p(\gamma)}{q_\gamma} \geq S(\rho) \geq 0 \quad (44)$$

In general $\mathbb{H}(\Gamma) \neq \mathbb{H}(p_\gamma)$ because $q_\gamma = \text{Tr}(Q_\gamma) \neq 1$.

Definition 41 (*POVM entropy*) The POVM entropy is the entropy $\mathbb{H}(\Gamma) \stackrel{\text{(def)}}{=} N - \mathbb{I}(\rho||\Gamma)$ corresponding to the information gain $\mathbb{I}(\rho||\Gamma)$ that can be extracted by the POVM.

In particular, assume that the POVM corresponds to a von Neumann measurement in a particular window of sample set $\Omega = \{\omega\}$ (see above Sect. 3.4.2). In the Hilbert space, let $|\omega\rangle$ be the basis of this window. Then, $\Gamma = \Omega$ and $Q_\omega = |\omega\rangle\langle\omega|$ so that $q_\omega = 1$ and $p(\omega) = \langle\omega|\rho|\omega\rangle$.

Definition 42 (*Window entropy*) The window entropy $\mathbb{H}(\Omega)$ is the entropy $N - \mathbb{I}(\rho||\Omega)$ corresponding to the information that can be extracted by a von Neumann measurement in the window.

Proposition 41 *The window entropy is the working entropy of the current window.*

$$\mathbb{H}(\Omega) = \mathbb{H}(w_\Lambda) \quad (45)$$

Proof Proceed to the reverse transcription of the current window into its probability space. Let w_Λ denote the working distribution, Definition (13). By definition, $p = w_\Lambda$. □

In standard quantum information, a POVM is called “information-complete” when the operators Q_γ , $\gamma \in \Gamma$ span the complete space $\mathcal{L}(\mathcal{H})$. Indeed, such a

measurement provides $|\Gamma| \geq d^2 - 1$ coefficients $p(\gamma)$ that allow the unique reconstruction of the density operator ρ and then the Bayesian probability distribution. This does not necessarily mean that the POVM entropy is equal to $S(\rho)$ because this information is encoded in a particular way, which can cause a bias not taken into account in Eq. (43) and then a loss of information (or an increase of entropy).

In general, a particular measurement is not information-complete and therefore the determination of the density operator requires independent measurements from additional POVMs.

7.4 Independent POVMs

Suppose that a POVM $\{Q_\gamma\}_{\gamma \in \Gamma}$, that we will refer to as (Γ) , is information-*incomplete* and consider the possibility to complement this POVM by another POVM.

The set of density operators $D(\mathcal{H}) \subset \mathcal{L}(\mathcal{H}) = \{\rho\}$ is a convex ensemble located in an *affine* subspace of real dimension $d^2 - 1$. Motivated by Ref. [25], it is helpful to consider rather the set of traceless Hermitian operators, $\{\mathbf{e}\}$ defined as

$$\mathbf{e} = \rho - \frac{1}{d} \mathbb{1}_d,$$

because this ensemble generates a *linear* vector space $\mathcal{E} \stackrel{\text{(def)}}{=} \text{Span}(\mathbf{e}) \subset \mathcal{L}(\mathcal{H})$ still of dimension $d^2 - 1$. This mapping $D(\mathcal{H}) \rightarrow \mathcal{E}$ can be extended to all operators of a POVM as follows. Consider the POVM (Γ) , $\{Q_\gamma\}_{\gamma \in \Gamma}$ and define $Q_\gamma \mapsto \mathbf{e}_\gamma$ as

$$q_\gamma = \text{Tr}(Q_\gamma) > 0 \quad ; \quad E_\gamma = \frac{1}{q_\gamma} Q_\gamma \in D(\mathcal{H}) \quad ; \quad \mathbf{e}_\gamma = E_\gamma - \frac{1}{d} \mathbb{1}_d \in \mathcal{E} \quad (46)$$

The POVM is then characterized by

$$\sum_{\gamma \in \Gamma} q_\gamma = d \quad ; \quad \sum_{\gamma \in \Gamma} q_\gamma E_\gamma = \mathbb{1}_d \quad ; \quad \sum_{\gamma \in \Gamma} q_\gamma \mathbf{e}_\gamma = 0 \quad (47)$$

At last, define a Hermitian inner product in \mathcal{E} as

$$\langle \mathbf{e}_1 \cdot \mathbf{e}_2 \rangle \stackrel{\text{(def)}}{=} \text{Tr}(\mathbf{e}_1^\dagger \mathbf{e}_2). \quad (48)$$

Let $Q < \mathbb{1}_d$ be an additional Hermitian positive operator. Let $q = \text{Tr}(Q) > 0$, $E_Q = (1/q)Q \in D(\mathcal{H})$ and $\mathbf{e}_Q = E_Q - (1/d)\mathbb{1}_d \in \mathcal{E}$. It turns out that Q is independent of the POVM if and only if \mathbf{e}_Q is orthogonal to every \mathbf{e}_γ . Indeed, assume that \mathbf{e}_Q is orthogonal to the subspace $\text{Span}\{\mathbf{e}_\gamma\}_{\gamma \in \Gamma} \subseteq \mathcal{E}$. We compute easily from Eqs. (46-48)

$$\forall \gamma \in \Gamma : \quad \langle \mathbf{e}_Q \cdot \mathbf{e}_\gamma \rangle = 0 \quad \iff \quad \frac{1}{qq_\lambda} \text{Tr}(QQ_\gamma) - \frac{1}{d} = 0$$

We have then

$$\forall \gamma \in \Gamma \quad \text{Tr}(Q Q_\gamma) = \frac{\text{Tr}(Q)\text{Tr}(Q_\gamma)}{d} \tag{49}$$

Conversely, if Eq. (49) holds, then \mathbf{e}_Q is orthogonal to every \mathbf{e}_γ .

