

# Equivalence of wave-particle duality to entropic uncertainty

Patrick J. Coles, Jędrzej Kaniewski, and Stephanie Wehner

*Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, 117543 Singapore*

Interferometers capture a basic mystery of quantum mechanics: a single particle can exhibit wave behavior, yet that wave behavior disappears when one tries to determine the particle’s path inside the interferometer. This idea has been formulated quantitatively as an inequality, e.g., by Englert and Jaeger, Shimony, and Vaidman, which upper bounds the sum of the interference visibility and the path distinguishability. Such wave-particle duality relations (WPDRs) are often thought to be conceptually inequivalent to Heisenberg’s uncertainty principle, although this has been debated. Here we show that WPDRs correspond precisely to a modern formulation of the uncertainty principle in terms of entropies, namely the min- and max-entropies. This observation unifies two fundamental concepts in quantum mechanics. Furthermore, it leads to a robust framework for deriving novel WPDRs by applying entropic uncertainty relations to interferometric models. As an illustration, we derive a novel relation that captures the coherence in a quantum beam splitter.

## INTRODUCTION

When Feynman discussed the two-path interferometer in his famous lectures [1], he noted that quantum systems (quantons) display the behavior of both waves and particles and that there is a sort of competition between seeing the wave behavior versus the particle behavior. That is, when the observer tries harder to figure out which path of the interferometer the quanton takes, the wave-like interference becomes less visible. This tradeoff is commonly called wave-particle duality (WPD). Feynman further noted that this is “a phenomenon which is impossible ... to explain in any classical way, and which has in it the heart of quantum mechanics. In reality, it contains the only mystery [of quantum mechanics].”

Many quantitative statements of this idea, so-called wave-particle duality relations (WPDRs), have been formulated [2–13]. Such relations typically consider the Mach-Zehnder interferometer for single photons, see Fig. 1. For example, a well-known formulation proven independently by Englert [2] and Jaeger et al. [3] quantifies the wave behavior by fringe visibility  $\mathcal{V}$ , and particle behavior by the distinguishability of the photon’s path,  $\mathcal{D}$ . (See below for precise definitions; the idea is that “waves” have a definite phase, while “particles” have a definite location, hence  $\mathcal{V}$  and  $\mathcal{D}$  respectively quantify how definite the phase and location are inside the interferometer.) They found the tradeoff:

$$\mathcal{D}^2 + \mathcal{V}^2 \leq 1 \quad (1)$$

which implies  $\mathcal{V} = 0$  when  $\mathcal{D} = 1$  (full particle behavior means no wave behavior) and vice-versa, and also treats the intermediate case of partial distinguishability.

It has been debated, particularly around the mid-1990’s [14–16], whether the WPD principle, closely related to Bohr’s complementarity principle [17], is equivalent to another fundamental quantum idea with no classical analog: Heisenberg’s uncertainty principle [18]. The latter states that there are certain pairs of observables, such as position and momentum or two orthogonal components of spin angular momentum, that cannot simultaneously be known or jointly measured. Likewise there

are many quantitative statements of this idea, known as uncertainty relations (URs) (see, e.g., [19–29]), and modern formulations typically use entropy instead of standard deviation as the uncertainty measure, so-called entropic uncertainty relations (EURs) [24]. This is because the standard deviation formulation suffers from trivial bounds when applied to finite-dimensional systems [21], whereas the entropic formulation not only fixes this weakness but also implies the standard deviation relation [22] and has relevance to information-processing tasks.

At present the debate regarding wave-particle duality and uncertainty remains unresolved, to our knowledge. Yet Feynman’s quote seems to suggest a belief that quantum mechanics has but one mystery and not two separate ones. In this article we confirm this belief by showing a quantitative connection between URs and WPDRs, demonstrating that URs and WPDRs capture the same underlying physics; see also [30, 31] for some partial progress along these lines. This may come as a surprise, since Englert [2] originally argued that (1) “does not make use of Heisenberg’s uncertainty relation in any form”. To be fair, the uncertainty relation that we show is equivalent to (1) was not known at the time of Englert’s paper, and was only recently discovered [25–29]. Specifically, we will consider EURs, where the particular entropies that are relevant to (1) are the so-called min- and max-entropies used in cryptography [32].

In what follows we provide a general framework for deriving and discussing WPDRs - a framework that is ultimately based on the entropic uncertainty principle. We illustrate our framework by showing that several different WPDRs from the literature are in fact particular examples of EURs. Making this connection not only unifies two fundamental concepts in quantum mechanics, but also implies that novel WPDRs can be derived simply by applying already-proven EURs. Indeed we use our framework to derive a novel WPDR for an exotic scenario involving a “quantum beam splitter” [33–36], where testing our WPDR would allow the experimenter to verify the beam splitter’s quantum coherence (see (17)).

We emphasize that the framework provided by EURs is highly robust, and entropies have well-characterized

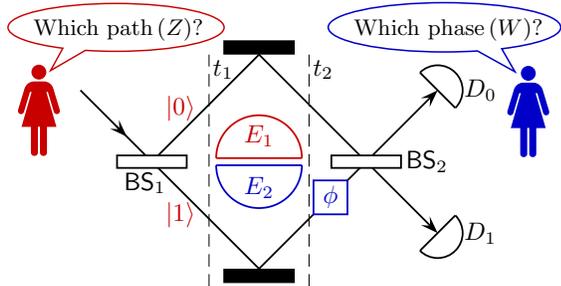


FIG. 1: Mach-Zehnder interferometer for single photons. Passing through the first beam splitter creates a superposition of which-path states,  $|0\rangle$  and  $|1\rangle$ , at time  $t_1$ , then the system interacts with an environment  $E = E_1 E_2$ . Finally at time  $t_2$  a phase shift  $\phi$  is applied to the lower arm and the two beams are recombined on a second beam splitter. (While this is the typical setup, our framework also allows  $E$  to play a more general role, e.g., being correlated to the photon before it enters the MZI.) Our complementary guessing game proceeds as follows. In one game (colored red) Alice tries to guess which of the two paths the photon took given that she has access to a portion of  $E$  denoted  $E_1$ , which could be, e.g., a gas of atoms whose internal states record information about the presence of a photon. In the other game (colored blue), one of two phases,  $\phi = \phi_0$  or  $\phi = \phi_0 + \pi$ , is randomly applied to the lower interferometer arm and Alice tries to guess  $\phi$  given that she has access to a different portion of  $E$  denoted  $E_2$ , which could be, e.g., the photon’s polarisation. We argue that WPDRs impose fundamental trade-offs on Alice’s ability to win these two games.

