Initial entanglement between detectors allows violating Heisenberg’s uncertainty relations

Antonio Di Lorenzo
E-mail: dilorenzo@infis.ufu.br
Universidade Federal de Uberlândia, Uberlândia, MG, Brazil

Abstract. Physics textbooks introduce an inequality derived by Kennard that concerns the impossibility of preparing a quantum state with well defined momentum and position. However, Heisenberg had formulated two different inequalities: one stating the impossibility of a simultaneous detection of position and momentum with arbitrary precision, and another one imposing a tradeoff between the precision on the measurement of a variable and the backaction on a subsequent measurement of a conjugated variable. Here, we explore the connections among these three inequalities, and we show that the latter can be violated, if the detectors are initially entangled. The results, besides being of fundamental interest, can be useful for building up an ideal momentum, or position, detector (i.e. one that introduces no noise in the measurement, besides the intrinsic statistical noise of the input state).

1. Introduction
Recently, there have been several experiments claiming to have violated Heisenberg’s uncertainty relation [1–4]. At first sight, this may seem as revolutionary, since the uncertainty principle is one of the pillars of quantum mechanics. However, there are two important considerations to make: (1) Heisenberg [5] formulated an inequality about the uncertainty of a position measurement and the disturbance in a subsequent, or simultaneous, momentum measurement. The inequality demonstrated in quantum mechanics textbooks is due to Kennard [6], who proved it shortly after Heisenberg’s semi–qualitative formulation, and it concerns the impossibility of preparing an ensemble of systems such that a measurement of momentum in a subensemble yields a variance \( \sigma_P^2 \) and a measurement of position in another subensemble, at the same time, yields a variance \( \sigma_Q^2 < \hbar^2 / 4 \sigma_P^2 \). (2) The inequality violated in the experiments concerns finite–dimensional systems, and was not formulated by Heisenberg. Instead, this inequality was retrofitted, following the reasoning: Heisenberg formulated an uncertainty relation for uncertainty in position and disturbance in a later measurement of momentum and another relation for the uncertainties in a joint measurement of momentum and position

\[
\varepsilon_Q \eta_{P|Q} \gtrsim \hbar, \quad \text{for sequential measurements}, \tag{1a}
\]
\[
\varepsilon_Q \eta_{P|Q} \gtrsim \hbar, \quad \text{for joint measurements}; \tag{1b}
\]

Kennard proved a rigorous inequality for the preparation variances

\[
\sigma_Q \sigma_P \geq \frac{\hbar}{2}; \tag{1c}
\]
Robertson and Schrödinger generalized Kennard’s result to finite–dimensional systems \[7,8\]

\[
\sigma_A \sigma_B \geq \frac{1}{2} |\langle [\hat{A}, \hat{B}] \rangle|; \tag{1d}
\]

therefore, Heisenberg would have endorsed the corresponding inequality for uncertainty in a variable and disturbance in the other variable, at least according to Ozawa \[9\]

\[
\varepsilon_A \eta_{B|A} \geq \frac{1}{2} |\langle [\hat{A}, \hat{B}] \rangle|, \tag{1e}
\]

\[
\varepsilon_{A|B} \varepsilon_{B|A} \geq \frac{1}{2} |\langle [\hat{A}, \hat{B}] \rangle|. \tag{1f}
\]

However, a simple inspection reveals that equations (1e) and (1f) cannot hold. Indeed, for a finite–dimensional system, both the uncertainty \(\varepsilon\) and the disturbance \(\eta\), whatever its definition is, must be finite. On the other hand, the precision of a measurement can be made, in principle, arbitrary small. Hence the products \(\varepsilon_A \eta_{B|A}\) and \(\varepsilon_{A|B} \varepsilon_{B|A}\) cannot have a lower bound. This consideration is to be contrasted with the original inequalities formulated in (1a) and (1b), for an infinite–dimensional Hilbert space. In this case, indeed, one can imagine that the disturbance \(\eta_{P|Q}\) or the uncertainty \(\varepsilon_{P|Q}\) diverge if the uncertainty \(\varepsilon_{Q|P}\) tends to zero, so that the inequalities make sense, and cannot be discarded a priori.

The goal of the present manuscript, therefore, is to verify whether the inequalities (1a) and (1b) can be violated. In order to do this, it is necessary to define the uncertainty of a measurement, and the disturbance of a subsequent or simultaneous measurement. It turns out that the latter definition is a very controversial subject \[10–13\]. Therefore, the violation of the Heisenberg inequality in the uncertainty–disturbance formulation depends fatally on one’s definition of disturbance. Here, we shall argue that our definition translates the intuitive notion of disturbance of a measurement. Furthermore, independently of the definition, we shall prove a somewhat surprising corollary: it is possible to make a very precise measurement of momentum by preceding it with an imprecise measurement of position. This results is uncontroversial, and it opens up a new possibility for realizing a precise measurement apparatus. So far, a precise apparatus should be prepared in a very sharp state of the pointer variable \(J\), i.e. a state very close to the ideal state \(|J = 0\rangle\). The result of the present manuscript, however, allows to make a very precise measurement, say, of momentum, even though the detector is not in such an initial state, provided that its output variable is (almost) perfectly anticorrelated to the input variable of the first measurement. In quantum mechanics, perfect anticorrelations can be achieved only by means of entanglement. Hence, our main result can be summarized as in the title of this contribution: Initial entanglement between detectors allows violating Heisenberg’s uncertainty relations.

A short account of the results presented here was given in Ref. \[12\]. Bullock and Busch have expanded the subject further \[14\]. Here, we shall present a more detailed derivation, in particular we shall justify the additional term due to the finite resolution of the readout that contributes to the uncertainty of the measurement. By making only the hypothesis of a uniform resolution of the detectors, we prove that the finite resolution contributes with additive terms to the cumulants of the observable outputs. We also include an appendix on the connection with the POVM approach that so far was published only in the pre–print version of \[12\].