To check the independence of the additional operator Q , construct a second POVM with two operators, $\{Q, \mathbb{1}_d - Q\}$. Assume that the system “lives” in the first POVM set, meaning that $\rho = \rho_r \in \text{Span}(Q_\gamma)_{\gamma \in \Gamma}$. Then, from linearity, Eq. (49) and $\text{Tr}(\rho_r) = 1$, the second measurement yields

$$p(Q) = \text{Tr}(\rho_r Q) = \frac{\text{Tr}(Q)\text{Tr}(\rho_r)}{d} = \frac{\text{Tr}(Q)}{d} = \text{Tr}\left(\frac{\mathbb{1}_d}{d} \times Q\right) \quad ; \quad p(\mathbb{1}_d - Q) = 1 - p(Q)$$

exhibiting the effective density operator $\rho_{\text{void}} = \mathbb{1}_d/d$ of a completely random system. Therefore $p(Q)$ is totally independent of the density matrix $\rho_r \in \text{Span}(Q_\gamma)_{\gamma \in \Gamma}$. Similarly, if the system lives in the second POVM set, $\rho = \rho_Q \in \text{Span}(Q, \mathbb{1}_d - Q)$ then the first POV-measurement yields

$$p(Q_\gamma) = \text{Tr}(\rho_Q Q_\gamma) = \text{Tr}\left(\frac{\mathbb{1}_d}{d} \times Q_\gamma\right)$$

and again the coefficients $p(Q_\gamma)$ are totally independent of the density matrix ρ_Q . We will refer to the two POVMs as mutually “independent”. More generally, consider two distinct POVMs, $\{Q_{\gamma_1}\}_{\gamma_1 \in \Gamma_1}$ and $\{Q_{\gamma_2}\}_{\gamma_2 \in \Gamma_2}$. For brevity, we say that a system defined by a density operator $\rho \in \mathcal{L}(\mathcal{H})$ “lives” in a POVM $\{Q_\gamma\}_{\gamma \in \Gamma}$ when $\rho \in \text{Span}\{Q_\gamma\}_{\gamma \in \Gamma}$.

Definition 43 (*Independent POVMs*) Two distinct POVMs, $\{Q_{\gamma_1}\}_{\gamma_1 \in \Gamma_1}$ and $\{Q_{\gamma_2}\}_{\gamma_2 \in \Gamma_2}$ are mutually independent if the measurement with one POVM when the system “lives” in the other POVM is identical to a measurement in a completely random state $\rho_{\text{void}} = \mathbb{1}_d/d$.

Proposition 42 *Two distinct POVMs, $\{Q_{\gamma_1}\}_{\gamma_1 \in \Gamma_1}$ and $\{Q_{\gamma_2}\}_{\gamma_2 \in \Gamma_2}$ are mutually independent if and only if*

$$\forall \gamma_1 \in \Gamma_1, \quad \forall \gamma_2 \in \Gamma_2 : \quad \text{Tr}(Q_{\gamma_1} Q_{\gamma_2}) = \frac{\text{Tr}(Q_{\gamma_1})\text{Tr}(Q_{\gamma_2})}{d} \tag{50}$$

Proof From Eq. (49) each \mathbf{e}_{γ_i} is orthogonal to every $\mathbf{e}_{\gamma_{3-i}}$ ($i = 1, 2$). □

Now, given that the two POVMs are independent, the information gains provided by the two measurements do not overlap. As a result the sum of the two information gains is still bounded by the total information, $N - S(\rho)$, stored in the system.

Proposition 43 (POVM entropic inequality) *Let $\Gamma_1 : \{Q_{\gamma_1}\}_{\gamma_1 \in \Gamma_1}$ and $\Gamma_2 : \{Q_{\gamma_2}\}_{\gamma_2 \in \Gamma_2}$ be two independent POVMs acting on a system in the state ρ . Then*

$$\mathbb{H}(\Gamma_1) + \mathbb{H}(\Gamma_2) \geq N + S(\rho) \geq N \tag{51}$$

Proof Proceed to the transformations $\mathbf{e} = \rho - \mathbb{1}_d/d$, $q_{\gamma_i} = \text{Tr}(Q_{\gamma_i})$, $E_{\gamma_i} = (1/q_{\gamma_i})Q_{\gamma_i} \in \mathcal{D}(\mathcal{H})$ and $\mathbf{e}_{\gamma_i} = E_{\gamma_i} - (1/d_{\gamma_i})\mathbb{1}_d \in \mathcal{E}$, where $i \in [[1, 2]]$ and $\gamma_i \in \Gamma_i$. Let $\mathcal{E}_i = \text{Span}_{\gamma_i \in \Gamma_i}(\mathbf{e}_{\gamma_i})$. The space \mathcal{E} splits into three mutually orthogonal subspaces, $\mathcal{E} = \mathcal{E}_1 \oplus \mathcal{E}_2 \oplus \mathcal{E}_0$. As a result, we have a unique decomposition $\mathbf{e} = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_0$. Define $\rho_i = \mathbf{e}_i + \mathbb{1}_d/d$. Then, still for $i \in [[1, 2]]$ and $\forall \gamma_i \in \Gamma_i$ we obtain successively by a straightforward computation

$$\begin{aligned} \langle \mathbf{e} \cdot \mathbf{e}_{\gamma_i} \rangle &= \langle (\mathbf{e}_0 + \mathbf{e}_1 + \mathbf{e}_2) \cdot \mathbf{e}_{\gamma_i} \rangle = \langle \mathbf{e}_i \cdot \mathbf{e}_{\gamma_i} \rangle \\ \text{Tr}\left[\left(\rho - \frac{\mathbb{1}_d}{d}\right)\left(\frac{Q_{\gamma_i}}{q_{\gamma_i}} - \frac{\mathbb{1}_d}{d}\right)\right] &= \text{Tr}\left[\left(\rho_i - \frac{\mathbb{1}_d}{d}\right)\left(\frac{Q_{\gamma_i}}{q_{\gamma_i}} - \frac{\mathbb{1}_d}{d}\right)\right] \\ \text{Tr}(\rho Q_{\gamma_i}) &= \text{Tr}(\rho_i Q_{\gamma_i}). \end{aligned}$$

so that $p(\gamma_i) = \text{Tr}(\rho Q_{\gamma_i})$ depends only on ρ_i . Therefore, the two information gains $\mathbb{I}_1 = \mathbb{I}(\rho \parallel \Gamma_1)$ and $\mathbb{I}_2 = \mathbb{I}(\rho \parallel \Gamma_2)$ are independent and the total information extracted by the two POVMs is the sum of the two information gains. This sum is trivially bounded by the storage capacity N of the register, and even by the actual information stored in the register $N - S(\rho)$, i.e. $\mathbb{I}_1 + \mathbb{I}_2 \leq N - S(\rho) \leq N$. In terms of entropy, $\mathbb{H}(\Gamma_i) = N - \mathbb{I}_i$, we obtain Eq. (51). \square

To our knowledge, the POVM inequality, Eq. (51), is new but the concept of “unbiased POVM” was previously defined by Kalev and Gour [26]. In standard quantum information, the inequality is rather expressed for von Neumann measurements. Independent POVMs are then particularized by independent von Neumann measurements in the so called “mutually unbiased bases”.

