Classical and Quantum Information Acquisition
Measurement and POVM

Gennaro Auletta
gennaro.auletta@gmail.com
www.gennaroauletta.net

International Conference on New Frontiers in Physics
Crete, 8–13 June

Outline

To measure is to acquire some piece of information about something and could be therefore also considered a kind of information exchange. An unjustified generalization of the classical theory of communication has produced two misunderstandings (or two unwarranted generalizations):

- There is already information selection at the start of the information exchange. This is true for controlled information exchange in certain conditions but it is not true for quantum–mechanical systems (which can be in an initial superposition state) but it is neither true for living systems, which in general try to extrapolate (or to guess) the vital meaning of an uncertain signal (Who is the sender? To which purpose?) once they receive a certain piece of information. The case of mimicry is sufficiently clear: an innocent signal may obscure a real danger.
- Given the previous assumption, the model that has been imposed is the match or mismatch between input and output. However, this does not correspond to the real situation of most information exchanges. This is the reason why in AI and PDP the so–called hidden units (bridging between inputs and outputs) were introduced. In the following I shall introduce an interface of this kind between input and output, classically represented by data and quantum–mechanically by the coupling between object system and apparatus. Seen in this perspective, classical and quantum measurement are less far away than it is customary to assume.

Classical situation: data

Classically, we have an unknown parameter k whose value we wish to know and some data d pertaining to a set D at our disposal. We never have direct access to systems or events (whose properties are described by k) but always to things or systems *through* data. These data can be represented by the position of the pointer of our measuring apparatus or simply by the impulse our sensory system has received, or even by the way we receive information about the position of the pointer through our sensory system. It does not matter how long this chain may be. The important point is a matter of principle: we can receive information about objects and events only conditionally from the data at our disposal.

Classical extrapolation

Obviously, once we have observed or acquired data, we must perform an information extrapolation that allows us to guess about the value of the parameter k. This is information selection. The probability that we select a response j having an event represented by an unknown parameter k (i.e. the probability that both event k and event j occur) is given by $p(j,k)$. Now, we may expand this probability by taking into account the data d that are somehow the interface between the source event k and our final selection event j:

$$p(j|k) = \sum_{d \in D} p(j|d)p(d|k), \tag{1}$$

where I have made use of a discrete case for the sake of simplicity and we are summing over all the data d pertaining to the set D.

By inserting the last equation into the know classical equation for the total probability we obtain:

$$p(j,k) = \sum_{d \in D} p(j|d)p(d|k)p(k) = \sum_{d \in D} p(j|d)p(d,k). \qquad (2)$$

Eq. (2) can be considered as a generalization of the well known formula

$$p(j) = \sum_{d \in D} p(j|d)p(d), \qquad (3)$$

and it reduces to the latter when $p(k) = 1$, i.e. when the event k occurs with certainty.

Faithfulness and Selection

It is important to stress that the two conditional probabilities $p(j|d)$ and $p(d|k)$ are quite different. This can be seen formally by the fact that in Eq. (1) we sum over the data d, which represents the conditioned result relative to k on the one hand and the condition for information extrapolation on the other. This means that the probability $p(d|k)$ represents how *faithful* our data are relative to k, that is, how reliable our apparatus (or our sensory system) is. Instead, the probability $p(j|d)$ represents our ability to *select a single* j able to interpret the parameter in the best way.

Bayesian probability

Probability (2) is Bayesian, since, by making use of the result

$$p(k|j) = \frac{p(j,k)}{p(j)}, \tag{4}$$

it is also true that

$$p(k|j) = p(k)\frac{\sum_{d\in D} p(j|d)p(d|k)}{p(j)} = p(k)\frac{p(j|k)}{p(j)}. \tag{5}$$

In other words, we can invert the kind of question we pose and try to guess the unknown parameter k conditionally on having detected j.
Now I shall show that the quantum case is not different.