2. Definitions
In the following, we define what we mean by ‘uncertainty’ and ‘disturbance’ in operational terms. Our definitions do not rely on any specific theory, so that they can be translated into different mathematical formulas depending on the underlying theory. In the next section, we shall provide the translation of these definitions in quantum mechanical terms.
2.1. Uncertainty
The uncertainty in a measurement is not in any way connected to the sample variance. A simple illustration is the following: suppose you are measuring the height of a large group of people. Say that the histogram is well approximated by a Gaussian with average $\bar{h} = 175$.cm and a variance $\sigma_h = 10$.cm. Let us consider a single person, whose height is reported by our apparatus as $h = 169.3$.cm. What is the uncertainty in the height of this person? Certainly not 10.cm! The uncertainty on an individual measurement depends on the resolution of the meter, and on how the meter interacted with the measured system. The uncertainty cannot be lower than the resolution, which for a typical tape is 1mm (if we exclude interpolation due to the human eye being a meter per se). For instance, was the 0 of the tape placed at the feet of the person, or by mistake a non-zero value was used? Was the tape held straight? And so on. So, how do we estimate the uncertainty in an individual measurement? Usually, we do by calibration. We
measure a large set of identically prepared systems, all of them having a value \([A_0 - \delta/2, A_0 + \delta/2]\), where \(\delta\) is in principle the best precision that we can achieve in the preparation, and we observe the outputs \(A\) of the apparatus. Unless our measuring apparatus is better than our preparation apparatus, we shall observe a distribution centered in \(\langle A \rangle\) with a spread \(\Delta > \delta\). The quadratic difference \(\varepsilon^2 = \Delta^2 - \delta^2\) is the statistical uncertainty of our apparatus. Usually, \(\delta\) is negligible compared to \(\Delta\). We note that in principle the uncertainty varies with the output \(h\).

We also introduce the systematic uncertainty (error, in the common terminology): it is the difference between the average value \(\langle A \rangle\) in the limit of a large number of observations \(N \to \infty\), and the ‘true’ value \(A_0\), \(U_{A_0} = |\langle A \rangle - A_0|\). This values is also known a priori, in principle.

2.2. Disturbance

The disturbance is the most controversial quantity to be defined. Since in quantum mechanics it is problematic to assume that a system possesses a property, unless it is prepared in an eigenstate of the corresponding observable, we avoid interpreting the disturbance as being a perturbation in an observable, but we interpret this concept as a disturbance of a measurement of an observable, i.e. we maintain an operational stance, in line with the Bohr’s approach to quantum mechanics, and we choose to speak only of results of observations.

We say that a measurement of a quantity \(B\) at a time \(t_B\) is disturbed by a measurement of a quantity \(A\) at time \(t_A\) if the variance increases or the average changes with respect to the case when the measurement of \(A\) is not performed. Let \(\langle B \rangle\) and \(\Delta_B^2\) the average value and the variance observed when the measurement of \(A\) is not performed, and \(\langle B \rangle_A\) and \(\Delta_B^2|A_0\) the corresponding values when \(A\) is measured.\(^1\) We shall also define the same values for the case when the measurement \(A\) is performed and the corresponding output is \(A_0\), \(\langle B \rangle_{A_0}\) and \(\Delta_B^2|A_0\).

We define the average statistical disturbance and the average systematic disturbance as \(\eta_{B|A}^2 = \Delta_{B|A}^2 - \Delta_B^2\) and \(D_{B|A} = |\langle B \rangle_A - \langle B \rangle|\). We define also the statistical disturbance and the systematic disturbance as \(\eta_{B|A_0}^2 = \Delta_{B|A_0}^2 - \Delta_B^2\) and \(D_{B|A_0} = |\langle B \rangle_{A_0} - \langle B \rangle|\).

It is worth noting that Busch, Lahti and Werner \([11]\) propose a different definition of disturbance. The statistics of the \(B\) measurement when the \(A\) measurement is performed is compared to the statistics of an ideal \(B\) measurement when \(A\) is not performed, \(\eta_{B|A}^2 = \Delta_{B|A}^2 - \Delta_{B|A}^{2 \text{(ideal)}}\). This formulation is in line with the interpretation that what is disturbed is the observable, not the measurement of the observable. Besides this difference in the definition, our results agree with those of \([11]\).

Our definition, on the other hand, should be taken with a grain of salt. Let us imagine a broken detector, which always yields the same output \(B_0\). This measurement would not be disturbed by any measurement performed before it, trivially, and the Heisenberg relation \((1a)\) would be violated. Our definition should be applied only to a ‘faithful’ measurement. By this term, we mean a measurement that yields different average outputs for different inputs, provided that the inputs are sufficiently far from each other (if their difference is below the resolving power of the measuring apparatus, of course the output will be the same).

3. Methods

In the theory of sequential or joint measurements it is usually assumed that the initial state of the system and of the two probes factorizes as

\[
\rho_0 = \rho_A \otimes \rho_B \otimes \rho_{\text{sys}}.
\]  \hspace{1cm} (2)

\(^1\) Clearly, causality implies that \(\langle B \rangle = \langle B \rangle_A\) and \(\Delta_B^2 = \Delta_{B|A}^2\) if the measurement of \(A\) is performed after \(B\) (or better, if the measurement \(B\) is either in the past light cone or outside the light cone of the measurement of \(A\)). Our definition, however, holds also if causality were violated.
However, we shall consider the general case when the probes are initially correlated, so that
the probes. If the probes are initially uncorrelated, then

\[ \rho_0 = \rho_{A,B} \otimes \rho_{sys}. \]  

We assume a von Neumann interaction, leading to the following time–evolution of the system
and the meters

\[ U = \exp [i \hat{P} \hat{\Phi}_B] \exp [i \hat{Q} \hat{\Phi}_A], \text{ for sequential measurements}, \]

\[ U = \exp [i \hat{P} \hat{\Phi}_B + i \hat{Q} \hat{\Phi}_A], \text{ for joint measurements}, \]

with \( \hat{\Phi}_\alpha \) representing the input variable of the detector \( \alpha \). We consider the readout states of
the detectors \( \hat{F}_A(\mu) \), described in the appendix, as Gaussian. It is convenient to work with
the characteristic function \( Z \) of the readouts, i.e. the Fourier transform of the joint probability
\( \mathcal{P}(\mu_A, \mu_B) \).

