Reconstructing quantum theory from operational principles – lecture notes

Lluís Masanes

June 20, 2017

Despite the success of quantum theory, there is still no consensus on what is this theory telling us about reality. One of the problems is that the standard postulates, in terms of Hilbert spaces and operators acting on them, do not have a clear operational or physical meaning. To overcome this situation, there has recently been an important effort to construct alternative formulations of quantum theory in terms of postulates which have a clear and direct operational meaning. These lecture notes include an exposition of some of these reformulations of quantum mechanics, and provide a review of the tools and concepts that are often used in this field. An effort has been make to lighten the mathematical content as much as possible. It is important to mention that, unfortunately, this is a partial presentation of the subject with very important omissions.

Contents

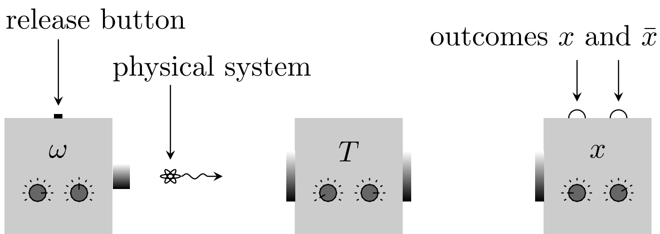
1 Post-classical probability theory

In classical probability theory, no matter how complex a system is, there is always a joint probability distribution which simultaneously describes the statistics of all the measurements that can be performed on a system. In other words, there exists a maximally informative measurement, of which all other measurements are functions. This fact no longer holds in quantum theory (QT) due to the existence of incompatible measurements. Motivated by this, Birkhoff and von Neumann generalized the formalism of classical probability theory to include incompatible measurements. This is nowadays called general probabilistic theories (GPTs), the convex operational framework, or Post-classical probability theory.

Recently, a lot of interest has been directed to the study of GPTs, with the double aim of reconstructing QT and exploring what lies beyond. This, in particular, led to the discovery that many features originally thought as specific to QT (such as for instance: Bell-inequality violation, no-cloning, monogamy of correlations, Heisenberg-type uncertainty relations, measurement-disturbance tradeoffs, and the possibility of secret key distribution), are common to most GPTs. In this light, the standard question “why does nature seem to be quantum instead of classical?” sounds less appropriate than asking “why QT instead of any other GPT”. A possible answer to this question is that any GPT different from QT violates at least one of the physically meaningful postulates that are introduced below.

1.1 Fundamental assumption

The formalism of GPTs can be derived or axiomatized in many different ways. In this notes we derive this formalism from the following fundamental informal assumption: *Experimentalists enjoy a classical reality where they can prepare experimental setups, perform mixtures of operations and measurements that produce a single outcome.*



Most GPTs will not be able to predict this classical reality, and hence, cannot be taken as universally valid. But this is not necessarily a problem.

1.2 States

In QT states (equivalence classes of preparations) are represented by density matrices. But, how can we represent states in theories that we do not yet know? The state of a system is represented by the probabilities of some reference measurement outcomes $x_1, \ldots, x_k$ which are called *fiducial*:

$$\omega_\Omega = \begin{bmatrix} p(x_1|\Omega) \\ p(x_2|\Omega) \\ \vdots \\ p(x_k|\Omega) \end{bmatrix} \in \mathcal{S} \subset \mathbb{R}^k , \tag{1}$$

where Ω denotes a preparation. This list of probabilities has to be minimal but contain sufficient information to predict the probability distribution of all measurements that can be in principle performed on the system. That is, for any outcome $\tilde{x}$, its probability $p(\tilde{x}|\Omega)$ is determined by the numbers $p(x_1|\Omega), p(x_2|\Omega), \ldots, p(x_k|\Omega)$. (Note that this is always possible since the list could contain the probabilities of all outcomes $\tilde{x}$; in particular, the list can be infinite $k = \infty$.) The number of fiducial outcomes k is equal to the dimension of the vector space spanned by the set of all mixed states $\mathcal{S}$, as otherwise one fiducial probability would be functionally related to the others, and the list not minimal. PICTURE.