statistical meanings. Note that current approaches to deriving WPDRs often involve brute force calculation of the quantities one aims to bound; there is no general, elegant method currently in use. Our approach simply involves judicious application of the relevant uncertainty relation. What’s more, we emphasize that uncertainty relations can be applied to interferometers in two different ways. One involves preparation uncertainty, which says that a quantum state cannot be prepared having low uncertainty for two complementary observables, and it turns out this is the principle relevant to the original presentation of (1) in [2]. The other involves measurement uncertainty, which says that two complementary observables cannot be jointly measured [7, 31], and we discuss why this principle is actually what was tested in some recent interferometry experiments [34, 37].

## RESULTS

### Framework

*Guessing games.*—We argue that a natural and powerful way to think of wave-particle duality is in terms of guessing games, and one’s ability to win such games is quantified by entropic quantities. Specifically we consider

*complementary* guessing games, where Alice is asked to guess one of two complementary observables - a modern paradigm for discussing the uncertainty principle. In the Mach-Zehnder interferometer (MZI), see Fig. 1, this corresponds to either guessing which path the photon took, or which phase was applied inside the interferometer. The which-path and which-phase observables are complementary and hence the uncertainty principle gives a fundamental restriction stating that Alice cannot be able to guess both observables.

*Binary interferometers.*—Our framework treats this complementary guessing game for binary interferometers. By *binary*, we mean any interferometer where there are only two interfering paths, i.e., all other paths are classically distinguishable (from each other and from the two interfering paths). In addition to the MZI, this includes as special cases, e.g., the Franson interferometer [38] (see Fig. 2) and the double slit interferometer (see Fig. 3). Note that binary interferometers go beyond interferometers with two physical paths. For example, in the Franson interferometer there are four possible paths but post-selecting on coincidence counts discards two of these paths, which are irrelevant to the interference anyway.

*Particle observable.*—Now we link wave and particle behavior to knowledge of complementary observables. In the case of particle behavior, the intuition is that particles have a well-defined spatial location, hence “particle-ness” should be connected to knowledge of the path inside interferometer. For binary interferometers, there may be more than two physical paths but only two of these are interfering. Hence we only consider the two-dimensional subspace associated with the two which-path states of interest, denoted  $|0\rangle$  and  $|1\rangle$ . This subspace can be thought of as an effective qubit, denoted  $Q$ , and the standard basis of this qubit:

$$\text{which-path: } Z = \{|0\rangle, |1\rangle\} \quad (2)$$

corresponds precisely to the which-path observable. For example, in the double slit (Fig. 3),  $|0\rangle$  and  $|1\rangle$  are the pure states that one would obtain at the slit exit from blocking the bottom and top slits respectively.

*Wave observable.*—Wave behavior is traditionally associated with having a large amplitude of intensity oscillations at the interferometer output. Indeed this has been quantified by the so-called *fringe visibility*, see (7), but to apply the uncertainty principle we need to relate wave behavior to an observable inside the interferometer. Classical waves (e.g., water waves) are often modelled as having a *well-defined phase* and being spatially *delocalized*. The analog in our context corresponds to the quanton being in an equally-weighted superposition of which-path states. Hence eigenstates of the “wave observable” should live in the  $XY$  plane of the Bloch sphere, so we consider observables on qubit  $Q$  (the interfering subspace) of the

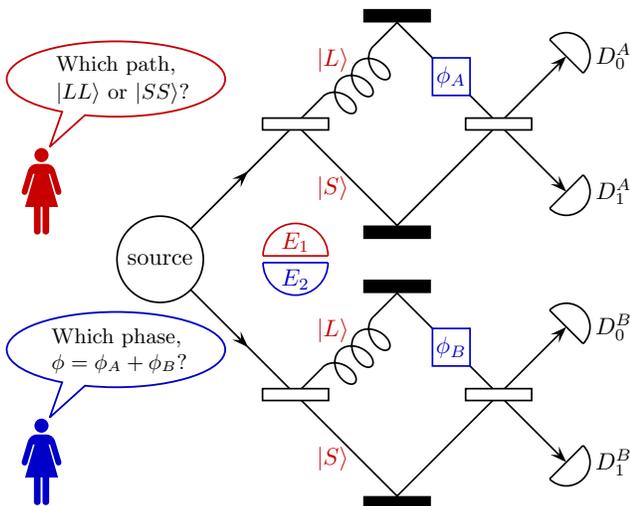


FIG. 2: In the Franson interferometer, the “quantum” consists of two photons. That is, a source produces time-energy entangled photons that each head separately towards a MZI that contains a long arm (depicted with extra loops) and a short arm. A simple model considers the four-dimensional Hilbert space associated with the four possible paths:  $|SS\rangle$ ,  $|SL\rangle$ ,  $|LS\rangle$ , and  $|LL\rangle$  with  $S$  = short path,  $L$  = long path. Two of these dimensions are post-selected away by considering only coincidence counts, i.e., the photons arriving at the same time is inconsistent with the  $|SL\rangle$  and  $|LS\rangle$  paths. The remaining paths,  $|0\rangle := |SS\rangle$  and  $|1\rangle := |LL\rangle$ , are indistinguishable in the special case of perfect visibility and they produce interference fringes as one varies  $\phi := \phi_A + \phi_B$ . Namely, the intensity of coincidence counts at detector pair  $(D_0^A, D_0^B)$  oscillates with  $\phi$ . Interaction with an environment system, or making the beam splitters asymmetric, may allow one to partially distinguish between  $|SS\rangle$  and  $|LL\rangle$ , and our entropic uncertainty framework can be applied to derive a tradeoff, e.g., of the form of (1). This tradeoff captures the idea that Alice can either guess which path ( $|SS\rangle$  vs.  $|LL\rangle$ ) or which phase ( $\phi = \phi_0$  vs.  $\phi = \phi_0 + \pi$ ), but she cannot do both (even if she extracts information from other systems  $E_1$  and  $E_2$ ).

form

$$\text{which-phase: } W = \{|w_{\pm}\rangle\}, \quad |w_{\pm}\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm e^{i\phi_0}|1\rangle). \quad (3)$$

In terms of the guessing game, guessing the value of the wave (or which-phase) observable corresponds to guessing whether a phase of  $\phi = \phi_0$  or  $\phi = \phi_0 + \pi$  was applied inside the interferometer (see, e.g., Fig 1). While  $\phi_0$  is a generic phase, its precise value will be singled out by the particular experimental setup. When the experimenter measures fringe visibility this corresponds to varying  $\phi_0$  to find the largest intensity contrast, and mathematically we model this by minimizing the uncertainty within the  $XY$  plane, see (4b).

