Quantum Errors and Disturbances: Response to Busch, Lahti and Werner

David Marcus Appleby

Centre for Engineered Quantum Systems, School of Physics, The University of Sydney,
Sydney, NSW 2006, Australia; marcus.appleby@sydney.edu.au; Tel.: +44-734-210-5857

Abstract: Busch, Lahti and Werner (BLW) have recently criticized the operator approach to the description of quantum errors and disturbances. Their criticisms are justified to the extent that the physical meaning of the operator definitions has not hitherto been adequately explained. We rectify that omission. We then examine BLW's criticisms in the light of our analysis. We argue that, although the BLW approach favour (based on the Wasserstein two-deviation) has its uses, there are important physical situations where an operator approach is preferable. We also discuss the reason why the error-disturbance relation is still giving rise to controversies almost a century after Heisenberg first stated his microscope argument. We argue that the source of the difficulties is the problem of interpretation, which is not so wholly disconnected from experimental practicalities as is sometimes supposed.

1. Introduction

The error-disturbance principle remains highly controversial almost a century after Heisenberg wrote the paper [1], which originally suggested it. It is remarkable that this should be so, since the disagreements concern what is arguably the most fundamental concept of all, not only in physics, but in empirical science generally: namely, the concept of measurement accuracy. Measuring instruments are not born equal. If one did not have a way to distinguish measurements that are in some sense "good" from measurements that are in some sense "bad", that is if one did not have what Busch *et al.* [2] call a "figure of merit", one would be forced to regard all measurements as being on the same footing. There would, in fact, be no reason to prefer numbers obtained using a state-of-the-art photon counter from those obtained using the less expensive, less demanding procedure of making a blind guess. Under such conditions, empirical science would be impossible. Since physics has actually made huge advances over the last century, it is obvious that, on a practical level, experimentalists have ways to distinguish good measurements from bad. However, those practical methods are not supported by an adequate understanding at the theoretical level.

It is worth asking why, given the fundamental importance of the problem, progress has been so slow. Although it is true that the problem is technically demanding, it appears to us that the main obstacle has always been, as it continues to be, conceptual. The classical concept of error involves a comparison between the measured value and the true value, as it existed before the measurement was made. The Bell–Kochen–Specker theorem [3–6], however, requires us to abandon the idea that a measurement ascertains the pre-existing value of a specified observable (aside from cases where there is a superselection rule). This is such a radical departure from classical ideas that Bell [7] suggested that "the field would be significantly advanced by banning [the word 'measurement'] altogether, in

favour for example of the word 'experiment' ". The question then arises: once the classical concept of measurement has gone up in smoke, what, if anything, is left of the classical concept of measurement accuracy? It will be seen that this is a special case of the more general question, which lies at the heart of all of the disputes about quantum foundations: once the classical concept of measurement has gone up in smoke, what, if anything, is left of the classical concept of empirically-accessible reality? The problem is consequently of a rather peculiar kind. Physics encompasses an enormous spectrum of problems, ranging from nuts-and-bolts problems, such as measuring a length precisely, to deep philosophical questions. The error-disturbance principle is unusual because it directly connects the two ends of the spectrum. On the one hand, it has, as we stressed above, an immediate, down-to-earth practical relevance. On the other hand, we would argue that one of the factors obstructing progress, the reason almost half a century elapsed before people started to come seriously to grips with the problem, was the obscurities of the Copenhagen interpretation. It thus provides a riposte to the suggestion that the interpretational issues are practically unimportant.

Although the connections with the interpretation problem are not the main point of this paper, they are part of the underlying motivation. It is therefore appropriate to say something about them in this introductory section (we shall give a more detailed discussion in a subsequent publication). Let us begin by observing that Heisenberg himself did not propose or even conjecture an error-disturbance principle. He did, of course, construct his famous microscope argument [1,8], which has suggested to many that he had in mind such a principle. However, that is based on a misunderstanding of the point of the microscope argument (that is, what *Heisenberg* saw as the point). That point emerges most clearly in Von Neumann's account [9], where it is made completely explicit that the function of the microscope argument is to give intuitive support to the inequality proved by Kennard [10] and Weyl [11] (the latter attributing the result to Pauli) and to its subsequent generalization by Robertson [12] and Schrödinger [13]. In his 1927 paper [1], Heisenberg was less explicit. At the time he wrote the paper, the Kennard–Pauli–Weyl proof was yet to come, and perhaps for that reason, he gave the microscope argument pride of place. However, he was using it to support his original prototype for the uncertainty principle, namely the order of magnitude estimate $p_1 q_1 \sim h$, where p_1, q_1 are the standard deviations of the p, q probability distributions scaled by a factor $\sqrt{2}$. There is no indication that he envisaged, in addition to this statement, an entirely different error-disturbance principle.

Nevertheless, although Heisenberg did not in fact propose an error-disturbance principle, one may feel that he *should* have done so, for it is strongly suggested by the considerations in his 1927 paper (to that extent, we agree with Busch *et al.* [14], that it is "latent" in what he says). Reflecting on the microscope experiment, it seems intuitively evident that the measurement of position really is (in some sense) less than perfectly accurate and that the electron really will (in some sense) be disturbed by the photon. The situation seems to be crying out for proper quantum mechanical analysis. Yet, it evidently did not seem that way to Heisenberg; nor, apparently, did it seem that way to most other people before the 1960s. During the period between 1927 and the 1965 paper of Arthurs and Kelly [15] which gave the first explicit model for a joint measurement of position and momentum, one finds various paraphrases and elaborations of the statements in Heisenberg's original paper, but we are not aware of any clear statement of the error-disturbance principle conceived of as a proposition distinct from the Kennard–Pauli–Weyl inequality or any recognition of the fact that a quantum mechanical definition of measurement accuracy is needed. The question arises: Why is it that Heisenberg and so many others failed to draw what seems to most people now the obvious conclusion from his uncertainty paper? The answer, we suggest, is that their understanding was obstructed by one of the features of the Copenhagen interpretation.

In the words of Bell [16], the Copenhagen interpretation divides the world "into speakable apparatus . . . that we can talk about . . . and unspeakable quantum system that we can not talk about" (ellipses in the original). Of course, the Copenhagen interpretation is not a sharply defined entity [17–19]). However, it appears to us that Bell's one-sentence summary does identify an idea that, in one form or another, is common to all the many variants. This idea has been hard to maintain

since the 1970s, when it was realized, in connection with the problem of gravity-wave detection, that the error-disturbance principle is relevant to highly accurate measurements of a *macroscopic* oscillator [20,21]. Such an oscillator is just as speakable as any other piece of laboratory apparatus; yet, at the same time, we need to analyse its behaviour quantum mechanically. However, in the early days of quantum mechanics, the unspeakability of quantum systems was accepted by almost everyone. Thinking of the quantum world as ineffable and beyond the reach of thought [22], forgetting that the quantum world is the one in front of our noses, encouraged the perception that quantum mechanical measurements are so utterly different from classical ones that no points of contact with classical concepts are possible. In particular, it encouraged the assumption that the classical concept of error cannot carry over to quantum mechanics in any shape or form. This, we would suggest, is why Heisenberg did not follow through on what now seems the obvious implication of his microscope argument and formulate an error-disturbance principle. He did not do so because he rejected the very idea of a quantum error or a quantum disturbance.

Corresponding to the idea that there are two different worlds, speakable and unspeakable, there is a widespread assumption that there are two kinds of measurement, classical and quantum. If highly accurate determinations of the centre-of-mass motion of a macroscopic object are to be treated as quantum measurements, then it is hard to see how one can consistently make such a distinction. Instead, one seems forced to the view that every measurement is a quantum measurement, measurements with a meter rule not excluded. To be sure, low precision measurements with a meter rule permit simplifying assumptions that cease to be valid as one increases the accuracy. However, that is purely a matter of practical convenience, not the signal of a fundamental difference of kind. In the case of kinematics, we continue to use the Newtonian theory when analysing low velocity motion, without taking this to mean that there is a fundamental difference of kind between the relativistic momentum of a space-ship travelling at near light speed and the Newtonian momentum of a train on the London underground. Similarly, in the case of measurements, we need a unified description.

In particular, we need a unified description of measurement errors. The statement, that the kind of sophisticated measurement on a macroscopic object that demands a quantum analysis is more accurate than a commonplace measurement with a meter rule, tacitly assumes that there is a single concept of accuracy applicable to both. Otherwise, we would not have the basis for a comparison. In the case of kinematics, the Newtonian definition of momentum is an approximation to the relativistic definition, valid for low velocities. In the same way, one would like an overarching quantum definition of error, which effectively reduces to the classical one in limiting cases. At first sight, this may seem impossible, since quantum mechanics requires us to drop the assumption that a measurement ascertains the pre-existing value of a specified observable. However, on further reflection, it will be seen that even on classical assumptions, one is never able to directly compare the measured value with the pre-existing true one. In classical physics, as in quantum physics, measured values are the only ones available. It follows that, although in classical principle the error is the difference between the measured value and the true one, in point of classical practice it must be possible to do everything using measured values only.

The purpose of this paper is to make a small beginning on the task of constructing a unified theory of measurement. We focus on Busch, Lahti and Werner's (BLW's) criticisms [2,23–26] of the operator approach [27–31] to the description of quantum errors and disturbances. Their criticisms raise some issues that are highly relevant to the above discussion and that need to be settled if we hope to make progress. It should be stressed that although our conclusion is that the operator approach is more useful than BLW allow, we are far from rejecting everything they say. In particular, we completely agree with them on what is, perhaps, the most essential point, that quantum errors and disturbances need to be defined operationally. Moreover, in defending the operator approach, it is no part of our intention to impugn the distributional approach they favour. No one would say that the RMS characterization of an ordinary uncertainty is either "better" or "worse" than an entropic

characterization. Rather, one has different quantitative measures, each of which has advantages and disadvantages; similarly here. The task is not to single out one particular approach as somehow canonical, but rather to achieve a clear understanding, at the basic conceptual level, of what is meant by the words "error" and "disturbance" in a quantum mechanical context and of the different ways of quantifying the concepts.

There are two versions of the operator approach (or O approach, as we will call it from now on). BLW's criticisms are largely directed against the state-dependent version proposed by Ozawa [28,29]. However, we had previously proposed a state-independent version [27]. Both versions are relevant to our discussion. In Section 2, we compare and contrast them.

Section 3 is the core of the paper. We begin with the classical concepts of error and disturbance. We show that there are at least two ways to reformulate them in a manner that does not involve a comparison with pre-existing values. We then show that the reformulated definitions have natural quantum generalizations, which we call the D and C definitions. The D and C errors are thus candidates for the overarching concept of measurement accuracy, which, as we argued above, is necessary if one wants to construct a unified theory of measurement, in which every measurement is seen as quantum. They also have an important bearing on BLW's criticism of the O approach. As BLW correctly observe, the O definitions are non-operational. However, the D and C definitions *are* operational. Moreover, the O quantities are upper bounds on the corresponding D and C quantities. This gives indirect operational meaning to the O quantities. Specifically, it means that if one of the O quantities is small, then there are at least two well-defined operational senses in which the measurement is accurate or non-disturbing. The situation when an O quantity is large is more problematic. In the state-independent case, it is possible that the smallness of the O error/disturbance is both necessary and sufficient for the measurement to be accurate/non-disturbing in a well-defined operational sense. However, we have not been able to prove this.

In Section 4, we analyse BLW's objections to the O approach in light of the foregoing. BLW contrast the operator approach with what they call a distributional approach. It is to be observed, however, that the D and C quantities are also defined distributionally. Since the O quantities owe their physical meaning to their connection with the D and C quantities, it follows that the O quantities are indirectly distributional. In short, the problem is not to decide between a distributional approach and some other, completely different approach. Rather, it is to decide between two different kinds of distributional approaches. As with all such questions, the answer is relative to the situation of interest. We show that there is at least one important class of physical problems for the which the D error and, by extension, the O error are clearly more appropriate than the definition that BLW favour, based on the Wasserstein two-deviation.

Finally, in the Appendix, we give a more careful proof of the error-disturbance and error-error relations than the one we presented in [27]. In that earlier paper, we skated over certain questions of domain and differentiability. We here take the opportunity to fill in the missing details.

2. The Operator Approach

In this section, we outline the operator characterization of quantum errors and disturbances. Our aim is purely descriptive. We justify the approach, and respond to the various criticisms that have been made of it, in subsequent sections.