For later convenience, we introduce the column vectors (indicated by horizontal lists within
square brackets) \( s = [q, p], \phi = [\phi_P, \phi_Q], j = [p, q] \). The components of the vectors are labeled
by an index \( \alpha \) taking values in the set \( \{P, Q\} \), so that \( s_P = q \) and \( s_Q = p \). We also introduce the vector \( S = [P, Q] \). With these definitions, e.g. the term \( pQ + qP \) can be written as \( S \cdot s \).

For a 1–dimensional system, the Moyal quasi–characteristic function \( M(q, p) \) [15] is defined as the Fourier transform of the Wigner quasiprobability \( W(P, Q) \) [16].

\[ M(q, p) = \int dPdQ e^{i(Pq + Qp)} W(P, Q). \]

We recall that \( M(q, 0) = Z(q) \) is the characteristic function for the probability distribution in \( K \), i.e. \( M(q, 0) = \int dP \exp(iPq)(P|0|P) \), and \( M(0, q) \) is the characteristic function for the probability in \( Q \). Let \( M_{sys}(s) \) the initial Moyal function of the system, and \( M_{det}(\phi, j) \) that of the probes. If the probes are initially uncorrelated, then \( M_{det}(\phi, j) = M_Q(\phi_Q, j_Q)M_P(\phi_P, j_P) \). However, we shall consider the general case when the probes are initially correlated, so that \( M_{det}(\phi, j) \neq M_Q(\phi_Q, j_Q)M_P(\phi_P, j_P) \). The variances and covariances of the probes can be expressed as logarithmic derivatives of their Moyal function:

\[ \delta_\alpha^2 = \langle \hat{J}_\alpha^2 \rangle - \langle \hat{J}_\alpha \rangle^2 = - \frac{\partial^2 \ln M_{det}(\phi, j)}{\partial \phi_\alpha^2} \bigg|_{\phi=j=[0,0]}, \]  

\[ \delta_\alpha^2 = \langle \hat{\Phi}_\alpha^2 \rangle - \langle \hat{\Phi}_\alpha \rangle^2 = - \frac{\partial^2 \ln M_{det}(\phi, j)}{\partial j_\alpha^2} \bigg|_{\phi=j=[0,0]}, \]  

\[ \xi = \langle \hat{\Phi}_P \hat{J}_Q \rangle - \langle \hat{\Phi}_P \rangle \langle \hat{J}_Q \rangle = - \frac{\partial^2 \ln M_{det}(\phi, j)}{\partial \phi_Q \partial j_P} \bigg|_{\phi=j=[0,0]}, \]

\[ \kappa = \langle \hat{\Phi}_Q \hat{J}_P \rangle - \langle \hat{\Phi}_Q \rangle \langle \hat{J}_P \rangle = - \frac{\partial^2 \ln M_{det}(\phi, j)}{\partial \phi_P \partial j_Q} \bigg|_{\phi=j=[0,0]}. \]

We defined as \( \delta_\alpha^2 \) the initial variance in the input variable \( \Phi_\alpha \) and as \( \delta_\alpha^2 \) the initial variance in
the output variable \( J_\alpha \). By Kennard relation, applied to the probes,

\[ \delta_\alpha \delta_\alpha \geq \frac{1}{2}. \]

We shall use the relation between the final Moyal function for system and probes in terms of
the initial ones [17, 18]

\[ M(s; \phi, j) = M_{sys}(s + \phi)M_{det}(\phi, j + 2\alpha_0 s + \alpha_r \phi), \]

The main idea of the present work is to question the preceding assumption. While it is
uncontroversial that the initial state of the system must be uncorrelated to that of the detectors
(otherwise, we would have a measurement before the interaction started), there is no compelling
reason to assume that the two detectors are initially uncorrelated. Therefore, we shall write, in
full generality,

\[ \rho_0 = \rho_{A,B} \otimes \rho_{sys}. \]
where
\[
\alpha_+ = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \alpha_- = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}, \quad \alpha_0 = \frac{\alpha_+ + \alpha_-}{2}.
\] (10)

The derivation of this equation is provided in the Supplemental Material of Ref. [18]. The symbol \( \tau \in \{+, -, 0\} \) specifies the order of the measurements: \( \tau = 0 \) means joint measurements, \( \tau = + \) means a position measurement followed by a momentum measurement, and \( \tau = - \) means the vice versa. We recall that the Moyal function in Eq. (9) provides a complete description of the system+probes, and it is equivalent to knowing the density matrix. The characteristic function of the system (\( s = 0 \)) and by putting \( j = 0 \) in Eq. (9). We note that \( \tau \) can be inferred directly from the experimental data.

References [17, 18], in deriving (9), assumed as readout states the projectors \( \hat{F}_{det}(\mu) = |J = \mu\rangle \langle J = \mu| \). In a realistic measurement with finite resolution, however, \( \hat{F}_{det}(\mu) = \int dJ p(\mu|J)\Pi(J) \), where \( p(\mu|J) \) is the probability of observing an output \( \mu \) given that the value of the pointer is \( J \). We recall that for a uniform resolution this latter probability depends only on the difference of its arguments, \( p(J|\mu) = \frac{1}{2\delta J} \). Here, we make only this hypothesis on the resolution probability, we do not assume e.g., that it is a Gaussian. The results of Refs. [17, 18] are then easily generalized thanks to the convolution theorem: the characteristic function, i.e. the Fourier transform of the joint probability of observing \( \mu = (\mu_P, \mu_Q) \) as the output,

\[
Z(\phi) = \int d\mu e^{i\mu \cdot \phi} \rho(\mu) = \int d\mu dJ e^{i\mu \cdot \phi} f(\mu - J)\Pi(J),
\] (11)

is obtained by introducing a factor \( z(\phi) = \int d\mu e^{i\mu \cdot \phi} f(\mu) \)

\[
Z(\phi) = z(\phi)M_{sys}(\phi)M_{det}(\phi, \alpha_\tau \phi).
\] (12)