## Entropic View

Our entropic view associates a kind of behavior with the availability of a kind of information, or lack of behavior with missing information, as follows:

$$\text{lack of particle behavior: } H_{\min}(Z|E_1) \quad (4a)$$

$$\text{lack of wave behavior: } \min_{W \in XY} H_{\max}(W|E_2) \quad (4b)$$

where  $H_{\min}$  and  $H_{\max}$  are the min- and max-entropies, defined below in (6), which are commonly used in quantum information theory,  $Z$  is the which-path observable in (2),  $W$  is the which-phase observable in (3) (whose uncertainty we optimize over the  $XY$  plane of the Bloch sphere), and  $E_1$  and  $E_2$  are some other quantum systems that contain information and measuring these systems may help to reveal the behavior (e.g.,  $E_1$  could be a which-path detector and  $E_2$  could be the quanton’s internal degree of freedom). Note that we use the same symbols ( $Z$ ,  $W$ , etc.) for the observables as for the random variables they give rise to. Full behavior (no behavior) of some kind corresponds to the associated entropy in (4) being zero (one). We formulate our general WPDR as

$$H_{\min}(Z|E_1) + \min_{W \in XY} H_{\max}(W|E_2) \geq 1. \quad (5)$$

This states that, for a binary interferometer, the sum of the ignorances about the particle and wave behaviors is lower bounded by 1 (i.e., 1 bit). Eq. (5) constrains Alice’s ability to win the complementary guessing game described above. If measuring  $E_1$  allows her to guess the quanton’s path, i.e., the min-entropy in (4a) is small, then even if she measures  $E_2$  she still will not be able to guess the quanton’s phase, i.e., the max-entropy in (4b) will be large (and vice-versa).

To be clear, (5) is explicitly an entropic uncertainty relation, and it has been exploited to prove the security of quantum cryptography [39]. The usefulness of (5) for cryptography is due to the clear operational meanings of the min- and max-entropies [32], which naturally express the monogamy of correlations as they give the distances to being uncorrelated ( $H_{\max}$ ) and being perfectly correlated ( $H_{\min}$ ). One can replace these entropies with the von Neumann entropy in (5) and the relation still holds; however, the min- and max-entropies give more refined statements about information processing since they are also applicable to finite numbers of experiments. From [32], the precise definitions of these entropies, for a generic classical-quantum state  $\rho_{XB}$ , are

$$H_{\min}(X|B) = -\log p_{\text{guess}}(X|B), \quad (6a)$$

$$H_{\max}(X|B) = \log p_{\text{secre}}(X|B), \quad (6b)$$

where all logarithms are base 2 in this article. Here,  $p_{\text{guess}}(X|B)$  denotes the probability for the experimenter to guess  $X$  correctly with the optimal strategy, i.e., with the optimally helpful measurement on system  $B$ . Also,

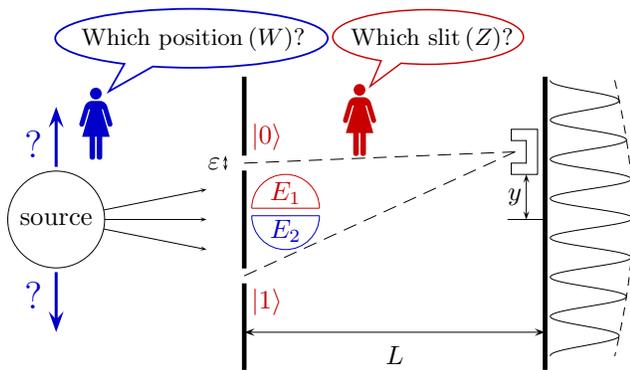


FIG. 3: For the double slit interferometer, the first game (colored red) involves Alice guessing which slit the quanton goes through, given that she can measure a system  $E_1$  that has interacted with, and hence may contain information about, the quanton. In the second game (colored blue), Bob randomly chooses the source’s vertical coordinate, and Alice tries to guess where the source was located, given some other system  $E_2$  and given where the quanton was finally detected. Note that the source’s location determines the relative phase between the which-slit states,  $|0\rangle$  and  $|1\rangle$ , and we assume Bob chooses one of two possible locations such that the relative phase is either 0 or  $\pi$ . Here, the state  $|0\rangle$  ( $|1\rangle$ ) is defined as the pure state at the slit exit that one would obtain from blocking the bottom (top) slit. Our framework provides a WPDR that constrains Alice’s ability to win these complementary games. Furthermore, one can reinterpret the probability to win the second game, for the case where  $E_2$  is trivial, in terms of the traditional fringe visibility. The latter quantifies the amplitude of intensity oscillations as one varies the detector location  $y$ . Note that varying  $y$  changes the relative path lengths from the slits to the detector and hence is analogous to applying a relative phase  $\phi$  between the two paths. [This assumes that the envelope function (dashed curve) associated with the interference pattern is flat over the range of  $y$  values considered, which is often the case when  $L$  is very large.] So the double slit fringe visibility is equivalent to the notion captured by  $\mathcal{V}$  in (7), where the detector is spatially fixed but a phase is varied, and we relate  $\mathcal{V}$  to our entropic measure of wave behavior in the Methods section.

$p_{\text{secre}}(X|B) = \max_{\sigma_B} F(\rho_{XB}, \mathbb{1} \otimes \sigma_B)^2$  quantifies the secrecy of  $X$  from  $B$ , as measured by the fidelity  $F$  of  $\rho_{XB}$  to a state that is completely uncorrelated.

