Anomalous Weak Values Without Post-Selection

Alastair A. Abbott, Ralph Silva, Julian Wechs, Nicolas Brunner, Cyril Branciard

Anomalous Weak Values Without Post-Selection

Alastair A. Abbott[1], Ralph Silva[2,3], Julian Wechs[1], Nicolas Brunner[2], and Cyril Branciard[1]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, Institut Néel, 38000 Grenoble, France

[2]Département de Physique Appliquée, Université de Genève, 1211 Genève, Switzerland

[3]Institute for Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland

September 2, 2019

A weak measurement performed on a pre- and post-selected quantum system can result in an average value that lies outside of the observable's spectrum. This effect, usually referred to as an "anomalous weak value", is generally believed to be possible only when a non-trivial post-selection is performed, i.e., when only a particular subset of the data is considered. Here we show, however, that this is not the case in general: in scenarios in which several weak measurements are sequentially performed, an anomalous weak value can be obtained without post-selection, i.e., without discarding any data. We discuss several questions that this raises about the subtle relation between weak values and pointer positions for sequential weak measurements. Finally, we consider some implications of our results for the problem of distinguishing different causal structures.

1 Introduction

All quantum measurements are subjected to a fundamental trade-off between information gain and disturbance of the measured system. In particular, one can perform weak measurements that provide little information but only weakly perturb the system. A particularly interesting situation arises when weak measurements are combined with post-selection [1]. This can be conveniently described within the von Neumann model of quantum measurements, where the quantum system to be measured is coupled via a joint unitary operation to another quantum system, the "pointer", which represents the measurement device. The measurement is then completed by performing a strong measurement of the pointer.

More formally, consider a system initially prepared (or pre-selected) in a pure state $|\psi\rangle$, and an observable $\hat{A}$ to be weakly measured on it. The system-pointer interaction is generated via a Hamiltonian of the form $\hat{H} = \gamma \hat{A} \otimes \hat{p}$, where $\hat{p}$ denotes the momentum operator acting on the pointer. The latter is initially in a state $|\varphi(0)\rangle$, which we shall take here to be a Gaussian wave packet centred at a position $x = 0$ with spread (i.e., standard deviation) σ. Assuming that we are in the weak measurement regime, with the coupling constant γ and interaction time Δt such that $g := \gamma \Delta t$ is small enough compared to the spread of the pointer, the global state after the coupling is given by

$$e^{-i\hat{H}\Delta t} |\psi\rangle |\varphi(0)\rangle \approx (\mathbb{1} - ig\hat{A}\hat{p}) |\psi\rangle |\varphi(0)\rangle \qquad (1)$$

(where tensor products are implicit, and taking $\hbar = 1$). For simplicity we will henceforth choose units so that $g = 1$; the strength of the measurement will then be controlled solely by the pointer spread σ, and the validity of the weak regime will depend only on this being sufficiently large.[1] Next, the system is post-selected onto the state $|\phi\rangle$ (e.g. via a strong projective measurement). The final state of the pointer is then (up to normalisation)

$$\begin{aligned} \langle\phi| (\mathbb{1} - i\hat{A}\hat{p}) |\psi\rangle |\varphi(0)\rangle &= \langle\phi|\psi\rangle (\mathbb{1} - iA^{\phi}_{\psi}\hat{p}) |\varphi(0)\rangle \\ &\approx \langle\phi|\psi\rangle \, e^{-iA^{\phi}_{\psi}\hat{p}} |\varphi(0)\rangle , \end{aligned} \qquad (2)$$

where

$$A^{\phi}_{\psi} := \frac{\langle\phi|\hat{A}|\psi\rangle}{\langle\phi|\psi\rangle} \qquad (3)$$

is the so-called *weak value* of the observable $\hat{A}$ given the pre-selection in the state $|\psi\rangle$ and post-selection in the state $|\phi\rangle$ [1]. The mean position of the pointer is thus displaced—via the displacement operator $e^{-iA^{\phi}_{\psi}\hat{p}}$, which generates the (possibly unnormalised) state $|\varphi(A^{\phi}_{\psi})\rangle = e^{-iA^{\phi}_{\psi}\hat{p}} |\varphi(0)\rangle$; see Appendix—to

$$\langle\hat{x}\rangle \approx \frac{\langle\varphi(A^{\phi}_{\psi})| \, \hat{x} \, |\varphi(A^{\phi}_{\psi})\rangle}{\langle\varphi(A^{\phi}_{\psi})|\varphi(A^{\phi}_{\psi})\rangle} = \mathrm{Re}(A^{\phi}_{\psi}). \qquad (4)$$

Notably, the real part of the weak value can become very large when the pre- and post-selected

A.A.A. and R.S. contributed equally to this work.

[1] A characteristic of the weak regime is that an individual measurement of the pointer position following the interaction with the system yields little information, since the pointer spread is much larger than the range of mean pointer positions one obtains (which scales with g). As only the ratio between these quantities is important, it is thus sufficient for our analysis to take $g = 1$. In the Appendix we present more precise statements of the conditions for the weak regime to be satisfied in the situations we consider throughout this paper.

states are almost orthogonal, i.e. $|\langle\phi|\psi\rangle| \ll 1$. In this case, the pointer is, on average, shifted by a large amount. Whenever $\mathrm{Re}(A^{\phi}_{\psi})$ is not in the interval $[\lambda_{\min}(\hat{A}), \lambda_{\max}(\hat{A})]$ (where $\lambda_{\min(\max)}(\hat{A}) = \min(\max)_k \lambda_k(\hat{A})$ and λ_k denotes the k^{th} eigenvalue of an observable), i.e. whenever it is outside of the (convex hull of the) spectrum of $\hat{A}$, the pointer's mean position thus moves beyond where it could have reached under simple weak measurements on an arbitrary preselected state without any post-selection. Indeed, in the absence of post-selection one has (now with exact equalities)

$$\begin{aligned}\langle\hat{x}\rangle &= \langle\psi|\langle\varphi(0)|\, e^{i\hat{A}\hat{p}}\,(\mathbb{1}\otimes\hat{x})\, e^{-i\hat{A}\hat{p}}\,|\psi\rangle\,|\varphi(0)\rangle \\ &= \langle\psi|\langle\varphi(0)|\left(\mathbb{1}\otimes\hat{x} + \hat{A}\otimes\mathbb{1}\right)|\psi\rangle\,|\varphi(0)\rangle \\ &= \langle\psi|\,\hat{A}\,|\psi\rangle \quad \in \ [\lambda_{\min}(\hat{A}), \lambda_{\max}(\hat{A})]. \end{aligned} \tag{5}$$

Note that $\langle\hat{x}\rangle$, both with and without post-selection, can be determined experimentally by performing sufficiently many measurements, despite the large variance of the pointer (indeed, to obtain a given accuracy the number of measurements required scales proportionally to σ^2).

The definition (3) of a weak value can be generalised to post-selections on a given result for any general quantum measurement [2, 3]. In particular, a trivial, deterministic measurement of the identity operator $\mathbb{1}$ amounts to performing no post-selection. This allows one to also consider a *weak value with no post-selection*, defined (see Appendix) as[2]

$$A^{\mathbb{1}}_{\psi} \coloneqq \langle\psi|\hat{A}|\psi\rangle\,. \tag{6}$$

With this definition, Eq. (5) gives $\langle\hat{x}\rangle = A^{\mathbb{1}}_{\psi} = \mathrm{Re}(A^{\mathbb{1}}_{\psi})$: we recover the same relation as in Eq. (4), although now $A^{\mathbb{1}}_{\psi}$ is restricted to lie in $[\lambda_{\min}(\hat{A}), \lambda_{\max}(\hat{A})]$ since here it is simply equal to the expectation value of $\hat{A}$.

The phenomenon of a weak value outside the spectrum of $\hat{A}$ is referred to as an "anomalous weak value" [1, 5, 6] since it conflicts with our classical intuition, which would lead us to expect $\langle\hat{x}\rangle$ to lie within the range of the spectrum of $\hat{A}$. This has been observed in many experiments [7–9], and appears to be directly linked to various (*a priori* unrelated) areas such as tunnelling times [10] and fast light propagation [11, 12]. In practice, anomalous weak values allow for the detection and precise estimation of very small physical effects [13–16], via a form of signal amplification, while they have also helped provide new insights on foundational aspects of quantum mechanics such as contextuality [17], counterfactual paradoxes [18], and the nature of the wave function [19, 20]. While astonishing at first sight, anomalous weak values can in fact be intuitively understood in terms of destructive interference of the pointer state, which occurs as a result of post-selection. With this in mind and given the rudimentary analysis above, it is rather natural to attribute the origin of anomalous weak values to the presence of post-selection; this opinion indeed seems to be widely shared in the community.

Here we show, however, that this is not the case in general, and that anomalous weak values can in fact be observed in the absence of post-selection and without discarding any outcomes. Specifically, we consider a situation in which two successive weak measurements are performed on a quantum system. The experiment thus involves two pointers, one associated to each weak measurement, which are measured jointly. Considering observables that are simply given by projectors, one expects to find the mean position of each pointer between 0 (the system's state being orthogonal to the projector) and 1 (the system's state being in the range of the projector). Yet, we will see that the average of the product of the pointer positions can become negative, something which cannot happen if the measurements are strong or classical. This may be understood in terms of the second measurement acting as an effective post-selection of the system, thus creating the desired interference. Importantly however, no data is discarded.

Below, after discussing in detail a simple example of this effect, we turn to the more general scenario of arbitrary sequences of weak measurements without post-selection, deriving bounds on how anomalous a weak value can be obtained in such a scenario and how these relate to the (jointly measured) pointer positions. We finish by discussing the use of anomalous weak values without post-selection to certify particular causal structures between measurements.