7.5 Mutually Unbiased Bases (MUB)

Mutually unbiased bases, first introduced by J. Swinger in 1960 [27] are extensively used in standard quantum information [25]. Let us first define precisely a pair of mutually unbiased bases Ω_1 and Ω_2 in the present model. Each basis Ω_i , of basic vectors $|\omega_i\rangle$, ($\omega_i \in \Omega_i$), ($i \in [[1, 2]]$), defines a von Neumann measurement i.e. a particular POVM, namely $\{|\omega_i\rangle\langle\omega_i|\}_{\omega_i \in \Omega_i}$.

Definition 44 (*Mutually unbiased bases (MUB) or mutually unbiased windows*) A pair of bases are mutually unbiased when they determine two independent von Neumann measurements.

This definition is not standard. Let us recover the standard definition by the following proposition:

Proposition 44 (MUB) *Two distinct windows of sample sets Ω_1 and Ω_2 and of basic vectors $|\omega_1\rangle$, ($\omega_1 \in \Omega_1$) and $|\omega_2\rangle$, ($\omega_2 \in \Omega_2$) are mutually unbiased if and only if*

$$\forall \omega_1 \in \Omega_1, \forall \omega_2 \in \Omega_2 \quad : \quad |\langle \omega_1 | \omega_2 \rangle|^2 = \frac{1}{d}. \quad (52)$$

Proof From Eq. (50) two von Neumann measurements are independent if and only if Eq. (52) holds. \square

Consider a pair of mutually unbiased bases, defining two independent von Neumann measurements. Then, Eq. (51) holds, with Ω_i standing for Γ_i , as

$$\mathbb{H}(\Omega_1) + \mathbb{H}(\Omega_2) \geq N + S(\rho) \geq N \quad (53)$$

We recover the well known entropic relations of standard quantum information theory in the special case of independent windows. The first bound, $N + S(\rho)$, corresponds to a special case of the Frank-Lieb's inequality [15] and the second bound, N , to the less tight Massen-Uffink's inequality [14]. Note the the present model provides an intuitive basis to these inequalities, usually regarded as somewhat technical and esoteric: Namely, whatever the measurement, it is impossible to recover more information than is stored in the memory. This is also the basis of the iconic uncertainty principle.

Finally, the totality of the information stored in the memory can be retrieved. Actually, this information is located in the principal window and is therefore perfectly classical.

8 Illustrative Example: One-Bit System

A one-bit system is a logical memory containing a maximum of one information bit. The issue is to extract this information.

There is not much to say about a classic one-bit system, it is a deterministic problem represented by a single Boolean variable, valid or invalid. Its entropy is zero and its Shannon information is therefore exactly 1 bit.

In contrast, a one-bit Bayesian system is a memory whose information content is specified by the state of the system. This state can be either a mixed state or a pure state, also called a qubit. It turns out that a qubit is actually a deterministic one-bit system, implemented by Bayesian inference.

8.1 Mixed One-Bit System

A mixed state corresponds to a one-bit memory without any particular constraint. The memory is investigated *via* a question-and-answer procedure. Let X_1 denote the Boolean variable representing the initial query. A remarkable property of the Bayesian representation is that this initial query only exposes one possible aspect of the system. At any rate, there are two classical states, namely, $\omega_1 = \bar{X}_1$ and $\omega_2 = X_1$, where \bar{X}_1 is the negation of X_1 . At this stage, there is no specific condition, so that the Bayesian prior information, say (Λ) , is initially void.

8.1.1 Source Observation Window

To address the Bayesian problem, we first construct a sample set $\Omega = \{\omega_1, \omega_2\}$ including the two initial classical states. Next, we build a real-valued probability space $\mathcal{P} = \text{Span}(\omega_1, \omega_2)$, of dimension $d = 2^N = 2$. This provides an initial view, called “source information window”.

Let $p = (p_1, p_2) \in \mathcal{P}$. To secure p_1, p_2 to be genuine probabilities, we have to introduce in the (initially void) prior (Λ) the relevant universal equations, Eqs. (2, 3, 4, etc.), limited here to the sole normalization equation. Therefore, the prior information (Λ) now reads $(\Lambda) : p_1 + p_2 = 1$. We will interpret p_1 and p_2 as conditional probabilities given (Λ) . Using the convention of Sect. (2.3)

$$p_1 = \mathbb{P}(-1) \stackrel{\text{(def)}}{=} \mathbb{P}(X_1 = 0|\Lambda) \quad \text{and} \quad p_2 = \mathbb{P}(1) \stackrel{\text{(def)}}{=} \mathbb{P}(X_1 = 1|\Lambda).$$

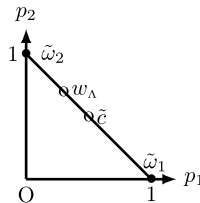
As a result, the unknowns p_1 and p_2 are the solutions of a LP system Eq. (7)

$$\begin{aligned} p_1 + p_2 &= 1 \\ \text{subject to } p &\geq 0 \end{aligned} \tag{54}$$

Each solution is a particular probability distribution $\mathbb{P}(\omega)$ on the sample set Ω . Actually, the normalization equation is implicit, so that the Bayesian formulation reduces to its simplest expression without any explicit constraint,

$$(\Lambda) : \quad \text{Assign a probability distribution } \mathbb{P}(\omega) \text{ on } \Omega = \{\omega_1, \omega_2\}$$

Let $\tilde{\omega}_1 = (1, 0)$ and $\tilde{\omega}_2 = (0, 1)$ denote the two deterministic solutions in \mathcal{P} . The LP system,



Equation (54), accepts not only the two classical deterministic distributions $\tilde{\omega}_1$ and $\tilde{\omega}_2$ but also a continuous set of solutions on their convex hull. The feasible solutions are located on a specific polytope \mathcal{W}_Λ , that is the line segment $[\tilde{\omega}_1, \tilde{\omega}_2]$ identical to the so-called “tautological simplex” of one variable \mathcal{W}_1 . The line itself is an affine 1-dimensional subspace P_Λ . The simplex vertices are $w_1 = \tilde{\omega}_1$ and $w_2 = \tilde{\omega}_2$. Therefore, the system is simplicial (Definition 9) and $\mathcal{W}_\Lambda = \mathcal{W}_1 = \text{conv}(\tilde{\omega}_1, \tilde{\omega}_2)$.