Consider a classical measurement of position. Let x_i, p_i be the position and momentum immediately before the measurement, and let x_f, p_f be their values immediately after it. Let μ_f be the final value of the pointer observable. Then, the error in the measurement of position is $\mu_f - x_i$, and the disturbance to the momentum is $p_f - p_i$ (classical physics does not, of course, require there to be a disturbance to the momentum, but such a disturbance is perfectly possible). On the level of formal analogy, it is natural to ask what happens if one replaces the classical variables in these expressions with the corresponding Heisenberg picture operators. Let $\mathcal{H}_s$ and $\mathcal{H}_a$ be the Hilbert spaces for the system and apparatus, respectively, and assume that the system + apparatus are initially in

the product state $\hat{\rho} \otimes \hat{\alpha}$, where $\hat{\rho}$ is the density matrix of the system and $\hat{\alpha}$ is the density matrix of the apparatus. Let $\hat{U} \colon \mathcal{H}_s \otimes \mathcal{H}_a \to \mathcal{H}_s \otimes \mathcal{H}_a$ be the unitary operator describing the measurement interaction, let

$$\hat{x}_i = \hat{x} \otimes I, \qquad\qquad \hat{p}_i = \hat{p} \otimes I, \qquad\qquad \hat{\mu}_i = I \otimes \hat{\mu} \qquad\qquad (1)$$

be the position, momentum and pointer Heisenberg picture observables immediately before the measurement interaction commences, and let

$$\hat{x}_f = U^\dagger \hat{x}_i U, \qquad\qquad \hat{p}_f = U^\dagger \hat{p}_i U, \qquad\qquad \hat{\mu}_f = U^\dagger \hat{\mu}_i U \qquad\qquad (2)$$

be the Heisenberg picture observables immediately after the interaction has finished. Formal analogy with the classical case then suggests that we define

$$\hat{\epsilon}_X = \hat{\mu}_f - \hat{x}_i, \qquad\qquad\qquad \hat{\delta}_P = \hat{p}_f - \hat{p}_i \qquad\qquad (3)$$

In Appleby [27,32–34] we also introduced the predictive error operator $\hat{\mu}_f - \hat{x}_f$. We shall not discuss it here since it does not give rise to conceptual difficulties. We refer to $\hat{\epsilon}_X$ (respectively, $\hat{\delta}_P$) as the error (respectively, disturbance) operator. We then obtain a numerical characterization of the error by defining

$$\Delta_e^{\hat{\rho}} x = \big(\mathrm{Tr}(\hat{\epsilon}_X^2 (\hat{\rho} \otimes \hat{\alpha})) \big)^{\frac{1}{2}} \qquad\qquad (4)$$

and a numerical characterization of the disturbance by defining

$$\Delta_d^{\hat{\rho}} p = \big(\mathrm{Tr}(\hat{\delta}_P^2 (\hat{\rho} \otimes \hat{\alpha})) \big)^{\frac{1}{2}} \qquad\qquad (5)$$

We label the quantities with a superscript $\hat{\rho}$ because, while the apparatus "ready" state $\hat{\alpha}$ is assumed to be always the same, the system state $\hat{\rho}$ can vary. The operators $\hat{\epsilon}_X$, $\hat{\delta}_P$ are unbounded, which means that the quantities $\Delta_e^{\hat{\rho}} x$, $\Delta_d^{\hat{\rho}} p$ are not defined for every state $\hat{\rho} \otimes \hat{\alpha}$. In the following, we will always assume that $\hat{\rho}$ is in the set of physical states $\mathcal{P}$ defined in the Appendix. If this is true, then, provided that $\hat{\alpha}$ is appropriately chosen, the expectation value $\mathrm{Tr}(M(\hat{\rho} \otimes \hat{\alpha}))$ is well defined and finite for every monomial M in $\hat{x}_i$, $\hat{x}_f$, $\hat{p}_i$, $\hat{p}_f$, $\hat{\mu}_i$, $\hat{\mu}_f$.

Of course, we have not yet justified the interpretation of $\Delta_e^{\hat{\rho}} x$ and $\Delta_d^{\hat{\rho}} p$ as an error and disturbance (beyond noting the formal analogy with classical physics, which, though suggestive, is clearly not sufficient to justify the proposal). We defer a proper justification to the next section and focus here on the question of whether there exists an error-disturbance relation expressible in terms of these quantities. In various special cases [15,32,33,35–38], one does indeed have

$$\Delta_e^{\hat{\rho}} x \, \Delta_d^{\hat{\rho}} p \geq \frac{\hbar}{2} \qquad\qquad (6)$$

analogous to the ordinary uncertainty relation $\Delta x \Delta p \geq \hbar/2$. However, as we showed in [27], it is easy to see that the inequality cannot be generally valid. In [27] we were concerned with joint measurements of x and p; however our argument also applies to single measurements of x only because any such measurement can be regarded as a joint measurement in which the momentum pointer does not interact with the system. Indeed, the example we gave was of precisely this kind. Eliminating the inessential reference to the momentum pointer we gave a simple model for the measurement process, in which the position pointer $\hat{\mu}$ is the position of a particle having

momentum $\hat{\pi}$ and in which the measurement rotates the system particle position onto the pointer particle position, so that

$$\hat{\mu}_f = \hat{x}_i \tag{7}$$

Such a rotation is effected by

$$\hat{U} = e^{-\frac{i\pi\hat{H}}{2\hbar}} \tag{8}$$

where

$$\hat{H} = \hat{x}_i\hat{\pi}_i - \hat{\mu}_i\hat{p}_i \tag{9}$$

(so, if $\hat{x}$, $\hat{\mu}$ were different components of the position of a single particle in three dimensions, $\hat{H}$ would be a component of the angular momentum operator). The fact that $\hat{\mu}_f = \hat{x}_i$ means that $\hat{\epsilon}_X = 0$. It is easy to see that $\hat{\delta}_P = -\hat{\pi}_i - \hat{p}_i$. Therefore, this is a measurement for which the error is zero, while the disturbance is finite for every physical state.

Although we are mainly concerned with the error-disturbance relation in this paper, it is worth noting that exactly the same argument shows [27] that the error-error relation

$$\Delta_e^{\hat{\delta}}x\Delta_e^{\hat{\delta}}p \geq \frac{\hbar}{2} \tag{10}$$

for a joint measurement of position and momentum cannot be valid in general. Indeed, consider a joint measurement in which the interaction of the particle with the position pointer is described by the unitary in Equation (8), while the momentum pointer just goes along for the ride, without interacting at all. One then has $\Delta_e^{\hat{\delta}}x = 0$ and $\Delta_e^{\hat{\delta}}p = \sqrt{\langle(\hat{\mu}_{P,i} - \hat{p}_i)^2\rangle}$ (where $\hat{\mu}_{P,i} = \hat{\mu}_{P,f}$ is the momentum pointer position). Even though the momentum is not really being measured at all, $\Delta_e^{\hat{\delta}}p$ is still finite for every physical state. Therefore, inequality (10) is violated for every physical state.

The fact that inequalities (6) and (10) are not generally valid was noted by us [27] and subsequently by Ozawa [28,29,39–41]; in the case of (10) also by Hall [42]. We, Ozawa and Hall responded to these facts by trying to find alternative inequalities that are generally valid. However, we, on the one hand, and Ozawa and Hall, on the other, were led in different directions. We begin by describing our approach to the problem, since this came first in point of time.

The essential point will emerge most clearly if we start with the violation of inequality (10) by the measurement described by Equation (8). For this measurement, it is not simply that the product $\Delta_e^{\hat{\delta}}x\Delta_e^{\hat{\delta}}p$ is less than $\hbar/2$ for a certain subset of initial states. The product is in fact strictly zero for every possible initial state. However, it would be rash to conclude from this that the measurement is in some sense "best possible". As we noted above, the momentum pointer does not interact with the system, which means that so far as momentum is concerned, the measurement is not only not highly accurate, it cannot properly be described as a measurement at all. It is true that $\Delta_e^{\hat{\delta}}p$ is small for a certain, highly specific set of initial states. However, that is not a reason for describing the measurement as accurate. Consider the following scenario:

> Alice goes to Bob's shop and buys what Bob says is a highly accurate ammeter. However, when she gets home, she finds that the needle is stuck at the 1 amp position. When she goes back to complain, Bob is unrepentant. He insists that the meter is indeed highly accurate provided one uses it to measure a 1 amp current.

Clearly, Alice will not be satisfied with this response. No more would she be satisfied with the claim that the interaction described by Equation (8) gives a highly accurate measurement of momentum.

This example shows that the smallness of the product $\Delta_e^{\hat{\rho}} x \Delta_e^{\hat{\rho}} p$ is not always the signature of a highly accurate joint measurement of position and momentum. Similar remarks apply to the product $\Delta_e^{\hat{\rho}} x \Delta_d^{\hat{\rho}} p$. Consider, for instance, a "measurement" for which $\hat{U}$ is the identity, so that there is no coupling whatever between system and apparatus. Here, $\Delta_d^{\hat{\rho}} p$ is zero for every possible initial state, while $\Delta_e^{\hat{\rho}} x$ is always finite and sometimes small. Yet, as in the broken ammeter example, it would be an abuse of language to describe this as a measurement of position that is always non-disturbing and sometimes highly accurate.

In [27], these considerations led us to look for replacements for the products $\Delta_e^{\hat{\rho}} x \Delta_e^{\hat{\rho}} p$, $\Delta_e^{\hat{\rho}} x \Delta_d^{\hat{\rho}} p$ whose smallness can unequivocally be regarded as the signature of a measurement that is in some sense "good". In the broken ammeter example, what makes Bob's claim absurd is the fact that an accurate classical ammeter is one for which the measured value is close to the true one, not just for one particular current, but for every current within a wide range. Applying the same principle to the quantum case suggests that we define the error by

$$\Delta_e x = \sup_{\hat{\rho} \in \mathcal{P}} \left(\Delta_e^{\hat{\rho}} x \right) \tag{11}$$

where $\mathcal{P}$ is the set of physical states, as defined in Appendix A. As we saw above, the smallness of $\Delta_e^{\hat{\rho}} x$ for some particular $\hat{\rho}$ is consistent with the apparatus being completely decoupled from the system, so that it is not really measuring anything. However, if $\Delta_e x$ is small, it means that $\Delta_e^{\hat{\rho}} x$ is small for every possible state, and we clearly are entitled to say that the measurement is highly accurate (taking into account the discussion in Section 3). Similar principles apply to the concept of disturbance. Consider, for instance, the measurement described by Equation (8), which rotates $\hat{\mu}$ onto $\hat{x}$. For this measurement, $\Delta_d^{\hat{\rho}} p$ will be small for certain special choices of $\hat{\rho}$ and $\hat{\alpha}$. However, it will typically be large. A medical procedure would not usually be described as non-invasive merely on the grounds that it can occasionally happen that the patient escapes almost intact; similarly here. We accordingly define the disturbance to be

$$\Delta_d p = \sup_{\hat{\rho} \in \mathcal{P}} \left(\Delta_d^{\hat{\rho}} p \right) \tag{12}$$

With these definitions, it can be shown

$$\Delta_e x \Delta_d p \geq \frac{\hbar}{2} \tag{13}$$

where we use the convention, here and elsewhere, that a product of the form $q \times \infty$ counts as infinite, even if $q = 0$. The quantity on the left-hand side should not be confused with $\sup_{\hat{\rho} \in \mathcal{P}} \left(\Delta_e^{\hat{\rho}} x \Delta_d^{\hat{\rho}} p \right)$. It is an open question whether the latter satisfies a similar inequality.

One can also prove a universally valid version of the error-error relation for a joint measurement of position and momentum

$$\Delta_e x \Delta_e p \geq \frac{\hbar}{2} \tag{14}$$

where $\Delta_e p$ is defined by taking the supremum of $\Delta_e^{\hat{\rho}} x$. In [27], we gave a proof of these relations, which glossed over some questions to do with domains of definition and differentiability. A completely rigorous proof is given in Appendix A below.

The quantities $\Delta_e x$, $\Delta_d p$ are not without interest, as we discuss below. However, they are not the appropriate definitions for a real measuring instrument. The demand that $\Delta_e x$ be small is the demand that $\Delta_e^{\hat{\rho}} x$ be small, not only when $\hat{\rho}$ is a wave-packet localized in the vicinity of the apparatus, but also when $\hat{\rho}$ is a wave-packet localized on the other side of the cosmic event-horizon. Clearly, this is not

a reasonable demand to make of a practical laboratory instrument, which is only designed to give accurate readings for a restricted set of input states. In [27], we accordingly proposed the following modified definitions:

$$\Delta_e^{\mathcal{R}} x = \sup_{\hat{\rho} \in \mathcal{R}} \left(\Delta_e^{\hat{\rho}} x \right) \tag{15}$$

$$\Delta_d^{\mathcal{R}} p = \sup_{\hat{\rho} \in \mathcal{R}} \left(\Delta_e^{\hat{\rho}} x \right) \tag{16}$$

where the supremum is now taken over a proper subset $\mathcal{R}$ of the set of physical states. We took $\mathcal{R}$ to be a set of physical states for which the mean values $\mathrm{Tr}(\hat{x}\hat{\rho})$, $\mathrm{Tr}(\hat{p}\hat{\rho})$ lie in a rectangular region of phase space with sides l_X, l_P and satisfying certain additional conditions. We then proved the inequalities

$$\Delta_e^{\mathcal{R}} x \Delta_d^{\mathcal{R}} p + \frac{\hbar}{l_X} \Delta_e^{\mathcal{R}} x + \frac{\hbar}{l_P} \Delta_d^{\mathcal{R}} p \geq \frac{\hbar}{2} \tag{17}$$

$$\Delta_e^{\mathcal{R}} x \Delta_e^{\mathcal{R}} p + \frac{\hbar}{l_X} \Delta_e^{\mathcal{R}} x + \frac{\hbar}{l_P} \Delta_e^{\mathcal{R}} p \geq \frac{\hbar}{2} \tag{18}$$

where we again use the convention that a product of the form $q \times \infty$ counts as infinite, even if $q = 0$. It will be observed that in the limit as l_X, $l_P \to \infty$, we recover inequalities (13) and (14). As with inequalities (13) and (14), the proof of inequalities (17) and (18), which we gave in [27], glossed over certain details. We give a completely rigorous proof in Appendix A below, where we also take the opportunity to strengthen the statement somewhat.

In practice one might not want to make a sharp distinction between states which are in the operating range of the instrument and ones which are not. There are various ways in which definitions (15), (16) might be modified to take this into account, but that would take us beyond the scope of this paper.

Let us now turn to the approach of Ozawa [28,29,40,41] and Hall [42]. In our approach, we replaced the state-dependent definitions $\Delta_e^{\hat{\rho}} x$, $\Delta_d^{\hat{\rho}} p$ with the quantities $\Delta_e x$, $\Delta_d p$ and $\Delta_e^{\mathcal{R}} x$, $\Delta_d^{\mathcal{R}} p$ and proved inequalities applying to those. Ozawa [28,29], by contrast, kept with the state-dependent definitions and showed

$$\Delta_e^{\hat{\rho}} x \Delta_d^{\hat{\rho}} p + \Delta^{\hat{\rho}} p \Delta_e^{\hat{\rho}} x + \Delta^{\hat{\rho}} x \Delta_d^{\hat{\rho}} p \geq \frac{\hbar}{2} \tag{19}$$

where $\Delta^{\hat{\rho}} x$, $\Delta^{\hat{\rho}} p$ are the ordinary uncertainties in the state $\hat{\rho}$. He also showed that [40,41], for a joint measurement of position and momentum,

$$\Delta_e^{\hat{\rho}} x \Delta_e^{\hat{\rho}} p + \Delta^{\hat{\rho}} p \Delta_e^{\hat{\rho}} x + \Delta^{\hat{\rho}} x \Delta_e^{\hat{\rho}} p \geq \frac{\hbar}{2} \tag{20}$$

It will be observed that these relations have a similar mathematical form to our inequalities (17) and (18). Hall [42] proved a relation similar to inequality (20). Other modifications and improvements have also been proved [43–45].