2 Illustrative example

To start with, let us consider a qubit system initially prepared in the state $|0\rangle$, undergoing a sequence of two weak von Neumann measurements of the projection observables $|\psi_j\rangle\langle\psi_j|$ $(j = 1, 2)$, where the states $|\psi_j\rangle$ and their orthogonal states $|\psi_j^{\perp}\rangle$ are defined as

$$\begin{aligned}|\psi_j\rangle &= \frac{1}{2}\,|0\rangle - (-1)^j\,\frac{\sqrt{3}}{2}\,|1\rangle\,, \\ |\psi_j^{\perp}\rangle &= \frac{\sqrt{3}}{2}\,|0\rangle + (-1)^j\,\frac{1}{2}\,|1\rangle\,. \end{aligned} \tag{7}$$

To each measurement is associated a pointer in the state $|\varphi_j(x_j)\rangle$, where x_j is the mean position of the pointer wavefunction. The two pointers are initially independent, and both centred at $x_j = 0$. The initial

[2] Note that the weak values $A^{\mathbb{1}}_{\psi}$ (in the absence of post-selection) and A^{ψ}_{ψ} (when one post-selects on the initial state, a situation called "re-selection" and studied in Ref. [4]) coincide. However, we emphasise that these correspond to different physical situations; in particular, without post-selection no data is discarded.

state of the system and pointers is therefore

$$|\Psi_0\rangle = |0\rangle\,|\varphi_1(0)\rangle\,|\varphi_2(0)\rangle\,. \tag{8}$$

Following the von Neumann measurement procedure described earlier with interaction Hamiltonians $\hat{H}_j = \gamma_j\,|\psi_j\rangle\langle\psi_j|\,\hat{p}_j$, the average post-measurement position of the corresponding pointer is (with appropriate units so that $\gamma_j\Delta t_j = 1$ as before) $x_j = 1$ if the state of the system is $|\psi_j\rangle$; if the state is $|\psi_j^\perp\rangle$ then the pointer does not move. The state of the system and pointers after the interaction with the first pointer is thus

$$\begin{aligned}|\Psi_1\rangle =& \left(|\psi_1\rangle\langle\psi_1|\,e^{-i\hat{p}_1} + |\psi_1^\perp\rangle\langle\psi_1^\perp|\,\mathbb{1}_1\right)\mathbb{1}_2\;|\Psi_0\rangle\\ =&\frac{1}{2}\,|\psi_1\rangle\,|\varphi_1(1)\rangle\,|\varphi_2(0)\rangle\\ &+\frac{\sqrt{3}}{2}\,|\psi_1^\perp\rangle\,|\varphi_1(0)\rangle\,|\varphi_2(0)\rangle\,.\end{aligned} \tag{9}$$

After interacting with the second pointer, it evolves to

$$\begin{aligned}|\Psi_2\rangle =& \left(|\psi_2\rangle\langle\psi_2|\,e^{-i\hat{p}_2} + |\psi_2^\perp\rangle\langle\psi_2^\perp|\,\mathbb{1}_2\right)\mathbb{1}_1\;|\Psi_1\rangle\\ =&-\frac{1}{4}\,|\psi_2\rangle\,|\varphi_1(1)\rangle\,|\varphi_2(1)\rangle\\ &+\frac{3}{4}\,|\psi_2\rangle\,|\varphi_1(0)\rangle\,|\varphi_2(1)\rangle\\ &+\frac{\sqrt{3}}{4}|\psi_2^\perp\rangle|\varphi_1(1)\rangle|\varphi_2(0)\rangle\\ &+\frac{\sqrt{3}}{4}|\psi_2^\perp\rangle|\varphi_1(0)\rangle|\varphi_2(0)\rangle\,.\end{aligned} \tag{10}$$

Tracing out the system, one finds that the joint pointer state is

$$\begin{aligned}\eta_{12} =& |\Phi_1^{(1)}\rangle\langle\Phi_1^{(1)}|\otimes|\phi_2(1)\rangle\langle\phi_2(1)|\\ &+|\Phi_1^{(0)}\rangle\langle\Phi_1^{(0)}|\otimes|\phi_2(0)\rangle\langle\phi_2(0)|\,,\end{aligned} \tag{11}$$

where

$$\begin{aligned}|\Phi_1^{(1)}\rangle &= \frac{1}{4}\big(\,|\varphi_1(1)\rangle - 3\,|\varphi_1(0)\rangle\,\big),\\ |\Phi_1^{(0)}\rangle &= \frac{\sqrt{3}}{4}\big(\,|\varphi_1(1)\rangle + |\varphi_1(0)\rangle\,\big)\end{aligned} \tag{12}$$

are (generally unnormalised) states of the first pointer. The norms of these states, and thus the weight of each state in the mixture η_{12}, depend on the strength of the first measurement through the overlap $\langle\varphi_1(1)|\varphi_1(0)\rangle$ of the corresponding pointer states.

Finally the positions of the pointers are measured. The quantity of interest is the average of the product of the pointer positions, i.e., the expectation value $\langle\hat{x}_1\otimes\hat{x}_2\rangle$. From Eq. (11), and using the facts that $\langle\varphi_2(1)|\hat{x}_2|\varphi_2(1)\rangle = 1$ and $\langle\varphi_2(0)|\hat{x}_2|\varphi_2(0)\rangle = 0$, we simply find that $\langle\hat{x}_1\otimes\hat{x}_2\rangle = \langle\Phi_1^{(1)}|\,\hat{x}_1\,|\Phi_1^{(1)}\rangle$.

Note that we have not yet specified the strength of either measurement. Considering Gaussian pointers with widths σ_j for each measurement, we find (see Appendix)

$$\langle\hat{x}_1\otimes\hat{x}_2\rangle = \frac{1}{16}\left(1 - 3e^{-\frac{1}{8\sigma_1^2}}\right). \tag{13}$$

Notice that this quantity depends on σ_1 but not on σ_2: the strength of the second measurement has no effect here and can be made as strong as one wishes.[3]

Since both observables being measured are projectors with spectra $\{0,1\}$, one would naturally expect an average value within the range $[0,1]$. Regardless of the strength of either measurement, each pointer, taken individually, indeed has an average position in $[0,1]$. Specifically (see Appendix), $\langle\hat{x}_1\rangle = 1/4$, which is independent of the strength of either measurement, while for the second pointer $\langle\hat{x}_2\rangle = \frac{1}{8}\left(5 - 3e^{-\frac{1}{8\sigma_1^2}}\right)$, which ranges from $\langle\hat{x}_2\rangle \approx 1/4$ when the first measurement is weak ($\sigma_1 \gg 1$), to $\langle\hat{x}_2\rangle \approx 5/8$ when it is strong ($\sigma_1 \ll 1$). In the latter regime, Eq. (13) gives $\langle\hat{x}_1\otimes\hat{x}_2\rangle \approx 1/16$, which is consistent with the above argument. However, if the first measurement is sufficiently weak, the average value can become negative; in the limit $\sigma_1 \to \infty$ we get

$$\langle\hat{x}_1\otimes\hat{x}_2\rangle \approx -\frac{1}{8}. \tag{14}$$

This pointer reading is anomalous in that it gives an average value outside of the natural range of $[0,1]$ that one would expect for the average of a product of two binary 0/1-valued measurements. As we will discuss in more detail below, this result can be linked to an anomalous weak value without post-selection, $(|\psi_2\rangle\langle\psi_2|\cdot|\psi_1\rangle\langle\psi_1|)_0^{\mathbb{1}} := \langle 0|\psi_2\rangle\langle\psi_2|\psi_1\rangle\langle\psi_1|0\rangle$ (see Eq. (19) below); specifically, we have here

$$\langle\hat{x}_1\otimes\hat{x}_2\rangle \approx \mathrm{Re}\left(\,\langle 0|\psi_2\rangle\langle\psi_2|\psi_1\rangle\langle\psi_1|0\rangle\,\right) = -\frac{1}{8}. \tag{15}$$

We emphasise that this anomalous value is obtained despite the absence of post-selection. This effect can nevertheless be understood intuitively by considering that the second measurement acts as an effective post-selection on $|\psi_2\rangle$, as the corresponding pointer moves only in this case. This becomes apparent upon rewriting the above weak value as

$$\langle 0|\psi_2\rangle\langle\psi_2|\psi_1\rangle\langle\psi_1|0\rangle = |\langle\psi_2|0\rangle|^2\,\frac{\langle\psi_2|\psi_1\rangle\langle\psi_1|0\rangle}{\langle\psi_2|0\rangle}, \tag{16}$$

which differs from the standard weak value $(|\psi_1\rangle\langle\psi_1|)_0^{\psi_2}$ for a post-selection on $|\psi_2\rangle$ only by the factor $|\langle\psi_2|0\rangle|^2$, which is the probability that the projection of $|0\rangle$ onto $|\psi_2\rangle$ is successful. As it turns out, this factor ensures in particular that the anomalous

[3] Note that this means that the variance of $\hat{x}_1\otimes\hat{x}_2$ can be made to be of the order of σ_1^2 (as opposed to $\sigma_1^2\sigma_2^2$; see Appendix), meaning less statistics are needed to estimate $\langle\hat{x}_1\otimes\hat{x}_2\rangle$ to a given precision than would otherwise be the case.

weak value without post-selection cannot be arbitrary large, a fact that we prove further below. For a sequence of two projection observables $\hat{A}$ and $\hat{B}$ (with eigenvalues 0 and 1), the above value of $-1/8$ for the real part is indeed the most anomalous value obtainable (see Appendix).

3 Analysis for arbitrary observables

In order to analyse more generally the phenomenon exhibited by the previous example, let us recall some facts about the sequential weak measurement framework within which it can be placed [21]. In this framework, a sequence of weak measurements is performed, each involving a coupling to a different pointer; these pointers can then be jointly measured following the sequence of interactions and, potentially, a post-selection on the system. This framework has received much interest recently from both practical [22–24] and foundational [4, 25, 26] viewpoints.

For the case of two sequential weak measurements, as in the example of the previous section, consider thus a system prepared in the pure state $|\psi\rangle$, which is subjected to a sequential weak measurement of the (generally noncommuting) observables $\hat{A}$ then $\hat{B}$, before being post-selected onto the state $|\phi\rangle$. The system-pointer interaction Hamiltonians are $\hat{H}_1 = \gamma_1 \hat{A}\hat{p}_1$ and $\hat{H}_2 = \gamma_2 \hat{B}\hat{p}_2$. We will choose again, for simplicity, the coupling constants and interaction times such that $\gamma_j \Delta t_j = 1$, and take Gaussian pointers initially in the states $|\varphi_1(0)\rangle$ and $|\varphi_2(0)\rangle$ with widths σ_1 and σ_2, which dictate the measurement strengths.

In analogy to Eq. (3), the *sequential weak value* $(BA)^{\phi}_{\psi}$ is defined,[4] for a system prepared in $|\psi\rangle$ and post-selected in $|\phi\rangle$, as [21]

$$(BA)^{\phi}_{\psi} := \frac{\langle\phi|\hat{B}\hat{A}|\psi\rangle}{\langle\phi|\psi\rangle}. \tag{17}$$

However, while the notion of an anomalous weak value for single (non-sequential) weak measurements is intimately linked to the pointer displacement (and even justified) by the relation $\langle\hat{x}\rangle = \mathrm{Re}(A^{\phi}_{\psi})$, the relationship between the mean pointer positions and $(BA)^{\phi}_{\psi}$ is more subtle for sequential weak measurements. In the presence of post-selection, it has instead been shown [21, 27] that

$$\langle\hat{x}_1 \otimes \hat{x}_2\rangle \approx \frac{1}{2}\big(\,\mathrm{Re}[(BA)^{\phi}_{\psi}] + \mathrm{Re}[A^{\phi}_{\psi}(B^{\phi}_{\psi})^{*}]\,\big) \tag{18}$$

within the weak regime (with large enough widths σ_1 and σ_2). This cautions that some care must be taken when linking (possibly anomalous) pointer positions to weak values.

The sequential weak value of Eq. (6) can be generalised to the case without post-selection, by defining, in a similar way to before, the *sequential weak value with no post-selection* as

$$(BA)^{\mathbb{1}}_{\psi} := \langle\psi|\hat{B}\hat{A}|\psi\rangle\,. \tag{19}$$

Indeed, this quantity has previously been considered in the study of time asymmetry in sequential weak measurements [25, 28] and their quasiprobabilistic interpretation [29]. Connecting this to the pointer positions, we prove in the Appendix that, contrary to Eq. (18) (which was obtained with post-selection), we recover here the direct relation[5]

$$\langle\hat{x}_1 \otimes \hat{x}_2\rangle \approx \mathrm{Re}[(BA)^{\mathbb{1}}_{\psi}], \tag{20}$$

as anticipated already in Eq. (15), which holds as long as the *first* measurement is sufficiently weak [29].