8.1.2 Simplicial Quantum State

The system, Eq. (54) defines a “mixed state” of rank $r = 2$ (i.e. with 2 vertices). The specific polytope \mathcal{W}_Λ is the tautological simplex. It is possible to single up a particular

solution, w_λ , called “working distribution”, by assigning a weight to each vertex of the simplex that is a discrete *contextual probability distribution*. Define

$$\Sigma_\lambda = \{\lambda_1, \lambda_2\} \quad \text{where } \lambda_1, \lambda_2 \geq 0 \quad \text{and } \lambda_1 + \lambda_2 = 1,$$

so that $w_\lambda = \lambda_1 \tilde{\omega}_1 + \lambda_2 \tilde{\omega}_2 \in \mathcal{W}_\lambda$. By convention, when the contextual distribution is not explicit, the default working distribution w_λ is the center of mass of the polytope, i.e. $\tilde{c} = (1/2)(\tilde{\omega}_1 + \tilde{\omega}_2)$.

The pair, $(w_\lambda, \mathcal{W}_\lambda)$, of a working distribution w_λ and a specific simplex, \mathcal{W}_λ , is termed “*simplicial quantum state*”.

8.1.3 Observable

In the source window, an observable is a function $Q : \Omega \rightarrow \mathbb{R}$. Let $Q(\omega) = q_\omega$. By definition, $q = (q_1, q_2) \in \mathcal{P}^*$, where \mathcal{P}^* is the dual of \mathcal{P} . The expectation of Q is defined with respect to the working distribution as

$$\langle Q \rangle = \langle qp \rangle|_{p=w_\lambda} = \langle qw_\lambda \rangle = \lambda_1 q_{\omega_1} + \lambda_2 q_{\omega_2}.$$

For instance, consider the particular observable $S_Z(\omega) = s_\omega$ defined as

$$S_Z(\omega_1) = 1 \quad ; \quad S_Z(\omega_2) = -1 \quad \text{i.e. } s = (1, -1) \in \mathcal{P}^*$$

We have

$$\langle S_Z \rangle = \langle sw_\lambda \rangle = \lambda_1 - \lambda_2$$

8.1.4 Entropy and Information

The entropy of the state is $\mathbb{H}(\Sigma_\lambda) = -\lambda_1 \log_2 \lambda_1 - \lambda_2 \log_2 \lambda_2$. This corresponds at the same time to the von Neumann entropy, the simplicial entropy and the working entropy and represents a Shannon information of $I = 1 - -\lambda_1 \log_2 \lambda_1$. Actually since the window is principal, this information can be entirely extracted from the window.

8.1.5 Other Observation Windows

We started from a unique Boolean variable, X_1 . Surprisingly enough, in a Bayesian framework, it is possible to pose the problem by using *other alternatives than X_1 and \bar{X}_1* , that is to consider *other observation windows*. These new alternatives are necessary in order to compute the expectation value of every relevant observable.

To change the observation window, let us introduce a new tool, namely, an auxiliary Hilbert space.

8.1.6 Transcription into an Auxiliary Hilbert Space \mathcal{H}

Indeed, to construct new observation windows, the fundamental novelty of quantum information is to transcribe the source window into an auxiliary Hilbert space \mathcal{H}

defined as the *complex* span of (ω_1, ω_2) . Let $(|1\rangle, |2\rangle)$ denote its basis vectors. Afterward, the new alternatives will be simply computed by changing this initial basis. The initial simplicial quantum state is transcribed as a standard *density operator* $\rho_\Lambda = \lambda_1|1\rangle\langle 1| + \lambda_2|2\rangle\langle 2|$, or

$$\rho_\Lambda = \lambda_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

The density operator is diagonal. As a result this source window is called *principal*.

Any observable in the source window $Q : \Omega \rightarrow \mathbb{R}$ is transcribed as the following diagonal operator

$$Q = \begin{bmatrix} q_{\omega_1} & 0 \\ 0 & q_{\omega_2} \end{bmatrix}$$

For instance, the observable S_Z is transcribed as

$$S_Z = \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{55}$$

where σ_3 is a Pauli matrix. Its expectation is computed by the so-called ‘‘Born rule’’, Eq. (35), as

$$\langle S_Z \rangle = \langle sw_\Lambda \rangle^{(\text{def})} = \text{Tr}(\rho_\Lambda S_Z) = \lambda_1 - \lambda_2$$

8.1.7 Changing the Window

To obtain new windows, we simply have to change the basis in \mathcal{H} . It turns out that the new corresponding probability problem can be simply retrieved by reverse transcription in the new basis. In general, the new density operator is no longer diagonal in the new basis, so that the new observation window is not principal but *twisted*. Let $|e_{i'}\rangle = (\alpha_{i',1}, \alpha_{i',2})^T$ for $i' = 1, 2$ be the expression of its eigenvectors in the new basis. The new expression $\rho'_\Lambda = \lambda_1|e_{1'}\rangle\langle e_{1'}| + \lambda_2|e_{2'}\rangle\langle e_{2'}|$ of the density operator is thus

$$\rho'_\Lambda = \lambda_1 \begin{bmatrix} \alpha_{1',1} \alpha_{1',1}^* & \alpha_{1',1} \alpha_{1',2}^* \\ \alpha_{1',2} \alpha_{1',1}^* & \alpha_{1',2} \alpha_{1',2}^* \end{bmatrix} + \lambda_2 \begin{bmatrix} \alpha_{2',1} \alpha_{2',1}^* & \alpha_{2',1} \alpha_{2',2}^* \\ \alpha_{2',2} \alpha_{2',1}^* & \alpha_{2',2} \alpha_{2',2}^* \end{bmatrix} = \begin{bmatrix} w'_1 & \rho'_{12} \\ \rho'_{21} & w'_2 \end{bmatrix}$$

and we have $\text{Tr}(\rho'_\Lambda) = \text{Tr}(\rho_\Lambda) = w'_1 + w'_2 = 1$. For example for $|e_{1'}\rangle = (\cos \theta, \sin \theta)^T$ and $|e_{2'}\rangle = (-\sin \theta, \cos \theta)^T$, we obtain

$$\rho'_\Lambda = \begin{bmatrix} \lambda_1 \cos^2 \theta + \lambda_2 \sin^2 \theta & (\lambda_1 - \lambda_2) \sin \theta \cos \theta \\ (\lambda_1 - \lambda_2) \sin \theta \cos \theta & \lambda_1 \sin^2 \theta + \lambda_2 \cos^2 \theta \end{bmatrix}.$$

To reverse transcribe into a new real-valued probability space \mathcal{P}' , use the eigenvectors $|e_{i'}\rangle \in \mathcal{H}$ to define the vectors $v'_i = (|\alpha_{i',1}|^2, |\alpha_{i',2}|^2)^T$ in \mathcal{P}' . In the example,

$w'_1 = \lambda_1 \cos^2 \theta + \lambda_2 \sin^2 \theta$, $w'_2 = \lambda_1 \sin^2 \theta + \lambda_2 \cos^2 \theta$, $v'_1 = (\cos^2 \theta, \sin^2 \theta)^T$ and $v'_2 = (\sin^2 \theta, \cos^2 \theta)^T$.