Preparation and Pre–measurement: Kets

I assume that the initial state of the apparatus is some ready–state $|A_0\rangle$ while the state of the object system is the superposition

$$|\psi\rangle = \sum_k c_k |k\rangle \,. \tag{6}$$

It is also in agreement with quantum mechanics to assume that the entanglement between apparatus and object system occurring during the pre–measurement step (at time t) is the result of a unitary transformation (I indeed remind that only the final step of selection or detection is not unitary):

$$\hat{U}_t \left(\sum_k c_k |k\rangle |A_0\rangle \right) = \sum_k c_k |k\rangle |a_k\rangle \ . \tag{7}$$

Preparation and Pre–measurement: Matrices

We may describe the initial state of the object system and apparatus in terms of the (initially factorized) density matrices:

$$\hat{\rho}^{A} = |A_0\rangle\langle A_0| \quad \text{and} \quad \hat{\rho}^{S} = |\psi\rangle\langle\psi| . \tag{8}$$

so that in matrix terms the pre–measurement stage can be described as

$$\hat{\rho}^{S}\hat{\rho}^{A} \mapsto \hat{U}_t\left(\hat{\rho}^{S}\hat{\rho}^{A}\right)\hat{U}_t^{\dagger} . \tag{9}$$

To be easy, in the case of a bidimensional system (with components m and n) we can write for the result of the pre measurement step:

$$\begin{aligned} \hat{U}_t\left(\hat{\rho}^{S}\hat{\rho}^{A}\right)\hat{U}_t^{\dagger} &= |c_m|^2 |m, a_m\rangle\langle m, a_m| + |c_n|^2 |n, a_n\rangle\langle n, a_n| \\ &\quad + c_m c_n^* |m, a_m\rangle\langle n, a_n| + c_m^* c_n |n, a_n\rangle\langle m, a_m| . \end{aligned} \tag{10}$$

It is understood that $|m\rangle$ and $|n\rangle$ are the eigenstates of the observable that is selected in the pre–measurement set–up.

A–priori Probabilities

Just before the detection, the probability distribution to read the value a_m of the apparatus observable will be simply given by

$$\wp(a_m) = \mathrm{Tr}_{\mathcal{A}}\left[|a_m\rangle\langle a_m|\mathrm{Tr}_{\mathcal{S}}\left(\hat{U}_t\hat{\rho}^{\mathcal{S}}\hat{\rho}^{\mathcal{A}}\hat{U}_t^{\dagger}\right)\right] = |c_m|^2 . \tag{11}$$

We can prove this quite easily. Indeed, computation of the partial trace

$$\mathrm{Tr}_{\mathcal{S}}\left(\hat{U}_t\hat{\rho}^{\mathcal{S}}\hat{\rho}^{\mathcal{A}}\hat{U}_t^{\dagger}\right) \tag{12}$$

will kill all of the system's states in (10). By applying the projector $\hat{P}_{a_m} = |\, a_m\rangle\,\langle a_m\,|$ to the previous result, we shall obtain

$$\hat{P}_{a_m}\mathrm{Tr}_{\mathcal{S}}\left(\hat{U}_t\hat{\rho}^{\mathcal{S}}\hat{\rho}^{\mathcal{A}}\hat{U}_t^{\dagger}\right) = |c_m|^2\,|\, a_m\rangle\,\langle a_m\,| , \tag{13}$$

and by finally tracing out the apparatus, we shall get the probability, which, in our case, is $|c_m|^2$.

Tracing out

Let us now apply to Eq. (11) the cyclic property of the trace, which for any three arbitrary observables tells us

$$\mathrm{Tr}\left[\hat{O}_1\hat{O}_2\hat{O}_3\right] = \mathrm{Tr}\left[\hat{O}_3\hat{O}_1\hat{O}_2\right] = \mathrm{Tr}\left[\hat{O}_2\hat{O}_3\hat{O}_1\right] , \qquad (14)$$

so that we obtain

$$\mathrm{Tr}_{\mathcal{A}}\left[|a_m\rangle\langle a_m|\mathrm{Tr}_{\mathcal{S}}\left(\hat{U}_t\hat{\rho}^{\mathcal{S}}\hat{\rho}^{\mathcal{A}}\hat{U}_t^{\dagger}\right)\right] = \mathrm{Tr}_{\mathcal{A}}\left[\mathrm{Tr}_{\mathcal{S}}\left(\hat{U}_t^{\dagger}|a_m\rangle\langle a_m|\hat{U}_t\hat{\rho}^{\mathcal{S}}\hat{\rho}^{\mathcal{A}}\right)\right] . \qquad (15)$$