This justifies that our earlier illustrative example could indeed be interpreted as yielding an anomalous weak value without post-selection. Crucially, although for a single measurement without post-selection $A^{\mathbb{1}}_{\psi}$ is simply the expectation value of $\hat{A}$, no such interpretation can be given to $(BA)^{\mathbb{1}}_{\psi}$: it is the weak value of a sequence of measurements and simply behaves *as if* it were the weak value of the operator $\hat{B}\hat{A}$ which, as already mentioned, is only Hermitian – and thus defines an observable – if $\hat{A}$ and $\hat{B}$ commute (see also footnote 4). In particular, this implies that unless $\hat{A}$ and $\hat{B}$ commute, $(BA)^{\mathbb{1}}_{\psi}$ need not be contained within the interval $[\Lambda_{\min}(\hat{A},\hat{B}), \Lambda_{\max}(\hat{A},\hat{B})]$, where $\Lambda_{\min(\max)}(\hat{A},\hat{B}) = \min(\max)_{k,\ell}\,\lambda_k(\hat{A})\lambda_\ell(\hat{B})$, as one one would naturally expect for the product of outcomes for a measurement of $\hat{A}$ then $\hat{B}$ [30, 31].

Nevertheless, as we noted after Eq. (16), the value of $(BA)^{\mathbb{1}}_{\psi}$ cannot be amplified arbitrarily. It is possible to place a more quantitive bound on the values that it can in fact take. Using the Cauchy-Schwartz inequality, we indeed have

$$\begin{aligned} |(BA)^{\mathbb{1}}_{\psi}| &= |\,\langle\psi|\hat{B}\hat{A}|\psi\rangle\,| \\ &\leq \sqrt{\langle\psi|\hat{A}^2|\psi\rangle\,\langle\psi|\hat{B}^2|\psi\rangle} \leq \|\hat{A}\|\,\|\hat{B}\|, \end{aligned} \tag{21}$$

(where $\|\cdot\|$ is the spectral norm). Thus, although the mean pointer position can show anomalous weak values without post-selection, the *magnitude* of the mean

[4] While this terminology is standard (see, e.g., Refs. [21, 27]), note that this should be read as the weak value for measuring $\hat{A}$ then $\hat{B}$ and *not* the weak value of $\hat{B}\hat{A}$, which indeed is not a valid observable in general since it may not be Hermitian. As Eq. (17) shows, this quantity nonetheless behaves *as if* it were the weak value of $\hat{B}\hat{A}$.

[5] As in the single measurement case, the weak values $(BA)^{1}_{\psi}$ (in the absence of post-selection) and $(BA)^{\psi}_{\psi}$ (in the case of "re-selection" [4]) coincide; cf. Footnote 2. Recall, however, that these correspond to different physical situations; crucially here, the mean pointer positions in Eqs. (20) and (18) differ in general between these scenarios. See Appendix for further discussion.

pointer position cannot be pushed outside what one can obtain using strong measurements.

The bound above implies in particular that for observables with symmetric spectra (with respect to 0), the real part of the weak value – and therefore the mean product of pointer positions, see Eq. (20) – cannot be anomalous; anomalous pointer positions are only obtained for observables with asymmetric spectra, such as projection observables. Nevertheless, one can also obtain complex weak values for observables with symmetric spectra. Take, for example, a system initially prepared in the $(+1)$-eigenstate $|0\rangle$ of the Pauli matrix $\hat{\sigma}_{\mathrm{Z}}$, on which a sequential weak measurement of the Pauli observables $\hat{\sigma}_{\mathrm{Y}}$ and $\hat{\sigma}_{\mathrm{X}}$ is performed. One thus obtains $(\sigma_{\mathrm{X}}\sigma_{\mathrm{Y}})^{\mathbb{1}}_0 = i$. The imaginary part of the weak value here can be detected by measuring the pointer momenta [16, 27] (see Appendix). Such complex anomalous weak values cannot be obtained without post-selection with only a single weak measurement or a sequence of weak measurements of commuting observables, and can thus themselves be considered anomalous in this sense.

4 More measurements

Eq. (21) might bound how anomalous a weak value can be without post-selection, but it is not generally tight. For two projection observables $\hat{A}$ and $\hat{B}$ (with eigenvalues ± 1), for example, it only implies a bound $\mathrm{Re}[(BA)^{\mathbb{1}}_{\psi}] \geq -1$; nevertheless, as we prove in the Appendix, the value of $-1/8$ obtained earlier for the real part of the weak value is the most negative value that one can obtain. Can one do better by considering longer sequences of successive weak measurements? Here we will see that this question has a subtle answer: the weak value itself can approach -1, but this will not mean the average product of the pointer positions does so as well.

For a sequence of n observables $\hat{A}_1, \ldots, \hat{A}_n$ to be measured weakly on the state $|\psi\rangle$ before a post-selection on $|\phi\rangle$, the sequential weak value is defined (following, e.g., Ref. [21]) as

$$(A_n \cdots A_1)^{\phi}_{\psi} := \frac{\langle\phi|\hat{A}_n \cdots \hat{A}_1|\psi\rangle}{\langle\phi|\psi\rangle}. \tag{22}$$

When no post-selection is performed, this can be generalised to

$$(A_n \cdots A_1)^{\mathbb{1}}_{\psi} := \langle\psi|\hat{A}_n \cdots \hat{A}_1|\psi\rangle\,, \tag{23}$$

in analogy to the cases discussed earlier. As we show in the Appendix, a similar bound to Eq. (21) can be derived, namely

$$|(A_n \cdots A_1)^{\mathbb{1}}_{\psi}| \;\leq\; \prod_{j=1}^{n} \|\hat{A}_j\|\,. \tag{24}$$

For n projection observables, this implies the bound $\mathrm{Re}[(A_n \cdots A_1)^{\mathbb{1}}_{\psi}] \geq -1$. As it turns out, it is possible to obtain an anomalous sequential weak value without post-selection approaching -1 and thus saturating this bound in the limit $n \to \infty$. To see this, take the initial state of the system to be $|\psi\rangle = |0\rangle$ and consider the sequence of n qubit projectors $\hat{A}_j = |a_j\rangle\langle a_j|$ with $|a_j\rangle = \cos(\frac{j\pi}{n+1})\,|0\rangle + \sin(\frac{j\pi}{n+1})\,|1\rangle$ for $j = 1, \ldots, n$. This sequence of weak measurements gives

$$(A_n \cdots A_1)^{\mathbb{1}}_{\psi} = -\left(\cos\tfrac{\pi}{n+1}\right)^{n+1} \xrightarrow{\;n\to\infty\;} -1. \tag{25}$$

Note that for $n = 2$ this coincides precisely with the explicit two-measurement example we began with.

As discussed above, for two sequential weak measurements in the absence of post-selection, the mean product of the pointer positions gives precisely the real part of the sequential weak value; see Eq. (20). However, for $n > 2$ measurements this direct relationship is broken and the mean product of the pointer positions corresponds instead to a mixture of sequential weak values for 2^{n-2} different permutations of the observables (see the Appendix for an explicit expression). For example, for $n = 3$, in the weak regime, we have [4, 25]

$$\langle\hat{x}_1 \otimes \hat{x}_2 \otimes \hat{x}_3\rangle \approx \frac{1}{2}\big(\mathrm{Re}[(A_3A_2A_1)^{\mathbb{1}}_{\psi}] + \mathrm{Re}[(A_2A_3A_1)^{\mathbb{1}}_{\psi}]\big). \tag{26}$$

The real part of $(A_3A_2A_1)^{\mathbb{1}}_{\psi}$ is thus not directly observed. However, as we show in the Appendix, its value (as well as the imaginary part) can nonetheless be deduced experimentally by measuring several different expectation values of the products of pointer positions and momenta [21].

Interestingly, by numerically minimising the mean product of the pointer positions for sequences of up to 5 projection observables, we were unable to obtain a value smaller than $-1/8$, and we conjecture that this is in fact the case for all n. Thus, although the weak value itself can be brought arbitrarily close to -1, it seems that additional sequential weak measurements may not lead to "more anomalous" pointer positions. This behaviour highlights oft-overlooked subtleties in the connection between anomalous weak values and pointer positions: for individual weak measurements, there is a direct correspondence between the pointer position and (the real part of) the weak value, and an anomalous weak value has an immediate physical relevance. For sequential weak measurements, a distinction must be made between anomalous weak values and anomalous pointer positions (with post-selection, this is already the case for two measurements; see Eq. (18) or Ref. [21]). It is the latter of these phenomena that arguably provides the more important anomaly, leading to average measurement results that lie outside the range of values one would expect to be attainable.

In contrast, this divergence between weak values and pointer positions for sequential weak measure-

ments means that, in general, it is more difficult to give a clear physical interpretation to sequential weak values, anomalous or not. Indeed, while some authors have argued that weak values for single weak measurements should be considered real properties of quantum states with direct physical meaning [32, 33], it is unclear whether such arguments are justified for sequential weak values [21] given the lack of examples of physical scenarios where they play a direct, crucial role.

5 Further Discussion

Compare the situation of a sequential weak measurement of two observables $\hat{A}$ and $\hat{B}$ with the alternative in which a bipartite system $|\psi_{ab}\rangle \in \mathcal{H}_a \otimes \mathcal{H}_b$ is prepared and $\hat{A}$ and $\hat{B}$ are weakly measured on the two different substituent systems. One can view this either as a measurement of the joint observable $\hat{A} \otimes \hat{B}$ (with two different pointers, one coupled to each observable) or a sequential measurement of the commuting observables $\hat{A} \otimes \mathbb{1}$ and $\mathbb{1} \otimes \hat{B}$. In the absence of any post-selection, one has $(A \otimes B)^{\mathbb{1}}_{\psi_{ab}} = \langle \psi_{ab} | \hat{A} \otimes \hat{B} | \psi_{ab} \rangle$ which, being just an expectation value, cannot lie outside the spectrum of the product observable $\hat{A} \otimes \hat{B}$. For tensor product measurements, an anomalous weak value is thus unobtainable without post-selection.

This observation raises some interesting implications. Consider for example a scenario in which two parties, Alice and Bob, each operate in a closed laboratory. Each receives a system, performs a weak measurement, and sends the resulting system out; they then come together to jointly measure their pointers. By repeating this many times (or on a large number of systems), they thus determine $\langle \hat{x}_1 \otimes \hat{x}_2 \rangle$. If Alice and Bob have no knowledge of their causal relationship, they could unknowingly be weakly measuring the same system at different times (either Alice then Bob, or vice versa), or measuring different parts of a (potentially entangled) bipartite system. By observing an anomalous weak value they can differentiate between these two scenarios.

The problem of distinguishing these two causal structures for quantum systems – the former is known as a *direct cause* relationship, while the later a *common cause* relationship, since any correlations must be due to a (possibly quantum) common cause – has been the subject of recent interest; see, e.g., Refs. [34–38]. An anomalous weak value thus provides a novel way to witness a direct causal relationship and distinguish between these cases. We leave it as an open question whether such a witness can be found whenever Alice and Bob are capable of signalling to each other; i.e., if whenever they are connected by a quantum channel of nonzero capacity they can always find a pair of observables to measure that would generate an anomalous weak value without post-selection.