The fact that (5) can be thought of as a WPDR, and furthermore that it encompasses the majority of WPDRs found in the literature for binary interferometers, is our main result.

## DISCUSSION

To illustrate this, we consider the celebrated MZI, shown in Fig. 1, since most literature WPDRs have been formulated for this interferometer. In the simplest case one sends in a single photon towards a 50/50 (i.e., symmetric) beam splitter,  $\text{BS}_1$ , which results in the state

$|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ , then a phase  $\phi$  is applied to the lower arm giving the state  $(|0\rangle + e^{i\phi}|1\rangle)/\sqrt{2}$ . Finally the two paths are recombined on a second 50/50 beam splitter  $\text{BS}_2$  and the output modes are detected by detectors  $D_0$  and  $D_1$ . Fringe visibility is then defined as

$$\text{fringe visibility: } \mathcal{V} := \frac{p_{\text{max}}^{D_0} - p_{\text{min}}^{D_0}}{p_{\text{max}}^{D_0} + p_{\text{min}}^{D_0}}, \quad (7)$$

where  $p^{D_0}$  is the probability for the photon to be detected at  $D_0$ ,  $p_{\text{max}}^{D_0} := \max_{\phi} p^{D_0}$  maximizes this probability over  $\phi$ , whereas  $p_{\text{min}}^{D_0} := \min_{\phi} p^{D_0}$ . In this trivial example one has  $\mathcal{V} = 1$ . However many more complicated situations, for which the analysis is more interesting, have been considered in the extensive literature; we now illustrate how these situations fall under the umbrella of our framework with some examples.

*$\mathcal{P}$ - $\mathcal{V}$  relation.*—As a warm-up, we begin with the simplest known WPDR, the predictability-visibility trade-off. Predictability  $\mathcal{P}$  quantifies the *prior* knowledge, given the experimental setup, about which path the photon will take inside the interferometer. More precisely,  $\mathcal{P} := 2p_{\text{guess}}(Z) - 1$  where  $p_{\text{guess}}(Z)$  is the probability of correctly guessing  $Z$ . Non-trivial predictability is typically obtained by choosing  $\text{BS}_1$  to be *asymmetric*. In such situations, the following bound holds [4, 5]:

$$\mathcal{P}^2 + \mathcal{V}^2 \leq 1. \quad (8)$$

This particularly simple example is a special case of Robertson’s uncertainty relation involving standard deviations [30, 31, 40, 41]. However, [41] argues that (8) is inequivalent to a family of EURs where the same (Rényi) entropy is used for both uncertainty terms, hence one gets the impression that entropic uncertainty is different from wave-particle duality. On the other hand, [41] did not consider the EUR involving the min- and max-entropies. For some probability distribution  $P = \{p_j\}$ , the unconditional min- and max-entropies are given by  $H_{\text{min}}(P) = -\log \max_j p_j$  and  $H_{\text{max}}(P) = 2 \log \sum_j \sqrt{p_j}$ . We find that (8) is equivalent to

$$H_{\text{min}}(Z) + \min_{W \in XY} H_{\text{max}}(W) \geq 1, \quad (9)$$

which is an EUR proved in the seminal paper by Maassen and Uffink [23], and corresponds to  $E_1$  and  $E_2$  in (5) being trivial. The entropies in (9) are evaluated for the state at any time while the photon is inside the interferometer. It is straightforward to see that  $H_{\text{min}}(Z) = -\log \frac{1+\mathcal{P}}{2}$  and in the Methods we prove that

$$\min_{W \in XY} H_{\text{max}}(W) = \log(1 + \sqrt{1 - \mathcal{V}^2}). \quad (10)$$

Plugging these relations into (9) gives (8).

*$\mathcal{D}$ - $\mathcal{V}$  relation.*—Let us move on to a more general and more interesting scenario where, in addition to prior which-path knowledge, one may obtain further knowledge *during* the experiment due to the interaction of the

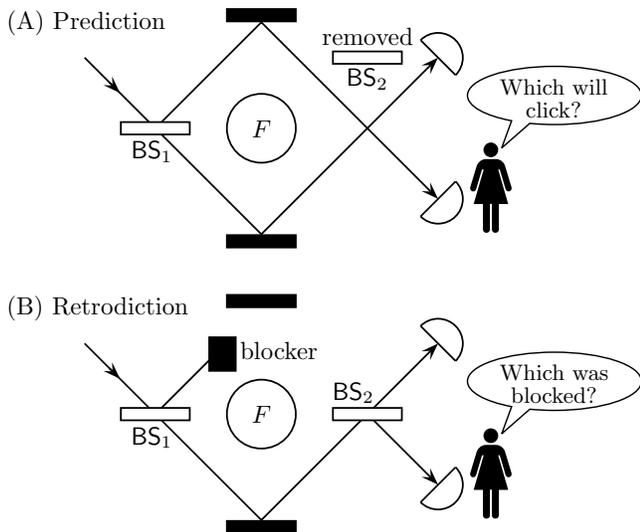


FIG. 4: Path prediction versus path retrodiction, in the MZI. (A) In the predictive scenario, the second beam splitter is removed and Alice tries to guess which detector will click. (B) In the retrodictive scenario, a blocker is randomly inserted into one of the interferometer arms and Alice tries to guess which arm was blocked (given the knowledge of which detector clicked).

photon with some environment  $F$ , which may act as a which-way detector. Most generally the interaction is given by a completely positive trace preserving (CPTP) map  $\mathcal{E}$ , with the input system being  $Q$  at time  $t_1$  and output systems being  $Q$  and  $F$  at time  $t_2$ , see Fig. 1. The final state is  $\rho_{QF}^{(2)} = \mathcal{E}(\rho_Q^{(1)})$ , where the superscripts (1) and (2) indicate the states at times  $t_1$  and  $t_2$ . We do not require  $\mathcal{E}$  to have any special form in order to derive our WPDR, so our treatment is general.