It is convenient to include the possibility that the system is present with certain probability $U(\omega) \in [0, 1]$, which by consistency, is equal to the sum of probabilities for all the outcomes of any measurement: $U(\omega) = \sum_i p(x^{(i)}|\omega)$. When the system is absent ($U(\omega) = 0$) the fiducial outcomes have zero probability, hence the corresponding state (1) is the null vector $\mathbf{0} \in \mathcal{S}$. The subset of normalized states $\mathcal{N} = \{\omega \in \mathcal{S} : U(\omega) = 1\}$ has dimension $k - 1$, and satisfies $\mathcal{S} = \mathrm{conv}(\mathcal{N} \cup \{\mathbf{0}\})$. PICTURE. By the rules of probability, the set of all the allowed states $\mathcal{S}$ is convex. Indeed, by preparing the state ω_1 with probability q and ω_2 with probability $1 - q$, we effectively prepare the mixed state $q\omega_1 + (1 - q)\omega_2$. The *pure states* of $\mathcal{S}$ are the extreme points of $\mathcal{N}$ (recall that $\mathbf{0}$ is also an extreme point, but it does not belong to $\mathcal{N}$). The other elements of $\mathcal{N}$ are called *mixed states*.

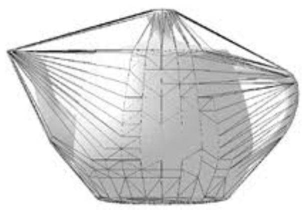
Note that, in QT, $\mathcal{S}$ is the set of density matrices, not the Hilbert space. As an instance, the fiducial outcomes for a qubit can be chosen to be

$$\omega = \begin{bmatrix} p(\sigma_x = 1) \\ p(\sigma_y = 1) \\ p(\sigma_z = 1) \\ p(\sigma_z = -1) \end{bmatrix} , \quad U(\omega) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \cdot \omega \ . \tag{2}$$

Note that the set of fiducial outcomes need not be unique, nor simultaneously measurable. Actually, only classical probability theory has all fiducial outcomes belonging to the same measurement:

$$\omega = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_k \end{bmatrix} , \quad U(\omega) = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \cdot \omega \ , \tag{3}$$

where $q_k \geq 0$ and $\sum_i q_i = 1$. In the formalism of GPTs every convex set can be seen as the state space $\mathcal{S}$ of an imaginary type of system, which in turn, allows for constructing multipartite states spaces which violate Bell inequalities more (or less) than QT. This illustrates the degree to which this formalism generalizes classical probability theory and QT, and allows us to catch a glimpse on the multitude of alternative theories that we are considering here.



1.3 Measurements

Let us denote the probability of obtaining any given outcome x when the system is in the state ω by $E_x(\omega) = p(x|\omega)$. (From now on, we identify preparations and states $\omega \equiv \Omega$.) Clearly, this function must satisfy $E_x(\omega) \subseteq [0, 1]$ for all $\omega \in \mathcal{S}$. Now, let us prove that $E_x : \mathbb{R}^k \to \mathbb{R}$ is linear. Suppose the system is prepared in state ω_1 with probability q and ω_2 with probability $(1-q)$. Then the relative frequency of any given outcome x should not depend on whether the actual preparation, ω_1 or ω_2, is forgotten before or after the measurement. This translates to

$$E_x\big(q\omega_1 + (1 - q)\omega_2\big) = qE_x(\omega_1) + (1 - q)E_x(\omega_2) \ ,$$

which together with $E_x(\mathbf{0}) = 0$ implies the linearity of E_x. Linear functions $E : \mathbb{R}^k \to \mathbb{R}$ satisfying $E_x(\mathcal{S}) \subseteq [0,1]$ are called *effects*. Sometimes it is convenient to use the scalar product notation $E \cdot \omega = E(\omega)$.

An r-outcome measurement is characterized by r effects $E_1, \ldots, E_r$ satisfying the normalization condition $E_1(\omega) + \cdots + E_r(\omega) = 1$ for all $\omega \in \mathcal{S}$. This can also be written in terms of the *unit effect* U as a vector equality $E_1 + \cdots + E_r = U$.

It is often understood that in classical probability theory and QT, all effects E are in principle measurable. However, this need not be the case in other GPTs, but it is sometimes taken as a postulate, and it can be seen as some kind of idealization.