Recently, there has also been substantial interest in quantum processes that are not consistent with any definite (possibly stochastic) causal ordering [39], and practical approaches to witness such "indefinite" causal orders have been developed [40, 41] and experimentally tested [42, 43]. It would be interesting to see whether indefinite causal orders may also be witnessed by, for example, producing anomalous weak values or pointer positions lying further outside the expected range than possible in a well-defined causal structure.

The ability for sequential weak measurements to lead to anomalous measurement results even in the absence of post-selection is thus interesting not only in clarifying the relation between post-selection and anomalous weak values, but also in that it raises interesting foundational questions. This sequential weak measurement scheme has, for example, already been used to study time asymmetry in quantum measurements [25, 28, 44] and the role of anomalous weak values in this context would be interesting to study further. One may expect anomalous weak values without post-selection – just as for those with post-selection – to shed light on other foundational results; for example, it would be interesting to see whether they can lead to proofs of contextuality as they do in the standard single weak measurement scenario [17].

Acknowledgements We thank Yelena Guryanova for fruitful discussions. We acknowledge financial support from the 'Retour Post-Doctorants' (ANR-13-PDOC-0026) and the 'Investissements d'avenir' (ANR-15-IDEX-02) programs of the French National Research Agency, and from the Swiss national science foundation (Starting grant DIAQ [grant No. 200021_169002], grant No. 200020_165843, and NCCR-QSIT).

References

[1] Y. Aharonov, D. Z. Albert, and L. Vaidman, How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100, Phys. Rev. Lett. **60**, 1351 (1988).

[2] H. M. Wiseman, Weak values, quantum trajectories, and the cavity-qed experiment on wave-particle correlation, Phys. Rev. A **65**, 032111 (2002).

[3] N. Brunner, A. Acín, D. Collins, N. Gisin, and V. Scarani, Optical telecom networks as weak quantum measurements with postselection, Phys. Rev. Lett. **91**, 180402 (2003).

[4] L. Diósi, Structural features of sequential weak measurements, Phys. Rev. A **94**, 010103(R) (2016).

[5] Y. Aharonov and L. Vaidman, Properties of a quantum system during the time interval between two measurements, Phys. Rev. A **41**, 11 (1990).

[6] Y. Aharonov and L. Vaidman, The two-state vector formalism: An updated review, in *Time in Quantum Mechanics*, Lecture Notes in Physics, Vol. 734, edited by R. S. Mayato, J. G. Muga, and Í. Egusquiza (Springer Verlag, Berlin, 2008) Chap. 13, pp. 339–447.

[7] N. W. M. Ritchie, J. G. Story, and R. G. Hulet, Realization of a measurement of a "weak value", Phys. Rev. Lett. **66**, 1107 (1991).

[8] K. J. Resch, J. S. Lundeen, and A. M. Steinberg, Experimental realization of the quantum box problem, Phys. Lett. A **324**, 125 (2004).

[9] G. J. Pryde, J. L. O'Brien, A. G. White, T. C. Ralph, and H. M. Wiseman, Measurement of quantum weak values of photon polarization, Phys. Rev. Lett. **94**, 220405 (2005).

[10] A. M. Steinberg, How much time does a tunneling particle spend in the barrier region? Phys. Rev. Lett. **74**, 2405 (1995).

[11] N. Brunner, V. Scarani, M. Wegmüller, M. Legré, and N. Gisin, Direct measurement of superluminal group velocity and signal velocity in an optical fiber, Phys. Rev. Lett. **93**, 203902 (2004).

[12] D. R. Solli, C. F. McCormick, R. Y. Chiao, S. Popescu, and J. M. Hickmann, Fast light, slow light, and phase singularities: A connection to generalized weak values, Phys. Rev. Lett. **92**, 043601 (2004).

[13] O. Hosten and P. Kwiat, Observation of the spin hall effect of light via weak measurements, Science **319**, 787 (2008).

[14] P. B. Dixon, D. J. Starling, A. N. Jordan, and J. C. Howell, Ultrasensitive beam deflection measurement via interferometric weak value amplification, Phys. Rev. Lett. **102**, 173601 (2009).

[15] N. Brunner and C. Simon, Measuring small longitudinal phase shifts: Weak measurements or standard interferometry? Phys. Rev. Lett. **105**, 010405 (2010).

[16] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd, Understanding quantum weak values: Basics and applications, Rev. Mod. Phys. **86**, 307 (2014).

[17] M. F. Pusey, Anomalous weak values are proofs of contextuality, Phys. Rev. Lett. **113**, 200401 (2014).

[18] Y. Aharonov, A. Botero, S. Popescu, B. Reznik, and J. Tollaksen, Revisiting hardy's paradox: counterfactual statements, real measurements, entanglement and weak values, Phys. Lett. A **301**, 130 (2002).

[19] J. S. Lundeen, B. Sutherland, A. Patel, C. Stewart, and C. Bamber, Direct measurement of the quantum wavefunction, Nature **474**, 188 (2011).

[20] S. Kocsis, B. Braverman, S. Ravets, M. J. Stevens, R. P. Mirin, L. K. Shalm, and A. M. Steinberg, Observing the average trajectories of single photons in a two-slit interferometer, Science **332**, 1170 (2011).

[21] G. Mitchison, R. Jozsa, and S. Popescu, Sequential weak measurement, Phys. Rev. A **76**, 062105 (2007).

[22] F. Piacentini, A. Avella, M. Levi, M. Gramegna, G. Brida, I. Degiovanni, E. Cohen, R. Lussana, F. Villa, A. Tosi, F. Zappa, and M. Genovese, Measuring incompatible observables by exploiting sequential weak values, Phys. Rev. Lett. **117**, 170402 (2016).

[23] Y. Kim, Y.-S. Kim, S.-Y. Lee, S.-W. Han, S. Moon, Y.-H. Kim, and Y.-W. Cho, Direct quantum process tomography via measuring sequential weak values of incompatible observables, Nat. Commun. **9**, 192 (2018).

[24] J.-S. Chen, M.-J. Hu, X.-M. Hu, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, and Y.-S. Zhang, Experimental realization of sequential weak measurements of arbitrary non-commuting pauli observables, Opt. Express **27**, 6089 (2019).

[25] D. Curic, M. C. Richardson, G. S. Thekkadath, J. Flórez, L. Giner, and J. S. Lundeen, Experimental investigation of measurement-induced disturbance and time symmetry in quantum physics, Phys. Rev. A **97**, 042128 (2018).

[26] D. Georgiev and E. Cohen, Probing finite coarse-grained virtual Feynman histories with sequential weak values, Phys. Rev. A **97**, 052102 (2018).

[27] K. J. Resch and A. M. Steinberg, Extracting joint weak values with local, single-particle measurements, Phys. Rev. Lett. **92**, 130402 (2004).

[28] A. Bednorz, K. Franke, and W. Belzig, Noninvasiveness and time symmetry of weak measurements, New J. Phys. **15**, 023043 (2013).

[29] A. Bednorz and W. Belzig, Quasiprobabilistic interpretation of weak measurements in mesoscopic junctions, Phys. Rev. Lett. **105**, 106803 (2010).

[30] J. S. Lundeen and C. Bamber, Procedure for direct measurement of general quantum states using weak measurement, Phys. Rev. Lett. **108**, 070402 (2012).

[31] G. Thekkadath, L. Giner, Y. Chalich, M. Horton, J. Banker, and J. Lundeen, Direct measurement of the density matrix of a quantum system, Phys. Rev. Lett. **117**, 120401 (2016).

[32] L. Vaidman, Weak-measurement elements of reality, Found. Phys. **26**, 895 (1996).

[33] L. Vaidman, A. Ben-Israel, J. Dziewior, L. Knips, M. Weißl, J. Meinecke, C. Schwemmer, R. Ber, and H. Weinfurter, Weak value beyond conditional expectation value of the pointer readings, Phys. Rev. A **96**, 032114 (2017).

[34] J. F. Fitzsimons, J. A. Jones, and V. Vedral, Quantum correlations which imply causation, Sci. Rep. **5**, 18281 (2015).

[35] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, and K. J. Resch, A quantum

advantage for inferring causal structure, Nat. Phys. **11**, 414 (2015).

[36] F. Costa and S. Shrapnel, Quantum causal modelling, New J. Phys. **18**, 063032 (2016).

[37] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, Quantum common causes and quantum causal models, Phys. Rev. X **7**, 031021 (2017).

[38] C. Giarmatzi and F. Costa, A quantum causal discovery algorithm, npj Quantum Inf. **4**, 17 (2018).

[39] O. Oreshkov, F. Costa, and Č. Brukner, Quantum correlations with no causal order, Nat. Commun. **3**, 1092 (2012).

[40] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, and Č. Brukner, Witnessing causal nonseparability, New. J. Phys. **17**, 102001 (2015).

[41] C. Branciard, Witnesses of causal nonseparability: an introduction and a few case studies, Sci. Rep. **6**, 26018 (2016).

[42] G. Rubino, L. A. Rozema, A. Feix, M. Araújo, J. M. Zeuner, L. M. Procopio, Č. Brukner, and P. Walther, Experimental verification of an indefinite causal order, Sci. Adv. **3**, e1602589 (2017).

[43] K. Goswami, C. Giarmatzi, M. Kewming, F. Costa, C. Branciard, J. Romero, and A. G. White, Indefinite causal order in a quantum switch, Phys. Rev. Lett. **121**, 090503 (2018).

[44] K. Franke, A. Bednorz, and W. Belzig, Time asymmetry in weak measurements, Phys. Scr. **T151**, 014013 (2012).

[45] Y. Aharonov and L. Vaidman, The two-state vector formalism of quantum mechanics, in *Time in Quantum Mechanics*, edited by J. G. Muga, R. Sala Mayato, and I. L. Egusquiza (Springer-Verlag, Berlin Heidelberg, 2002) Chap. 13, pp. 369–412.

A Weak measurements with Gaussian pointers

The weak measurement regime can in general be defined for any type of von Neumann measurement interaction scheme, as introduced in the main text, by comparing the various parameters that describe it: the strength of the measurement interaction (γ), the time of the interaction (Δt), the width of the measurement pointer (σ), the eigenspectrum of the observable $\hat{A}$ being measured as well as the weak values to be considered. The relation between weak values and pointer positions can then be obtained by taking the appropriate limits.

As the point of our paper is to analyse specific cases of anomalous weak values and anomalous pointer positions, for simplicity we choose a specific form for the pointer states, namely, Gaussian states. By a Gaussian pointer, we mean a pointer whose state $|\varphi(a)\rangle$ is described by a Gaussian wavefunction as follows:

$$|\varphi(a)\rangle \coloneqq \int_{-\infty}^{\infty} \mathrm{d}x \, \left(\frac{1}{2\pi\sigma^2}\right)^{1/4} \exp\left[-\frac{(x-a)^2}{4\sigma^2}\right] |x\rangle \,, \tag{A1}$$

where $\{|x\rangle\}_x$ is a continuous eigenbasis of the pointer position $\hat{x}$. For $a \in \mathbb{R}$, $|\varphi(a)\rangle$ is properly normalised; for a complex value of a, its norm is $e^{\frac{\mathrm{Im}(a)^2}{4\sigma^2}}$. The mean position of the pointer in the state $|\varphi(a)\rangle$ (possibly after renormalisation) is $\langle\hat{x}\rangle = \mathrm{Re}(a)$ and its variance is $\langle\hat{x}^2\rangle - \langle\hat{x}\rangle^2 = \sigma^2$, while the mean value of the momentum operator $\hat{p} = -i\frac{\partial}{\partial x}$ (taking $\hbar = 1$) is $\langle\hat{p}\rangle = \frac{\mathrm{Im}(a)}{2\sigma^2}$ ($= 0$ for $a \in \mathbb{R}$). Prior to the measurement, we always start with $a = 0$. Note that an operator of the form $e^{-i\alpha\hat{p}}$ (for any $\alpha \in \mathbb{C}$) acts as a displacement operator, such that $e^{-i\alpha\hat{p}}\,|\varphi(a)\rangle = |\varphi(a+\alpha)\rangle$.