By exception, when $v'_1 = v'_2$, the new window is “blind” and $w'_1 = w'_2 = 1/2$. For example, this is obtained for $\theta = \pi/4$, $|e_{1'}\rangle = (1/\sqrt{2})(1, 1)$ and $|e_{2'}\rangle = (1/\sqrt{2})(-1, 1)$.

Otherwise, the new simplex is the affine segment $[v'_1, v'_2]$ and the new working distribution is $w' = (w'_1, w'_2)^T$. Finally, this defines a new sampling set Ω' . Although the system is basically classical, Bayesian inference leads to a *twisted* observation window because the basis vectors are correlated and no longer independent.

Obviously, the old observables $\Omega \rightarrow \mathbb{R}$, represented by diagonal Hermitian operators in \mathcal{H} will change accordingly but will no longer remain diagonal. Therefore, they *cannot* be reverse-transcribed in the new window because they are still defined on $\Omega \neq \Omega'$. By contrast, the new window matches different observables, inaccessible from the old window, $\Omega' \rightarrow \mathbb{R}$ which became diagonal in the new window. Nevertheless, all observables can always be computed in the Hilbert space from any observation window because each observable is expressed as an operator, whether diagonal or not, acting on the Hilbert space \mathcal{H} .

8.2 Qubit, Pure One-Bit State

We define a qubit as a pure state in a 1-bit LP system. As a result, a qubit represents a deterministic distribution. In a principal window, the qubit is simply a classical state!

Assume now that the source window is not necessarily principal. Define a covector $a_\theta = (a_{\theta,\omega_1}, a_{\theta,\omega_2})$ in \mathcal{P}^* depending on a setting θ associated with an observable, A_θ , so that $A_\theta(p) = a_{\theta,\omega_1} \times p_1 + a_{\theta,\omega_2} \times p_2$. Without loss in generality for feasible LP problems, we can choose the following formulation of a_θ

$$a_\theta = (a_{\theta,\omega_1}, a_{\theta,\omega_2}) = (\sin^2 \theta/2, -\cos^2 \theta/2),$$

The qubit is the unique solution of the Bayesian problem Eq. (8)

$$(\theta) : \quad \text{Assign } \mathbb{P} \text{ subject to } \langle A_\theta \rangle = 0$$

The rank of the LP system is $m = d = 2$ and the solution is $w_\theta = (\cos^2 \theta/2, \sin^2 \theta/2)$. The quantum state $(w_\theta, \mathcal{W}_\theta)$ is thus characterized by the isolated vertex w_θ and $\mathcal{W}_\theta = \{w_\theta\}$.

8.2.1 Observable

Consider an observable $Q(\omega) = q_\omega$ in the source window. The quantum expectation is defined as,

$$\langle Q \rangle = \langle qw_\theta \rangle = q_{\omega_1} w_{\theta,1} + q_{\omega_2} w_{\theta,2} = q_{\omega_1} \cos^2 \theta/2 + q_{\omega_2} \sin^2 \theta/2$$

Specifically, the expectation of the observable $S_Z = \sigma_3 = (1, -1)$, Eq. (55), is $\langle S_Z \rangle = \cos^2 \theta/2 - \sin^2 \theta/2 = \cos \theta$.

8.2.2 Transcription into an Auxiliary Hilbert Space \mathcal{H}

Different transcriptions depending on a gauge parameter are equivalent. Consider a gauge labelled ϕ . With this gauge, construct an auxiliary Hilbert space. The quantum state is transcribed as the rank 1 density operator $\rho_{\theta,\phi} = |a\rangle\langle a|$ with the Gleason’s vector $|a\rangle$ (Definition 28),

$$|a\rangle = e^{i\phi} \sqrt{w_{\theta,1}} \cdot |1\rangle + e^{-i\phi} \sqrt{w_{\theta,2}} \cdot |2\rangle = e^{i\phi/2} \cos \theta/2 \cdot |1\rangle + e^{-i\phi/2} \sin \theta/2 \cdot |2\rangle$$

as

$$\rho_{\theta,\phi} = |a\rangle\langle a| = \frac{1}{2} \begin{bmatrix} 1 + \cos \theta & e^{-i\phi} \sin \theta \\ e^{i\phi} \sin \theta & 1 - \cos \theta \end{bmatrix}$$

The gauge parameter expresses an exact symmetry of the Hilbert space with respect to a rotation. There is also an antiunitary gauge generated by complex conjugation, that is simply here $\phi \mapsto -\phi$. Note that the formulation is classical for $\theta = 0$.

8.2.3 Entropy and Information

The entropy of the Shannon distribution is zero. As a result, the Shannon information is 1 bit. The totality of this information is extracted by a measurement.

9 Discussion

The current model is of course expected to have repercussions in different fields, starting with physics, considered the science of observing the world based on reasoning [28].

9.1 Spin-off in Physics

Unquestionably, quantum information is the basis of quantum physic.

Actually, quantum physics identifies the universe with a Hilbert space. However, a major problem is that its state is usually represented by a wave vector, that is a *pure state*. In the present framework, a pure state describes an information register set to zero (with a particular Boolean gauge), so that the complexity of the world would be just an artifact only due to a sophisticated observation window. Therefore, paradoxically, for the universe to have a non-trivial content, a

pure state is excluded and *only a mixed state, that is a probabilistic mixture of pure states is acceptable*.

But now we have to explain how standard physics can be so efficient despite problematic prerequisites. The solution could be that the behavior of the standard wave vector is actually a witness, characterizing the symmetry of the system, i.e. its gauge group. In turn, from proposition (8), this group itself characterizes the quantum state of the system. This ties in with Steven Weinberg's deep intuition [29], namely "specifying Nature's symmetry group may be all we need to say about the physical world".

Another objection is that the present model only deals with logical concepts and can therefore provide only a bare landscape of the world, free from any specific ontological or "ontic" ingredient. Perhaps this is not so essential, especially since genuine ontological elements are undoubtedly unimaginable and therefore unfalsifiable, whereas the candidate "beables", whether fire, aether, epicycles, points, vectors, strings or branes are highly problematic or at best purely phenomenological models. This suggests circumventing any specific ontology and adopt the concept "It from bit", in agreement with the celebrated Wheeler's doctrine. This means that abstract information is the ultimate ingredient while deliberately ignoring any ontological significance.