The path distinguishability is defined by  $\mathcal{D} := 2p_{\text{guess}}(Z|F) - 1$ , where  $p_{\text{guess}}(Z|F)$  is the probability for correctly guessing the photon's path  $Z$  at time  $t_2$  given that the experimenter performs the optimally helpful measurement on  $F$ . We find that (1) is equivalent to

$$H_{\min}(Z|F) + \min_{W \in XY} H_{\max}(W) \geq 1, \quad (11)$$

where the entropy terms are evaluated for the state  $\rho_{QF}^{(2)}$ , which corresponds to  $E_1 = F$  and  $E_2$  being trivial in (5). First, it is obvious from the operational meaning of the conditional min-entropy (6a) that we have  $H_{\min}(Z|F) = -\log \frac{1+\mathcal{D}}{2}$ , and second we use our result (10) to rewrite (11) as (1). As emphasized in [2], we note that (1) and its entropic form (11) do not require  $\text{BS}_1$  to be symmetric. Hence  $\mathcal{D}$  accounts for both the prior  $Z$  knowledge associated with the asymmetry of  $\text{BS}_1$  as well as the  $Z$  information gained from  $F$ .

*Preparation vs. measurement uncertainty.*—The above analysis shows that (1) and (8) correspond to applying the *preparation uncertainty relation* at time  $t_2$  (just be-

fore the photon reaches  $\text{BS}_2$ ). Preparation uncertainty restricts one's ability to predict the outcomes of *future* measurements of complementary observables. Thus, to experimentally measure  $\mathcal{P}$  or more generally  $\mathcal{D}$ , the experimenter *removes*  $\text{BS}_2$  and sees how well he/she can guess which detector clicks, see Fig. 4A. Of course, to then measure  $\mathcal{V}$ , the experimenter reinserts  $\text{BS}_2$  to close the interferometer. We emphasize that this procedure falls into the general framework of *preparation uncertainty*.

On the other hand, uncertainty relations can be applied in a conceptually different way. Instead of two complementary output measurements and a fixed input state, consider a fixed output measurement and two complementary sets of input states. Namely consider the input ensembles from (2) and (3), now labeled as  $Z_i = \{|0\rangle, |1\rangle\}$  and  $W_i = \{|w_{\pm}\rangle\}$ , where  $i$  stands for “input”, to indicate the physical scenario of a sender inputting states into a channel. Imagine this as a retrodictive guessing game, where Bob controls the input and Alice has control over both  $F$  and the detectors. Bob chooses one of the ensembles and flips a coin to determine which state from the ensemble he will send, and Alice's goal is to guess Bob's coin flip outcome. Assuming  $\text{BS}_1$  is 50/50, the two  $Z_i$  states are generated by Bob blocking the opposite arm of the interferometer, as in Fig. 4B, while the  $W_i$  states are generated by applying a phase (either  $\phi_0$  or  $\phi_0 + \pi$ ) to the lower arm.

It may not be common knowledge that this scenario leads to a different class of WPDRs, therefore we illustrate the difference in Fig. 4. For clarity, we refer to  $\mathcal{D}$  introduced above as output distinguishability, whereas in the present scenario we use the symbol  $\mathcal{D}_i$  and call this quantity input distinguishability, defined by

$$\mathcal{D}_i := 2p_{\text{guess}}(Z_i|F)_{D_0} - 1, \quad (12)$$

where  $p_{\text{guess}}(Z_i|F)_{D_0}$  is Alice's probability to correctly guess Bob's  $Z_i$  state given that she has access to  $F$  and she knows that detector  $D_0$  clicked at the output. Likewise we define the notion of *input* visibility  $\mathcal{V}_i$  via:

$$\mathcal{V}_i := \max_{W \in XY} [2p_{\text{guess}}(W_i)_{D_0} - 1] \quad (13)$$

which quantifies how well Alice can determine  $W_i$  given that she knows  $D_0$  clicked.

Now the uncertainty principle says there is a tradeoff: if Alice can guess the  $Z_i$  states well then she cannot guess the  $W_i$  states well, and vice-versa. In other words, Alice's measurement apparatus, the apparatus to the right of the dashed line labeled  $t_1$  in Fig. 1, cannot *jointly measure* Bob's  $Z$  and  $W$  observables. EURs involving von Neumann entropy have previously been applied to the joint measurement scenario [27, 42], we do the same for the min- and max-entropies to obtain (see Methods for details)

$$\mathcal{D}_i^2 + \mathcal{V}_i^2 \leq 1, \quad (14)$$

which can now be applied to a variety of situations.

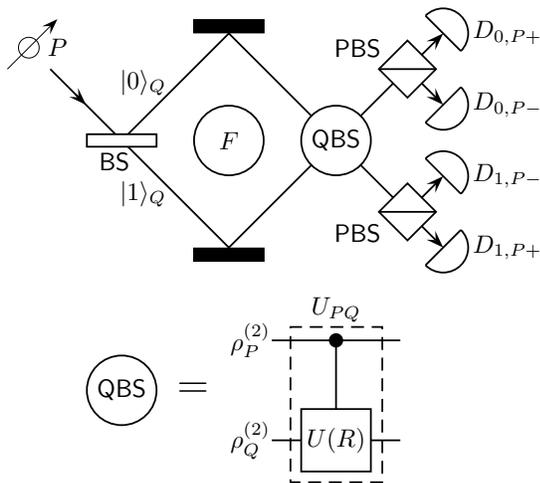


FIG. 5: In the quantum beam splitter (QBS) scenario, the second beam splitter is in a superposition of “absent” and “present”, as determined by the polarization state  $\rho_P^{(2)}$  at time  $t_2$ . The QBS can be modelled as a controlled-unitary,  $U_{PQ} = |H\rangle\langle H|_P \otimes \mathbb{1}_Q + |V\rangle\langle V|_P \otimes U(R)$ , where  $U(R)$  is the unitary on  $Q$  associated with an asymmetric beam splitter with reflection probability  $R$ . Polarization-resolving detectors (PBS = polarizing beam splitter) on the output modes help to reveal the “quantumness” of the QBS.