Postulate (No restriction). *In any system, all its effects are in principle measurable.*

1.4 Transformations

Transformations are implemented by controlling the dynamics of a system (for example, adjusting some external fields). A transformation can be represented by a map $T : \mathcal{S} \to \mathcal{S}$ which, for the same reason as outcome probabilities E, has to be linear (i.e. a matrix). Sometimes there are pairs of (physical) transformations T, T' whose composition leaves the system unaffected, independently of its initial state $TT' = \mathbb{1}$. In this case we say that these transformations are reversible $T' = T^{-1}$. (We stress that here we refer to physical transformations. If T has a mathematical inverse which is not a physical transformation, then T is not reversible.) The set of reversible transformations forms a (compact) group of real matrices $\mathcal{G}$. A widely-used postulate, sometimes called “reversibility”, is the following.

Postulate (Transitivity). *For every pair of pure states $\omega_1, \omega_2 \in \mathcal{S}$ there is a reversible transformation $T \in \mathcal{G}$ taking one state onto the other $\omega_2 = T\omega_1$.*

The set of transformations generated by continuous-time dynamics form a connected Lie group; and the elements of the corresponding Lie algebra are the possible “Hamiltonians” of the system. (Note that, in general, these Hamiltonians will have have nothing to do with Hermitian matrices).

1.5 Multi-partite systems

Two systems A and B constitute a composite system AB if a measurement for A together with a measurement for B uniquely specify a measurement for AB, independently of the temporal ordering. The fact that subsystems are systems themselves implies that each global state ω_{AB} has well-defined reduced states ω_A and ω_B which do not depend on which transformations and measurements are performed on the other subsystem. This is often referred to as no-signaling, but here we take it as part of the definition of bipartite system. The next proposition establish a relationship between the local and the global fiducial outcomes.

Proposition 1. *If $x_1, \ldots, x_{k_A}$ and $y_1, \ldots, y_{k_B}$ are fiducial outcomes of A and B, then the joint AB state can be written as*

$$\omega_{AB} = \begin{bmatrix} \lambda_{AB} \\ \eta_{AB} \end{bmatrix} \in \mathbb{R}^{k_{AB}} , \tag{4}$$

with the locally-tomographic part is

$$\lambda_{AB} = \begin{bmatrix} p(x_1, y_1) \\ p(x_1, y_2) \\ \vdots \\ p(x_{k_A}, y_{k_B}) \end{bmatrix} \in \mathbb{R}^{k_A} \otimes \mathbb{R}^{k_B} , \tag{5}$$

and the holistic part

$$\eta_{AB} = \begin{bmatrix} p(z_1) \\ \vdots \\ p(z_{h_{AB}}) \end{bmatrix} \in \mathbb{R}^{h_{AB}} \ . \tag{6}$$

Holistic effects can only be measured when the two subsystems are jointly addressed. If E_x and E_y are effects of A and B, then their joint probability is given by the standard tensor-product rule

$$p(x,y) = (E_x \otimes E_y) \cdot \lambda_{AB} \ . \tag{7}$$

If U_A and U_B are the unit effects of A and B, then the reduced states are

$$\omega_A = (\mathbb{1}_A \otimes U_B) \cdot \lambda_{AB} \ , \tag{8}$$

$$\omega_B = (U_A \otimes \mathbb{1}_B) \cdot \lambda_{AB} \ . \tag{9}$$

The global dimension is given by $k_{AB} = k_A k_B + h_{AB}$.

Proof. We start by proving that the $k_A k_B$ joint fiducial probabilities $p(x_i, y_j)$ determine any other joint probability $p(x,y)$. If the system B is measured first, giving outcome y_j, then the system A is in the state determined by the fiducial probabilities $p(x_i|y_j) = p(x_i,y_j)/p(y_j)$, and the single-system probability rule can be applied $p(x|y_j) = \sum_i E_x^i \, p(x_i|y_j)$. Multiplying by $p(y_j)/p(x)$ and using Bayes' rule gives

$$p(y_j|x) = \sum_i E_x^i \, p(x_i,y_j)/p(x) \ . \tag{10}$$

By using the freedom in the ordering of measurements, we can interpret $p(y_j|x)$ as the state of the system B once the system A has been measured giving outcome x, and the single-system probability rule can be applied again: $p(y_j|x) = \sum_j E_y^j \, p(y_j|x) = \sum_{i,j} E_x^i \, E_y^j \, p(x_i,y_j)/p(x)$. Multiplying both sides of this equality by $p(x)$ gives (7).