Let us clarify here the conditions that define the standard weak measurement regime, under which the approximations of Eqs. (1)–(2) of the main text are valid. As everywhere in the paper, we choose units such that $g = \gamma\Delta t = 1$; as we noted in Footnote 1, this effectively fixes the scale for the pointer movement, while the weak regime in general depends on the spread of the pointer with respect to this scale, i.e., on σ/g. By considering the spectral decomposition $\hat{A} = \sum_k a_k\,|a_k\rangle\langle a_k|$ of the observable under consideration and the completeness relation $\mathbb{1} = \sum_k |a_k\rangle\langle a_k|$ (with $\{|a_k\rangle\}_k$ an orthonormal basis of the system Hilbert space), one can write $e^{-i\hat{A}\otimes\hat{p}} = \sum_k |a_k\rangle\langle a_k| \otimes e^{-ia_k\hat{p}}$ and $\mathbb{1} - i\hat{A}\otimes\hat{p} = \sum_k |a_k\rangle\langle a_k| \otimes (\mathbb{1} - ia_k\hat{p})$, so that the difference between the left and right hand sides of Eq. (1) is

$$\sum_k \langle a_k|\psi\rangle\,|a_k\rangle \otimes |\delta_k\rangle \quad \text{with} \quad |\delta_k\rangle \coloneqq e^{-ia_k\hat{p}}\,|\varphi(0)\rangle - (\mathbb{1} - ia_k\hat{p})\,|\varphi(0)\rangle = |\varphi(a_k)\rangle - (\mathbb{1} - ia_k\hat{p})\,|\varphi(0)\rangle \,. \tag{A2}$$

The approximation of Eq. (1) is valid if for each k (for which $|\langle a_k|\psi\rangle|$ is non-negligible), the norm of $|\delta_k\rangle$ is

small enough (compared e.g. to that of the lhs of Eq. (1), which is 1). Using Eq. (A1), one finds

$$\langle\delta_k|\delta_k\rangle = 2\Big(1-e^{-\frac{a_k^2}{8\sigma^2}}\Big)+\frac{1}{4}\frac{a_k^2}{\sigma^2}\Big(1-2e^{-\frac{a_k^2}{8\sigma^2}}\Big) = \frac{3}{64}\Big(\frac{a_k}{\sigma}\Big)^4+O\Big[\Big(\frac{a_k}{\sigma}\Big)^6\Big], \tag{A3}$$

which is indeed small if

$$\sigma \gg |a_k| \quad \forall\, k. \tag{A4}$$

Similarly, the difference between the two lines of Eq. (2) is (ignoring the common prefactor $\langle\phi|\psi\rangle$)

$$e^{-i\,A_\psi^\phi\,\hat{p}}\,|\varphi(0)\rangle - (\mathbb{1}-i\,A_\psi^\phi\,\hat{p})\,|\varphi(0)\rangle = |\varphi(A_\psi^\phi)\rangle - (\mathbb{1}-i\,A_\psi^\phi\,\hat{p})\,|\varphi(0)\rangle\,, \tag{A5}$$

which is also small if

$$\sigma \gg |A_\psi^\phi|. \tag{A6}$$

Thus the weak regime is valid whenever the two conditions (A4) and (A6) are fulfilled.

B Relating weak values to pointer positions and momenta

B.1 The case of a single weak measurement

It will be useful to relate here the formula for a weak value to the pointer position and momentum in a more general setting than that considered in the main text, where the initial state is not considered *a priori* to be pure, and where post-selection is conditioned on a given result of an arbitrary Positive-Operator Valued Measure (POVM) measurement on the system, rather than a projective measurement. In such a setting, Eq. (3) can be generalised to

$$A_\rho^E := \frac{\mathrm{Tr}(E\hat{A}\rho)}{\mathrm{Tr}(E\rho)}, \tag{A7}$$

which now defines the weak value of the observable $\hat{A}$, given the pre-selection in the state ρ and post-selection by the POVM element E. This definition was first proposed explicitly in Ref. [2] (although earlier alluded to in Ref. [45]), and shown to be indeed the natural generalisation of the standard definition (3) of a weak value. Note already that this definition indeed reduces to Eq. (3) if the preparation is a pure state $|\psi\rangle$ (i.e. for $\rho=|\psi\rangle\langle\psi|$) and the post-selection is a projection onto another pure state $|\phi\rangle$ (for $E=|\phi\rangle\langle\phi|$). It also allows one to define a weak value with no post-selection by taking a trivial POVM element $E=\mathbb{1}$, which indeed reduces to the definition of Eq. (6) in the case of a pure state $\rho=|\psi\rangle\langle\psi|$ – and which simply coincides here, in the case of a single observable, with the expectation value of $\hat{A}$; note in particular that, contrary to a general weak value, the weak value with no post-selection is linear in the pre-selected state.

In the von Neumann measurement scenario that we consider here, one thus prepares the density matrix ρ, weakly measures $\hat{A}\,(=\sum_k a_k\,|a_k\rangle\langle a_k|)$ (with a pointer in a Gaussian state as described above), and finally post-selects an outcome corresponding to the POVM element E. The initial density matrix of the system and pointer is given by

$$\varrho = \rho\otimes|\varphi(0)\rangle\langle\varphi(0)| = \sum_{k\ell}|a_k\rangle\langle a_k|\,\rho\,|a_\ell\rangle\langle a_\ell|\otimes|\varphi(0)\rangle\langle\varphi(0)|\,. \tag{A8}$$

Under the interaction between the system and the pointer, the joint state evolves to

$$e^{-i\hat{A}\hat{p}}\,\varrho\,e^{+i\hat{A}\hat{p}} = \sum_{k\ell}|a_k\rangle\langle a_k|\,\rho\,|a_\ell\rangle\langle a_\ell|\otimes e^{-ia_k\hat{p}}\,|\varphi(0)\rangle\langle\varphi(0)|\,e^{+ia_\ell\hat{p}} \tag{A9}$$

$$= \sum_{k\ell}|a_k\rangle\langle a_k|\,\rho\,|a_\ell\rangle\langle a_\ell|\otimes|\varphi(a_k)\rangle\langle\varphi(a_\ell)|\,. \tag{A10}$$

Due to the post-selection upon E, the state of the pointer is then projected onto the (unnormalised) state

$$\eta = \mathrm{Tr}_S\left((E\otimes\mathbb{1})\left(e^{-i\hat{A}\hat{p}}\,\varrho\,e^{+i\hat{A}\hat{p}}\right)\right) = \sum_{k\ell}\mathrm{Tr}\left(E\,|a_k\rangle\langle a_k|\,\rho\,|a_\ell\rangle\langle a_\ell|\right)|\varphi(a_k)\rangle\langle\varphi(a_\ell)| \tag{A11}$$

(where Tr_S is the partial trace over the state of the system).

The expectation value of the position of the pointer, given that the post-selection was successful, is

$$\langle \hat{x} \rangle = \frac{\mathrm{Tr}\,(\hat{x}\,\eta)}{\mathrm{Tr}\,(\eta)} = \frac{\sum_{k\ell} \mathrm{Tr}\left(E\,|a_k\rangle\!\langle a_k|\,\rho\,|a_\ell\rangle\!\langle a_\ell|\right)\langle\varphi(a_\ell)|\hat{x}|\varphi(a_k)\rangle}{\sum_{mn} \mathrm{Tr}\left(E\,|a_m\rangle\!\langle a_m|\,\rho\,|a_n\rangle\!\langle a_n|\right)\langle\varphi(a_n)|\varphi(a_m)\rangle}. \tag{A12}$$

Evaluating the expressions in the fraction above for the Gaussian pointer of Eq. (A1) (with $a_k, a_\ell, a_m, a_n \in \mathbb{R}$), and taking the weak regime approximation in which $\sigma \gg |a_k - a_\ell|$, $\sigma \gg |a_m - a_n|$, one finds, to the lowest order,

$$\langle\varphi(a_\ell)|\hat{x}|\varphi(a_k)\rangle = \frac{a_k + a_\ell}{2} e^{-\frac{(a_k-a_\ell)^2}{8\sigma^2}} \approx \frac{a_k + a_\ell}{2}, \qquad \langle\varphi(a_n)|\varphi(a_m)\rangle = e^{-\frac{(a_m-a_n)^2}{8\sigma^2}} \approx 1, \tag{A13}$$

so that we are left with

$$\langle \hat{x} \rangle \approx \frac{\sum_{k\ell} \mathrm{Tr}\left(E\,|a_k\rangle\!\langle a_k|\,\rho\,|a_\ell\rangle\!\langle a_\ell|\right)(a_k + a_\ell)/2}{\sum_{mn} \mathrm{Tr}\left(E\,|a_m\rangle\!\langle a_m|\,\rho\,|a_n\rangle\!\langle a_n|\right)} = \frac{\left(\mathrm{Tr}(E\hat{A}\rho) + \mathrm{Tr}(E\rho\hat{A})\right)/2}{\mathrm{Tr}(E\rho)} = \mathrm{Re}\left(\frac{\mathrm{Tr}(E\hat{A}\rho)}{\mathrm{Tr}(E\rho)}\right), \tag{A14}$$

where we used the spectral decomposition of $\hat{A}$ and the cyclic property of the trace, together with the fact that E, $\hat{A}$ and ρ are Hermitian and thus $\mathrm{Tr}(EP\hat{A}) = \mathrm{Tr}((EP\hat{A})^\dagger)^*$. Recalling the generalised definition (A7) of a weak value, we thus find that the mean position of the pointer (when the post-selection is successful) is, as in Eq. (4),

$$\langle \hat{x} \rangle \approx \mathrm{Re}(A^E_\rho). \tag{A15}$$

Similarly, by using the fact that

$$\langle\varphi(a_\ell)|\hat{x}^2|\varphi(a_k)\rangle = \left(\sigma^2 + \left(\frac{a_k + a_\ell}{2}\right)^2\right) e^{-\frac{(a_k-a_\ell)^2}{8\sigma^2}} \approx \sigma^2 + \left(\frac{a_k + a_\ell}{2}\right)^2 \tag{A16}$$

one may verify that the variance of the pointer position is

$$(\Delta\hat{x})^2 = \langle\hat{x}^2\rangle - \langle\hat{x}\rangle^2 \approx \sigma^2 + \frac{1}{2}\left[\mathrm{Re}\big((A^2)^E_\rho\big) + \frac{\mathrm{Tr}(E\hat{A}\rho\hat{A})}{\mathrm{Tr}(E\rho)}\right] - \big[\mathrm{Re}(A^E_\rho)\big]^2, \tag{A17}$$

which, for a weak enough measurement is dominated by σ^2. Note that for pure pre- and post-selected states (i.e. for $\rho = |\psi\rangle\!\langle\psi|$ and $E = |\phi\rangle\!\langle\phi|$) one has $\frac{\mathrm{Tr}(E\hat{A}\rho\hat{A})}{\mathrm{Tr}(E\rho)} = \frac{\langle\phi|\hat{A}|\psi\rangle\langle\psi|\hat{A}|\phi\rangle}{\langle\phi|\psi\rangle\langle\psi|\phi\rangle} = |A^\phi_\psi|^2$. In the absence of post-selection (i.e., $E = \mathbb{1}$) on the other hand, when the weak value is equal to the expectation value, one has simply $(\Delta\hat{x})^2 = \sigma^2 + (\Delta\hat{A})^2$ (where $(\Delta\hat{A})^2$ is the variance of the observable $\hat{A}$ for the state ρ).