*Quantum BS<sub>2</sub>.*—As an interesting application of (14), we consider the scenario proposed in [33] and implemented in [34–36], where the photon’s polarization  $P$  acts as a control system to determine whether or not BS<sub>2</sub> appears in the photon’s path and hence whether the interferometer is open or closed, see Fig. 5. Since  $P$  can be prepared in an arbitrary input state  $\rho_P^{(2)}$ , such as a superposition, this effectively means that BS<sub>2</sub> is a “quantum beam splitter”, i.e., it can be in a quantum superposition of being absent or present. The interaction coupling  $P$  to  $Q$  is modelled as a controlled unitary as in Fig. 5. In this case the two visibilities are equivalent (see Methods)

$$\mathcal{V}_i = \mathcal{V} = 2|\kappa|\sqrt{R(1-R)}\langle V|\rho_P^{(2)}|V\rangle \quad (15)$$

where we assume the dynamics are path-preserving, i.e.,  $\mathcal{E}_Q(|0\rangle\langle 0|) = |0\rangle\langle 0|$  and  $\mathcal{E}_Q(|1\rangle\langle 1|) = |1\rangle\langle 1|$ , where  $\mathcal{E}_Q = \text{Tr}_F \circ \mathcal{E}$  is the reduced channel on  $Q$ , which implies that  $\mathcal{E}_Q(|0\rangle\langle 1|) = \kappa|0\rangle\langle 1|$ , i.e., off-diagonal elements get scaled by a complex number  $\kappa$  with  $|\kappa| \leq 1$ . In (15),  $\mathcal{V}$  is evaluated for any pure state input  $\rho_Q^{(1)}$  from the  $XY$  plane of the Bloch sphere (e.g.,  $|+\rangle$ ). Now we apply (14) to this scenario and use (15) to obtain:

$$\mathcal{D}_i^2 + \mathcal{V}^2 \leq 1, \quad (16)$$

which extends a recent result in [13] to the case where  $F$  is non-trivial. This general treatment includes the special case where  $\rho_P^{(2)} = |V\rangle\langle V|$ , corresponding to a closed interferometer with an asymmetric BS<sub>2</sub>. Ref. [37] experimentally tested this special case. However, [37] did not

remark that their experiment actually tested a relation different from (1), namely they tested a special case of (16).

Similarly, [34] tested (16) rather than (1), but they allowed  $\rho_P^{(2)}$  to be in a superposition. At first sight this seems to test the WPDR in the case of a quantum beam splitter, but it turns out that neither the visibility  $\mathcal{V}$  nor the distinguishability  $\mathcal{D}_i$  depends on the phase coherence in  $\rho_P^{(2)}$  and hence the data could be simulated by a classical mixture of BS<sub>2</sub> being absent or present. Nevertheless, our framework provides a WPDR that captures the coherence in  $\rho_P^{(2)}$ , by conditioning on the polarisation  $P$  at the interferometer output (see Methods). For example, defining the polarization-enhanced distinguishability,  $\mathcal{D}_i^P := 2p_{\text{guess}}(Z_i|FP)_{D_0} - 1$ , which corresponds to choosing  $E_1 = FP$ , we obtain the novel WPDR:

$$(\mathcal{D}_i^P)^2 + \mathcal{V}^2 \leq 1, \quad (17)$$

which captures the beam splitter’s coherence (see [49] for elaboration) and could be tested with the setup in [34].

*Non-trivial E<sub>2</sub>.*—The above examples use the environment solely to enhance the particle behavior. To give a corresponding example for wave behavior, i.e., where system  $E_2$  in (5) is non-trivial, the main result of [11] is a WPDR for the case when the environment  $F$  is measured (after it has interacted with the quanton) and the resulting information is used to enhance the fringe visibility. This scenario is called *quantum erasure* since the goal is to erase the which-path information stored in the environment, to recover full visibility. This falls under our framework by taking  $E_2$  to be the classical output of the measurement on the environment. For elaboration see [49], where we also cast the main results of Refs. [10] and [12] within our framework.

*Conclusions.*—We have unified the wave-particle duality principle and the entropic uncertainty principle, showing that WPDRs are EURs in disguise. We leave it for future work to extend this connection to multiple interference pathways [6]. The framework presented here can be applied universally to binary interferometers. Our framework makes it clear how to formulate novel WPDRs by simply applying known EURs to novel interferometer models, and these new WPDRs will likely inspire new interferometry experiments. We note that all of our relations also hold if one replaces both min- and max-entropy with the well-known von Neumann entropy. Alternatively, one can use smooth entropies [29, 39], and the resulting smooth WPDRs may find application in the security analysis of interferometric quantum key distribution [43], which often exploits the Franson setup (Fig. 2).

## METHODS

We emphasize that our treatment, in what follows, will be for a generic binary interferometer. We will first discuss our general treatment, then we will specialize to the predictive and retrodictive scenarios (see Fig. 4).

*Origin of general WPDR.*—It is known that the min- and max-entropies satisfy the uncertainty relation [29]:

$$H_{\min}(Z|E_1) + H_{\max}(W|E_2) \geq 1, \quad (18)$$

for any tripartite state  $\rho_{AE_1E_2}$  where  $A$  is a qubit and  $Z$  and  $W$  are mutually unbiased bases on  $A$ . Noting that the which-path and which-path observables in (2) and (3) are mutually unbiased (for all  $\phi_0$  in (3), i.e., for all  $W$  in the  $XY$  plane) gives our general WPDR in (5).

*Complementary guessing game.*—The operational interpretation of (5) in terms of the complementary guessing game described, e.g., in Figs. 1-3 can be seen clearly as follows. While the min-entropy is related to the guessing probability via (6a), we establish a similar relation for the max-entropy. First we prove [49] that, for a general classical-quantum state  $\rho_{XB} = \sum_j |j\rangle\langle j| \otimes \sigma_B^j$  where  $X$  is binary,

$$H_{\max}(X|B) = \log \left( 1 + 2 \left\| \sqrt{\sigma_B^0} \sqrt{\sigma_B^1} \right\|_1 \right), \quad (19)$$

where the 1-norm is  $\|M\|_1 = \text{Tr} \sqrt{M^\dagger M}$ . Next we show [49], for any positive semi-definite operators  $M$  and  $N$ ,

$$\|M - N\|_1^2 + 4 \left\| \sqrt{M} \sqrt{N} \right\|_1^2 \leq (\text{Tr} M + \text{Tr} N)^2. \quad (20)$$