Next we show that the set of vectors λ_{AB} span the full locally-tomographic subspace $\mathbb{R}^{k_A} \otimes \mathbb{R}^{k_B}$. This implies that the degrees of freedom in λ_{AB} are not redundant. In QT, the only vectors η_{AB} which have pure states as marginals $\omega_A \in \mathcal{S}_A$, $\omega_B \in \mathcal{S}_B$, are product $\eta_{AB} = \omega_A \otimes \omega_B$. The same proof technique applies in this more general formalism. This implies that the set of all vectors η_{AB} must contain all product states, otherwise there would be a state in $\mathcal{S}_A$ or $\mathcal{S}_B$ which is not the marginal of any state in $\mathcal{S}_{AB}$. Next, note that by minimality, $\mathcal{S}_A$ contains k_A linearly independent vectors, and analogously for $\mathcal{S}_B$. The tensor products of all these vectors forms a set of $k_A k_B$ linearly independent vectors in $\mathbb{R}^{k_A k_B}$ as claimed. $\square$

The above can be extended to multi-partite systems by systematically performing bi-partitions. The following postulate is often assumed, and it restores the familiar tensor-product structure of multi-partite state spaces.

Postulate (Tomographic locality). *The state of a composite system is completely characterized by the correlations of measurements on the individual components. (Equivalently: $k_{AB} = k_A k_B$.)*

QT satisfies tomographic locality. For example, any 2-qubit state is characterized by the joint probabilities

$$\omega_{AB} = \begin{pmatrix} p(U,U) & p(U,\sigma_x=1) & p(U,\sigma_y=1) & p(U,\sigma_z=1) \\ p(\sigma_x=1,U) & p(\sigma_x=1,\sigma_x=1) & p(\sigma_x=1,\sigma_y=1) & p(\sigma_x=1,\sigma_z=1) \\ p(\sigma_x=1,U) & p(\sigma_x=1,\sigma_x=1) & p(\sigma_x=1,\sigma_y=1) & p(\sigma_x=1,\sigma_z=1) \\ p(\sigma_y=1,U) & p(\sigma_y=1,\sigma_x=1) & p(\sigma_y=1,\sigma_y=1) & p(\sigma_y=1,\sigma_z=1) \\ p(\sigma_z=1,U) & p(\sigma_z=1,\sigma_x=1) & p(\sigma_z=1,\sigma_y=1) & p(\sigma_z=1,\sigma_z=1) \end{pmatrix} . \tag{11}$$

However, in QT restricted to the real numbers we have

$$\omega_{AB} = \begin{pmatrix} p(U,U) & p(U,\sigma_x=1) & p(U,\sigma_z=1) & \\ p(\sigma_x=1,U) & p(\sigma_x=1,\sigma_x=1) & p(\sigma_x=1,\sigma_z=1) & \\ p(\sigma_x=1,U) & p(\sigma_x=1,\sigma_x=1) & p(\sigma_x=1,\sigma_z=1) & \\ p(\sigma_z=1,U) & p(\sigma_z=1,\sigma_x=1) & p(\sigma_z=1,\sigma_z=1) & \\ & & & p(\sigma_y=1,\sigma_y=1) \end{pmatrix} . \quad (12)$$

Hence, tomographic locality does not hold. Also, note that since $p(\sigma_y=1,U)$ nor $p(U,\sigma_y=1)$ are not well-defined, the outcome $p(\sigma_y=1,\sigma_y=1)$ cannot be understood as the correlation between local outcomes.

2 Information-processing properties

Does information play a significant role in the foundations of physics? Information is the abstraction that allows us to refer to the states of systems when we choose to ignore the systems themselves. This is only possible in very particular frameworks, like in classical probability theory or QT, or more generally, whenever there exists an information unit such that the state of any system can be reversibly encoded in a sufficient number of such units. How can we interpret that the state spaces of a 1/2-spin particle and a photon are identical?