For definiteness, we consider \( \tau = + \), i.e. a measurement of position is made first, and a measurement of momentum follows. Then, \( Z(\phi) = z(\phi)M_{sys}(\phi)M_{det}(\phi, [0, \phi_P]) \). Notice the appearance of the term \( [0, \phi_P] \) as the second argument of \( M_{det} \). By taking the second derivatives of \( \ln (Z) \), we obtain the statistical variances of the readout

\[
\Delta_Q^2 = -\frac{\partial^2 \ln Z(\phi)}{\partial \phi_Q^2} \bigg|_{\phi=0} = \sigma_Q^2 + \delta_Q^2 + \delta_Q^2,
\] (13)

\[
\Delta_{P,Q}^2 = -\frac{\partial^2 \ln Z(\phi)}{\partial \phi_P^2} \bigg|_{\phi=0} = \sigma_P^2 + \delta_P^2 + \delta_P^2 + \tilde{\delta}_Q^2 + 2\kappa.
\] (14)

Since the finite resolution enters with a multiplicative term, we find that its contribution to the cumulants is simply of adding a term. The last two terms in Eq. (14) arise because of the derivation rule for a function \( g(u) = f(u, u) \)

\[
\frac{d^2g(u)}{du^2} = \left[ \frac{\partial^2 f(u, v)}{\partial u^2} + \frac{\partial^2 f(u, v)}{\partial v^2} + 2\frac{\partial^2 f(u, v)}{\partial u \partial v} \right]_{v=u},
\] (15)

with \( f(u, v) \rightarrow \ln \{M_{det}([\phi_P, 0], [0, jQ])\} \).

If the \( Q \) measurement were not performed, the statistical variance in the \( P \) measurement would be

\[
\Delta_P^2 = \sigma_P^2 + \delta_P^2 + \delta_P^2.
\] (16)

Therefore the statistical disturbance is

\[
\eta_{P,Q}^2 \equiv \Delta_{P,Q}^2 - \Delta_P^2 = \delta_Q^2 + 2\kappa.
\] (17)
Here, $\kappa$ is the initial covariance between the input variable $\Phi_Q$ of the first probe and the output variable $J_P$ of the second probe.

The statistical noise for the position measurement is

$$\epsilon_Q^2 \equiv \Delta_Q^2 - \sigma_Q^2 = \delta_Q^2 + \delta_Q^2, \quad (18)$$

while for the momentum measurement

$$\epsilon_P^2 \equiv \Delta_P^2 - \sigma_P^2 = \delta_P^2 + \delta_P^2 + \tilde{\delta}_Q^2 + 2\kappa \quad (19)$$

We can make the systematic error $D_Q = \langle J_Q \rangle$ and disturbance $D_{P|Q} = \langle J_P \rangle + \langle \Phi_Q \rangle$ vanish, by applying two appropriate translations to the initial state of the probe (since we are looking for minimal disturbance and error, it is sensible to assume that the probes are not biased).

4. Results and applications

Inspection of (17) reveals that the square of the disturbance can not only be arbitrary small, but even negative. This means that the first measurement actually did not disturb the second one, but rather it helped make it sharper. At the same time, while the disturbance reaches zero, the error (or uncertainty) in the first measurement does not diverge. Therefore, the error–disturbance relation can be violated. In order for $\eta_P^2$ to become negative, it is necessary that the correlator $\kappa$ defined in (7d) becomes negative. The lowest disturbance, and therefore the lowest error in the second measurement, is achieved when the input variable $\Phi_Q$ of the first probe is perfectly anti–correlated to the output variable $J_P$ of the second probe.

On the other hand, multiplying the errors in (18) and (19) yields

$$\epsilon_Q^2 \epsilon_P^2 = (\delta_Q^2 + \delta_Q^2)(\delta_P^2 + \delta_P^2 + \tilde{\delta}_Q^2 + 2\kappa) \geq \delta_Q^2 \delta_P^2 + \Phi_Q \geq \frac{1}{4}, \quad (20)$$

because of Kennard relation applied to the variable $\hat{J}_Q$ and $\hat{J}_P + \hat{\Phi}_Q$, which are canonically conjugated.

In summary, we found that the error–disturbance relation can be violated, with our definition of disturbance, if the detectors have initial correlations, while the error–error relation cannot be violated. We note that what we define as the error is equivalent to the disturbance of [11, 14], and therefore, using this alternative definition of disturbance, one would conclude that the error–disturbance relation cannot be violated. Irrespectively of the definition, there is a useful consequence of our result: it is possible to make a measurement of momentum infinitely sharp by preceding it with a position measurement. This is because the former measurement perturbs the momentum by a quantity $\Phi_Q$, the write–in variable of the position probe. If this variable is perfectly anti–correlated to the readout variable $J_P$ of the second probe, the outcome of the latter will be automatically corrected. Now, in quantum mechanics, two variables can be perfectly anti–correlated only if the two probes are in a maximally entangled state. Specifically, they need to be in an Einstein–Podolsky–Rosen state [19].

Acknowledgments

This work was performed as part of the Brazilian Instituto Nacional de Ciência e Tecnologia para a Informação Quântica (INCT–IQ) and it was supported by Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG).

We are referring to the original formulation with continuous variables of the so–called EPR paradox, not to the more commonly known discrete version of Bohm and Aharonov [20].
Appendix A. Description of the measurement
In a measurement, a meter interacts with a quantum system. What sets apart the meter from the system is its ability to cause a sensation in a human being, if not directly, through a chain of amplifications and further interactions with the visible electromagnetic field. Here, the effect of this von Neumann chain [21] is accounted for by considering the readout state of the probe to be a mixed state.