Combining (20) with (19), and using the well-known formula  $\|\sigma_B^0 - \sigma_B^1\|_1 = 2p_{\text{guess}}(X|B) - 1$ , gives

$$H_{\max}(X|B) \leq \log \left( 1 + \sqrt{1 - (2p_{\text{guess}}(X|B) - 1)^2} \right). \quad (21)$$

Now one can define generic measures of particle and wave behavior directly in terms of the guessing probabilities:

$$\mathcal{D}_g := 2p_{\text{guess}}(Z|E_1) - 1, \quad (22)$$

$$\mathcal{V}_g := \max_{W \in XY} [2p_{\text{guess}}(W|E_2) - 1] \quad (23)$$

for some arbitrary quantum systems  $E_1$  and  $E_2$ , and rearrange (5) into the traditional form for WPDRs:

$$\mathcal{D}_g^2 + \mathcal{V}_g^2 \leq 1. \quad (24)$$

This operationally-motivated relation, which follows directly from (5), clearly imposes a restriction on Alice's ability to win the complementary guessing game, since  $\mathcal{D}_g$  and  $\mathcal{V}_g$  are defined in terms of the winning probabilities. Below we show that  $\mathcal{V}_g$  becomes the fringe visibility when  $E_2$  is discarded.

*Predictive WPDRs.*—We now elaborate on our framework for deriving predictive WPDRs. Let us denote the quanton's spatial degree of freedom as  $S$ , which includes the previously mentioned  $Q$  as a subspace. At time  $t_2$  (see, e.g., Fig. 1) - the time just before a phase  $\phi$  is applied and the interferometer is closed -  $S$  and its environment  $E$  are in some state  $\rho_{SE}^{(2)}$ , where again  $E = E_1E_2$  is a generic bipartite system. The preparation is arbitrary, i.e., we need not specify what happened at earlier times, such as what the system's state was at time  $t_1$  (prior

to the interaction between  $S$  and  $E$ ). While in general a binary interferometer may have more than two paths, all but two of these are non-interfering (by definition), hence we only consider the two-dimensional subspace associated with the two which-path states of interest, denoted  $|0\rangle$  and  $|1\rangle$ . This subspace, defined by the projector  $\Pi := |0\rangle\langle 0| + |1\rangle\langle 1|$ , can be thought of as an effective qubit system  $Q$ . (Note that  $Q = S$  in the MZI.) Without loss of generality, we project the state  $\rho_{SE}^{(2)}$  onto this subspace and denote the resulting (renormalized) state as

$$\rho_{QE}^{(2)} = (\Pi \otimes \mathbb{1}) \rho_{SE}^{(2)} (\Pi \otimes \mathbb{1}) / \text{Tr}(\Pi \rho_S^{(2)}). \quad (25)$$

Experimentally this corresponds to post-selecting on the interfering portion of the data. To derive predictive WPDRs, we apply (5) to the state  $\rho_{QE}^{(2)}$  in (25), where we associate the subsystems  $E_1$  and  $E_2$  of  $E$  with the particle and wave terms respectively.

For example this approach gives the WPDRs discussed in [2], Eqs. (1) and (8). To show this we must prove (10), which relates our entropic measure of wave behavior in (4b) to fringe visibility, and we now do this for generic binary interferometers. We remark that one can take (7) as a generic definition for fringe visibility, where the label  $D_0$  is arbitrary, i.e., it corresponds to some arbitrary detector. For generic binary interferometers, there is a phase shift  $\phi$  applied just after time  $t_2$ , as depicted in Fig. 1. Let  $U_\phi = |0\rangle\langle 0| + e^{i\phi}|1\rangle\langle 1|$  denote the unitary associated with this phase shift, and note that we only need to specify the action of  $U_\phi$  on the  $Q$  subspace since the state  $\rho_{QE}^{(2)}$  lives in this subspace.

Finally the quanton is detected somewhere, i.e., system  $S$  is measured and a detector  $D_0$  clicks. This measurement is a positive operator valued measure (POVM)  $C = \{C_0, C_1, \dots\}$  on the larger space, system  $S$  rather than the subspace  $Q$  (e.g., think of the double slit case, where the detection screen performs a position measurement on  $S$ ). We associate the POVM element  $C_0$  with the event of detector  $D_0$  clicking. To prove (10), we need to restrict the form of  $C_0$ . We show that (10) holds so long as  $C_0$  is *unbiased* with respect to the which-path basis  $Z$  on the subspace  $Q$ . Fortunately this condition is satisfied for all three types of interferometers in Figs. 1, 2, and 3. More precisely, it is satisfied for the MZI provided  $\text{BS}_2$  is 50/50, for the Franson case provided both  $\text{BS}_2$  (the second beam splitters in Fig. 2) are 50/50, and for the double slit for some limiting choice of experimental parameters such as large  $L$  in Fig. 3. We now state a general lemma that applies to all of these interferometers.

**Lemma 1.** Consider a binary interferometer where  $\tilde{C}_0 := \Pi C_0 \Pi$  denotes the projection of POVM element  $C_0$  onto the interfering subspace ( $Q$ ). Suppose  $\tilde{C}_0$  is proportional to a projector projecting onto a state from the  $XY$  plane of the Bloch sphere of  $Q$ , i.e.,

$$\tilde{C}_0 = q |w_+\rangle\langle w_+| \quad (26)$$

for some  $0 < q \leq 1$ , where  $|w_+\rangle$  is given by (3) for some

arbitrary phase  $\phi_0$ . Then it follows that

$$\min_{W \in XY} H_{\max}(W) = \log(1 + \sqrt{1 - \mathcal{V}^2}), \quad (27)$$

where  $\mathcal{V}$  is given by (7), and  $H_{\max}(W)$  is evaluated for the state  $\rho_Q^{(2)} = \text{Tr}_E(\rho_{QE}^{(2)})$ .