The reason for the displacement of the projector $\hat{P}_{a_m}$ is that the partial trace on $\mathcal{S}$ (i.e. $\mathrm{Tr}_{\mathcal{S}}$) by definition does not act on the apparatus. For the reason that this partial trace does not act on $\hat{\rho}^{A}$, we can rewrite last equation as

$$\mathrm{Tr}_{\mathcal{A}}\left[\mathrm{Tr}_{\mathcal{S}}\left(\hat{U}_t^{\dagger}|a_m\rangle\langle a_m|\hat{U}_t\hat{\rho}^{\mathcal{S}}\hat{\rho}^{\mathcal{A}}\right)\right] = \mathrm{Tr}_{\mathcal{A}}\left[\mathrm{Tr}_{\mathcal{S}}\left(\hat{U}_t^{\dagger}|a_m\rangle\langle a_m|\hat{U}_t\hat{\rho}^{\mathcal{S}}\right)\hat{\rho}^{\mathcal{A}}\right] . \qquad (16)$$

Effects

The expression

$$\mathrm{Tr}_{\mathcal{S}}\left(\hat{U}_t^{\dagger}|a_m\rangle\langle a_m|\hat{U}_t\hat{\rho}^{\mathcal{S}}\right)=\hat{E}(a_m) \tag{17}$$

is a projection-like operator that selects the outcome $|a_m\rangle$ but does not satisfy the requirement of orthogonality that is typical of projectors. Indeed, for any couple of projectors pertaining to the same set (expressing kets of the same orthonormal set) we have:

$$\hat{P}_j\hat{P}_k=\hat{P}_k\hat{P}_j=0\ . \tag{18}$$

The operator $\hat{E}(a_m)$ is called *effect* and is the quintessence of the Positive–Operator–Valued–Measure.

POVM

Effects allow us to write the above probability (11) as

$$\wp(a_m) = \mathrm{Tr}_{\mathcal{A}} \left[\hat{E}(a_m) \hat{\rho}^{\mathcal{A}} \right], \tag{19}$$

which is formally similar to the traditional expression

$$\wp(k) = \mathrm{Tr} \left[\hat{P}_k \hat{\rho} \right] \ . \tag{20}$$

Note that the probability $\wp(a_m)$ is a conditional one. Starting from the definition (17), we compute now the effect explicitly:

$$\begin{aligned} \hat{E}(a_m) &= \mathrm{Tr}_{\mathcal{S}} \left(\hat{U}_t^\dagger | a_m \rangle \langle a_m | \hat{U}_t \hat{\rho}^{\mathcal{S}} \right) \\ &= \left\langle \psi \left| \left(\hat{U}_t^\dagger | a_m \rangle \langle a_m | \hat{U}_t \, | \psi \rangle \, \langle \psi | \right) \right| \psi \right\rangle \\ &= \left\langle \psi \left| \hat{U}_t^\dagger \right| a_m \right\rangle \left\langle a_m \left| \hat{U}_t \right| \psi \right\rangle \\ &= \hat{\vartheta}^\dagger(a_m) \hat{\vartheta}(a_m). \end{aligned} \tag{21}$$

Amplitude Operators

The expressions

$$\hat{\vartheta}(a_m) = \langle a_m | \hat{U}_t | \psi \rangle \quad \text{and} \quad \hat{\vartheta}^\dagger(a_m) = \left\langle \psi \middle| \hat{U}_t^\dagger \middle| a_m \right\rangle . \tag{22}$$

are not probability amplitudes because the involved unitary operators represents the coupling of the apparatus and the system, whereas the above ket and bra belong to system's and the apparatus' Hilbert space only, respectively. Therefore they are *operators*. In particular, the amplitude operator $\hat{\vartheta}(a_m)$ describes all steps of the measurement of a given observable:

- *Preparation* of the initial state of the system (input $|\psi\rangle$),
- Unitary evolution (coupling or *premeasurement*) of the apparatus together with the object system (the bridge provided by $\hat{U}_t$), and
- *Detection* by the apparatus (output $\langle a_m|$).