We can justify this assumption in the following way: There are many microscopic states that correspond to the same perceived readout \( J \) of a detector, let them \( |J,m\rangle \), with \( m \) a collective quantum number. The probability of observing an output \( J \) is therefore

\[
\mathcal{P}(J) = \sum_m \langle J, m| \otimes \langle s| U(\hat{\rho}_{sys} \otimes \hat{\rho}_{det}) U^\dagger |J, m \rangle \otimes |s \rangle = \text{Tr}\{(\mathbb{1} \otimes \hat{\Pi}_{det}(J))U(\hat{\rho}_{sys} \otimes \hat{\rho}_{det})U^\dagger\},
\]

where \( U \) is the time-evolution operator, and \( \hat{\Pi}_{det}(J) \) is the projection operator on the subspace spanned by \( |J, m\rangle \) when \( m \) varies for fixed \( J \). Next, we consider that there is always some finite resolution in the observation and that the observation may fail, both these processes being characterized by a conditional probability \( p(\mu |J) \) of observing a value \( \mu \) when the actual value is \( J \). Strictly speaking, the observed \( \mu \) should take discrete values, however, since we are more familiar with integrals than with sums, it is often treated as a continuous variable.

The probability of observing \( \mu \) is therefore

\[
\mathcal{P}(\mu) = \int dJ p(\mu |J) \text{Tr}\{(\mathbb{1} \otimes \hat{\Pi}_{det}(J))U(\hat{\rho}_{sys} \otimes \hat{\rho}_{det})U^\dagger\},
\]

\[
\mathcal{P}(\mu) = \text{Tr}\{(\mathbb{1} \otimes \hat{F}_{det}(\mu))U(\hat{\rho}_{sys} \otimes \hat{\rho}_{det})U^\dagger\},
\]

where \( \hat{F}_{det}(\mu) = \int dJ p(\mu |J) \hat{\Pi}_{det}(J) \) is a family of positive operators describing the readout states. These must satisfy the normalization \( \int d\mu \hat{F}_{det}(\mu) = \mathbb{1} \), with \( d\mu \) a Lebesgue-Stieltjes measure, corresponding to a discrete or continuous distribution of outputs \( \mu \).

The readout states \( \hat{F}_{det}(\mu) \) are not normalized, but in general \( \text{Tr}[\hat{F}_{det}(\mu)] = pr(\mu) \neq 1 \). The quantities \( pr(\mu) \) are the priors [22], and they are not necessarily one, unless \( \hat{F}_{det}(\mu) \) is a rank–1 projector. Instead, they represent the unnormalized probability of obtaining an outcome \( \mu \) in absence of any other information. They are not a conventional probability, since

\[
\int d\mu pr(\mu) = \int d\mu \text{Tr}[\hat{F}_{det}(\mu)] = \text{Tr}\mathbb{1} = d,
\]

with \( d \) the dimension of the Hilbert space of the meter (which could be infinite). However, the only relevant quantity are the ratios \( pr(\mu_1)/pr(\mu_2) \) representing relative probabilities.

Furthermore, it follows from their definition that the readout states are classical, i.e. \( [\hat{F}_{det}(\mu_1), \hat{F}_{det}(\mu_2)] = 0 \), \( \forall \mu_1, \mu_2 \). This hypothesis allows to individuate one, or more, privileged basis, \( |J\rangle \), the one that diagonalizes at once all \( \hat{F}_{det}(\mu) \). The existence of such a basis can also be justified by the decoherence approach [23, 24]. The \( \hat{F}_{det}(\mu) \) are labeled by the average of the operator \( \hat{J} = \int dJJ |J\rangle \langle J| \) and they have a spread \( \delta'(\mu) \) in \( J \), i.e.,

\[
\mu = \text{Tr}_{det}\{\hat{J}\hat{\rho}_{det}(\mu)\},
\]

\[
\delta'^2(\mu) = \text{Tr}_{det}\{\hat{J}^2\hat{\rho}_{det}(\mu)\} - \left(\text{Tr}_{det}\{\hat{J}\hat{\rho}_{det}(\mu)\}\right)^2,
\]

with \( \hat{\rho}_{det}(\mu) = \hat{F}_{det}(\mu)/pr(\mu) \) the normalized readout states. The spread \( \delta'(\mu) \) is but the resolution of the probe, and in principle it can be different for different outputs.
For instance, we could choose
\[ \hat{F}_{\text{det}}(\mu_n) = \int_{\mu_n - \delta_n/2}^{\mu_n + \delta_n/2} dJ |J\rangle \langle J| \]  
(A.7) and let the readout \( \mu \) take discrete values \( \mu_n \) spaced by \( (\delta_n + \delta_{n+1})/2 \) from one another (then \( d\mu \) is a distribution formed by a sum of Dirac deltas). In this case the priors are \( p_r(\mu_n) = \delta_n' \).

In general, we can write
\[ \hat{F}_{\text{det}}(\mu) = \int dJ p(\mu|J) |J\rangle \langle J|, \]  
(A.8) with \( p(\mu|J) \) a probability distribution for \( \mu \). The priors are thus \( p_r(\mu) = \delta' \). Eq. (A.5) is satisfied for \( p(\mu|J) = f(\mu - J) \), where \( f(\mu) \) is a probability distribution having zero average and spread \( \delta' \). An important example is given by the Gauss-Laplace distribution,
\[ \hat{F}_{\text{det}}(\mu) = \int_{-\infty}^{+\infty} dJ \exp \left[ -\frac{(J - \mu)^2}{2\delta^2} \right] |J\rangle \langle J|, \]  
(A.9) with \( \mu \in \mathbb{R} \) and uniform priors \( p_r(\mu) = 1 \).