The reader should not conclude from our earlier discussion that we have any objection to the state-dependent definitions employed by Ozawa, Hall and others. Asking whether a state-independent definition is better than a state-dependent one is like asking whether a hammer is better than a screw-driver. The answer to all such questions, concerning the suitability of a tool, is relative to the use to which it is put. The fact that Bob, in the broken ammeter example, makes an inappropriate use of it does not invalidate the idea that the classical error is the difference between the measured value and the true one. The same applies here. It is true that there exist quantum analogues of the broken ammeter—processes that do not properly count as a measurement for which the state-dependent error is small. Nevertheless, the state-dependent error has a well-defined physical meaning (as we discuss in Section 3), and this makes it a potentially useful tool. State-independent

definitions, such as the ones proposed by ourselves or BLW [2,23], have the advantage that they supply what BLW call an overall figure of merit; while state-dependent definitions, if not handled with care, can lead to unreasonable conclusions. However, as Rozema *et al.* [46] point out, state-independent definitions have the disadvantage that they are insensitive to fine, state-dependent details, which can be important. The state-dependent error can be used to analyse those details. It is to be observed, furthermore, that the state-dependent quantities $\Delta_{\mathrm{e}}^{\hat{\rho}}x$, $\Delta_{\mathrm{d}}^{\hat{\rho}}p$ are the limits of $\Delta_{\mathrm{e}}^{\mathcal{R}}x$, $\Delta_{\mathrm{d}}^{\mathcal{R}}p$ as $\mathcal{R}$ is shrunk to a single point. If one takes the view that use of $\Delta_{\mathrm{e}}^{\hat{\rho}}x$, $\Delta_{\mathrm{d}}^{\hat{\rho}}p$ is in all circumstances inappropriate, then it is hard to see how one can avoid taking the view that the use of $\Delta_{\mathrm{e}}^{\mathcal{R}}x$, $\Delta_{\mathrm{d}}^{\mathcal{R}}p$ is also inappropriate when $\mathcal{R}$ is very small. Which raises the question: "Just how large has $\mathcal{R}$ got to be in order for the use of $\Delta_{\mathrm{e}}^{\mathcal{R}}x$, $\Delta_{\mathrm{d}}^{\mathcal{R}}p$ to be justified?" It is difficult to see how the answer can be other than arbitrary. It appears to us that such discussions are fruitless and that the solution to the quandary "state-dependent or state-independent?" is not to regard it as a quandary. Instead of making a once-and-for-all choice, we are free to use either or both, in a manner adjusted to the question of interest. We return to the points raised in this paragraph at the end of Section 4.

So far from being rivals, Ozawa's inequalities and ours are closely related. Let $\mathcal{R}$ be any region satisfying the condition of Theorem 4 in the Appendix. If we take the supremum on both sides of Ozawa's inequality (19), we obtain the relation

$$\Delta_{\mathrm{e}}^{\mathcal{R}}x\Delta_{\mathrm{d}}^{\mathcal{R}}p + K_x\Delta_{\mathrm{e}}^{\mathcal{R}}x + K_p\Delta_{\mathrm{d}}^{\mathcal{R}}p \geq \frac{\hbar}{2} \tag{21}$$

where

$$K_x = \sup_{\hat{\rho}\in\mathcal{R}}(\Delta^{\hat{\rho}}p), \qquad\qquad K_p = \sup_{\hat{\rho}\in\mathcal{R}}(\Delta^{\hat{\rho}}x) \tag{22}$$

This is weaker than our inequality (17) if l_{X}, l_{P} are large, but stronger if they and the region $\mathcal{R}$ are small. It is probably fair to say that that an experimenter will never be committed to the proposition that the system state is *precisely* $\hat{\rho}$. The statement, that the length of a rod is 1.2 m, will not usually be taken to mean that the length is 1.2 m to infinitely many decimal places. Rather, it will be taken to mean that the length is in the interval $[1.15, 1.25]$. A similar principle applies to statements regarding the wave function. This is true even in a Bayesian approach [47], where quantum states are interpreted as gambling commitments (the fact that gambling commitments are not, in practice, expressed in terms of arbitrarily large integers means that state space has to be coarse-grained in this approach). It follows that a so-called state-dependent definition of error or disturbance is really a state-independent one for which the region $\mathcal{R}$ is very small. If $\mathcal{R}$ is sufficiently small one will want to use Ozawa's inequalities, but if it is larger one will want to use ours (provided $\mathcal{R}$ satisfies the condition of Theorem 4).

Although $\Delta_{\mathrm{e}}^{\hat{\rho}}x$ will, in practice, only be small for a restricted set of states, the limiting situation, when it becomes zero for all $\hat{\rho} \in \mathcal{P}$, is still conceptually important. It can be shown (Appleby [34], Ozawa [48], Busch [49]) that the condition $\Delta_{\mathrm{e}}x = 0$ is both necessary and sufficient for the distribution of measured values to be $\langle x|\hat{\rho}|x\rangle$. No real measuring instrument could have precisely this distribution of measured values for every input state $\hat{\rho}$; in particular, it cannot do so for states such that the support of $\langle x|\hat{\rho}|x\rangle$ is not compact (in this connection, it may be worth remarking that the x and p space wave-functions cannot both have compact support, meaning that at least one of the two distributions $\langle x|\hat{\rho}|x\rangle$, $\langle p|\hat{\rho}|p\rangle$ must be practically unrealizable). Nevertheless, the idea that $\langle x|\hat{\rho}|x\rangle$ is the probability distribution for a measurement of position has played a fundamental role in physical thinking ever since Born [50,51] first proposed it (in connection with the momentum distribution). There is no problem here, provided we understand the proposal to be, not that $\langle x|\hat{\rho}|x\rangle$ is an operational distribution (one corresponding to an actual measurement), but that it is the canonical or target distribution to which an operational distribution may conform more or less well.

A similar result can be proved for joint measurements minimizing the product $\Delta_e x \Delta_e p$: namely, that the product is minimized if and only if the distribution of measured values is the Husimi function (Appleby [34], Werner [52], Busch *et al.* [14]). In Appleby [53], we extended the analysis to measurements of angular momentum and showed that a determination of spin-direction is optimal if and only if the distribution of measured values is $\langle \mathbf{n} | \hat{\rho} | \mathbf{n} \rangle$, where $|\mathbf{n}\rangle$ is a suitably-normalized SU(2) coherent state.

3. Physical Interpretation of the Operator Definitions

We now come to the problem of interpreting the quantities defined in the last section. Quantum mechanics forces us to drop the classical assumption that a measurement ascertains the pre-existing value of a specified observable [3–6]. Even if one postulates that the observable measured does have a pre-existing value, that value must typically differ from the value found by measurement. In the Bohm theory, for example, the result of a measurement of velocity is usually quite different from the postulated pre-existing velocity [54–57]. Classically, the error is usually defined in terms of the difference between the measured value and the pre-existing true one. It might consequently seem that, in abandoning the idea that measurements ascertain pre-existing values, we are obliged also to abandon the concept of experimental error (in the Introduction, we argued that that is exactly how it did seem to, for example, Heisenberg). We begin by showing that that is not the case. Specifically, we describe a classical model for which the classical error can be defined in a way that does not involve a comparison with pre-existing values. We then show that this alternative definition naturally carries over to quantum mechanics.

The example we consider is that of a one-dimensional classical gas. Let x and p be the position and momentum of a particular particle in this gas, and let $\lambda(dx dp)$ be the phase space probability measure. Suppose we measure x. Let μ_f be the pointer position after the measurement. We assume that the measurement process is stochastic and is described by a transition kernel $\chi(d\mu_f | x, p)$, such that the expectation value of a function $f(x, p, \mu_f)$ is given by (see, for example, Cinlar [58])

$$\langle f \rangle^\lambda = \int \left(\int f(x, p, \mu_f) \chi(d\mu_f | x, p) \right) \lambda(dx dp) \tag{23}$$

The superscript λ is to serve as a reminder that λ is arbitrary, unlike χ, which characterizes the measurement interaction and is therefore fixed. Define

$$\sigma_{ce}(x, p) = \left(\int (\mu_f - x)^2 \chi(d\mu_f | x, p) \right)^{\frac{1}{2}} \tag{24}$$

It will be seen that $\sigma_{ce}(x, p)$ is the RMS difference between the measured value and the pre-existing true one when λ is concentrated on the single point (x, p). We then define the classical error by

$$\Delta_{ce} x = \sup_{x, p \in \mathbb{R}} \left(\sigma_{ce}(x, p) \right) \tag{25}$$

Of course, this definition is open to the same objection as the quantity $\Delta_e x$ defined in the last section; namely, that it is likely to be infinite for a realistic model. However, this need not detain us, because we are not interested in the model for its own sake, but only as a conceptual bridge, which will take us from classical intuition to a reasonable quantum mechanical definition of measurement error. Now, let

$$\bar{x}^\lambda = \int x \lambda(dx dp), \qquad\qquad \Delta^\lambda x = \left(\int (x - \bar{x}^\lambda)^2 \lambda(dx dp) \right)^{\frac{1}{2}} \tag{26}$$

be the mean and standard deviation relative to the measure λ. Then, by an application of the Cauchy–Schwartz inequality,

$$
\begin{aligned}
\sqrt{\langle (\mu_f - \bar{x}^\lambda)^2 \rangle^\lambda} &= \left(\langle (\mu_f - x)^2 \rangle^\lambda + \langle (x - \bar{x}^\lambda)^2 \rangle^\lambda + 2 \langle (x - \bar{x}^\lambda)(\mu_f - x) \rangle^\lambda \right)^{\frac{1}{2}} \\
&\leq \Delta^\lambda x + \sqrt{\langle (\mu_f - x)^2 \rangle^\lambda} \\
&\leq \Delta^\lambda x + \Delta_{ce} x
\end{aligned}
\tag{27}
$$

Thus, the classical error bounds the increase in the RMS deviation of the pointer position from the initial state mean as compared to the initial state standard deviation. Note that $\sqrt{\langle (\mu_f - \bar{x})^2 \rangle^\lambda}$, $\Delta^\lambda x$ are λ-dependent, but $\Delta_{ce} x$ is not. The inequality is actually tight. To see this, choose a sequence (x_n, p_n), such that $\sigma_{ce}(x_n, p_n) \to \Delta_{ce} x$, and let λ_n be the measure concentrated on the point (x_n, p_n). Then

$$
\sqrt{\langle (\mu_f - \bar{x}^{\lambda_n})^2 \rangle^{\lambda_n}} - \Delta^{\lambda_n} x \to \Delta_{ce} x
\tag{28}
$$

Therefore

$$
\Delta_{ce} x = \sup_{\lambda \in \mathcal{M}} \left(\sqrt{\langle (\mu_f - \bar{x}^\lambda)^2 \rangle^\lambda} - \Delta^\lambda x \right)
\tag{29}
$$

where $\mathcal{M}$ is the set of all phase-space probability measures. This gives us an alternative formula for the classical error.

We can derive a similar formula for the classical disturbance. Let p_f be the momentum after the measurement and $\xi(dp_f|x,p)$ the transition kernel, such that

$$
\langle f \rangle^\lambda = \int \left(\int f(p_f, x, p) \xi(dp_f|x,p) \right) \lambda(dxdp)
\tag{30}
$$

Define the classical disturbance by

$$
\Delta_{cd} p = \sup_{x,p \in \mathbb{R}} \left(\int (p_f - p)^2 \xi(dp_f|x,p) \right)^{\frac{1}{2}}
\tag{31}
$$

Then, by an argument similar to the one above, we find

$$
\Delta_{cd} p = \sup_{\lambda \in \mathcal{M}} \left(\sqrt{\langle (p_f - \bar{p}^\lambda)^2 \rangle^\lambda} - \Delta^\lambda p \right)
\tag{32}
$$

We are now free to throw away the ladder and take Equations (29) and (32) to be the *definitions* of $\Delta_{ce} x$, $\Delta_{cd} p$. These alternative definitions do not involve a direct comparison between the measured value and the pre-existing one. Consequently, they do not involve the expectation values of products of pairs of variables like μ_f and x, which, in a quantum mechanical context, become non-commuting operators. Instead, they are framed in terms of the moments of probability distributions, which are also defined in quantum mechanics. They therefore generalize. Just as we can classically, so in quantum mechanics, we can define the error and disturbance in terms of the increase in an RMS deviation from an initial state mean

$$
\Delta_{De} x = \sup_{\hat{\rho} \in \mathcal{P}} \left(\Delta_{De}^{\hat{\rho}} x \right), \qquad \Delta_{De}^{\hat{\rho}} x = \sqrt{\mathrm{Tr}\left((\hat{\mu}_f - \bar{x}^{\hat{\rho}})^2 (\hat{\rho} \otimes \hat{\alpha}) \right)} - \Delta^{\hat{\rho}} x
\tag{33}
$$

$$
\Delta_{Dd} p = \sup_{\hat{\rho} \in \mathcal{P}} \left(\Delta_{Dd}^{\hat{\rho}} p \right), \qquad \Delta_{Dd}^{\hat{\rho}} p = \sqrt{\mathrm{Tr}\left((\hat{p}_f - \bar{p}^{\hat{\rho}})^2 (\hat{\rho} \otimes \hat{\alpha}) \right)} - \Delta^{\hat{\rho}} p
\tag{34}
$$

where we employ the notations of the last section, together with

$$\bar{x}^{\hat{\rho}} = \mathrm{Tr}(\hat{x}\hat{\rho}), \qquad\qquad\qquad \bar{p}^{\hat{\rho}} = \mathrm{Tr}(\hat{p}\hat{\rho}) \qquad\qquad (35)$$

We may also define

$$\Delta_{\mathrm{De}}^{\mathcal{R}} x = \sup_{\hat{\rho}\in\mathcal{R}}\left(\Delta_{\mathrm{De}}^{\hat{\rho}} x\right), \qquad\qquad \Delta_{\mathrm{Dd}}^{\mathcal{R}} p = \sup_{\hat{\rho}\in\mathcal{R}}\left(\Delta_{\mathrm{Dd}}^{\hat{\rho}} p\right) \qquad\qquad (36)$$

We will refer to these as the D definitions ("D" for "maximal increase in the RMS deviation from the initial state mean"). They are important because they show that the Bell–Kochen–Specker theorem is not, as it might seem, an insuperable obstacle blocking the path from the original classical intuition to a satisfactory quantum generalization. On the contrary, if the concepts are appropriately formulated, there is complete continuity between classical and quantum in this regard. However, although the D definitions are valid and useful, they should not be regarded as canonical. In the first place, there are other classical definitions that also have natural quantum generalizations (as we will see in the next paragraph). In the second place, there is no reason to make classical physics the arbiter. There may be useful quantum definitions that are not the generalization of any classical concept.