2.1 Number of perfectly distinguishable states

We say that the states $\omega_1, \dots, \omega_C \in \mathcal{S}$ are *perfectly distinguishable* if there exists a C-outcome measurement $E_1, \dots, E_C$ such that

$$E_i \cdot \omega_j = \delta_{ij} \ , \quad (13)$$

where δ_{ij} is a Kronecker delta. Note that perfectly distinguishable states ω_j must be normalized. We say that a state space $\mathcal{S}$ has *capacity* C if the largest set of perfectly distinguishable states has size C. We say that $E_1, \dots, E_C$ is a complete measurement if it perfectly distinguishes C states in $\mathcal{S}$ and C is the capacity of $\mathcal{S}$. In QT the capacity is equal to the dimension of the associated Hilbert space, and the perfectly-distinguishable states are associated to an orthogonal basis.

Postulate (Multiplicativity of capacity). *In any bipartite system $C_{AB} = C_A C_B$.*

Loosely speaking, the following postulate states that what characterizes a type of system is its capacity for carrying information, and nothing else. This requirement provides a lot of structure to the theory, since it imposes that $\mathcal{S}_C$ is contained in $\mathcal{S}_{C+1}$, and so on, providing a sort of onion-like structure. In order to articulate the following postulate, we need to express it in terms of the subsets of normalized states $\mathcal{N}_C \subset \mathcal{S}_C$.

Postulate (Subspaces). *Let $\mathcal{N}_C$ be a state space with capacity C and $E_1, \dots, E_C$ a complete measurement. Any state space $\mathcal{N}_{C-1}$ with capacity $C-1$ is equivalent to*

$$\mathcal{N}_{C-1} \equiv \{\omega \in \mathcal{N}_C : E_1(\omega) + \cdots + E_{C-1}(\omega) = 1\} \ . \quad (14)$$

Without the following postulate it would be impossible to estimate a state without making additional assumptions (e.g. bounds on the energy).

Postulate (Finite dimensionality). *The number of parameters k that characterize the state of a system with finite C is finite ($k \leq \infty$).*

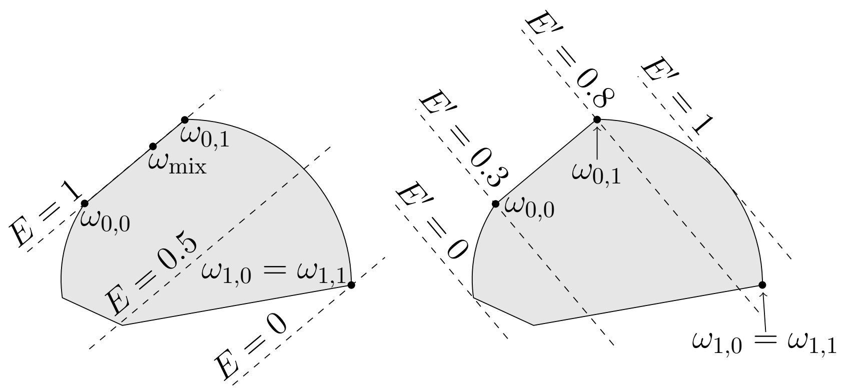


Figure 1: **No Simultaneous Encoding.** This figure shows that there cannot be mixed states in the boundary of $\mathcal{N}_{C=2}$. If there is one, say ω_{mix}, then this boundary contains a non-trivial face (left figure). Since all effects are observable, we can decode a with the effect E, which gives probability one for all states inside that facet, and probability zero for some other state(s). By encoding $(a, a') = (0,0), (0,1)$ in two different states inside that face we can perfectly retrieve a through E, while still getting some partial information about a' with another effect E' (right figure).

2.2 Information with exclusive access

The capacity C of a system quantifies the amount of information that can be retrieved from a system, but not the one that can be encoded. Let us consider the situation where Alice simultaneously encodes two variables a and a' in a given system, and sends it to Bob, who will only retrieve one of the two. But which one Bob decodes is unknown to Alice. Depending on its interest, a or a', Bob is allowed to make a different measurement on the system that Alice has sent him, obtaining his bests guesses b, b' for a, a' respectively. The following postulate is for single systems what Information Causality is for bipartite correlations.