The conditional state of the system for given \( \mu \) is
\[ \hat{\rho}_{\text{sys}|\mu} = \mathcal{P}(\mu)^{-1} \text{Tr}_{\text{det}} \left( (1 \otimes \hat{F}_{\text{det}}(\mu)) U (\hat{\rho}_{\text{sys}} \otimes \hat{\rho}_{\text{det}}) U^\dagger \right). \]  
(A.10) Equivalent formulas arise when treating postselected weak measurement [25, 26], but with the difference that \( U \) is expanded in a perturbation series and the \( \delta \)s of the system and the probe are reversed. Let us introduce the preparation basis \( |I\rangle \), the one in which \( \hat{\rho}_{\text{det}} = \int dI w(I)|I\rangle \langle I| \). We can rewrite the conditional state of the system
\[ \hat{\rho}_{\text{sys}|\mu} = \frac{\int dJdIM_{J,I}(\mu)\hat{\rho}_{\text{sys}}M_{J,I}^\dagger(\mu)}{\int dJdI\text{Tr}_{\text{sys}} \left\{ E_{J,I}(\mu)\hat{\rho}_{\text{sys}} \right\}}. \]  
(A.11) We defined the generalized operations
\[ M_{J,I}(\mu) = \sqrt{p(\mu|J)} w(I)|J\rangle \langle J|, \]  
(A.12) and the generalized effects \( E_{J,I}(\mu) = M_{J,I}^\dagger(\mu)M_{J,I}(\mu) \). Both \( M_{J,I}(\mu) \) and \( E_{J,I}(\mu) \) are operators on the Hilbert space of the system alone. If we do not make the hypothesis of classical readout, Eq. (A.12) becomes \( M_{J,I}(\mu) = \sqrt{p(\mu|J)} w(I)|J \rangle \langle J| \), where \( |\mu : J\rangle \) is the basis that diagonalizes \( \hat{F}_{\text{det}}(\mu) \).

The measurement can also be described in the language of positive-operator valued measures (POVM) [27–32]. The linear operations
\[ \mathcal{I}_D(\hat{\rho}_{\text{sys}}) \equiv \int_D d\mu \int dJdIM_{J,I}(\mu)\hat{\rho}_{\text{sys}}M_{J,I}^\dagger(\mu), \]  
(A.13) where \( D \) is a measurable domain in the space of possible outcomes (i.e. an element of a \( \sigma \)-algebra in the Kolmogorov formalism), are called \textit{instruments} if they preserve the non-negativity of \( \rho \). In the POVM formalism, the instruments are introduced as the basic objects describing the measurement. Besides preserving the nonnegativity of the density operator, they must also satisfy [27, 33]
(i) \( \mathcal{I}_\emptyset = 0 \), i.e., the measurement always gives an outcome (notice, however, that the outcome is not necessarily a number, but it could consist, e.g., in the observation that the apparatus has failed, etc.).

(ii) \( \mathcal{I}_{\cup D_j} = \sum_j \mathcal{I}_{D_j} \) for any family of disjoint elements of the \( \sigma \)-algebra.

(iii) \( \text{Tr}_{\text{sys}}[\mathcal{I}_{\Omega} (\rho_{\text{sys}})] = 1 \), where \( \Omega \) is the whole set of possible outcomes.

The instruments defined in Eq. (A.13) automatically satisfy the three above properties. An interesting question is whether any instrument can be written as in Eq. (A.13).

The instruments describe the change of the density matrix one assigns to the system, conditioned on the information that the outcome of the measurement was in \( D \). In other words, the new density matrix is

\[
\rho_{\text{sys}|D} = \frac{\mathcal{I}_D (\rho_{\text{sys}})}{P(D|\rho_{\text{sys}})},
\]

with the normalization \( P(D|\rho_{\text{sys}}) \) being the probability of obtaining an outcome in \( D \),

\[
P(D|\rho_{\text{sys}}) = \text{Tr}_{\text{sys}} [E(D)\rho_{\text{sys}}],
\]

\[
E(D) = I_D^+(1).
\]

The family of conjugate operators \( I_D^+ \) is defined univocally by

\[
\text{Tr}_{\text{sys}}[\rho I_D^+(\sigma)] = \text{Tr}_{\text{sys}}[I_D(\rho)\sigma].
\]

When both \( \hat{F}_{\text{det}} (\mu) \) and \( \hat{\rho}_{\text{det}} \) are pure states, then only one term survives in the double sum of Eq. (A.13) and the instruments are called Lüders instruments. The operators \( E(\mu) \) can be inferred from the observed probabilities \( \mathcal{P}(\mu) \) by changing the preparation \( \rho_{\text{sys}} \). Recovering \( E_{J,I}(\mu) \), however, is a nontrivial task, and requires the knowledge of the measurement apparatus, as different \( M_{J,I}(\mu) \) correspond to the same instrument. If \( E_{J,I}(\mu) \) are known, the operations are determined modulo a unitary operator: \( M_{J,I}(\mu) = V_{J,I}(\mu) E_{J,I}^{1/2}(\mu) \), where \( E_{J,I} \) is univocally defined as a positive operator. In other words, we consider the eigenvalues of the positive operator \( E_{J,I}(\mu) \), take their arithmetic square root, and define \( E_{J,I}^{1/2}(\mu) \) as the operators having the same eigenstates as \( E_{J,I}(\mu) \), and as eigenvalues the said arithmetic root. The additional operations \( V_{J,I}(\mu) \) represent an unwanted and avoidable feedback on the system. For instance, in a Stern-Gerlach apparatus, the beam exiting the magnetic field gradient could pass through a uniform magnetic field, which rotates the spin, so that a second measurement would not confirm the result of the first one. Ideally, \( V_{J,I}(\mu) = 1 \). Only by making subsequent measurements on the system, is it possible to check whether the measurement process is introducing this unwanted feedback.