We arrive at another natural generalization of classical ideas if we consider measurements on a pair of correlated particles. Suppose we have two particles with positions $\hat{x}_A$, $\hat{x}_B$ and momenta $\hat{p}_A$, $\hat{p}_B$ and suppose we measure $\hat{x}_B$. Suppose that the unitary operator describing the measurement interaction is of the form $\hat{I}\otimes\hat{U}$, where $\hat{U}$ acts on $\mathcal{H}_B\otimes\mathcal{H}_{\mathrm{ap}}$ in the product $\mathcal{H}_A\otimes\mathcal{H}_B\otimes\mathcal{H}_{\mathrm{ap}}$ ($\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}_{\mathrm{ap}}$ being respectively the Hilbert spaces of particles A, B and the apparatus). Let $\hat{\mu}_B$ be the pointer position. Classically, it would be natural to define the error to be the maximal increase in the correlation $\langle(\mu_{B,\mathrm{f}} - x_A)^2\rangle$ as compared to $\langle(x_{B,\mathrm{i}} - x_A)^2\rangle$ and the disturbance to be the maximal increase in the correlation $\langle(p_{B,\mathrm{f}} - p_A)^2\rangle$ as compared to $\langle(p_{B,\mathrm{i}} - p_A)^2\rangle$. The point to notice here is that $[\hat{\mu}_{B,\mathrm{f}},\hat{x}_A] = [\hat{x}_{B,\mathrm{i}},\hat{x}_A] = [\hat{p}_{B,\mathrm{f}},\hat{p}_A] = [\hat{p}_{B,\mathrm{i}},\hat{p}_A] = 0$. Therefore, the classical definitions are expressed in terms of the moments of probability distributions, which are also defined quantum mechanically. They therefore generalize to

$$\Delta_{\mathrm{Ce}} x = \sup_{\hat{\rho}\in\mathcal{P}}\left(\Delta_{\mathrm{Ce}}^{\hat{\rho}} x\right), \qquad\qquad \Delta_{\mathrm{Cd}} p = \sup_{\hat{\rho}\in\mathcal{P}}\left(\Delta_{\mathrm{Cd}}^{\hat{\rho}} p\right) \qquad\qquad (37)$$

where

$$\Delta_{\mathrm{Ce}}^{\hat{\rho}} x = \sup_{\hat{\rho}_{AB}\in\mathcal{P}_{AB}}\left(\sqrt{\mathrm{Tr}\big((\hat{\mu}_{B,\mathrm{f}} - \hat{x}_A)^2(\hat{\rho}_{AB}\otimes\hat{\alpha})\big)} - \sqrt{\mathrm{Tr}\big((\hat{x}_{B,\mathrm{i}} - \hat{x}_A)^2(\hat{\rho}_{AB}\otimes\hat{\alpha})\big)}\right) \qquad (38)$$

$$\Delta_{\mathrm{Cd}}^{\hat{\rho}} p = \sup_{\hat{\rho}_{AB}\in\mathcal{P}_{AB}}\left(\sqrt{\mathrm{Tr}\big((\hat{p}_{B,\mathrm{f}} - \hat{p}_A)^2(\hat{\rho}_{AB}\otimes\hat{\alpha})\big)} - \sqrt{\mathrm{Tr}\big((\hat{p}_{B,\mathrm{i}} - \hat{p}_A)^2(\hat{\rho}_{AB}\otimes\hat{\alpha})\big)}\right) \qquad (39)$$

$\mathcal{P}_{AB}$ being the set of physical states $\hat{\rho}_{AB}$, such that $\mathrm{Tr}_A(\hat{\rho}_{AB}) = \hat{\rho}$. We may also define

$$\Delta_{\mathrm{Ce}}^{\mathcal{R}} x = \sup_{\hat{\rho}\in\mathcal{R}}\left(\Delta_{\mathrm{Ce}}^{\hat{\rho}} x\right), \qquad\qquad \Delta_{\mathrm{Cd}}^{\mathcal{R}} p = \sup_{\hat{\rho}\in\mathcal{R}}\left(\Delta_{\mathrm{Cd}}^{\hat{\rho}} p\right) \qquad\qquad (40)$$

We refer to these quantities as the C definitions ("C" for "correlation").

Let us now turn to the definitions in Section 2, which we will refer to as the O definitions ("O" for "operator"). The commutators $[\hat{\mu}_{\mathrm{f}},\hat{x}_{\mathrm{i}}]$ and $[\hat{p}_{\mathrm{f}},\hat{p}_{\mathrm{i}}]$ are typically non-zero, so the O quantities are typically not generalizations of the corresponding classical quantities, as Busch and co-workers have

stressed [24,49]. The O quantities do, however, impose bounds on the D and C quantities, and this gives them an indirect physical interpretation. We have

$$\Delta_e^{\hat{\rho}} x = \left(\text{Tr}\Big(\big((\hat{\mu}_f - \bar{x}^{\hat{\rho}})^2 - \{(\hat{\mu}_f - \bar{x}^{\hat{\rho}}), (\hat{x}_i - \bar{x}^{\hat{\rho}})\} + (\hat{x}_i - \bar{x}^{\hat{\rho}})^2 \big) (\hat{\rho} \otimes \hat{\alpha}) \Big) \right)^{\frac{1}{2}} \tag{41}$$

from which it follows

$$\Delta_{De}^{\hat{\rho}} x \leq \Delta_e^{\hat{\rho}} x \leq \Delta_{De}^{\hat{\rho}} x + 2\Delta^{\hat{\rho}} x \tag{42}$$

for all $\hat{\rho} \in \mathcal{P}$. Similarly

$$\Delta_{Dd}^{\hat{\rho}} x \leq \Delta_d^{\hat{\rho}} x \leq \Delta_{Dd}^{\hat{\rho}} x + 2\Delta^{\hat{\rho}} x \tag{43}$$

$$\Delta_{Ce}^{\hat{\rho}} x \leq \Delta_e^{\hat{\rho}} x \leq \Delta_{Ce}^{\hat{\rho}} x + 2\Delta^{\hat{\rho}} x \tag{44}$$

$$\Delta_{Cd}^{\hat{\rho}} x \leq \Delta_d^{\hat{\rho}} x \leq \Delta_{Cd}^{\hat{\rho}} x + 2\Delta^{\hat{\rho}} x \tag{45}$$

Taking suprema, we deduce

$$\Delta_{De}^{\mathcal{R}} x \leq \Delta_e^{\mathcal{R}} x \leq \Delta_{De}^{\mathcal{R}} x + 2 \sup_{\hat{\rho} \in \mathcal{R}} \left(\Delta^{\hat{\rho}} x \right) \tag{46}$$

$$\Delta_{Dd}^{\mathcal{R}} x \leq \Delta_d^{\mathcal{R}} x \leq \Delta_{Dd}^{\mathcal{R}} x + 2 \sup_{\hat{\rho} \in \mathcal{R}} \left(\Delta^{\hat{\rho}} x \right) \tag{47}$$

$$\Delta_{Ce}^{\mathcal{R}} x \leq \Delta_e^{\mathcal{R}} x \leq \Delta_{Ce}^{\mathcal{R}} x + 2 \sup_{\hat{\rho} \in \mathcal{R}} \left(\Delta^{\hat{\rho}} x \right) \tag{48}$$

$$\Delta_{Cd}^{\mathcal{R}} x \leq \Delta_d^{\mathcal{R}} x \leq \Delta_{Cd}^{\mathcal{R}} x + 2 \sup_{\hat{\rho} \in \mathcal{R}} \left(\Delta^{\hat{\rho}} x \right) \tag{49}$$

When $\mathcal{R} = \mathcal{P}$, these inequalities reduce to

$$\Delta_{De} x \leq \Delta_e x, \qquad \Delta_{Dd} x \leq \Delta_d x, \qquad \Delta_{Ce} x \leq \Delta_e x, \qquad \Delta_{Cd} x \leq \Delta_d x \tag{50}$$

We also have the following constraints on the relative sizes of the D and C quantities:

$$\left| \Delta_{Ce}^{\hat{\rho}} x - \Delta_{De}^{\hat{\rho}} x \right| \leq 2\Delta^{\hat{\rho}} x, \qquad\qquad \left| \Delta_{Cd}^{\hat{\rho}} x - \Delta_{Dd}^{\hat{\rho}} x \right| \leq 2\Delta^{\hat{\rho}} x \tag{51}$$

$$\left| \Delta_{Ce}^{\mathcal{R}} x - \Delta_{De}^{\mathcal{R}} x \right| \leq 2 \sup_{\hat{\rho} \in \mathcal{R}} \left(\Delta^{\hat{\rho}} x \right), \qquad\qquad \left| \Delta_{Cd}^{\mathcal{R}} x - \Delta_{Dd}^{\mathcal{R}} x \right| \leq 2 \sup_{\hat{\rho} \in \mathcal{R}} \left(\Delta^{\hat{\rho}} x \right) \tag{52}$$

These inequalities mean, among other things, that the O quantities are upper bounds on the corresponding D and C quantities.

Our discussion raises some important questions. If $\sup_{\hat{\rho} \in \mathcal{R}} \left(\Delta^{\hat{\rho}} x \right)$ is large, then the above inequalities are consistent with one of the O quantities being large, while the corresponding D and C quantities are both small. They also leave open the possibility that, in the case when $\sup_{\hat{\rho} \in \mathcal{R}} \left(\Delta^{\hat{\rho}} x \right)$ is large, one of the D quantities is large, while the corresponding C quantity is small, or *vice versa*. One would like to know if these possibilities are actually realized.

Korzekwa *et al.* [59] answer the first of these questions for the case of the state-dependent disturbances. Consider two non-commuting observables $\hat{R}$, $\hat{S}$ on a finite dimensional Hilbert space. Suppose that the system is initially in an eigenstate of $\hat{R}$, which is not also an eigenstate of $\hat{S}$, and suppose that one makes a von Neumann measurement of $\hat{R}$. Then, the D and C disturbances are both zero, while the O disturbance is non-zero.

Busch [60] gives an example that shows that it is possible for the state-dependent D and C errors to be zero, while the state-dependent O error is non-zero. Unlike Korzekwa *et al.*'s example, it is

rather artificial (it is a quantum version of the broken-ammeter scenario); however, it is enough to establish the point of principle. Suppose the system and pointer particles are both spin-1/2 particles and that the measured observable and pointer observables are the $\hat{\sigma}_z$ operators for their respective particles. Suppose that the initial system + apparatus state is $|\psi\rangle \otimes |\psi\rangle$ and that $\hat{U} = \hat{I}$. Then, it is easily seen that the state-dependent O error is the ordinary uncertainty of $\hat{\sigma}_z$ in the state $|\psi\rangle$, while the state-dependent D and C errors are zero.

We can use a modification of this example to show that it is possible for the state-dependent D quantities to be zero while the state-dependent C quantities are non-zero. Let everything be as in the last paragraph, except that the system + apparatus are in the maximally-mixed state $(1/4)\hat{I} \otimes \hat{I}$. Then, the D error is zero, while the C error is $\sqrt{2}$ (the supremum in Equation (38) being achieved for the maximally-entangled state $\hat{\rho}_{AB} = |\Psi\rangle\langle\Psi|$ with $|\Psi\rangle = (1/\sqrt{2}(|+\rangle \otimes |+\rangle + |-\rangle \otimes |-\rangle)$, where $|\pm\rangle$ are the eigenstates of $\hat{\sigma}_z$). To show that the same is true of the D and C disturbances, continue to assume that system+apparatus are in the maximally-mixed state, but take the evolution operator $\hat{U}$ to be $\hat{\sigma}_y \otimes \hat{I}$. Then, the state-dependent D disturbance to the observable $\hat{\sigma}_x$ is zero, while the state-dependent C disturbance is $\sqrt{2}$ (the supremum in Equation (39) being achieved for the maximally-entangled state $\hat{\rho}_{AB} = |\Psi\rangle\langle\Psi|$ with $|\Psi\rangle = (1/\sqrt{2}(|+\rangle \otimes |+\rangle + |-\rangle \otimes |-\rangle)$, where $|\pm\rangle$ are the eigenstates of $\hat{\sigma}_x$).

Of course, the last three examples (unlike the example of Korzekwa *et al.*) are somewhat artificial. It would be interesting to see if the conclusion continues to hold for more realistic measuring processes. Furthermore, we have not addressed the more challenging and, to our mind, more interesting question, of what can be said in the state-*independent* case. This requires further investigation.

The D and C quantities have a direct, operational interpretation as errors and disturbances. The smallness of one of these quantities is both necessary and sufficient for the measurement to be accurate or non-disturbing in a well-defined, operational sense. By contrast, the interpretation of the O quantities, as we have presented it here, is indirect: their meaning comes from the fact that they supply various bounds on the D and C quantities. Moreover, although the smallness of an O quantity is sufficient, we have not been able to show that it is necessary for the measurement to be accurate or non-disturbing in a well-defined sense. In the case of the state-independent quantities, it is possible that, with more work, one could establish necessitys as well. If that were so, it would mean, in effect, that the state-independent O quantities were fully operational characterizations of the error and disturbance.