On these bases, let us outline a purely information model.

Towards new foundations of physics: The world could be reconstructed from a number of observations, expressed for convenience in terms of binary Boolean variables regarded as discrete degrees of freedom. From proposition (8), the state of the universe could be represented by a gauge group whose invariant observables express symmetry, thus joining Klein's Erlangen program in mathematics [30] (see e.g. J.-B. Zuber [31]). This implies that the universe is described by a mixed state. Next, we would have to find the origin of most concepts in the corpus of information theory. For example, the concept of cosmic time could be related to the total entropy of the universe. Incidentally, the need for mixed states could solve some paradoxes, like the information loss in black holes. Finally, the method paves the way to a huge field of investigation obviously outside the scope of this article.

9.2 Beyond Physics

The Bayesian approach is likely to be powerful in all other area of reasoning.

The first application concerns Data Science. It provides an explanation of the speedup of both Bayesian computation and quantum computing. This explains especially the efficiency of neural networks which are implicit Bayesian calculations. This efficiency ultimately rests on the unique ability of real numbers to perform optimization unlike discrete implementations.

Beyond, this suggests that the Church-Turing principle may not be the end of the history and that Bayesian inference could be a more powerful tool than the Turing machine to conceive universal computation as previously suggested, but only for quantum computation, by Deutsch [32].

Bayesian inferences could even have spin-off in *pure mathematics* because the means of deducing mathematics from logic could include Bayesian inference and

not only deduction. Leopold Kronecker is famous for having declared that “God made the integers, all else is the work of man.” [33] One step further, one could assert that “God made logical rules, all else is the work of man.” At last, more punctually, quantum information could explain the hitherto unknown link [34] between the theory of potential and probability [35].

More unexpected for quantum physicists, though suspected by David Bohm and Basil Hiley [36], other sciences including *soft sciences* already benefit from this approach. Applications have been described, e.g. in cognition and decision making [37–39], psychology [40, 41], social science [42] or grammatical language [43]. Beyond cognition, other emblematic examples could be found in biology, e.g. in both the immune system and immunotherapy and even in evolution theory.

10 Conclusion

Our goal was to propose an interpretation of quantum formalism. Although it is a long-standing issue, whose origins can be traced back to von Neumann [22], the foundations of quantum mechanics have remained elusive, giving rise to questioning and discomfort [44]. The probabilistic “Born interpretation” aroused the Einstein’s famous sentence, “I, at any rate, am convinced that He does not throw dice” [45]. Later, in a celebrated lecture [46], R. Feynman gave his equally famous verdict, “I think I can safely say that nobody understands quantum mechanics”. Let us finally quote the striking Jaynes’ opinion: “A standard of logic that would be considered a psychiatric disorder in other fields, is the accepted norm in quantum theory” [7].

To address this discomfort, countless approaches have been devised. Some authors tried to circumvent the conventional logic. Others attempted to reinterpret the experimental results. Finally, some simply denied the existence of a problem. In a tasty paper, updated in 2002 [47], Christopher Fuchs enumerated with humor a number of “religions”: “The Bohmians [48], the Consistent Historians [49], the Transactionalists [50], the Spontaneous Collapseans [51], the Einselectionists [52], the Contextual Objectivists [53], the outright Everettians [54, 55], and many more beyond that”. Recent approaches try to derive quantum logic from *ad hoc* information-theoretic extra principles assumed “reasonable” or, following Spekkens [56], propose frameworks claimed “operational” [57–62], based on the compatibility with specific information processing tasks. Epistemic approaches propose generalized probabilistic theories (GPT) comparing quantum and classical probabilities [63, 64]. Specifically, new frameworks aim to identify the additional axioms needed to derive the quantum formalism from probabilistic constraints, e.g. from “information causality” or from entropy [65, 66]. Another appealing approach inspired from thermodynamics is to use an entropic method of inference [67]. Eventually, a more direct way is to compare quantum states with Bayesian states of knowledge [68–71].

In the present paper, we abstain from introducing any extra axiom and we support the approaches based on Bayesian inference theory.

Although using quantum terminology when appropriate, we have basically dealt with classical information in a classical memory, but at the end, we obtain the exact apparatus of quantum information. This means that quantum information as such

is nothing but information itself and therefore independent of any physical content. Our major conclusion, as sketched in Sect. (1.2), is somewhat baffling: Quantum information is simply classical information processed by Bayesian inference theory.

As far as quantum formalism itself is concerned, the current model is the first to logically *deduce* from information theory its fundamental characteristics, almost always posited from the outset as seemingly arbitrary postulates: Why is the theory probabilistic? Why is the theory linear? Where does the Hilbert space come from? Also, most of the emblematic paradoxes, such as entanglement, contextuality, non-signaling correlation, measurement problem, no-cloning theorem etc., find a perfectly rational explanation. At last the controversial concept of Shannon information conveyed by a wave vector, or stored in the system is clarified.

Beyond physics, quantum information appears as a multipurpose technique for analyzing a system of logical constraints, in line with classical information. Whereas classical information is the universal tools of *logic*, quantum information in the universal tool of *inference*. This is perhaps the most important conclusion of this article.