One may also consider measuring the expectation value of the *momentum* of the pointer instead, conditioned again on a successful post-selection:

$$\langle \hat{p} \rangle = \frac{\mathrm{Tr}\,(\hat{p}\,\eta)}{\mathrm{Tr}\,(\eta)} = \frac{\sum_{k\ell} \mathrm{Tr}\left(E\,|a_k\rangle\!\langle a_k|\,\rho\,|a_\ell\rangle\!\langle a_\ell|\right)\langle\varphi(a_\ell)|\hat{p}|\varphi(a_k)\rangle}{\sum_{mn} \mathrm{Tr}\left(E\,|a_m\rangle\!\langle a_m|\,\rho\,|a_n\rangle\!\langle a_n|\right)\langle\varphi(a_n)|\varphi(a_m)\rangle}. \tag{A18}$$

The relevant quantity for the pointer states is, in the weak regime approximation,

$$\langle\varphi(a_\ell)|\hat{p}|\varphi(a_k)\rangle = \frac{1}{2\sigma^2}\,\frac{a_k - a_\ell}{2i} e^{-\frac{(a_k-a_\ell)^2}{8\sigma^2}} \approx \frac{1}{2\sigma^2}\,\frac{a_k - a_\ell}{2i}, \tag{A19}$$

from which (together with Eq. (A13)) we find that the expectation of the pointer's momentum is

$$\langle \hat{p} \rangle \approx \frac{1}{2\sigma^2}\frac{\sum_{k\ell} \mathrm{Tr}\left(E\,|a_k\rangle\!\langle a_k|\,\rho\,|a_\ell\rangle\!\langle a_\ell|\right)(a_k - a_\ell)/2i}{\sum_{mn} \mathrm{Tr}\left(E\,|a_m\rangle\!\langle a_m|\,\rho\,|a_n\rangle\!\langle a_n|\right)} = \frac{1}{2\sigma^2}\frac{\left(\mathrm{Tr}(E\hat{A}\rho) - \mathrm{Tr}(E\rho\hat{A})\right)/2i}{\mathrm{Tr}(E\rho)} = \frac{1}{2\sigma^2}\,\mathrm{Im}\left(\frac{\mathrm{Tr}(E\hat{A}\rho)}{\mathrm{Tr}(E\rho)}\right), \tag{A20}$$

that is,

$$\langle \hat{p} \rangle \approx \frac{1}{2\sigma^2}\,\mathrm{Im}(A^E_\rho). \tag{A21}$$

The expectation value of the momentum is thus directly linked here to the imaginary part of the weak value. Using Eqs. (A15) and (A21), one may therefore recover both the real and imaginary parts of A^E_ρ from the expectation values of the pointer's position and momentum (in a regime where σ is large enough to ignore higher order terms, but not so small as to render the term that remains above, with the pre-factor $\frac{1}{2\sigma^2}$, unmeasurable).

Note that, unlike the expression for $\langle\hat{x}\rangle$, the above expression for $\langle\hat{p}\rangle$ depends explicitly on the width of the pointer. However, by using the fact that

$$\langle\varphi(a_\ell)|\hat{p}^2|\varphi(a_k)\rangle = \frac{1}{4\sigma^4}\left(\sigma^2 - \left(\frac{a_k - a_\ell}{2}\right)^2\right) e^{-\frac{(a_k-a_\ell)^2}{8\sigma^2}} \approx \frac{1}{4\sigma^4}\left(\sigma^2 - \left(\frac{a_k - a_\ell}{2}\right)^2\right), \tag{A22}$$

the error in the pointer momentum is seen to scale similarly, and so the relative error scales in the same way as that of the pointer position. Indeed, one finds

$$(\Delta\hat{p})^2 = \langle\hat{p}^2\rangle - \langle\hat{p}\rangle^2 \approx \frac{1}{4\sigma^4}\left(\sigma^2 - \frac{1}{2}\left[\mathrm{Re}\big((A^2)^E_\rho\big) - \frac{\mathrm{Tr}(E\hat{A}\rho\hat{A})}{\mathrm{Tr}(E\rho)}\right] - \left[\mathrm{Im}(A^E_\rho)\right]^2\right). \tag{A23}$$

B.2 Two sequential weak measurements

Let us now turn to the sequential measurement of two observables $\hat{A}$ and $\hat{B}$. The sequential weak value $(BA)^\phi_\psi$ for a system prepared in the pure state $|\psi\rangle$ and post-selected in $|\phi\rangle$, defined in Eq. (17), was introduced in Ref. [21]. As in the previous section, one may consider a more general setting where the initial state ρ is not necessarily assumed to be pure, and the post-selection is conditioned on a general POVM element E. In such a case, the definition of Eq. (17) can, similarly to Eq. (A7), naturally be generalised to

$$(BA)^E_\rho \coloneqq \frac{\mathrm{Tr}(E\hat{B}\hat{A}\rho)}{\mathrm{Tr}(E\rho)}. \tag{A24}$$

One can indeed verify that one recovers Eq. (17) for $\rho = |\psi\rangle\langle\psi|$ and $E = |\phi\rangle\langle\phi|$. As before, this definition also allows one to define a sequential weak value with no post-selection by taking the trivial POVM element $E = \mathbb{1}$, as in Eq. (19) for the case of a pure state $\rho = |\psi\rangle\langle\psi|$. Again, and contrary to a general sequential weak value, the sequential weak value with no post-selection is linear in the pre-selected state; note, however, that it no longer coincides with an expectation value, as in general the product $\hat{B}\hat{A}$ is not Hermitian, and thus does not define a valid observable.

We consider here a sequential von Neumann measurement scenario where two separate Gaussian pointers (labelled by the subscripts $j = 1, 2$) are used to measure $\hat{A}\,(= \sum_k a_k\,|a_k\rangle\langle a_k|)$ and $\hat{B}\,(= \sum_m b_m\,|b_m\rangle\langle b_m|)$ on a system prepared in the state ρ and post-selected on a POVM element E. Similarly to the analysis in the previous section, the final (unnormalised) state of the two pointers after the post-selection is given by (with implicit identity operators)

$$\begin{aligned}\eta &= \mathrm{Tr}_S\left(E\, e^{-i\hat{B}\hat{p}_2} e^{-i\hat{A}\hat{p}_1}\left(\rho\otimes|\varphi_1(0)\rangle\langle\varphi_1(0)|\otimes|\varphi_2(0)\rangle\langle\varphi_2(0)|\right) e^{+i\hat{A}\hat{p}_1} e^{+i\hat{A}\hat{p}_2}\right)\\ &= \sum_{k\ell mn}\mathrm{Tr}\left(E\;|b_m\rangle\langle b_m|a_k\rangle\langle a_k|\,\rho\,|a_\ell\rangle\langle a_\ell|b_n\rangle\langle b_n|\right)|\varphi_1(a_k)\rangle\langle\varphi_1(a_\ell)|\otimes|\varphi_2(b_m)\rangle\langle\varphi_2(b_n)|\,.\end{aligned} \tag{A25}$$

Using the weak regime approximations of Eq. (A13), for both the weak measurement of $\hat{A}$ and of $\hat{B}$, we find that the expectation value of the product of the pointer positions, given that the post-selection was successful, is

$$\begin{aligned}\langle\hat{x}_1\otimes\hat{x}_2\rangle = \frac{\mathrm{Tr}\left(\hat{x}_1\otimes\hat{x}_2\,\eta\right)}{\mathrm{Tr}\left(\eta\right)} &\approx \frac{\sum_{k\ell mn}\mathrm{Tr}\left(E\;|b_m\rangle\langle b_m|a_k\rangle\langle a_k|\,\rho\,|a_\ell\rangle\langle a_\ell|b_n\rangle\langle b_n|\right)(a_k+a_\ell)(b_m+b_n)/4}{\mathrm{Tr}(E\rho)}\\ &= \frac{1}{4}\frac{\mathrm{Tr}(E\hat{B}\hat{A}\rho) + \mathrm{Tr}(E\rho\hat{A}\hat{B}) + \mathrm{Tr}(E\hat{A}\rho\hat{B}) + \mathrm{Tr}(E\hat{B}\rho\hat{A})}{\mathrm{Tr}(E\rho)}\\ &= \frac{1}{2}\left[\mathrm{Re}\left(\frac{\mathrm{Tr}(E\hat{B}\hat{A}\rho)}{\mathrm{Tr}(E\rho)}\right) + \mathrm{Re}\left(\frac{\mathrm{Tr}(E\hat{A}\rho\hat{B})}{\mathrm{Tr}(E\rho)}\right)\right].\end{aligned} \tag{A26}$$

For pure pre- and post-selected states (i.e. for $\rho = |\psi\rangle\langle\psi|$ and $E = |\phi\rangle\langle\phi|$), one has $\frac{\mathrm{Tr}(E\hat{A}\rho\hat{B})}{\mathrm{Tr}(E\rho)} = \frac{\langle\phi|\hat{A}|\psi\rangle\langle\psi|\hat{B}|\phi\rangle}{\langle\phi|\psi\rangle\langle\psi|\phi\rangle} = A^\phi_\psi(B^\phi_\psi)^*$, so that one recovers Eq. (18). Note however that Eq. (18) does *not* hold for the generalised weak values $(BA)^E_\rho$, A^E_ρ and B^E_ρ; the correct generalisation is instead given by the equation above.