Finally, let us note that there is no reason to assume that our analysis is complete. The O quantities may capture other operationally identifiable features of the measurement, which the D and C quantities both miss.

4. Response to Criticisms

We now consider BLW's critique of the O definitions (also see Busch *et al.* [49] and Korzekwa *et al.* [59]). BLW contrast the O approach with what they call a distributional approach. They argue that, although the O approach has its uses in certain special cases, the version of the distributional approach based on the Wasserstein two-deviation is, in general, greatly preferable. In addressing their criticisms, let us begin by observing that the D and C definitions are themselves distributional definitions. Moreover, although the O quantities are not defined distributionally, their physical interpretation (as given in Section 3) depends on the fact that they supply various bounds on the corresponding D and C quantities. Therefore, the distinction between operator and distributional approaches is less clear-cut than it may initially appear. The problem is not really to decide between a distributional approach and some other completely different approach; rather, it is to decide between alternative versions of the distributional approach. As with all such problems, the answer is dependent on the situation of interest. In the following, it is certainly not our intention to suggest that the O definitions are preferable to BLW's definitions in every situation. We only argue

that there is a physically-important class of situations in which the D definitions, and consequently, the O definitions, are preferable.

It should be observed that BLW's criticisms are directly almost entirely against Ozawa's state-dependent version of the O approach. Concerning our state-independent version, they go so far as to say, on p. 1278 of [24], that the quantity $\Delta_e x$ is a "reliable indicator of the presence or absence of differences between the target and approximator observables". However, they qualify that statement by adding that it is a "curiosity" that that should be so, since "the error interpretation of the state-dependent quantities used for its determination [*i.e.*, the determination of $\Delta_e x$] is not generally applicable". We hope that our discussion in previous sections will have gone some way towards clarifying the interpretation of $\Delta_e x$ and that our discussion in this section, of BLW's criticisms of the state-dependent O definitions (on which, as they correctly observe, the interpretation of $\Delta_e x$, $\Delta_d p$ depend) will go further in that direction. Nevertheless, we will leave open what is perhaps the most important question, whether there exist processes, which are highly accurate (respectively, non-disturbing) as judged by any reasonable operational criterion, for which $\Delta_e x$ (respectively, $\Delta_d p$) is large.

BLW accept that the O definitions give valid characterizations of the error (respectively, disturbance) under conditions where the observables $\hat{x}_i$, $\hat{\mu}_f$ (respectively, $\hat{p}_i$, $\hat{p}_f$) commute. However, in cases where these observables do not commute, they argue that $\hat{x}_i$, $\hat{\mu}_f$ (respectively, $\hat{p}_i$, $\hat{p}_f$) are not jointly measurable and, consequently, that the interpretation of $\hat{\epsilon}_X$, $\hat{\delta}_P$ as error and disturbance operators is ungrounded. This objection would be justified if we were relying on a naive, purely formal analogy with the classical expressions $\langle (\mu_f - x_i)^2 \rangle$ and $\langle (p_f - p_i)^2 \rangle$. However, since we are actually relying on the fact that the O quantities bound the D and C quantities and since the definitions of the latter are just as operational as BLW's own definitions, there is no problem here.

BLW go on to substantiate their criticisms by giving examples of measurements where the O error is zero, even though the distribution of measured values is quite different from the initial state distribution. We will here confine ourselves to their Example 7 in [24]. The reader will easily perceive that a suitably modified version of our discussion applies to their Examples 8, 9 and 10 in [24] (also to Example 5 in [49]). The example is of a measurement of position in which the POVM (positive operator valued measure) describing the distribution of measured values is the spectral measure of the shifted oscillator Hamiltonian

$$\hat{H}' = \hat{x} + \alpha \left(\hat{H} - \frac{\hbar\omega}{2} \right), \qquad\qquad \hat{H} = \frac{1}{2m}\hat{p}^2 + \frac{m\omega^2}{2}\hat{x}^2 \qquad (53)$$

and in which the initial system state is the ground state of $\hat{H}$. It is easily verified that $\Delta_e^{\hat{\rho}} x = 0$. On the other hand, it can be seen from Figure 1 that the probability distributions for $\hat{x}_i$ and $\hat{\mu}_f$ are very different. In particular, the distribution for $\hat{x}_i$ is continuous, whereas that for $\hat{\mu}_f$ is discrete. BLW take this to mean that the measurement is highly inaccurate and that the O definition of error is correspondingly misleading. They are right to the extent that there are applications (tomography, for example) for which this measurement would be very ill-suited. However, the purpose of a measurement is not always to accurately reproduce the initial state probability distribution. That is obviously the case in classical physics. Consider, for instance, measurements using a digital ammeter. Here, too, the initial state probability distribution is continuous, while the distribution of measured values is discrete. However, this would not usually be seen as a reason for preferring an analogue meter. The same is true in quantum physics: there are situations where one is only concerned with certain specific features of the distribution of measured values, its detailed shape being otherwise unimportant. Consider, for instance, a state discrimination problem where Bob is promised one of a finite set of N non-overlapping wave-packets localized within the intervals $(n - 1/2, n + 1/2)$, $n = 1, \ldots, N$, and he has to decide which particular wave-packet Alice has sent. In a situation like this, the only important probability is the probability of Bob misidentifying the state that Alice sent. The probability distribution of measured values is irrelevant, except in so far as it has consequences

for this failure probability. In particular, there is no reason to prefer a measurement for which the distribution of measured values is continuous. Indeed, it is easily seen that there is a measurement for which the possible pointer values are $0, 1, \ldots, N$ (with zero corresponding to an input state whose support is disjoint from the interval $(1/2, N + 1/2)$) and for which the distribution of measured values is consequently discrete, but having zero failure probability. It is also easy to see that there are measurements with continuous distributions of measured values, more closely approximating the initial state probability distribution in the sense of the Wasserstein two-deviation, but having failure probability greater than zero. Broadly and qualitatively speaking, what one wants in this situation is that the quantity $\langle (\hat{\mu}_f - \bar{x})^2 \rangle$ be as small as possible. A measurement like the one described in BLW's Example 7 satisfies this requirement. The distributions depicted in Figure 1 are indeed very different. However, they have exactly the same mean and variance. Consequently, $\langle (\hat{\mu}_f - \bar{x})^2 \rangle$ is not enlarged at all as compared to the initial state variance. This is one of the pieces of information conveyed by the statement that $\Delta_e^{\hat{\rho}} x = 0$ (see inequality (50)), which is not misleading at all, *provided* it is correctly understood. By contrast, the Wasserstein two-deviation would cause one to prefer, to the measurement depicted in Figure 1, one for which the second distribution was a smeared out version of the first—even though this is likely to be worse for Bob's particular purposes.

(a)

(b)
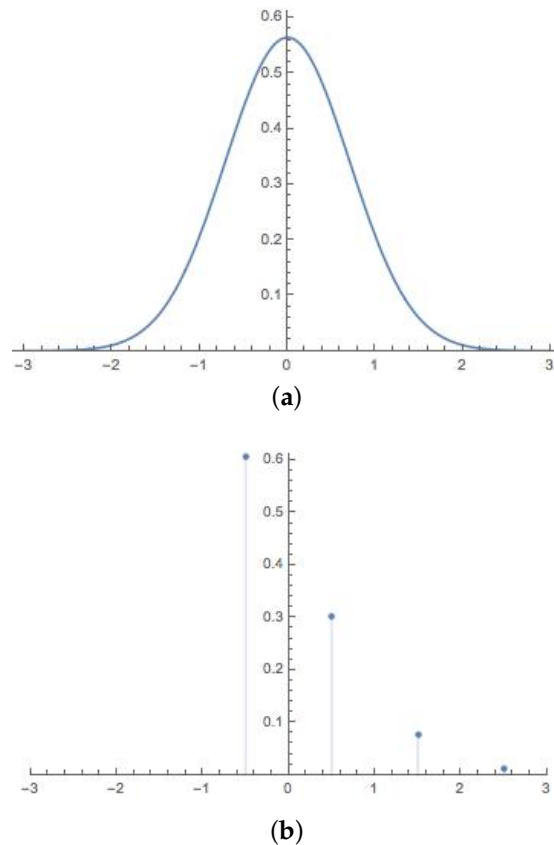
Figure 1. Probability distributions for $\hat{x}_i$ (**a**) and $\hat{\mu}_f$ (**b**) in Busch, Lahti and Werner's (BLW) Example 7 (units chosen so that $\hbar = m/\alpha = \alpha\omega = 1$).

Similarly with the disturbance: in a situation where one is interested in the deviation from the initial state mean, but not in any other feature of the probability distribution, then the D definition, and consequently the O definition of disturbance, will be more useful than the one based on the Wasserstein two-deviation.

It is seldom, if ever, the case, that a single figure of merit captures every potentially relevant feature of a piece of technology. Suppose one is buying a car. If one wants a vehicle that can drive

very fast round a carefully prepared track one will choose one figure of merit; if, on the other hand, one wants a vehicle suitable for conveying a family of six to the beach, one will choose another, quite different figure of merit. Similarly with quantum measurements.

In their Examples 7–10, BLW criticize the O definitions on the grounds that the O error can be zero in situations where the initial state and final pointer distributions are very different. In Examples 4 and 6 of Busch *et al.* [49] and Example 3 of Busch [60], the authors make the opposite point, that the O error can be large in situations where the initial state and final pointer distributions are identical; a fact that they regard as an evident defect of the operator approach. Their argument is based on the principle that a perfectly accurate measurement is one which perfectly reproduces the initial state probability distribution. To see that the principle is not generally valid, consider the following scenario:

> Alice lives in a city where 50% of the population is infected with HIV. She is worried that she may have it, so she goes to her doctor Bob to be tested. Bob pulls a coin out of his pocket and tosses it. He then puts on a grave face and says "I am sorry, I have bad news for you." Alice is outraged, on the grounds that this is not a proper test. Bob, however, insists that it is a proper test. After all, it has the same probability distribution. What more can she want?

This is a classical example. One can easily construct a quantum example. Suppose, for instance, that Alice and Bob are two students who want to perform a test of the Bell inequalities. Unfortunately, they cannot afford state-of-the-art photon counters, so they decide that Alice will toss a fair coin at her station and that Bob will independently toss another fair coin at his. On the principle adopted in Examples 4 and 6 of [49] and Example 3 of [60], these are perfectly accurate measurements. However, they will, of course, fail to reveal any correlations between the two particles.

Outside of the three examples under discussion, Busch and his co-workers adopt a state-independent version of the principle, according to which a measurement is perfectly accurate if it perfectly reproduces the initial state distribution *for every initial state*. The phrase in italics makes a crucial difference, as can be seen from the following modified version of the doctor scenario (originally suggested by Poulin [61]):

> Alice takes 10 cities, in each of which the incidence of HIV is different. She then takes a sample of 100 people from each of these cities and presents them to Bob for testing; without, however, telling Bob which patient comes from which city. It turns out the proportion of positive test results for each city coincides with the actual proportion of HIV-infected people in that city. Alice concludes that, whatever it is that Bob is doing, it probably deserves to be considered a test.

Similarly with the state-independent version of the Busch *et al.* principle: if a measurement reproduces the initial state distribution for *every* choice of state, then it is very plausible to argue that it is, in some sense, highly accurate. Calculation confirms that impression. In particular, it is easily seen that a measurement for which the state-independent Wasserstein two-error is zero will successfully reveal the correlations in a Bell experiment.

However, in the examples under discussion, Busch and his co-workers adopt a state-*dependent* version of the principle. Like the state-dependent version of the operator approach, this version of the principle can easily lead to unreasonable conclusions (*cf.* the broken-ammeter scenario in Section 2). To show that their objection is not valid, we will focus on Example 3 in [60]. The extension to the other two examples will, we hope, be apparent. We have already discussed this example at the end of Section 3 (specializing to the case of a spin measurement). As we noted there, the O error is non-zero if the initial system state is not an eigenstate of $\hat{\sigma}_z$. On the other hand, the fact that the initial system and apparatus states are the same and the fact that the system and apparatus do not interact means that the distribution of measured values is identical to the initial state probability distribution of the

measured observable. Busch argues on the basis of this that the measurement is perfectly accurate. The fact that the example is a quantum version of the first doctor scenario may make one suspicious of this conclusion. Busch is, of course, well-aware that the set-up envisaged is not the kind of thing anyone would normally call a measuring apparatus and, indeed, he explicitly draws attention to the fact by comparing it to a broken clock. His point is that a broken clock actually is right twice a day, and he thinks that the same applies to his example. To see that there is an important difference between the two cases, consider a situation where the measured particle is one of a maximally-entangled pair (this differs from the situation we considered in Section 3, where we took it that the measured particle was initially in a pure state). In that case, the C error is non-zero and equal to the O error.

It is (to say the least) questionable whether the process just described counts as a measurement at all. Yet, not only the state-dependent Wasserstein two-error, but also the state-dependent D and C errors are zero. That is not a weakness of the definitions: In all three cases, the fact that the error is zero is a well-defined operational statement, which happens to be true—just as Bob's statement, in the broken-ammeter scenario, happens to be true. It does, however, illustrate the limitations of state-dependent definitions. We argued in Section 2 that state-dependent definitions have their uses. However, they need to be used with caution. In particular, a state-dependent error is not a figure of merit: its smallness does not, by itself, mean that a measurement is in any sense "good".

At this point, we ought to stress that, although their arguments are, as it seems to us, invalid, the point that Busch and his co-workers are trying to establish, that there are measurements that are highly accurate as judged by any reasonable operational criterion, but for which the O error is large, could be right. In Section 3, we showed that smallness of the D or C quantities is both necessary and sufficient for a measurement to be accurate or non-disturbing in a well-defined, operational sense. However, in the case of the state-independent O quantities, we only established sufficiency. In the state-*dependent* case, Korzekwa *et al.* [59] have shown that there are processes that are completely non-disturbing as judged by any reasonable operational criterion, but for which the state-dependent O disturbance is non-zero (see the discussion at the end of Section 3). However, it remains an open question whether the same is true of the state-dependent O error. The more challenging and, to our mind more important question, of what can be said regarding the state-*independent* O errors and disturbances, also remains open.