This fully corresponds to the classically case previous examined. In both cases, we pass information from the past to the future through some current connection.

System and Apparatus

I assume that both the system $\mathcal{S}$ and the apparatus $\mathcal{A}$ are represented by two–level systems. In particular, I choose the operator $\hat{\sigma}_z^{\mathcal{S}}$ for the observable of $\mathcal{S}$ and $\hat{\sigma}_x^{\mathcal{A}}$ for the observable of the apparatus. The system $\mathcal{S}$ is initially prepared in a superposition of the two eigenstates of $\hat{\sigma}_z^{\mathcal{S}}$, which are the spin–up and the spin–down state, respectively, given by

$$|\uparrow\rangle_{\mathcal{S}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad |\downarrow\rangle_{\mathcal{S}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} , \tag{23}$$

so that

$$|\psi(0)\rangle_{\mathcal{S}} = c_{\uparrow} |\uparrow\rangle_{\mathcal{S}} + c_{\downarrow} |\downarrow\rangle_{\mathcal{S}} . \tag{24}$$

The apparatus is initially in the z spin–down state

$$|\phi(0)\rangle_{\mathcal{A}} = |\downarrow\rangle_{\mathcal{A}} , \tag{25}$$

so that if, after the interaction, the system is in $|\downarrow\rangle_{\mathcal{S}}$, the state of $\mathcal{A}$ remains unchanged. Otherwise, it will become $|\uparrow\rangle_{\mathcal{A}}$. The interaction Hamiltonian can then be explicitly written as

$$\hat{H}_{\mathcal{SA}} = \varepsilon \left(1 + \hat{\sigma}_z^{\mathcal{S}}\right) \hat{\sigma}_x^{\mathcal{A}} , \tag{26}$$

where ε is some coupling function.

Eigekets

The first step we need to make in order to calculate the action of the unitary operator $\hat{U}_t$ on the initial state $|\Psi(0)\rangle_{\mathcal{SA}}$, is to diagonalize the matrix $\hat{H}_{\mathcal{SM}}$. The observable $\hat{\sigma}_z^{\mathcal{S}}$ is already diagonal with respect to the basis states (23). In fact, we have

$$\hat{\sigma}_z^{\mathcal{S}} \left|\uparrow\right\rangle_{\mathcal{S}} = \left|\uparrow\right\rangle_{\mathcal{S}} \quad \text{and} \quad \hat{\sigma}_z^{\mathcal{S}} \left|\downarrow\right\rangle_{\mathcal{S}} = - \left|\downarrow\right\rangle_{\mathcal{S}} \ . \tag{27}$$

On the other hand, the two eigenvalues of $\hat{\sigma}_x^{\mathcal{A}}$ are ± 1 and the eigenkets are given by

$$\left|\uparrow\right\rangle_x^{\mathcal{A}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \left|\downarrow\right\rangle_x^{\mathcal{A}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} , \tag{28}$$

respectively. In terms of the z spin–up and spin–down states, these eigenkets are

$$\left|\uparrow\right\rangle_x^{\mathcal{A}} = \frac{1}{\sqrt{2}} \left(\left|\uparrow\right\rangle_{\mathcal{A}} + \left|\downarrow\right\rangle_{\mathcal{A}} \right) , \qquad \left|\downarrow\right\rangle_x^{\mathcal{A}} = \frac{1}{\sqrt{2}} \left(\left|\uparrow\right\rangle_{\mathcal{A}} - \left|\downarrow\right\rangle_{\mathcal{A}} \right) . \tag{29}$$

Evolving

Now, in order to find the time evolution of the quantum state of the compound system, we only need to write its initial state in terms of the eigenkets (23) and (28) of $\hat{H}_{\mathcal{SA}}$

$$\begin{aligned} |\Psi(t)\rangle_{\mathcal{SA}} &= \hat{U}_t \, |\Psi(0)\rangle_{\mathcal{SA}} \\ &= e^{-\frac{\imath}{\hbar} t\varepsilon \left(1+\hat{\sigma}_z^{\mathcal{S}}\right) \hat{\sigma}_x^{\mathcal{A}}} \left[\left(c_\uparrow \, |\uparrow\rangle_{\mathcal{S}} + c_\downarrow \, |\downarrow\rangle_{\mathcal{S}} \right) |\downarrow\rangle_{\mathcal{M}} \right] , \end{aligned} \tag{30}$$