References

1. Shannon, C. E.: A mathematical theory of communication, Bell System Technical Journal, **27** 379–423, 656–715. (1948), <http://pespmc1.vub.ac.be/books/Shannon-TheoryComm.pdf>
2. Szilard, L.: On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings, Zeitschrift fuer Physik, **53** 840–856, (1929). <http://sns.ias.edu/~tlusty/courses/InfoInBioParis/Papers/Szilard1929.pdf>
3. Brillouin, L.: Science and Information Theory, Academic Press, New-York, (1956). <http://www.physics.mcgill.ca/~delrio/courses/phys559/lectures%20and%20notes/Brillouin-Information-Theory-Ch1-4.pdf>
4. Wheeler, J.A.: Information, physics, quantum: the search for links. In: Zurek, W. (ed.) Complexity, entropy, and the physics of information, pp. 3–28. Addison-Wesley, Redwood City (1990)
5. Landauer, R.: The physical nature of information, Physics Letters A, **217** 188–193, (1996). http://cqi.inf.usi.ch/qic/64_Landauer_The_physical_nature_of_information.pdf
6. Jaynes, E. T.: Probability Theory: the logic of science, Cambridge University Press, Cambridge, UK, (2003). <http://omega.math.albany.edu:8008/JaynesBook.html>
7. Jaynes, E. T.: Clearing up mysteries, in: J. Skilling (Ed.), Maximum entropy and Bayesian methods, Kluwer, Dordrecht, (1989), pp. 1–29. <http://bayes.wustl.edu/etj/articles/cmystery.pdf>
8. Rovelli, C.: Relational quantum mechanics. Int. J. Theor. Phys. **35**, 1637–1678 (1996). <https://doi.org/10.1007/BF02302261>
9. Caves, C.M., Fuchs, C.A., Schack, R.: Quantum probabilities as Bayesian probabilities. Phys. Rev. A **65**(2), 22305 (2002). [arXiv:quant-ph/0106133](https://arxiv.org/abs/quant-ph/0106133)
10. Mermin, N. D.: Qbism puts the scientist back into science, Nature, **507** 421–422, (2014). http://www.nature.com/polopoly_fs/1.14912!/menu/main/topColumns/topLeftColumn/pdf/507421a.pdf
11. Fuchs, C. A.: QBism, The Perimeter of Quantum Bayesianism (2010). [arXiv:1003.5209](https://arxiv.org/abs/1003.5209)
12. Boyd, S., Vandenberghe, L.: Convex optimization, Cambridge University Press, Cambridge, (2004). http://www.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf
13. Cox, R. T.: Probability, frequency, and reasonable expectation, American Journal of Physics, **14** 1–13, (1946). <http://jimbeck.caltech.edu/summerlectures/references/ProbabilityFrequencyReasonableExpectation.pdf>
14. Maassen, H., Uffink, J.B.M.: Generalized entropic uncertainty relations. Phys. Rev. Lett. **60**, 1103–1106 (1988)
15. Frank, R.L., Lieb, E.H.: Entropy and the uncertainty principle. Annales Henri Poincaré **13**, 1711–1717 (2012). [arXiv:1109.1209](https://arxiv.org/abs/1109.1209)
16. Yeung, R.W.: Information theory and network coding. Springer-Verlag, New York (2008)
17. Murty, K.G.: Linear programming. Wiley, New York (1983)

18. Caticha, A.: Lectures on probability, entropy, and statistical physics, In: 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, (2008). [arXiv:0808.0012](https://arxiv.org/abs/0808.0012)
19. Feldmann, M.: New loophole for the Einstein-Podolsky-Rosen paradox. *Found. Phys. Lett.* **8**(1), 41–53 (1995). <https://doi.org/10.1007/BF02187530>. [arXiv:quant-ph/9904051](https://arxiv.org/abs/quant-ph/9904051)
20. Barrett, J., Linden, N., Massar, S., Pironio, S., Popescu, S., Roberts, D.: Nonlocal correlations as an information-theoretic resource. *Phys. Rev. A* **71**(2), 022101 (2005). [arXiv:quant-ph/0404097](https://arxiv.org/abs/quant-ph/0404097)
21. Popescu, S., Rohrlich, D.: Quantum nonlocality as an axiom. *Found. Phys.* **24**, 379 (1994). [arXiv:quant-ph/9508009](https://arxiv.org/abs/quant-ph/9508009)
22. Birkhoff, G., von Neumann, J.: The logic of quantum mechanics, *Annals of Mathematics*, **37** 823, (1936). <http://www.jstor.org/stable/1968621>
23. Jaynes, E. T.: Information theory and statistical mechanics. II, *Phys. Rev.* **108**(2) 170–190, (1957). <http://bayes.wustl.edu/etj/articles/theory.2.pdf>
24. Wigner, E. P.: Normal form of antiunitary operators, *J. Math. Phys.* **1**(5) 409–413, (1960). https://link.springer.com/chapter/10.1007/978-3-662-02781-3_38
25. Durt, T., Englert, B.-G., Bengtsson, I., Życzkowski, K.: On mutually unbiased bases. *Int. J. Quantum Info.* **8**, 535–640 (2010). [arXiv:1004.3348](https://arxiv.org/abs/1004.3348)
26. Kalev, A., Gour, G.: Mutually unbiased measurements in finite dimensions, *New Journal of Physics*, **16**(5), 053038 (2014). [arXiv:1401.2706](https://arxiv.org/abs/1401.2706). <http://stacks.iop.org/1367-2630/16/i=5/a=053038>
27. Schwinger, J.: Unitary Operator Bases, *Proc. Nat. Acad. Sci. USA.* **46**, 560 (1960). <http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=222876&blobtype=pdf>
28. Wiener, N.: The role of the observer, *Philosophy of Science*, **3**(3), 307–319 (1936). <http://www.jstor.org/stable/184668>
29. Feynman, R., Weinberg, S.: Elementary particle physics and the laws of physics. *Gen. Relativ. Gravitat.* **33**, 615–616 (1986)
30. Klein, F.: A comparative review of recent researches in geometry, *Gesammelte mathematische Abhandlungen*, **1**, 460–497 (1931). [arXiv:0807.3161](https://arxiv.org/abs/0807.3161). <http://gdz.sub.uni-goettingen.de/dms/load/img/?PPN=PPN243240503&DMDID=dmdlog45>
31. Zuber, J.-B.: Invariances in physics and group theory, In: Conference “Lie and Klein; the Erlangen program and its impact on mathematics and physics”, Strasbourg, Sept. 2012, (2013). [arXiv:1307.3970](https://arxiv.org/abs/1307.3970)
32. Deutsch, D.: Quantum theory, the Church-Turing principle and the universal quantum computer, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **400**(1818), 97–117 (1985). https://people.eecs.berkeley.edu/~christos/classics/Deutsch_quantum_theory.pdf
33. Gray, J.: *Plato’s Ghost: the modernist transformation of mathematics*. Princeton University Press, Princeton (2008)
34. Bouleau, N.: La jonction entre la théorie du potentiel et les probabilités, *Cahiers du séminaire d’histoire des mathématiques*, **8**, 43–66 (1987). http://www.numdam.org/article/CSHM_1987__8__43_0.pdf
35. Meyer, P.-A., Dellacherie, C.: *Probabilité et potentiel*, vol. 1 and 2, Hermann, Paris, 1975, (2008)
36. Hiley, B. J.: Process and the implicate order: their relevance to quantum theory and mind. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.357.4976&rep=rep1&type=pdf>
37. Aerts, D., Gabora, L., Sozzo, S., Veloz, T.: Quantum structure in cognition: Fundamentals and applications, In: V. Privman, V. Ovchinnikov (Eds.), *Proceedings of the Fifth International Conference on Quantum, Nano and Micro Technologies (ICQMN 2011)*, International Academy, Research and Industry Association, (2011), pp. 57–62. [arXiv:1104.3344](https://arxiv.org/abs/1104.3344)
38. Busemeyer, J.R., Bruza, P.D.: *Quantum models of cognition and decision*. Cambridge University Press, Cambridge (2012)
39. Bruza, P., Busemeyer, J., Gabora, L.: Introduction to the Special Issue on Quantum Cognition (2013). [arXiv:1309.5673](https://arxiv.org/abs/1309.5673)
40. Dzhafarov, E.N., Kujala, J.V.: Selectivity in probabilistic causality: where psychology runs into quantum physics. *J. Math. Psychol.* **56**, 54–63 (2012). [arXiv:1110.2388](https://arxiv.org/abs/1110.2388)
41. Dzhafarov, E.N., Kujala, J.V.: Quantum entanglement and the issue of selective influences in psychology: an overview. *Lecture Notes Comput. Sci.* **7620**, 184–195 (2012). [arXiv:1209.0041](https://arxiv.org/abs/1209.0041)
42. Haven, E., Khrennikov, A.: *Quantum Soc. Sci.* Cambridge University Press, Cambridge (2013)