Similar calculations as above also lead to

$$\langle \hat{p}_1 \otimes \hat{x}_2 \rangle \approx \frac{1}{2\sigma_1^2} \frac{1}{2} \left[\mathrm{Im} \left(\frac{\mathrm{Tr}(E\hat{B}\hat{A}\rho)}{\mathrm{Tr}(E\rho)} \right) + \mathrm{Im} \left(\frac{\mathrm{Tr}(E\hat{A}\rho\hat{B})}{\mathrm{Tr}(E\rho)} \right) \right], \tag{A27}$$

$$\langle \hat{x}_1 \otimes \hat{p}_2 \rangle \approx \frac{1}{2\sigma_2^2} \frac{1}{2} \left[\mathrm{Im} \left(\frac{\mathrm{Tr}(E\hat{B}\hat{A}\rho)}{\mathrm{Tr}(E\rho)} \right) - \mathrm{Im} \left(\frac{\mathrm{Tr}(E\hat{A}\rho\hat{B})}{\mathrm{Tr}(E\rho)} \right) \right], \tag{A28}$$

$$\langle \hat{p}_1 \otimes \hat{p}_2 \rangle \approx -\frac{1}{4\sigma_1^2\sigma_2^2} \frac{1}{2} \left[\mathrm{Re} \left(\frac{\mathrm{Tr}(E\hat{B}\hat{A}\rho)}{\mathrm{Tr}(E\rho)} \right) - \mathrm{Re} \left(\frac{\mathrm{Tr}(E\hat{A}\rho\hat{B})}{\mathrm{Tr}(E\rho)} \right) \right]. \tag{A29}$$

The real and imaginary parts of the weak value $(BA)_\rho^E = \frac{\mathrm{Tr}(E\hat{B}\hat{A}\rho)}{\mathrm{Tr}(E\rho)}$ are thus not directly given by the mean values of the pointer positions and momenta (conditioned on a successful post-selection), as observed previously in Refs. [21, 27], but can still easily be recovered by combining the mean values above as follows:

$$\langle \hat{x}_1 \otimes \hat{x}_2 \rangle - 4\sigma_1^2\sigma_2^2 \, \langle \hat{p}_1 \otimes \hat{p}_2 \rangle \approx \mathrm{Re}[(BA)_\rho^E], \qquad 2\sigma_1^2 \, \langle \hat{p}_1 \otimes \hat{x}_2 \rangle + 2\sigma_2^2 \, \langle \hat{x}_1 \otimes \hat{p}_2 \rangle \approx \mathrm{Im}[(BA)_\rho^E]. \tag{A30}$$

Nevertheless, if no post-selection is made ($E = \mathbb{1}$), then the two summands in Eq. (A26) and in Eq. (A27) are equal, and one again directly obtains

$$\langle \hat{x}_1 \otimes \hat{x}_2 \rangle \approx \mathrm{Re}[(BA)_\rho^{\mathbb{1}}], \qquad \langle \hat{p}_1 \otimes \hat{x}_2 \rangle \approx \frac{1}{2\sigma_1^2} \mathrm{Im}[(BA)_\rho^{\mathbb{1}}] \tag{A31}$$

(as in Eq. (20) for $\langle \hat{x}_1 \otimes \hat{x}_2 \rangle$). The example of an imaginary anomalous weak value without post-selection using Pauli observables described in the main text, for which $(\hat{\sigma}_\mathrm{X}\hat{\sigma}_\mathrm{Y})_0^{\mathbb{1}} = i$, can, from Eq. (A31), thus be observed by measuring $\langle \hat{p}_1 \otimes \hat{x}_2 \rangle$.

Moreover, these expressions hold as long as the weak regime is applicable for the first measurement, irrespective of the strength of the second [28, 29]. To see this, note that, in the absence of post-selection, Eq. (A25) reduces to

$$\eta = \sum_{k\ell m} \langle b_m|a_k\rangle\langle a_k| \, \rho \, |a_\ell\rangle\langle a_\ell|b_m\rangle \, |\varphi_1(a_k)\rangle\langle\varphi_1(a_\ell)| \otimes |\varphi_2(b_m)\rangle\langle\varphi_2(b_m)| \, . \tag{A32}$$

Eq. (A13) thus holds with *exact equalities* for the terms in $\mathrm{Tr}(\hat{x}_1 \otimes \hat{x}_2 \, \eta)$ and $\mathrm{Tr}(\hat{p}_1 \otimes \hat{x}_2 \, \eta)$ corresponding to the second pointer, and the weak regime approximation is thus only required for the first measurement. This also justifies further the claim that the second measurement can be seen as performing an effective post-selection, since it can be taken to be arbitrarily strong.

One can also calculate the variances of these quantities using the weak regime approximations given in Eqs. (A16) and (A22), although the explicit expressions become somewhat complicated and are of insufficient interest to reproduce fully here. We note simply that, in general, the variance of $\hat{x}_1 \otimes \hat{x}_2$ is dominated by $\sigma_1^2\sigma_2^2$ in the weak regime so that more statistics are needed to determine the average product of pointer positions than for a single weak measurement. However, in the absence of post-selection one has $(\Delta(\hat{x}_1 \otimes \hat{x}_2))^2 \approx \sigma_1^2\sigma_2^2 + \sigma_1^2 \, \langle \hat{B}^2 \rangle + \sigma_2^2 \, \langle \hat{A}^2 \rangle + O[(\Lambda_\mathrm{max} - \Lambda_\mathrm{min})^2]$ (with $\Lambda_{\mathrm{max/min}} = \Lambda_{\mathrm{max/min}}(\hat{A}, \hat{B})$ as defined in the main text), which again is valid as long as the weak regime is applicable for the first measurement. Thus, by making the second measurement sufficiently strong (i.e., σ_2 small) the variance can be made to scale as σ_1^2, justifying our comment in Footnote 3.

Note that in some specific cases, the expectation values above can also be calculated exactly. For instance, in the first illustrative example (with no post-selection) introduced in the main text, we had

$$\langle \hat{x}_1 \otimes \hat{x}_2 \rangle = \langle \Phi_1^{(1)}| \, \hat{x}_1 \, |\Phi_1^{(1)}\rangle \, \langle \varphi_2(1)| \, \hat{x}_2 \, |\varphi_2(1)\rangle + \langle \Phi_1^{(0)}| \, \hat{x}_1 \, |\Phi_1^{(0)}\rangle \, \langle \varphi_2(0)| \, \hat{x}_2 \, |\varphi_2(0)\rangle = \langle \Phi_1^{(1)}| \, \hat{x}_1 \, |\Phi_1^{(1)}\rangle \, . \tag{A33}$$

Using the explicit forms of Eq. (12) for $|\Phi_1^{(1)}\rangle$ together with Eq. (A1) for the first Gaussian pointer, one finds

$$\langle \hat{x}_1 \otimes \hat{x}_2 \rangle = \frac{1}{16}\big(1 - 3 \, \langle \varphi_1(1)|\hat{x}_1|\varphi_1(0)\rangle - 3 \, \langle \varphi_1(0)|\hat{x}_1|\varphi_1(1)\rangle \big) = \frac{1}{16} \left(1 - 3e^{-\frac{1}{8\sigma_1^2}} \right), \tag{A34}$$

as in Eq. (13). This is in indeed consistent with Eq. (A31) in the weak regime limit where $\sigma_1 \gg 1$ (see Eq. (15)).

The explicit forms of Eqs. (9), (11) and (12) together with Eq. (A1) also allow one to calculate the average positions of the two pointers independently. Specifically, one can calculate $\langle \hat{x}_1 \rangle$ directly from the state after the

first measurement, Eq. (9), as the second measurement does not interact with the first pointer at all. Observing that only the first term from Eq. (9) contributes (because $\langle\varphi_1(0)|\,\hat{x}_1\,|\varphi_1(0)\rangle = 0$), we therefore have

$$\langle\hat{x}_1\rangle = \frac{1}{4}\,\langle\psi_1|\psi_1\rangle\,\langle\varphi_1(1)|\hat{x}_1|\varphi_1(1)\rangle\,\langle\varphi_2(0)|\varphi_2(0)\rangle = \frac{1}{4}\,. \tag{A35}$$

For the second pointer, since $\langle\varphi_2(0)|\,\hat{x}_2\,|\varphi_2(0)\rangle = 0$, we find from Eqs. (11), (12) and (A1) that

$$\langle\hat{x}_2\rangle = \langle\Phi_1^{(1)}|\Phi_1^{(1)}\rangle\,\langle\varphi_2(1)|\hat{x}_2|\varphi_2(1)\rangle = \langle\Phi_1^{(1)}|\Phi_1^{(1)}\rangle = \frac{1}{8}\left(5 - 3e^{-\frac{1}{8\sigma_1^2}}\right) \in \left[\frac{1}{4}, \frac{5}{8}\right], \tag{A36}$$

as claimed in the main text.

B.3 Generalisation to n sequential weak measurements

Considering now n sequential weak measurements, the previous definition of a weak value, for a system prepared in the (possibly mixed) state ρ and post-selected on a POVM element E can be generalised to

$$(A_n\cdots A_1)_\rho^E \coloneqq \frac{\mathrm{Tr}(E\hat{A}_n\cdots\hat{A}_1\rho)}{\mathrm{Tr}(E\rho)}. \tag{A37}$$

For $\rho = |\psi\rangle\langle\psi|$ and $E = |\phi\rangle\langle\phi|$, this indeed reduces to the definition (22) introduced in Ref. [21]. Without post-selection ($E = \mathbb{1}$), this simplifies to $(A_n\cdots A_1)_\rho^{\mathbb{1}} \coloneqq \mathrm{Tr}(\hat{A}_n\cdots\hat{A}_1\rho)$, which in turn reduces to Eq. (23) for a pure pre-selected state $\rho = |\psi\rangle\langle\psi|$ – and which, once again (and contrary to a general sequential weak value), is linear in the pre-selected state.[6]

The previous calculations for the mean values of products of positions and momenta can easily be generalised to n sequential weak measurements. For two complementary subsets $\mathcal{X}$ and $\mathcal{P}$ of $\{1,\ldots,n\}$, Eqs. (A26)–(A29) generalise to (see also Ref. [21])

$$\Big\langle \bigotimes_{i\in\mathcal{X}}\hat{x}_i \bigotimes_{j\in\mathcal{P}}\hat{p}_j \Big\rangle \approx (-i)^{|\mathcal{P}|}\Big(\prod_{j\in\mathcal{P}}\frac{1}{2\sigma_j^2}\Big)\frac{1}{2^n}\frac{\sum_{s_1,\ldots,s_n\in\{0,1\}}(-1)^{\sum_{j\in\mathcal{P}}s_j}\,\mathrm{Tr}(E\hat{A}_n^{1-s_n}\cdots\hat{A}_1^{1-s_1}\rho\hat{A}_1^{s_1}\cdots\hat{A}_n^{s_n})}{\mathrm{Tr}(E\rho)}. \tag{A38}$$

After a couple of lines of algebraic manipulations (using in particular the fact that $\mathrm{Tr}(EQ)^* = \mathrm{Tr}(EQ^\dagger)$ for any matrix Q of the form $\hat{A}_n^{1-s_n}\cdots\hat{A}_1^{1-s_1}\rho\hat{A}_1^{s_1}\cdots\hat{A}_n^{s_n}$, and relabelling certain summation indices $s_j \leftrightarrow 1-s_j$), this can be written as

$$\begin{aligned}\Big\langle \bigotimes_{i\in\mathcal{X}}\hat{x}_i \bigotimes_{j\in\mathcal{P}}\hat{p}_j \Big\rangle \approx& (-1)^{\lfloor\frac{|\mathcal{P}|}{2}\rfloor}\Big(\prod_{j\in\mathcal{P}}\frac{1}{2\sigma_j^2}\Big)\frac{1}{2^{n-1}}\\ &\times\sum_{s_2,\ldots,s_n\in\{0,1\}}(-1)^{\sum_{j\in\mathcal{P}\setminus\{1\}}s_j}\;{}^{\mathrm{Re}}\!/_{\mathrm{Im}}\Big[\frac{\mathrm{Tr}(E\hat{A}_n^{1-s_n}\cdots\hat{A}_2^{1-s_2}\hat{A}_1\rho\hat{A}_2^{s_2}\cdots\hat{A}_n^{s_n})}{\mathrm{Tr}(E\rho)}\Big],\end{aligned} \tag{A39}$$

with ${}^{\mathrm{Re}}\!/_{\mathrm{Im}}$ to be replaced by Re if $|\mathcal{P}|$ is even, and by Im if $|\mathcal{P}|$ is odd.