where $|\Psi(0)\rangle_{\mathcal{SA}} = |\psi(0)\rangle_{\mathcal{S}} \, |\phi(0)\rangle_{\mathcal{A}}$ and $\hat{U}_t = e^{-\frac{\imath}{\hbar} t\varepsilon \left(1+\hat{\sigma}_z^{\mathcal{S}}\right) \hat{\sigma}_x^{\mathcal{A}}}$. The action of the Hamiltonian onto its eigenkets simply returns the corresponding eigenvalues. Therefore, since

$$|\downarrow\rangle_{\mathcal{A}} = \frac{1}{\sqrt{2}} \left(|\uparrow\rangle_x^{\mathcal{A}} - |\downarrow\rangle_x^{\mathcal{A}} \right) , \tag{31}$$

we have

$$\begin{aligned} |\Psi(t)\rangle_{\mathcal{SA}} &= \frac{1}{\sqrt{2}} \Big(c_\uparrow e^{-\frac{2\imath}{\hbar} t\varepsilon} \, |\uparrow\rangle_{\mathcal{S}} \, |\uparrow\rangle_x^{\mathcal{A}} + c_\downarrow \, |\downarrow\rangle_{\mathcal{S}} \, |\uparrow\rangle_x^{\mathcal{A}} \\ &\quad - c_\uparrow e^{+\frac{2\imath}{\hbar} t\varepsilon} \, |\uparrow\rangle_{\mathcal{S}} \, |\downarrow\rangle_x^{\mathcal{A}} - c_\downarrow \, |\downarrow\rangle_{\mathcal{S}} \, |\downarrow\rangle_x^{\mathcal{A}} \Big) , \end{aligned} \tag{32}$$

where the exponential $e^{-\frac{\imath}{\hbar} t\varepsilon \left(1+\hat{\sigma}_z^{\mathcal{S}}\right) \hat{\sigma}_x^{\mathcal{A}}}$ is 1 when the eigenvalue of $\hat{\sigma}_z^{\mathcal{S}}$ is -1, i.e. for the state $|\downarrow\rangle_{\mathcal{S}}$.

Fine–Tuning

Substituting expressions (29) into Eq. (32), we obtain

$$|\Psi(t)\rangle_{\mathcal{SA}} = \frac{1}{\sqrt{2}}\left(-c_{\uparrow}\frac{2\imath}{\sqrt{2}}\sin\frac{2t\varepsilon}{\hbar}\,|\uparrow\rangle_{\mathcal{S}}\,|\uparrow\rangle_{\mathcal{A}} + \sqrt{2}c_{\downarrow}\,|\downarrow\rangle_{\mathcal{S}}\,|\downarrow\rangle_{\mathcal{A}} + c_{\uparrow}\frac{2\imath}{\sqrt{2}}\cos\frac{2t\varepsilon}{\hbar}\,|\uparrow\rangle_{\mathcal{S}}\,|\downarrow\rangle_{\mathcal{A}}\right) . \tag{33}$$

Choosing now

$$\frac{2t\varepsilon}{\hbar} = \frac{\pi}{2}\,, \quad \text{or} \quad t = \frac{\pi\hbar}{4\varepsilon}\,, \tag{34}$$

i.e. by fine–tuning the interaction time, we finally obtain

$$|\Psi(t)\rangle_{\mathcal{SA}} = +c_{\downarrow}\,|\downarrow\rangle_{\mathcal{S}}\,|\downarrow\rangle_{\mathcal{A}} - \imath c_{\uparrow}\,|\uparrow\rangle_{\mathcal{S}}\,|\uparrow\rangle_{\mathcal{A}}\ , \tag{35}$$

which is the required coupling between the system and the meter.