We are interested in particular to the case of two sequential measurements. A probe measuring \( \hat{A} \) couples first to the system, then, after the first probe decouples, a second probe measuring \( \hat{B} \) couples. The time-evolution operator factorizes then as \( U = U_B U_A \), where \( U_A \) (respectively, \( U_B \)) acts on the Hilbert space of the system and the probe \( A \) (respectively, \( B \)). As we are making two measurements, it is natural to assume that the readout states factor, \( \hat{F}_{\text{det}} (\mu) = \hat{F}_A (\mu_A) \hat{F}_B (\mu_B) \), each factor acting on a different Hilbert space. A fundamental consideration is that, if we assume as well that the initial state of the two probes factors as \( \rho_{\text{det}} = \rho_A \otimes \rho_B \), i.e. if we assume no initial correlations between the probes, then the probability of observing the outcomes \( \mu_A, \mu_B \) can be written

\[
\mathcal{P}(\mu_A, \mu_B) = \text{Tr}_{\text{sys}} \{ I_{\mu_B} \{ I_{\mu_A} (\rho_{\text{sys}}) \} \}
\]

and the conditional state of the system is

\[
\rho_{\text{sys}|\mu} \propto I_{\mu_B} \{ I_{\mu_A} (\rho_{\text{sys}}) \},
\]
where the composing instruments are defined by

\[ \mathcal{I}_{\mu_k}^k(\rho) = \int dI_k dJ_k M^k_{I_kJ_k}(\mu_k)\rho M^k_{I_kJ_k}(\mu_k), \quad (A.20) \]

\[ M^k_{I,J}(\mu) = \sqrt{p(\mu|J)w(J|\mu)}, \quad (A.21) \]

However, if we do assume initial correlations, then the total instrument cannot be written as a composition of two instruments, contrary to what is commonly assumed (see e.g. [34]), and we can only associate one set of instrument, defined on the probability space \( M_A \times M_B \).

Finally, we note that the marginal probabilities are

\[ \mathcal{P}(\mu_A) = \int d\mu_B \text{Tr}\left\{ [\mathbb{1} \otimes \hat{F}_A(\mu_A) \otimes \hat{F}_B(\mu_B)] U_A U_B(\rho_{\text{sys}} \otimes \rho_{\text{det}}) U^\dagger_A U^\dagger_B \right\} \]

\[ = \text{Tr}_{\text{sys},A}\{[\mathbb{1} \otimes F_A(\mu_A)] U_A(\rho_{\text{sys}} \otimes \rho_A) U^\dagger_A \}, \quad \text{with } \rho_A = \text{Tr}_B(\rho_{\text{det}}) \quad (A.22) \]

\[ \mathcal{P}(\mu_B) = \int d\mu_A \text{Tr}\left\{ [\mathbb{1} \otimes \hat{F}_A(\mu_A) \otimes \hat{F}_B(\mu_B)] U_A U_B(\rho_{\text{sys}} \otimes \rho_{\text{det}}) U^\dagger_A U^\dagger_B \right\} \]

\[ = \text{Tr}_{\text{sys},B}\{[\mathbb{1} \otimes F_B(\mu_B)] U_B\rho_{\text{sys},B} U^\dagger_B \}, \quad \text{with } \rho_{\text{sys},B} = \text{Tr}_A[U_A(\rho_{\text{sys}} \otimes \rho_{\text{det}}) U^\dagger_A]. \quad (A.23) \]

Notice that the correlations are swapped: from correlations between \( A \) and \( B \), they become correlations between \( B \) and the system, after tracing out the first probe. Applying Bayes’ rule, we see that neither conditional probability

\[ \mathcal{P}(\mu_B|\mu_A) = \frac{\mathcal{P}(\mu_A,\mu_B)}{\mathcal{P}(\mu_A)} \]

\[ = \frac{\text{Tr}[E(\mu_A,\mu_B)\rho_{\text{sys}}]}{\text{Tr}_{\text{sys},A}\{[\mathbb{1} \otimes F_A(\mu_A)] U_A(\rho_{\text{sys}} \otimes \rho_A) U^\dagger_A \}}, \quad (A.24) \]

\[ \mathcal{P}(\mu_A|\mu_B) = \frac{\mathcal{P}(\mu_A,\mu_B)}{\mathcal{P}(\mu_B)} \]

\[ = \frac{\text{Tr}[E(\mu_A,\mu_B)\rho_{\text{sys}}]}{\text{Tr}_{\text{sys},B}\{[\mathbb{1} \otimes F_B(\mu_B)] U_B\rho_{\text{sys},B} U^\dagger_B \}}, \quad (A.25) \]

admits a conditional state \( \rho_{\text{sys}}|\mu_A \) or \( \rho_{\text{sys}}|\mu_B \) such that

\[ \mathcal{P}(\mu_B|\mu_A) = \text{Tr}_{\text{sys},B}\{[\mathbb{1} \otimes F_B(\mu_B)] U_B \rho_{\text{sys},B|\mu_A} U^\dagger_B \}, \quad (A.26) \]

\[ \mathcal{P}(\mu_A|\mu_B) = \text{Tr}_{\text{sys},A}\{[\mathbb{1} \otimes F_A(\mu_A)] U_A \rho_{\text{sys}|\mu_B} U^\dagger_A \}, \quad (A.27) \]

unless \( \rho_{\text{det}} = \rho_A \otimes \rho_B \). This is because, for a given output \( \mu_A \) (respectively, \( \mu_B \)), the system becomes correlated with the probe \( B \) (respectively, \( A \)).

References

[1] Erhart J, Sponar S, Sulyok G, Badurek G, Ozawa M and Hasegawa Y 2012 Experimental demonstration of a universally valid error–disturbance uncertainty relation in spin measurements Nat. Phys. 8 185–189 URL http://dx.doi.org/10.1038/nphys2194

[2] Rozema L A, Darabi A, Mahler D H, Hayat A, Soudagar Y and Steinberg A M 2012 Violation of Heisenberg’s measurement-disturbance relationship by weak measurements Phys. Rev. Lett. 109(10) 100404 URL http://link.aps.org/doi/10.1103/PhysRevLett.109.100404

[3] Baek S Y, Kaneda F, Ozawa M and Edamatsu K 2013 Experimental violation and reformulation of the Heisenberg’s error-disturbance uncertainty relation Scientific Reports 3 2221