5. Conclusions

In the Introduction, we argued that we need to develop a unified theory of measurement, in which classical measurements are seen as a limiting case of quantum measurements, rather in the way that Newtonian kinematics is a limiting case of relativistic kinematics. In particular, we need an overarching quantum mechanical concept of measurement accuracy that effectively reduces to the classical one in special cases, such as measurements with a meter rule. We argued that, contrary to initial appearances, the Bell-KStheorem is not a major obstacle.

We made a start on the problem by showing that there are at least two ways to reformulate the classical definitions of error and disturbance in a way that does not involve a comparison with pre-existing values. The reformulated definitions have natural quantum generalizations, which we called the D and C definitions. The D and C definitions are examples of quantum definitions that reduce to the classical concept in special cases. They also bound the O quantities introduced in [27–29]. They thereby give physical meaning to the O quantities.

We then turned to BLW's criticisms of the O definitions. We argued that one should not expect there to be a single, canonical way of quantifying the concepts of measurement accuracy and disturbance. The answer to the question "what is the most appropriate quantitative definition?" is always relative to the physical problem of interest. We specified a class of problems for which the D definitions, and consequently, the O definitions, are more appropriate than BLW's definitions based on the Wasserstein two-deviation.

Our analysis raises a number of questions that might be interesting to investigate. Firstly, in the state-independent case, one would like to know (1) whether the smallness of the O quantities is necessary and sufficient for the corresponding D and C quantities both to be small and (2) whether the smallness of the D quantities is necessary and sufficient for the corresponding C quantities to be small. Secondly, we have seen that there are physical problems for which the D quantities are better-suited than the ones based on the Wasserstein two-deviation. One would like to know if the same is true of the C quantities. Thirdly, it would be interesting to see if the O quantities capture any other operationally well-defined feature of the measurement, in addition to those captured by the D and C quantities. Finally, it would be interesting to see if one can prove error-disturbance and error-error relations expressed in terms of the D and C quantities.

Appendix A. Proof of Inequalities (13), (14), (17) and (18)

In [27], we gave a proof of inequalities (13), (14), (17) and (18), which skated over certain questions of domain and differentiability. Our proof has since been criticized for its lack of rigour. We here take the opportunity to fill in the missing details. We incidentally strengthen the statement of inequalities (17) and (18) slightly.

The problem we face is that the operators $\hat{x}$, $\hat{p}$ are not defined on the whole Hilbert space. Specifically, $|\psi\rangle$ is in the domain of $\hat{x}$ if and only if the function $x^2|\langle x|\psi\rangle|^2$ is integrable. The domain of $\hat{p}$ is defined similarly. Expressed more intuitively: $|\psi\rangle$ is in the domain of $\hat{x}$ if and only if $\langle x^2\rangle < \infty$. We take the view that states for which that is not true are never realized in a real, Earthbound laboratory experiment. To put it more succinctly: they are unphysical.

One's first thought may be that one can define the domain of physical states to be the set of all $|\psi\rangle$, such that $\langle x|\psi\rangle$ (respectively, $\langle p|\psi\rangle$) is zero for all x, such that $|x| > B_X$ (respectively, all p, such that $|p| > B_P$), for suitably large positive constants B_X, B_P. However, the set of such states is empty (because, if $\langle x|\psi\rangle$ is zero outside the interval $[-B_X, B_X]$, then its Fourier transform is analytic, by Schwartz's extension of the Paley–Wiener theorem [see, for example, Treves [62]]). Nevertheless, although the theory forces at least one of the wave-functions $\langle x|\psi\rangle$, $\langle p|\psi\rangle$ to have an infinite tail, nothing observable under ordinary laboratory conditions can depend on it. It is not possible that a currently performable laboratory experiment can give rise to a state in which there is a significant probability of the momentum being greater than $10^{(10^{10})}$ kg·m·s^{-1}, and even if it were possible, one would not use non-relativistic quantum mechanics to describe it; nor is it possible to produce states for which there is a significant probability of $|x|$ being greater than $10^{(10^{10})}$ m. We need to give quantitative expression to this point, that the infinite tails are physically irrelevant. We accordingly take the view (inspired by the rigged Hilbert space formulation of quantum mechanics [63–65]) that the set of physical pure states $\mathcal{P}_0$ consists of those states $|\psi\rangle$ for which the position space wave function $\langle x|\hat{\rho}|y\rangle$ is (a)C^∞ and (b) rapidly decreasing at infinity in the sense that

$$\sup_{x\in\mathbb{R}}\left(|1+|x|^2)^n \left|\frac{\partial^m}{\partial x^m}\langle x|\psi\rangle\right|\right) < \infty \tag{A1}$$

for every pair of non-negative integers n, m (in other words, $\langle x|\psi\rangle$ is a test function for the space of tempered distributions [62]). Note that this is equivalent to requiring that the momentum space wave function is C^∞ and rapidly decreasing at infinity. Note also that $\mathcal{P}_0$ is in the domain of every monomial in $\hat{x}$ and $\hat{p}$.

At first sight, this definition may appear arbitrary. The reader may allow that it is reasonable to impose some restriction on the behaviour at infinity, but wonder why we make this particular choice. Indeed, the requirement is much stronger than we need for our present purposes. We make

the definition nonetheless because actually it is not arbitrary, as can be seen by using a von Neumann lattice [9,66,67]. For some appropriate scale-length λ, let $|n, m\rangle$ be the coherent state with wave function

$$\langle x|n, m\rangle = \left(\frac{1}{\pi\lambda^2}\right)^{\frac{1}{4}} e^{-\frac{1}{2\lambda^2}(x-n\lambda)^2 + \frac{2\pi mi}{\lambda}x} \tag{A2}$$

Then, the set $\{|n, m\rangle : n, m \in \mathbb{Z}\}$ with one point removed is a basis. Choose some suitably enormous integer N, and let $\hat{P}$ be the projector onto the finite dimensional subspace spanned by the set $\{|n, m\rangle : -N \leq n, m \leq N\}$. Then, for any state $|\psi\rangle$ that is relevant to a real laboratory experiment, the quantity $\|(1 - \hat{P})|\psi\rangle\|)$ will be negligible. Consequently, predictions obtained using the state $|\psi\rangle$ will be experimentally indistinguishable from ones obtained using the state

$$|\psi_r\rangle = \frac{1}{\|\hat{P}|\psi\rangle\|}\hat{P}|\psi\rangle \tag{A3}$$

Without loss of predictive power, we may therefore replace $|\psi\rangle$ with $|\psi_r\rangle$. The fact that $|\psi_r\rangle$ is a finite linear combination of coherent states means that it belongs to $\mathcal{P}_0$. Of course, $\mathcal{P}_0$ also includes states like $|n, m\rangle$ with n, m both much larger than N, which are certainly not relevant to ordinary, Earthbound laboratory experiments (being localized outside the cosmic event horizon). The point is only that every pure state that is experimentally relevant is empirically indistinguishable from a state in $\mathcal{P}_0$.

Finally, we need to define $\mathcal{P}$, the set of physical density matrices. For each non-negative integer m and real β, define norms

$$N_{m,\beta}(|\psi\rangle) = \sup_{x \in \mathbb{R}} \left((1 + |x|^2)^\beta \left|\frac{\partial^m}{\partial x^m}\langle x|\psi\rangle\right|\right)$$
$$\tilde{N}_{m,\beta}(|\psi\rangle) = \sup_{p \in \mathbb{R}} \left((1 + |p|^2)^\beta \left|\frac{\partial^m}{\partial p^m}\langle p|\psi\rangle\right|\right) \tag{A4}$$

Now, let

$$\hat{\rho} = \sum_n \xi_n |n\rangle\langle n| \tag{A5}$$

be an arbitrary density matrix with eigenvectors $|n\rangle$. We define $\mathcal{P}$ to consist of those $\hat{\rho}$ for which $|n\rangle \in \mathcal{P}_0$ for all n and for which

$$\sup_n \left(N_{m,\beta}(|n\rangle)\right) < \infty \qquad\qquad \sup_n \left(\tilde{N}_{m,\beta}(|n\rangle)\right) < \infty \tag{A6}$$

for all m, β. Note that it is enough to demand that one set of suprema is finite, since the finiteness of the other is then automatic. Note also that in the case when the spectrum of $\hat{\rho}$ has degeneracies, the finiteness of the suprema does not depend on the particular choice of eigenvectors. Finally, let us remark that for the technical purposes of this Appendix, it would be enough to require that the suprema are finite for the particular case $m = 0$, $\beta = 3/2$.

This definition is justified by the fact that no experimentally relevant density matrix can be distinguished empirically from a state in $\mathcal{P}$. Indeed, let $\hat{\rho}$ be an experimentally relevant density matrix, and let $\hat{P}$ be the projector onto the first N eigenstates. Choose N so that $\text{Tr}((1 - \hat{P})\hat{\rho})$ is smaller than some suitably tiny number. No practicable experiment can distinguish between $\hat{\rho}$ and

$(1/(\text{Tr}(\hat{P}\hat{\rho})))\hat{P}\hat{\rho}\hat{P}$. By the argument we used to justify the definition of $\mathcal{P}_0$, the state $(1/(\text{Tr}(\hat{P}\hat{\rho})))\hat{P}\hat{\rho}\hat{P}$ is in turn empirically indistinguishable from one of the form

$$\hat{\rho}_0 = \sum_n \xi_n |n\rangle\langle n| \tag{A7}$$

where the states $|n\rangle$ are a finite orthonormal set in $\mathcal{P}_0$. The fact that the set is finite means $\hat{\rho}_0 \in \mathcal{P}$.

The proof of the main theorem depends on three technical lemmas. Define displacement operators

$$\hat{D}_{xp} = e^{i(p\hat{x} - x\hat{p})} \tag{A8}$$

and, for each $|\psi\rangle \in \mathcal{P}_0$, let

$$|\psi_{xp}\rangle = \hat{D}_{xp}|\psi\rangle \tag{A9}$$

It is easily seen that $|\psi_{xp}\rangle \in \mathcal{P}_0$ for all x, p.

Lemma 1. *For all $|\psi\rangle \in \mathcal{P}_0$, the function $|\psi_{xp}\rangle$ is differentiable in the sense that*

$$\left\| \frac{1}{\epsilon}\left(|\psi_{x+\epsilon,p}\rangle - |\psi_{x,p}\rangle\right) + \frac{i}{\hbar}\left(\hat{p} - \frac{1}{2}p\right)|\psi_{x,p}\rangle \right\| \leq \epsilon B(4 + p^2)\tilde{N}_{0,\frac{3}{2}}(|\psi\rangle) \tag{A10}$$

$$\left\| \frac{1}{\epsilon}\left(|\psi_{x,p+\epsilon}\rangle - |\psi_{x,p}\rangle\right) - \frac{i}{\hbar}\left(\hat{x} - \frac{1}{2}x\right)|\psi_{x,p}\rangle \right\| \leq \epsilon B(4 + x^2)N_{0,\frac{3}{2}}(|\psi\rangle) \tag{A11}$$

for all x, p, all $\epsilon > 0$, where B is a fixed positive constant independent of $|\psi\rangle$.

Proof. It is easily seen that

$$\left| \frac{1}{\epsilon}\left(\langle p'|\psi_{x+\epsilon,p}\rangle - \langle p'|\psi_{x,p}\rangle\right) + \frac{i}{\hbar}\left(p' - \frac{1}{2}p\right)\langle p'|\psi_{x,p}\rangle \right|^2$$
$$\leq \frac{\epsilon^2 C}{16\hbar^4}(2p' - p)^4 \left|\langle p' - p|\psi\rangle\right|^2 \tag{A12}$$

where

$$C = \sup_{u \in \mathbb{R}} \left(\frac{(\cos u - 1)^2}{u^4} + \frac{(\sin u - u)^2}{u^4} \right) \tag{A13}$$

Hence

$$\left\| \frac{1}{\epsilon}\left(|\psi_{x+\epsilon,p}\rangle - |\psi_{x,p}\rangle\right) + \frac{i}{\hbar}\left(\hat{p} - \frac{1}{2}p\right)|\psi_{x,p}\rangle \right\|^2$$
$$\leq \frac{3\pi\epsilon^2 C}{128\hbar^2}(4 + p^2)^2 \left(\tilde{N}_{0,\frac{3}{2}}(|\psi\rangle)\right)^2 \tag{A14}$$

The second inequality is proved in the same way. $\square$

Lemma 2. *For all $|\psi\rangle \in \mathcal{P}_0$, $|\psi_{xp}\rangle$, $\hat{x}|\psi_{xp}\rangle$, $\hat{p}|\psi_{xp}\rangle$ are uniformly continuous on every compact subset of $\mathbb{R}^2$. Specifically, let $\mathcal{C}$ be such a set. Then*

$$\left\| |\psi_{x_1,p_1}\rangle - |\psi_{x_2,p_2}\rangle \right\| \leq \epsilon_x B_1 \tilde{N}_{0,\frac{3}{2}}(|\psi\rangle) + \epsilon_p B_1 N_{0,\frac{3}{2}}(|\psi\rangle) \tag{A15}$$

$$\left\| \hat{x}|\psi_{x_1,p_1}\rangle - \hat{x}|\psi_{x_2,p_2}\rangle \right\| \leq \epsilon_x B_2 \left(1 + \tilde{N}_{0,\frac{3}{2}}(|\psi\rangle) + \tilde{N}_{0,\frac{3}{2}}(\hat{x}|\psi\rangle)\right)$$

$$+ \epsilon_p B_2 \left(N_{0,\frac{3}{2}}(|\psi\rangle) + N_{0,\frac{3}{2}}(\hat{x}|\psi\rangle) \right) \tag{A16}$$

$$\left\| \hat{p}|\psi_{x_1,p_1}\rangle - \hat{p}|\psi_{x_2,p_2}\rangle \right\| \le \epsilon_x B_3 \left(\tilde{N}_{0,\frac{3}{2}}(|\psi\rangle) + \tilde{N}_{0,\frac{3}{2}}(\hat{p}|\psi\rangle) \right)$$
$$+ \epsilon_p B_3 \left(1 + N_{0,\frac{3}{2}}(|\psi\rangle) + N_{0,\frac{3}{2}}(\hat{p}|\psi\rangle) \right) \tag{A17}$$

for all (x_1, p_1), $(x_2, p_2) \in \mathcal{C}$, *where* $\epsilon_x = |x_1 - x_2|$, $\epsilon_p = |p_1 - p_2|$, *and where the* B_j *are positive constants that depend on* $\mathcal{C}$, *but not on* $|\psi\rangle$.