Matrices

This formalism can be easily translated in the density–matrix formalism above.:

$$\hat{\rho}^{\mathcal{S}} = |\psi(0)\rangle_{\mathcal{S}}\langle\psi(0)| \quad \text{and} \quad \hat{\rho}^{\mathcal{A}} = |\phi(0)\rangle_{\mathcal{A}}\langle\phi(0)| \,. \tag{36}$$

The probability to get e.g. the spin-up state of the apparatus is:

$$\begin{aligned} \wp(\uparrow_{\mathcal{A}}) &= \mathrm{Tr}_{\mathcal{A}}\left[|\uparrow\rangle_{\mathcal{A}}\langle\uparrow| \,\mathrm{Tr}_{\mathcal{S}}\left(\hat{U}_t\hat{\rho}^{\mathcal{S}}\hat{\rho}^{\mathcal{A}}\hat{U}_t^{\dagger}\right)\right] \\ &= \mathrm{Tr}_{\mathcal{A}}\left(|\uparrow\rangle_{\mathcal{A}}\langle\uparrow| \,|c_{\uparrow}|^2\right) \\ &= |c_{\uparrow}|^2, \end{aligned} \tag{37}$$

whilst the relative effect is

$$\begin{aligned} \hat{E}(\uparrow_{\mathcal{A}}) &= \mathrm{Tr}_{\mathcal{S}}\left(\hat{U}_t^{\dagger}\,|\uparrow\rangle_{\mathcal{A}}\langle\uparrow|\,\hat{U}_t\hat{\rho}^{\mathcal{S}}\right) \\ &= {}_{\mathcal{S}}\langle\psi(0)|\left(\hat{U}_t^{\dagger}\,|\uparrow\rangle_{\mathcal{A}}\langle\uparrow|\,\hat{U}_t\,|\psi(0)\rangle_{\mathcal{S}}\langle\psi(0)|\right)|\psi(0)\rangle_{\mathcal{S}} \\ &= {}_{\mathcal{S}}\langle\psi(0)|\,\hat{U}_t^{\dagger}\,|\uparrow\rangle_{\mathcal{A}}\langle\uparrow|\,\hat{U}_t\,|\psi(0)\rangle_{\mathcal{S}}\,, \end{aligned} \tag{38}$$

where

$$\hat{\vartheta}^{\dagger} = {}_{\mathcal{S}}\langle\psi(0)|\,\hat{U}_t^{\dagger}\,|\uparrow\rangle_{\mathcal{A}} \quad \text{and} \quad \hat{\vartheta} = {}_{\mathcal{A}}\langle\uparrow|\,\hat{U}_t\,|\psi(0)\rangle_{\mathcal{S}}\,. \tag{39}$$

Preparation

Let us now come back to the conceptual problems. The *preparation* of the object system in a certain state is therefore the first step. With such an operation we have ensured that there is an initial source of variety. We do not need to assume that at this stage we have already selected message (as it is often thought in classical communication theory). Since the same state can be selected by different preparation procedures, we may operationally define the state as an equivalence class of preparation procedures.

Pre–measurement

Coupling between the object system and apparatus or *pre–measurement*. The purpose of this operation is to select a specific observable. Indeed, different observables (especially non–commuting ones) require different experimental set-ups. It is true that we can express the state of the two systems (i.e. apparatus and object system) involved here in different bases which correspond each time to a different observable (basis degeneracy). However, we should not mix this mathematical formalism with the operational issue connected with a selected experimental set up. In the latter case, a change of observable implies a change of set up too, what means a different kind of pre–measurement. Since the same observable can be selected through different pre–measurement set–ups, we can define the observable as an equivalence class of pre–measurements.

Detection

Finally, the *detection* event, thanks to the previous coupling and the previous preparation, will allow us to assign a certain property to the system. Events in themselves only happen and it is only thanks to such a coupling (the chosen experimental set up) that they are able to tell something about the object system. Nevertheless, the specificity of any event is to constitute the only irreversible step in the whole measurement procedure. We can define a property as an equivalence class of detections: thanks to quantum theory and given a certain event we are able to assign a property. However, we are bale to do that in presence of different detection events.

Conclusion

The new challenge for the future is to understand how such a formalism can help us to deal in new a generalized way with the information exchanges in our universe.

Auletta, G., Fortunato, M., and Parisi, G., *Quantum Mechanics*, Cambridge, University Press, 2009.

Auletta, G. and Torcal, L., From Wave–Particle to Features–Event Complementarity, *International Journal of Theoretical Physics* **50** (2011):3654–68.