43. Clark, S., Coecke, B., Grefenstette, E., Pulman, S., Sadrzadeh, M.: A quantum teleportation inspired algorithm produces sentence meaning from word meaning and grammatical structure (2013). [arXiv:1305.0556](#)
44. Ionicioiu, R.: Quantum mechanics: Knocking at the gates of mathematical foundations, In: I. Parvu, G. Sandu, I. Toader (Eds.), *Romanian Studies in Philosophy of Science*, Vol. 313 of Boston Studies in the Philosophy and History of Science, Springer, (2015). [arXiv:1506.04511](#)
45. Einstein, A., Born, M.: *The Born-Einstein letters*. Walker and Company, New-York (1971)
46. Feynman, R.: *Lessons on Probability and Uncertainty in quantum mechanics*, Cornell Messenger Lecture Archive, (2011)
47. Fuchs, C.A.: Quantum mechanics as quantum information (and only a little more). In: Khrennikov, A. (ed.) *Quantum Theory: Reconsideration of Foundations*, pp. 463–543. Växjö University Press, Växjö, Sweden (2002). [arXiv:quant-ph/0205039](#)
48. Articles, In: J. T. Cushing, A. Fine, S. Goldstein (Eds.), *Bohmian Mechanics and Quantum Theory: An Appraisal*, Kluwer, Dordrecht, (1996)
49. Griffiths, R.B., Omnès, R.: Consistent histories and quantum measurements. *Phys. Today* **52**(8), 26–31 (1999)
50. Cramer, J.G.: An overview of the transactional interpretation of quantum mechanics. *Int. J. Theor. Phys.* **27**, 227–236 (1988)
51. Ghirardi, G. C., Pearle, P.: Dynamical reduction theories: Changing quantum theory so the statevector represents reality, In: M. F. A. Fine, L. Wessels (Eds.), *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Philosophy of Science Association, East Lansing, MI, (1990)
52. Zurek, W.H.: Decoherence, Einselection and the existential interpretation (The Rough Guide). *Phil. Trans. R. Soc. Lond. A* **356**, 1793–1821 (1998). [arXiv:quant-ph/9805065](#)
53. Grangier, P.: Contextual objectivity: a realistic interpretation of quantum mechanics (2000). [arXiv:quant-ph/0012122](#)
54. Deutsch, D.: *The fabric of reality: the science of parallel universes and its implications*. Allen Lane, New York (1997)
55. Vaidman, L.: The many-worlds interpretation of quantum mechanics, In: E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, Stanford University, Stanford, CA, (2002). <http://plato.stanford.edu/entries/qm-manyworlds/>
56. Spekkens, R.W.: Contextuality for preparations, transformations, and unsharp measurements. *Phys. Rev. A* **71**, 052108 (2005). [arXiv:quant-ph/0406166](#)
57. Chiribella, G., D’Ariano, G.M., Perinotti, P.: Informational derivation of quantum theory. *Phys. Rev. A* **84**(1), 012311 (2011). [arXiv:1011.6451](#)
58. Hardy, L.: Reconstructing quantum theory (2013). [arXiv:1303.1538](#)
59. Moeller, M. P., Masanes, L.: Information-theoretic postulates for quantum theory (2012). [arXiv:1203.4516](#)
60. Masanes, L., Mueller, M.P., Augusiak, R., Perez-Garcia, D.: A digital approach to quantum theory. *Proc. Nat. Acad. Sci. USA* **110**(41), 16373 (2013). [arXiv:1208.0493](#)
61. Chiribella, G., D’Ariano, G. M., Perinotti, P.: Quantum from principles, In: R. Spekkens, G. Chiribella (Eds.), *Quantum theory: informational foundations and foils*, Springer Verlag, in press. [arXiv:1506.00398](#)
62. Oeckl, R.: A local and operational framework for the foundations of physics. [arXiv:1610.09052](#)
63. Pitowsky, I.: Quantum mechanics as a theory of probability (2005). [arXiv:quant-ph/0510095](#)
64. Janotta, P., Hinrichsen, H.: Generalized probability theories: what determines the structure of quantum theory? *J. Phys. A Math. General* **47**(32), 323001 (2014). [arXiv:1402.6562](#)
65. Pawłowski, M., Paterek, T., Kaszlikowski, D., Scarani, V., Winter, A., Żukowski, M.: Information causality as a physical principle. *Nature* **461**(7267), 1101–1104 (2009)
66. Barnum, H., Barrett, J., Clark, L.O., Leifer, M., Spekkens, R., Stepanik, N., Wilce, A., Wilke, R.: Entropy and information causality in general probabilistic theories. *N. J. Phys.* **12**, 033024 (2010). [arXiv:0909.5075](#)
67. Caticha, A.: *Entropic dynamics: mechanics without mechanism* (2017). [arXiv:1704.02663](#)
68. Fuchs, C. A., Schack, R.: *Quantum-Bayesian Coherence: The No-Nonsense Version* (2013). [arXiv:1301.3274](#)
69. Leifer, M. S., Spekkens, R. W.: Formulating Quantum Theory as a Causally Neutral Theory of Bayesian Inference (2011). [arXiv:1107.5849](#)
70. De Raedt, H., Katsnelson, M. I., Michielsen, K.: Quantum theory as the most robust description of reproducible experiments (2013). [arXiv:1303.4574](#)

71. Fuchs, C.A., Mermin, N.D., Schack, R.: An introduction to QBism with an application to the locality of quantum mechanics. *Am. J. Phys.* **82**, 749–754 (2014). [arXiv:1311.5253](https://arxiv.org/abs/1311.5253)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.