Proof. The first inequality is a straightforward consequence of Lemma 1 and the inequalities

$$\left\| \left(\hat{x} - \frac{1}{2}x \right) |\psi_{x,p}\rangle \right\|^2 \le \left(N_{0,\frac{3}{2}}(|\psi\rangle) \right)^2 \frac{\pi}{32}(4 + 3x^2) \tag{A18}$$

$$\left\| \left(\hat{p} - \frac{1}{2}p \right) |\psi_{x,p}\rangle \right\|^2 \le \left(\tilde{N}_{0,\frac{3}{2}}(|\psi\rangle) \right)^2 \frac{\pi}{32}(4 + 3p^2) \tag{A19}$$

To prove the second inequality, let $|\phi\rangle = \hat{x}|\psi\rangle$. Then

$$\hat{x}|\psi_{xp}\rangle = |\phi_{xp}\rangle + x|\psi_{xp}\rangle \tag{A20}$$

implying

$$\left\| \hat{x}|\psi_{x_1,p_1}\rangle - \hat{x}|\psi_{x_2,p_2}\rangle \right\|$$
$$\le \left\| |\phi_{x_1,p_1}\rangle - |\phi_{x_2,p_2}\rangle \right\| + x_1 \left\| \psi_{x_1,p_1}\rangle - |\psi_{x_2,p_2}\rangle \right\| + |x_1 - x_2| \tag{A21}$$

The proof now reduces to an application of the first inequality. The last inequality is proved in the same way. $\square$

    Let the initial apparatus state be

$$\hat{\alpha} = \sum_{n=1}^{n_a} \lambda_n |\phi_n\rangle\langle\phi_n| \tag{A22}$$

for some set of positive numbers λ_n and orthonormal set $|\phi_n\rangle$. We argue on the same physical grounds adduced in the first few paragraphs of this Appendix that the quantities $\mathrm{Tr}\left((\hat{\rho} \otimes \hat{\alpha})\hat{\epsilon}_x^2\right)$, $\mathrm{Tr}\left(\hat{\rho} \otimes \hat{\alpha})\hat{\delta}_p^2\right)$ are well defined for all $\hat{\rho} \in \mathcal{P}$. Finally, for given positive real numbers l_X, l_P, define $\mathcal{C}_{l_X,l_P}$ to be the phase-space box consisting of all x, p, such that $-\frac{l_X}{2} \le x \le \frac{l_X}{2}$, $-\frac{l_P}{2} \le x \le \frac{l_P}{2}$.

Lemma 3. *Let* $\hat{\rho}$ *be any element of* $\mathcal{P}$, *and let* $\hat{\rho}_{xp} = \hat{D}_{xp}\hat{\rho}\hat{D}_{xp}^\dagger$. *Let* $\hat{A}$ *be any self-adjoint operator, such that* $\mathrm{Tr}(\hat{\rho}_{xp} \otimes \hat{\alpha})[\hat{p}, \hat{A}])$, $\mathrm{Tr}(\hat{\rho}_{xp} \otimes \hat{\alpha})[\hat{x}, \hat{A}])$ *are defined, and* $\mathrm{Tr}\left((\hat{\rho}_{xp} \otimes \hat{\alpha})\hat{A}^2\right)$, $\mathrm{Tr}\left((\hat{\rho}_{xp} \otimes \hat{\alpha})\hat{x}\hat{A}^2\hat{x}\right)$, $\mathrm{Tr}\left((\hat{\rho}_{xp} \otimes \hat{\alpha})\hat{p}\hat{A}^2\hat{p}\right)$ *are both defined and bounded on* $\mathcal{C}_{l_X,l_P}$. *Then,* $\mathrm{Tr}\left((\hat{\rho}_{xp} \otimes \hat{\alpha})\hat{A}\right)$ *is differentiable on* $\mathcal{C}_{l_X,l_P}$, *and*

$$\frac{\partial}{\partial x}\left(\mathrm{Tr}(\hat{\rho}_{xp} \otimes \hat{\alpha})\hat{A} \right) = \frac{i}{\hbar}\mathrm{Tr}(\hat{\rho}_{xp} \otimes \hat{\alpha})[\hat{p}, \hat{A}]) \tag{A23}$$

$$\frac{\partial}{\partial p}\left(\mathrm{Tr}(\hat{\rho}_{xp} \otimes \hat{\alpha})\hat{A} \right) = -\frac{i}{\hbar}\mathrm{Tr}(\hat{\rho}_{xp} \otimes \hat{\alpha})[\hat{x}, \hat{A}]) \tag{A24}$$

Moreover, the derivatives are uniformly continuous on $\mathcal{C}_{l_X,l_P}$.

Proof. We have

$$\hat{\rho} = \sum_{n=1}^{n_s} \xi_n |\psi_n\rangle\langle\psi_n| \tag{A25}$$

for some set of positive numbers ξ_n and orthonormal vectors $|\psi_n\rangle \in \mathcal{P}_0$. Define

$$f_{n,m}(x,p) = \left\| \hat{A} |\psi_{n,x,p} \otimes \phi_m\rangle \right\| \tag{A26}$$

$$g_{n,m}(x,p) = \left| \hat{A}\hat{x} |\psi_{n,x,p} \otimes \phi_m\rangle \right\| \tag{A27}$$

$$h_{n,m}(x,p) = \left\| \hat{A}\hat{p} |\psi_{n,x,p} \otimes \phi_m\rangle \right\| \tag{A28}$$

$$C_1 = \sup_{x,p \in \mathcal{C}_{l_X,l_P}} \left(\mathrm{Tr}\big((\hat{\rho}_{xp} \otimes \hat{\alpha}) \hat{A}^2 \big) \right) \tag{A29}$$

$$C_2 = \sup_{x,p \in \mathcal{C}_{l_X,l_P}} \left(\mathrm{Tr}\big((\hat{\rho}_{xp} \otimes \hat{\alpha}) \hat{x} \hat{A}^2 \hat{x} \big) \right) \tag{A30}$$

$$C_3 = \sup_{x,p \in \mathcal{C}_{l_X,l_P}} \left(\mathrm{Tr}\big((\hat{\rho}_{xp} \otimes \hat{\alpha}) \hat{p} \hat{A}^2 \hat{p} \big) \right) \tag{A31}$$

$$C_4 = \sup_n \big(N_{0,\frac{3}{2}}(|\psi_n\rangle) \big), \qquad\qquad \tilde{C}_4 = \sup_n \big(\tilde{N}_{0,\frac{3}{2}}(|\psi_n\rangle) \big) \tag{A32}$$

$$C_5 = \sup_n \big(N_{0,\frac{3}{2}}(\hat{x}|\psi_n\rangle) \big), \qquad\qquad \tilde{C}_5 = \sup_n \big(\tilde{N}_{0,\frac{3}{2}}(\hat{x}|\psi_n\rangle) \big) \tag{A33}$$

$$C_6 = \sup_n \big(N_{0,\frac{3}{2}}(\hat{p}|\psi_n\rangle) \big), \qquad\qquad \tilde{C}_6 = \sup_n \big(\tilde{N}_{0,\frac{3}{2}}(\hat{p}|\psi_n\rangle) \big) \tag{A34}$$

It follows from Lemmas 1 and 2 that

$$\left| \frac{1}{\epsilon} \big(\langle \psi_{n,x+\epsilon,p} \otimes \phi_m | \hat{A} | \psi_{n,x+\epsilon,p} \otimes \phi_m \rangle - \langle \psi_{n,x,p} \otimes \phi_m | \hat{A} | \psi_{n,x,p} \otimes \phi_m \rangle \big) \right.$$
$$\left. - \frac{i}{\hbar} \langle \psi_{n,x,p} \otimes \phi_m | [\hat{p}, \hat{A}] | \psi_{n,x,p} \otimes \phi_m \rangle \right|$$
$$\leq 2\epsilon \tilde{C}_4 B(4+p^2) f_{n,m}(x,p) + \frac{\epsilon}{\hbar} B_1 \tilde{C}_4 \left(h_{n,m}(x,p) + \frac{|p|}{2} f_{n,m}(x,p) \right) \tag{A35}$$

It follows from the Cauchy–Schwartz inequality that

$$\sum_{n,m} \xi_n \lambda_m f_{n,m}(x,p) \leq \sqrt{C_1}, \qquad\qquad \sum_{n,m} \xi_n \lambda_m h_{n,m}(x,p) \leq \sqrt{C_3} \tag{A36}$$

Consequently

$$\left| \frac{1}{\epsilon} \Big(\mathrm{Tr}\big((\hat{\rho}_{x+\epsilon,p} \otimes \hat{\alpha}) \hat{A} \big) - \mathrm{Tr}\big((\hat{\rho}_{x,p} \otimes \hat{\alpha}) \hat{A} \big) \Big) - \frac{i}{\hbar} \mathrm{Tr}\big((\hat{\rho}_{x,p} \otimes \hat{\alpha})[\hat{p}, \hat{A}] \big) \right|$$
$$\leq 2\epsilon \sqrt{C_1} \tilde{C}_4 B(4+p^2) + \frac{\epsilon}{\hbar} B_1 \tilde{C}_4 \left(\sqrt{C_3} + \frac{|p|}{2} \sqrt{C_1} \right)$$
$$\to 0 \tag{A37}$$

as $\epsilon \to 0$, which establishes Equation (A23). Equation (A24) is proved similarly.
Lemmas 1 and 2 also imply

$$\left| \langle \psi_{n,x+\epsilon_1,p+\epsilon_2} \otimes \phi | [\hat{p}, \hat{A}] | \psi_{n,x+\epsilon_1,p+\epsilon_2} \rangle - \langle \psi_{n,x,p} \otimes \phi | [\hat{p}, \hat{A}] | \psi_{n,x,p} \rangle \right|$$
$$\leq 2 h_{n,m}(x,p) B_1 \big(\epsilon_1 \tilde{C}_4 + \epsilon_2 C_4 \big)$$
$$+ 2 f_{n,m}(x,p) B_3 \big(\epsilon_1 (\tilde{C}_4 + \tilde{C}_6) + \epsilon_2 (1 + C_4 + C_6) \big) \tag{A38}$$

The uniform continuity of the x derivative now follows by another application of the Cauchy–Schwartz inequality. The uniform continuity of the p derivative is proved similarly. $\square$

We are now ready to prove our main result.

Theorem 4. *Let $\mathcal{R}$ be a subset of $\mathcal{P}$ containing at least one state ρ, such that*

$$\hat{\rho}_{xp}, \quad \frac{1}{\mathrm{Tr}(\hat{\rho}_{xp}\hat{x}^2)}\hat{x}\hat{\rho}_{xp}\hat{x}, \quad \frac{1}{\mathrm{Tr}(\hat{\rho}_{xp}\hat{p}^2)}\hat{p}\hat{\rho}_{xp}\hat{p} \quad \in \mathcal{R} \tag{A39}$$

for all $(x, p) \in \mathcal{C}_{l_X, l_P}$. Then

$$\Delta_{\mathrm{e}}^{\mathcal{R}}x\Delta_{\mathrm{d}}^{\mathcal{R}}p + \frac{\hbar}{l_X}\Delta_{\mathrm{e}}^{\mathcal{R}}x + \frac{\hbar}{l_P}\Delta_{\mathrm{d}}^{\mathcal{R}}p \geq \frac{\hbar}{2} \tag{A40}$$

for every measurement of position, and

$$\Delta_{\mathrm{e}}^{\mathcal{R}}x\Delta_{\mathrm{e}}^{\mathcal{R}}p + \frac{\hbar}{l_X}\Delta_{\mathrm{e}}^{\mathcal{R}}x + \frac{\hbar}{l_P}\Delta_{\mathrm{e}}^{\mathcal{R}}p \geq \frac{\hbar}{2} \tag{A41}$$

for every joint measurement of position and momentum.