Generalising Eq. (A30), the real and imaginary parts of the sequential weak value $(A_n\cdots A_1)_\rho^E$ defined above are then recovered by combining the mean values $\big\langle\bigotimes_{i\in\mathcal{X}}\hat{x}_i\bigotimes_{j\in\mathcal{P}}\hat{p}_j\big\rangle$ as follows:

$$\sum_{\mathcal{P}:|\mathcal{P}|\,{}^{\text{even}}\!/_{\text{odd}}}(-1)^{\lfloor\frac{|\mathcal{P}|}{2}\rfloor}\Big(\prod_{j\in\mathcal{P}}2\sigma_j^2\Big)\Big\langle\bigotimes_{i\in\mathcal{X}}\hat{x}_i\bigotimes_{j\in\mathcal{P}}\hat{p}_j\Big\rangle \;\approx\; {}^{\mathrm{Re}}\!/_{\mathrm{Im}}\big[(A_n\cdots A_1)_\rho^E\big], \tag{A40}$$

where the sum is over the 2^{n-1} subsets $\mathcal{P}$ of $\{1,\ldots,n\}$ either such that $|\mathcal{P}|$ is even (for the real part), or such that $|\mathcal{P}|$ is odd (for the imaginary part), and with $\mathcal{X} = \{1,\ldots,n\}\setminus\mathcal{P}$.

In the case with no post-selection ($E = \mathbb{1}$) and when $n \notin \mathcal{P}$, Eq. (A39) can be further simplified (using the cyclic property of the trace) to

$$\begin{aligned}\Big\langle \bigotimes_{i\in\mathcal{X}}\hat{x}_i \bigotimes_{j\in\mathcal{P}}\hat{p}_j \Big\rangle \approx& (-1)^{\lfloor\frac{|\mathcal{P}|}{2}\rfloor}\Big(\prod_{j\in\mathcal{P}}\frac{1}{2\sigma_j^2}\Big)\frac{1}{2^{n-2}}\\ &\times\sum_{\substack{s_2,\ldots,\\ s_{n-1}\in\{0,1\}}}(-1)^{\sum_{j\in\mathcal{P}\setminus\{1\}}s_j}\;{}^{\mathrm{Re}}\!/_{\mathrm{Im}}\big[\,\mathrm{Tr}(\hat{A}_2^{s_2}\cdots\hat{A}_{n-1}^{s_{n-1}}\hat{A}_n\hat{A}_{n-1}^{1-s_{n-1}}\cdots\hat{A}_2^{1-s_2}\hat{A}_1\rho)\big].\end{aligned} \tag{A41}$$

[6] Note that, as seen previously for one and two observables (cf. Footnotes 2 and 5), $(A_n\cdots A_1)_\psi^{\mathbb{1}} = (A_n\cdots A_1)_\psi^\psi$; that is, for a pure pre-selected state $|\psi\rangle$, performing no post-selection gives the same weak value as post-selecting on the pre-selected state, or "re-selecting" [4]. However, we emphasise that, as seen in Eqs. (18) and (20), the mean pointer positions will nonetheless differ in these physically distinct scenarios. Note also that this property does not generalise to mixed states.

Noting that $\mathrm{Tr}(\hat{A}_2^{s_2}\cdots\hat{A}_{n-1}^{s_{n-1}}\hat{A}_n\hat{A}_{n-1}^{1-s_{n-1}}\cdots\hat{A}_2^{1-s_2}\hat{A}_1\rho) = (A_2^{s_2}\cdots A_{n-1}^{s_{n-1}}A_nA_{n-1}^{1-s_{n-1}}\cdots A_2^{1-s_2}A_1)_\rho^{\mathbb{1}}$, one can see that $\langle\bigotimes_{i\in\mathcal{X}}\hat{x}_i\bigotimes_{j\in\mathcal{P}}\hat{p}_j\rangle$ is obtained in Eq. (A41) as the real or imaginary part of a combination of 2^{n-2} sequential weak values with no post-selection, for 2^{n-2} different permutations of the observables. The expression above allows one, in the case where $n=3$, $\mathcal{P}=\emptyset$ and $\rho=|\psi\rangle\langle\psi|$, to recover Eq. (26) of the main text. As for the case with two sequential weak measurements without post-selection, this relation between pointer expectation values and sequential weak values holds even if the final measurement is arbitrarily strong, and the weak regime need only be applicable for the first $n-1$ measurements [28, 29].

When $n\in\mathcal{P}$, $\langle\bigotimes_{i\in\mathcal{X}}\hat{x}_i\bigotimes_{j\in\mathcal{P}}\hat{p}_j\rangle$ vanishes to the first order in the absence of post-selection and the right-hand side of Eq. (A39) becomes 0. It follows that for $E=\mathbb{1}$, one can sum only over subsets $\mathcal{P}$ of $\{1,\ldots,n-1\}$ in Eq. (A40) to recover the weak value $(A_n\cdots A_1)_\rho^E$.

Finally, when $\mathcal{P}=\emptyset$ and $E=\mathbb{1}$, $\langle\hat{x}_1\otimes\cdots\otimes\hat{x}_n\rangle$ can be expressed in a particularly nice form as the nested anti-commutator [4, 29]

$$\langle\hat{x}_1\otimes\cdots\otimes\hat{x}_n\rangle \approx \frac{1}{2^{n-1}}\,\mathrm{Tr}[\{A_1,\{A_2,\ldots,\{A_{n-1},A_n\}\ldots\}\}\,\rho]. \tag{A42}$$

C Proof that $\mathrm{Re}[(BA)_\psi^{\mathbb{1}}]\geq -1/8$ for two projection observables $\hat{A}$ and $\hat{B}$ (with eigenvalues $0,1$)

Suppose the two observables that are sequentially weakly measured on some initial state $|\psi\rangle$ are projectors, $\hat{A}$ and $\hat{B}$. Define then the (normalised) states $|\alpha\rangle \coloneqq \frac{\hat{A}|\psi\rangle}{\sqrt{\langle\psi|\hat{A}|\psi\rangle}}$ and $|\beta\rangle \coloneqq \frac{\hat{B}|\psi\rangle}{\sqrt{\langle\psi|\hat{B}|\psi\rangle}}$ (in the case where $\langle\psi|\hat{A}|\psi\rangle=0$ or $\langle\psi|\hat{B}|\psi\rangle=0$, one can define $|\alpha\rangle$ or $|\beta\rangle$ to be any state orthogonal to $|\psi\rangle$), such that $\langle\alpha|\psi\rangle=\sqrt{\langle\psi|\hat{A}|\psi\rangle}\geq 0$, $\langle\beta|\psi\rangle=\sqrt{\langle\psi|\hat{B}|\psi\rangle}\geq 0$, $\hat{A}\,|\psi\rangle=\langle\alpha|\psi\rangle\,|\alpha\rangle$ and $\hat{B}\,|\psi\rangle=\langle\beta|\psi\rangle\,|\beta\rangle$. We then have:

$$\mathrm{Re}[(BA)_\psi^{\mathbb{1}}]=\mathrm{Re}[\langle\psi|\hat{B}\hat{A}|\psi\rangle]=\langle\alpha|\psi\rangle\,\langle\beta|\psi\rangle\,\mathrm{Re}[\langle\beta|\alpha\rangle]. \tag{A43}$$

Now, using the AM-GM and Cauchy-Schwartz inequalities,

$$\langle\alpha|\psi\rangle\,\langle\beta|\psi\rangle\leq\left(\frac{\langle\alpha|\psi\rangle+\langle\beta|\psi\rangle}{2}\right)^2=\left(\left(\frac{\langle\alpha|+\langle\beta|}{2}\right)|\psi\rangle\right)^2\leq\left\|\frac{|\alpha\rangle+|\beta\rangle}{2}\right\|^2=\frac{1+\mathrm{Re}[\langle\beta|\alpha\rangle]}{2}. \tag{A44}$$

Hence, either $\mathrm{Re}[\langle\beta|\alpha\rangle]\geq 0$ in Eq. (A43), in which case (recalling that $\langle\alpha|\psi\rangle\,\langle\beta|\psi\rangle\geq 0$) one has $\mathrm{Re}[(BA)_\psi^{\mathbb{1}}]\geq 0\,(\geq -\frac{1}{8})$, or $\mathrm{Re}[\langle\beta|\alpha\rangle]<0$ and one then has (using the fact that $\frac{1+x}{2}x\geq-\frac{1}{8}$ for all x)

$$\mathrm{Re}[(BA)_\psi^{\mathbb{1}}]\geq\frac{1+\mathrm{Re}[\langle\beta|\alpha\rangle]}{2}\,\mathrm{Re}[\langle\beta|\alpha\rangle]\geq-\frac{1}{8}\,, \tag{A45}$$

which concludes the proof.

Note, furthermore, that by the linearity of the weak value with no post-selection with respect to the pre-selected state, $(BA)_\rho^{\mathbb{1}}=\sum_i q_i(BA)_{\psi_i}^{\mathbb{1}}$ for a mixed state $\rho=\sum_i q_i\,|\psi_i\rangle\langle\psi_i|$, which implies that the bound above also holds for $(BA)_\rho^{\mathbb{1}}$; that is, $\mathrm{Re}[(BA)_\rho^{\mathbb{1}}]\geq -1/8$ for any two projection observables $\hat{A}$ and $\hat{B}$ and any mixed state ρ.

D Proof of Eq. (24)

To bound the magnitude of the sequential weak value with no post-selection for n measurements, let us write (using the Cauchy-Schwartz inequality in the first line):

$$\begin{aligned}|(A_n\cdots A_1)_\psi^{\mathbb{1}}|^2 \;=\; |\,\langle\psi|\hat{A}_n\cdots\hat{A}_1|\psi\rangle\,|^2 \;&\leq\; \langle\psi|\psi\rangle\,\langle\psi|(\hat{A}_n\cdots\hat{A}_1)^\dagger\hat{A}_n\cdots\hat{A}_1|\psi\rangle\\ &=\langle\psi|\hat{A}_1\cdots\hat{A}_{n-1}\hat{A}_n^2\hat{A}_{n-1}\cdots\hat{A}_1|\psi\rangle\\ &\leq\|\hat{A}_n\|^2\,\langle\psi|\hat{A}_1\cdots\hat{A}_{n-1}^2\cdots\hat{A}_1|\psi\rangle\;\leq\;\cdots\;\leq\;\prod_{j=1}^n\|\hat{A}_j\|^2.\end{aligned} \tag{A46}$$

We further note that by using the linearity of the weak value with no post-selection and the triangle inequality, it is easy to see that Eq. (24) also holds for a preparation in any mixed state $\rho = \sum_i q_i \, |\psi_i\rangle\langle\psi_i|$:

$$|(A_n \cdots A_1)^{\mathbb{1}}_{\rho}| \;=\; \big|\sum_i q_i (A_n \cdots A_1)^{\mathbb{1}}_{\psi_i}\big| \;\leq\; \sum_i q_i |(A_n \cdots A_1)^{\mathbb{1}}_{\psi_i}| \;\leq\; \big(\underbrace{\textstyle\sum_i q_i}_{1}\big) \prod_{j=1}^{n} \|\hat{A}_j\| \,. \tag{A47}$$