[4] Weston M M, Hall M J W, Palsson M S, Wiseman H M and Pryde G J 2013 Experimental test of universal complementarity relations Phys. Rev. Lett. 110(22) 220402 URL http://link.aps.org/doi/10.1103/PhysRevLett.110.220402
[5] Heisenberg W 1927 Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik Zeitschr. f. Phys. 43(3-4) 172–198 URL http://dx.doi.org/10.1007/BF01397280
[6] Kennard E H 1927 Zur Quantenmechanik einfacher Bewegungstypen Zeitschr. f. Phys. 44(4) 326–352 URL http://dx.doi.org/10.1007/BF01391200
[7] Robertson H P 1929 The uncertainty principle Phys. Rev. 34(1) 163–164 URL http://link.aps.org/doi/10.1103/PhysRev.34.163
[8] Schrödinger E 1930 Zum Heisenbergschen Unschärfeprinzip Sitzungsber. Preuss. Akad. Wiss., Phys.-Math. Kl. 19 296–303 [About Heisenberg uncertainty relation, (A. Angelow and M. C. Batoni transl.), Bulg. J. Phys. 26, 193–203 (1999)]
[9] Ozawa M 2004 Quantum measurements, instruments, and uncertainty principle Interdiscip. Inform. Sci. 10 69–81 URL http://dx.doi.org/10.4036/iis.2004.69
[10] Ozawa M 2003 Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement Phys. Rev. A 67(4) 042105 URL http://link.aps.org/doi/10.1103/PhysRevA.67.042105
[11] Busch P, Lahti P and Werner R F 2013 Proof of Heisenberg’s error-disturbance relation Phys. Rev. Lett. 111(16) 160405 URL http://link.aps.org/doi/10.1103/PhysRevLett.111.160405
[12] Di Lorenzo A 2013 Correlations between detectors allow violation of the Heisenberg noise-disturbance principle for position and momentum measurements Phys. Rev. Lett. 110(12) 120403 URL http://link.aps.org/doi/10.1103/PhysRevLett.110.120403
[13] Distler J and Paban S 2013 Uncertainties in successive measurements Phys. Rev. A 87(6) 062112 URL http://link.aps.org/doi/10.1103/PhysRevA.87.062112
[14] Bullock T J and Busch P 2014 Focusing in Arthurs-Kelly-type joint measurements with correlated probes Phys. Rev. Lett. 113(12) 120401 URL http://link.aps.org/doi/10.1103/PhysRevLett.113.120401
[15] Moyal J E 1949 Quantum mechanics as a statistical theory Math. Proc. Cambr. Phil. Soc. 45 99–124 URL http://dx.doi.org/10.1017/S0305004100000487
[16] Wigner E 1932 On the quantum correction for thermodynamic equilibrium Phys. Rev. 40(5) 749–759 URL http://link.aps.org/doi/10.1103/PhysRev.40.749
[17] Di Lorenzo A 2011 Strong correspondence principle for joint measurement of conjugate observables Phys. Rev. A 83(4) 042104 URL http://link.aps.org/doi/10.1103/PhysRevA.83.042104
[18] Di Lorenzo A 2013 Sequential measurement of conjugate variables as an alternative quantum state tomography Phys. Rev. Lett. 110(1) 010404 URL http://link.aps.org/doi/10.1103/PhysRevLett.110.010404
[19] Einstein A, Podolsky B and Rosen N 1935 Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. 47(10) 777–780 URL http://link.aps.org/doi/10.1103/PhysRev.47.777
[20] Bohm D and Aharonov Y 1957 Discussion of experimental proof for the paradox of Einstein, Rosen, and Podolsky Phys. Rev. 108 1070–1076 URL http://link.aps.org/doi/10.1103/PhysRev.108.1070
[21] von Neumann J 1932 Mathematische Grundlagen der Quantenmechanik (Berlin: Springer) [Mathematical Foundations of Quantum Mechanics (Princeton University Press, Princeton, 1996)]
[22] Jaynes E 2003 Probability Theory: The Logic of Science (Cambridge University Press) ISBN 9780521592710 URL http://books.google.com.br/books?id=tTN4HuUNXjgC
[23] Zurek W H 1981 Pointer basis of quantum apparatus: Into what mixture does the wave packet collapse? Phys. Rev. D 24(6) 1516–1525 URL http://link.aps.org/doi/10.1103/PhysRevD.24.1516

[24] Zurek W H 1982 Environment-induced superselection rules Phys. Rev. D 26(8) 1862–1880 URL http://link.aps.org/doi/10.1103/PhysRevD.26.1862

[25] Aharonov Y, Albert D Z and Vaidman L 1988 How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100 Phys. Rev. Lett. 60(14) 1351–1354 URL http://link.aps.org/doi/10.1103/PhysRevLett.60.1351

[26] Di Lorenzo A 2012 Full counting statistics of weak-value measurement Phys. Rev. A 85(3) 032106 URL http://link.aps.org/doi/10.1103/PhysRevA.85.032106

[27] Davies E B 1976 Quantum Theory of Open Systems (London: Academic Press)

[28] Helstrom C W 1976 Quantum Detection and Estimation Theory (London: Academic Press)

[29] Holevo A S 1982 Probabilistic and statistical aspects of quantum theory (Amsterdam: North Holland)

[30] Kraus K 1983 States, effects, and operations: fundamental notions of quantum theory (Springer-Verlag) ISBN 9780387127323

[31] Ludwig G 1983 Foundations of Quantum Mechanics I (Springer London, Limited) ISBN 9783540168843

[32] Busch P, Grabowski M and Lahti P J 1995 Operational Quantum Physics (Berlin: Springer) second corrected printing

[33] Davies E and Lewis J 1970 An operational approach to quantum probability Comm. Math. Phys. 17(3) 239–260 URL http://dx.doi.org/10.1007/BF01647093

[34] Carmeli C, Heinosaari T and Toigo A 2011 Sequential measurements of conjugate observables J. Phys. A: Math. Gen. 44 285304 URL http://stacks.iop.org/1751-8121/44/i=28/a=285304