Proof. To prove the first relation, observe that it is automatic if either of the quantities $\Delta_{\mathrm{e}}^{\mathcal{R}}x$, $\Delta_{\mathrm{d}}^{\mathcal{R}}p$ is infinite. Suppose, on the other hand, they are both finite. It is easily seen that

$$[\hat{\epsilon}_{\mathrm{x}}, \hat{\delta}_{\mathrm{p}}] = -i\hbar - [\hat{x}_{\mathrm{i}}, \hat{\delta}_{\mathrm{p}}] + [\hat{p}_{\mathrm{i}}, \hat{\epsilon}_{\mathrm{x}}]. \tag{A42}$$

Let $\hat{\rho}$ be any state in $\mathcal{R}$ satisfying condition (A39). It is easily seen that $\mathrm{Tr}(\hat{\rho}_{xp}\hat{x}^2)$, $\mathrm{Tr}(\hat{\rho}_{xp}\hat{p}^2)$ are bounded. Therefore, we can apply Lemma 3 with $\hat{A} = \hat{\epsilon}_{\mathrm{x}}, \hat{\delta}_{\mathrm{p}}$ to deduce

$$\mathrm{Tr}\big((\hat{\rho}_{xp} \otimes \hat{\alpha})[\hat{\epsilon}_{\mathrm{x}}, \hat{\delta}_{\mathrm{p}}]\big) = -i\hbar(1 + \boldsymbol{\nabla} \cdot \mathbf{v}) \tag{A43}$$

where $\boldsymbol{\nabla} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial p} \end{pmatrix}$ and

$$\mathbf{v} = \begin{pmatrix} \mathrm{Tr}\big((\hat{\rho}_{xp} \otimes \hat{\alpha})\hat{\epsilon}_{\mathrm{x}}\big) \\ \mathrm{Tr}\big((\hat{\rho}_{xp} \otimes \hat{\alpha})\hat{\delta}_{\mathrm{p}}\big) \end{pmatrix} \tag{A44}$$

Since $\boldsymbol{\nabla} \cdot \mathbf{v}$ is continuous, it is integrable. Hence

$$\begin{aligned}
\Delta_{\mathrm{e}}^{\mathcal{R}}x\Delta_{\mathrm{d}}^{\mathcal{R}}p &\geq \frac{\hbar}{2}\sup_{x,p\in\mathcal{C}_{l_X,l_P}}(1 + \boldsymbol{\nabla} \cdot \mathbf{v}) \\
&\geq \frac{\hbar}{2}\left(1 - \frac{1}{l_X l_P}\left|\int_{\mathcal{C}_{l_X,l_P}}\boldsymbol{\nabla} \cdot \mathbf{v}\, dx dp\right|\right) \\
&= \frac{\hbar}{2}\left(1 - \frac{1}{l_X l_P}\left|\int_{\mathcal{B}_{l_X,l_P}}\mathbf{n} \cdot \mathbf{v}\, ds\right|\right) \\
&\geq \hbar\left(\frac{1}{2} - \frac{1}{l_X}\Delta_{\mathrm{e}}^{\mathcal{R}}x - \frac{1}{l_P}\Delta_{\mathrm{d}}^{\mathcal{R}}p\right)
\end{aligned} \tag{A45}$$

where $\mathcal{B}_{l_X,l_P}$ is the boundary of $\mathcal{C}_{l_X,l_P}$ and $\mathbf{n}$ is the outward-pointing normal. The second inequality is proved in the same way, starting from the commutation relation

$$[\hat{\epsilon}_{\mathrm{x}}, \hat{\epsilon}_{\mathrm{p}}] = -i\hbar - [\hat{x}_{\mathrm{i}}, \hat{\epsilon}_{\mathrm{p}}] + [\hat{p}_{\mathrm{i}}, \hat{\epsilon}_{\mathrm{x}}] \tag{A46}$$

$\square$

Inequalities (13), (14) are proved by specializing to the case $\mathcal{R} = \mathcal{P}$ and taking the limit as $l_X, l_P \to \infty$.

References and Notes

1. Heisenberg, W. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik* **1927**, *43*, 172–198. (In German)
2. Busch, P.; Lahti, P.; Werner, R.F. Measurement uncertainty relations. *J. Math. Phys.* **2014**, *55*, 042111.
3. Bell, J.S. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.* **1966**, *38*, 447–452.
4. Kochen, S.; Specker, E. The problem of hidden variables in quantum mechanics. *J. Math. Mech.* **1967**, *17*, 59–87.
5. Clifton, R.; Kent, A. Simulating quantum mechanics by non-contextual hidden variables. *Proc. Roy. Soc. Lond. A* **2000**, *456*, 2101–2114.
6. Appleby, D.M. The Bell-Kochen-Specker Theorem. *Stud. Hist. Philos. Mod. Phys.* **2005**, *36*, 1–28, doi:10.1016/j.shpsb.2004.05.003.
7. Bell, J.S. On the impossible pilot wave. *Found. Phys.* **1982**, *12*, 989–999.
8. Heisenberg, W. *The Physical Principles of the Quantum Theory*; University of Chicago Press: Chicago, IL, USA, 1930.
9. Von Neumann, J. *Mathematische Grundlagen der Quantenmechanik*; Springer: Berlin, Germany, 1932; English Translation Published as *Mathematical Foundations of Quantum Mechanics*; Beyer, R.T., Translator; Princeton University Press: Princeton, NJ, USA, 1955.
10. Kennard, E. Zur quantenmechanik einfacher bewegungstypen. *Zeitschrift für Physik* **1927**, *44*, 326–352. (In German)
11. Weyl, H. *Gruppentheorie und Quantenmechanik*; Hirzel, 1928; English Translation of the Revised Second Edition Published as *The Theory of Groups and Quantum Mechanics*; Robertson, H.P., Translator; Dover Publications: New York, NY, USA, 1950.
12. Robertson, H.P. The uncertainty principle. *Phys. Rev.* **1929**, *34*, 163–164.
13. Schrödinger, E. Um Heisenbergschen Unschärfeprinzip. *Sitzungsberichte der Preussischen Akademie der Wissenschaften Physikalisch-Mathematische Klasse* **1930**, *14*, 296–303. (In German)
14. Busch, P.; Heinonen, T.; Lahti, P. Heisenberg's uncertainty principle. *Phys. Rep.* **2007**, *452*, 155–176.
15. Arthurs, E.; Kelly, J.L. On the simultaneous measurement of a pair of conjugate observables. *Bell Syst. Tech. J.* **1965**, *44*, 725–729.
16. Bell, J.S. Speakable and unspeakable in quantum mechanics. In *Speakable and Unspeakable in Quantum Mechanics*; Cambridge University Press: Cambridge, UK, 1987; pp. 169–172.
17. Howard, D. Who invented the "Copenhagen interpretation"? A study in mythology. *Philos. Sci.* **2004**, *71*, 669–682.
18. Camilleri, K. *Heisenberg and the Interpretation of Quantum Mechanics*; Cambridge University Press: Cambridge, UK, 2009.
19. Faye, J. Copenhagen interpretation of quantum mechanics. In *The Stanford Encyclopedia of Philosophy*; Zalta, E.N., Ed.; Stanford University: Stanford, CA, USA, 2014.
20. Braginsky, V.B.; Vorontsov, Y.I.; Thorne, K.S. Quantum nondemolition measurements. *Science* **1980**, *209*, 547–557.
21. Braginsky, V.B.; Khalili, F.Y. *Quantum Measurement*; Cambridge University Press: Cambridge, UK, 1992.
22. Plotnitsky, A. *Niels Bohr and Complementarity: An Introduction*; Springer Briefs in Physics; Springer: Berlin, Germany, 2013.
23. Busch, P.; Lahti, P.; Werner, R.F. Proof of Heisenberg's error-disturbance relation. *Phys. Rev. Lett.* **2013**, *111*, 160405.
24. Busch, P.; Lahti, P.; Werner, R.F. Colloquium: Quantum root-mean-square error and measurement uncertainty relations. *Rev. Mod. Phys.* **2014**, *86*, 1261–1281.
25. Busch, P.; Lahti, P.; Werner, R.F. Measurement uncertainty: Reply to critics. **2014**, arXiv:1402.3102. Available online: http://arxiv.org/abs/1402.3102 (accessed on 29 April 2016).
26. Busch, P.; Lahti, P.; Werner, R.F. Comment on "Experimental Test of Error-Disturbance Uncertainty Relations by Weak Measurement". **2014**, arXiv:1403.0367. Available online: http://arxiv.org/abs/1403.0367 (accessed on 29 April 2016).
27. Appleby, D.M. Error principle. *Int. J. Theor. Phys.* **1998**, *37*, 2557–2572.

28.	Ozawa, M. Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement. *Phys. Rev. A* **2003**, *67*, 042105.

29.	Ozawa, M. Physical content of Heisenberg's uncertainty relation: Limitation and reformulation. *Phys. Lett. A* **2003**, *318*, 21–29.

30.	Ozawa, M. Disproving Heisenberg's error-disturbance relation. **2013**, arXiv:1308.3540. Available online: http://arxiv.org/abs/1308.3540 (accessed on 29 April 2016).

31.	Ozawa, M. Heisenberg's uncertainty relation: Violation and reformulation. In Proceedings of the Second International Symposium on Emergent Quantum Mechanics, Vienna, Austria, 3–6 October 2013; Volume 504, p. 012024.

32.	Appleby, D.M. Concept of experimental accuracy and simultaneous measurements of position and momentum. *Int. J. Theor. Phys.* **1998**, *37*, 1491–1509.

33.	Appleby, D.M. Maximal accuracy and minimal disturbance in the Arthurs-Kelly simultaneous measurement process. *J. Phys. A* **1998**, *31*, 6419–6436.

34.	Appleby, D.M. Optimal joint measurements of position and momentum. *Int. J. Theor. Phys.* **1999**, *38*, 807–825.

35.	Arthurs, E.; Goodman, M.S. Quantum correlations: A generalized Heisenberg uncertainty relation. *Phys. Rev. Lett.* **1988**, *60*, 2447–2449.

36.	Yuen, H.P. Generalized quantum measurements and approximate simultaneous measurements of noncommuting observables. *Phys. Lett. A* **1982**, *91*, 101–104.

37.	Ishikawa, S. Uncertainty relations in simultaneous measurements for arbitrary observables. *Rep. Math. Phys.* **1991**, *29*, 257–273.

38.	Ozawa, M. Quantum limits of measurements and uncertainty principle. In Proceedings of the Quantum Aspects of Optical Communications: Proceedings of a Workshop, Paris, France, 26–28 November 1990; Volume 378.

39.	Ozawa, M. Position measuring interactions and the Heisenberg uncertainty principle. *Phys. Lett. A* **2002**, *299*, 1–7, doi:10.1016/S0375-9601(02)00659-X.

40.	Ozawa, M. Uncertainty principle for quantum instruments and computing. *Int. J. Quantum Inf.* **2003**, *1*, 569–588.

41.	Ozawa, M. Uncertainty relations for joint measurements of noncommuting observables. *Phys. Lett. A* **2004**, *320*, 367–374.

42.	Hall, M.J.W. Prior information: How to circumvent the standard joint-measurement uncertainty relation. *Phys. Rev. A* **2004**, *69*, 052113.

43.	Weston, M.M.; Hall, M.J.W.; Palsson, M.S.; Wiseman, H.M.; Pryde, G.J. Experimental test of universal complementarity relations. *Phys. Rev. Lett.* **2013**, *110*, 220402.

44.	Branciard, C. Deriving tight error-trade-off relations for approximate joint measurements of incompatible quantum observables. *Phys. Rev. A* **2014**, *89*, 022124.

45.	Ozawa, M. Error-disturbance relations in mixed states. **2014**, arXiv:1404.3388. Available online: http://arxiv.org/abs/1404.3388 (accessed on 29 April 2016).

46.	Rozema, L.A.; Mahler, D.H.; Hayat, A.; Steinberg, A.M. A note on different definitions of momentum disturbance. *Quantum Stud. Math. Found.* **2015**, *2*, 17–22.

47.	Fuchs, C.A.; Schack, R. Quantum-Bayesian Coherence. *Rev. Mod. Phys.* **2013**, *85*, 1693–1715.

48.	Ozawa, M. Uncertainty relations for noise and disturbance in generalized quantum measurements. *Ann. Phys.* **2004**, *311*, 350–416.

49.	Busch, P.; Heinonen, T.; Lahti, P. Noise and disturbance in quantum measurement. *Phys. Lett. A* **2004**, *320*, 261–270.

50.	Born, M. Zur Quantenmechanik der Stoßorgänge. *Zeitschrift für Physik* **1926**, *37*, 863–867. (In German). Published in English translation as "On the quantum mechanics of collisions" in Wheeler and Zurek [51].

51.	Wheeler, J.A.; Zurek, W.H. (Eds.) *Quantum Theory and Measurement*; Princeton University Press: Princeton, NJ, USA, 1983.

52.	Werner, R.F. The uncertainty relation for joint measurement of postion and momentum. *Quantum Inf. Comput.* **2004**, *4*, 546–562.

53.	Appleby, D.M. Optimal measurements of spin direction. *Int. J. Theor. Phys.* **2000**, *39*, 2231–2252.

54. Bohm, D.; Hiley, B. *The Undivided Universe: An Ontological Interpretation of Quantum Theory*; Routledge: Abingdon, UK, 1993.

55. Holland, P.R. *The Quantum Theory of Motion*; Cambridge University Press: Cambridge, UK, 1993.

56. Appleby, D.M. Generic Bohmian trajectories of an isolated particle. *Found. Phys.* **1999**, *29*, 1863–1883.

57. Appleby, D.M. Bohmian trajectories post-decoherence. *Found. Phys.* **1999**, *29*, 1885–1916.

58. Çinlar, E. *Probability and Stochastics*; Graduate Texts in Mathematics No. 261; Springer: Berlin, Germany, 2011.

59. Korzekwa, K.; Jennings, D.; Rudolph, T. Operational constraints on state-dependent formulations of quantum error-disturbance trade-off relations. *Phys. Rev. A* **2014**, *89*, 052108.

60. Busch, P. Quantum rms error and Heisenberg's error-disturbance relation. In Proceedings of the Wigner 111—Colourful and Deep Scientific Symposium, Budapest, Hungary, 11–13 November 2013.

61. Poulin, D. Université de Sherbrooke, Québec, Canada. Personal Communication, 2014.

62. Treves, F. *Topological Vector Spaces, Distributions and Kernels*; Academic Press: Cambridge, MA, USA, 1967.

63. Gel'fand, I.M.; Vilenkin, N.Y. *Generalized Functions: Applications of Harmonic Analysis*; Academic Press: Cambridge, MA, USA, 1964; Volume 4.

64. Böhm, A. *The Rigged Hilbert Space and Quantum Mechanics*; Lecture Notes in Physics, No. 78; Springer: Berlin, Germany, 1978.

65. Böhm, A.; Gadella, M. *Dirac Kets, Gamow Vectors and Gel'fand Triplets*; Lecture Notes in Physics, No. 348; Springer: Berlin, Germany, 1989.

66. Perelomov, A.M. On the completeness of a system of coherent states. *Theor. Math. Phys.* **1971**, *6*, 156–164.

67. Bargmann, V.; Butera, P.; Girardello, L.; Klauder, J.R. On the completeness of the coherent states. *Rep. Math. Phys.* **1971**, *2*, 221–228.