Postulate (No Simultaneous Encoding). *If a system with capacity C is used to perfectly encode a C-valued classical variable, then it cannot simultaneously encode any further information:*

$$P(b|a, a') = \delta_{ba} \quad \Rightarrow \quad P(b'|a, a' = 0) = P(b'|a, a' = 1) \ . \tag{15}$$

This postulate is satisfied by classical probability theory and QT. It implies that any state space with $C = 2$ has no boundary states which are not extremal (See proof in Figure 1).

3 Reconstructions of quantum theory

Here we go.

3.1 Quantum theory from 5 reasonable axioms

The following was obtained by Lucien Hardy, and has become a very influential contribution.

Postulate (Tomographic locality). $k_{AB} = k_A k_B$.

Postulate (Multiplicativity of capacity). $C_{AB} = C_A C_B$.

Postulate (Continuous Transitivity). *For every pair of pure states $\omega_1, \omega_2 \in \mathcal{S}$ there is a reversible transformation $T \in \mathcal{G}_{connected}$ taking one state onto the other $\omega_2 = T\omega_1$.*

Postulate (Subspaces). *Let $\mathcal{N}_C$ be a state space with capacity C and $E_1, \dots, E_C$ a complete measurement. Any state space $\mathcal{N}_{C-1}$ with capacity $C-1$ is equivalent to*

$$\mathcal{N}_{C-1} \equiv \{\omega \in \mathcal{N}_C : E_1(\omega) + \cdots + E_{C-1}(\omega) = 1\} \ . \tag{16}$$

The above postulates imply that k_C, the dimension of $\mathcal{S}_C$, is of the form

$$k_C = C^r \ , \tag{17}$$

for some positive integer r, which depends on the theory. In classical probability theory $r = 1$ and in QT we have $r = 2$.

Postulate (Simplicity). *The dimension of $\mathcal{S}_C$ is the minimal one compatible with the other axioms (minimal r).*

Hardy proves that the only theory satisfying all these postulates is QT. Hence, any other probabilistic theory must violate at least one of these postulates. An interesting research direction consists of relaxing some of these postulates (e.g. Simplicity) and obtain modifications of quantum theory, against which quantum theory could be experimentally contrasted. It is interesting to mention that, if instead of Continuous Transitivity we consider just Transitivity, then the only theory satisfying the above postulates is classical probability theory.

3.2 A derivation of QT from physical requirements

Several authors, including Hardy himself, have expressed a dislike for Simplicity. The following reconstruction avoids this postulate. This derivation of QT was obtained by Markus Müller and the myself, and it was very influenced by previous work from Borivoje Dakic and Caslav Brukner.

Postulate (Finite dimensionality). *k_2 is finite.*

Postulate (Tomographic locality). *$k_{AB} = k_A k_B$.*

Postulate (Continuous Transitivity). *For every pair of pure states $\omega_1, \omega_2 \in \mathcal{S}$ there is a reversible transformation $T \in \mathcal{G}_{connected}$ taking one state onto the other $\omega_2 = T\omega_1$.*

Postulate (Subspaces). *Let $\mathcal{N}_C$ be a state space with capacity C and $E_1, \dots, E_C$ a complete measurement. Any state space $\mathcal{N}_{C-1}$ with capacity $C-1$ is equivalent to*

$$\mathcal{N}_{C-1} \equiv \{\omega \in \mathcal{N}_C : E_1(\omega) + \cdots + E_{C-1}(\omega) = 1\} \ . \tag{18}$$

Postulate (No restriction). *All effects in $\mathcal{S}_2$ are in principle measurable.*

Here we have again that, if instead of Continuous Transitivity we consider Transitivity, then classical probability theory also satisfies the postulates. Some of the above postulates (e.g. No restriction) can be relaxed and the expense of making its statement a bit more involved.

4 Final remarks

One can be concerned with the logical minimality of the postulates and the possibility of hidden redundancies. For example, the Subspaces postulate seems like a really strong statement. One way to avoid unnecessarily strong postulates is to follow a classificational approach, where one systematically classifies all GPTs compatible with a small set of postulates. Once this is done, this classification will reveal which postulates can be additionally imposted in order to single out QT or whatever is the goal. This approach has been taken in, for instance, arXiv:1610.04859 and arXiv:1208.0493.