*Proof.* In what follows it should be understood that probabilities and expectation values are evaluated for the state  $\rho_Q^{(2)}$ . Suppose that  $\tilde{W}$  is optimal in the sense that  $\max_{W \in XY} \Pr(w_+) = \Pr(\tilde{w}_+)$  where  $\Pr(w_{\pm}) := \langle w_{\pm} | \rho_Q^{(2)} | w_{\pm} \rangle$ . Then we have

$$\min_{W \in XY} H_{\max}(W) = \log \left( 1 + \sqrt{1 - \langle \sigma_{\tilde{W}} \rangle^2} \right) \quad (28)$$

where we denote Pauli operators by  $\sigma_W := |w_+\rangle\langle w_+| - |w_-\rangle\langle w_-|$ , and  $\langle \sigma_{\tilde{W}} \rangle = \Pr(\tilde{w}_+) - \Pr(\tilde{w}_-)$ .

The probability for  $D_0$  to click is

$$p^{D_0} = \text{Tr}(C_0 U_{\phi} \rho_Q^{(2)} U_{\phi}^{\dagger}) = \text{Tr}(U_{\phi}^{\dagger} \tilde{C}_0 U_{\phi} \rho_Q^{(2)}) \quad (29)$$

and maximising this over  $\phi$  gives

$$p_{\max}^{D_0} = q \max_{W \in XY} \Pr(w_+) = q \Pr(\tilde{w}_+). \quad (30)$$

Now, due to the geometry of the Bloch sphere, we have  $p_{\min}^{D_0} = \Pr(\tilde{w}_-)$ . Thus,  $p_{\max}^{D_0} + p_{\min}^{D_0} = q$  and  $p_{\max}^{D_0} - p_{\min}^{D_0} = q \langle \sigma_{\tilde{W}} \rangle$ . This gives  $\mathcal{V} = \langle \sigma_{\tilde{W}} \rangle$ , completing the proof.  $\square$

*Retrodictive WPDRs.*—While we saw that the predictive approach allowed for any preparation but required complementary output measurements, the opposite is true in the retrodictive case, i.e., the form of the output measurement is arbitrary while we require complementary preparations. The input ensembles  $Z_i = \{|0\rangle, |1\rangle\}$  and  $W_i = \{|w_+\rangle, |w_-\rangle\}$  can be generated by performing the relevant measurements on a reference qubit  $Q'$  that is initially entangled to the quanton  $S$ . Associating state ensembles with measurements on a reference system is a useful trick, e.g., for deriving (14). Thus, at time  $t_1$  (just after the quanton enters the interferometer, see Fig. 1) we introduce a qubit  $Q'$  that is maximally entangled to the interfering subspace ( $Q$ ) of  $S$ , denoted by the state  $\bar{\rho}_{Q'S}^{(1)} = |\Phi\rangle\langle\Phi|$  with  $|\Phi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ . The dynamics after time  $t_1$  is modelled as a quantum operation  $\mathcal{A}$ , defined in [44] as a completely positive, trace non-increasing map, that maps  $S \rightarrow E_1 E_2$ . The output of  $\mathcal{A}$  does not contain  $S$  because the quanton is eventually detected by a detector, at which point we no longer need a quantum description of the quanton's spatial degree of freedom; we only care where it was detected. The map  $\mathcal{A}$  corresponds to a particular detection event; for concreteness say that detector  $D_0$  clicking is the associated event. The probability for this event is the trace of the state after the action of  $\mathcal{A}$ , and renormalizing gives the final state

$$\bar{\rho}_{Q'E_1 E_2}^{D_0} := \frac{(\mathcal{I} \otimes \mathcal{A})(\bar{\rho}_{Q'S}^{(1)})}{\text{Tr}[(\mathcal{I} \otimes \mathcal{A})(\bar{\rho}_{Q'S}^{(1)})]}. \quad (31)$$

Our framework applies the uncertainty relation (5) to the state  $\bar{\rho}_{Q'E_1 E_2}^{D_0}$  to derive retrodictive WPDRs.

For example, this covers the scenario from the Discussion where  $\mathcal{A}$  involves two sequential steps. First  $S$  interacts with an environment  $F$  inside the interferometer between times  $t_1$  and  $t_2$ , which corresponds to a channel  $\mathcal{E}$  mapping  $S$  to  $SF$ . Second, the quanton is detected at the interferometer output, say at detector  $D_0$ , modelled as a map  $\mathcal{B}(\cdot) = \text{Tr}_S[C_0(\cdot)]$  acting on  $S$ , where  $C_0$  is the POVM element associated with detector  $D_0$  clicking. Hence we choose  $\mathcal{A} = \mathcal{B} \circ \mathcal{E}$ . Applying (5) to this case while choosing  $E_1 = F$  and  $E_2$  to be trivial gives

$$H_{\min}(Z|F)_{\bar{\rho}} + \min_{W \in XY} H_{\max}(W)_{\bar{\rho}} \geq 1, \quad (32)$$

where the subscript  $\bar{\rho}$  means evaluating on the state in (31). Note that measuring  $Z$  on system  $Q'$  corresponds to sending the states  $\{|0\rangle, |1\rangle\}$  with equal probability through the interferometer, and similarly for  $W$  (with an inconsequential complication of taking the transpose of the  $W$  basis states). Realizing this, the first and second terms in (32) map onto  $\mathcal{D}_i$  and  $\mathcal{V}_i$  respectively:

$$H_{\min}(Z|F)_{\bar{\rho}} = 1 - \log(1 + \mathcal{D}_i), \quad (33)$$

$$\min_{W \in XY} H_{\max}(W)_{\bar{\rho}} = \log(1 + \sqrt{1 - \mathcal{V}_i^2}). \quad (34)$$

Hence (32) becomes (14).

It remains to show that  $\mathcal{V}_i$  appearing in (14) can be replaced by  $\mathcal{V}$  for many cases of interest, such as the QBS case. We do this in the following lemma, where the proof is given in [49] and is similar to the proof of Lemma 1.

**Lemma 2.** Consider any binary interferometer with an unbiased input, i.e., where the state at time  $t_1$  is unbiased with respect to the which-path basis (of the form  $|\psi_Q^{(1)}\rangle = (|0\rangle + e^{i\phi}|1\rangle)/\sqrt{2}$ ). Let  $\mathcal{E}_S = \text{Tr}_F \circ \mathcal{E}$  be the channel describing the quanton's interaction with  $F$  inside the interferometer, and let  $\mathcal{G}(\cdot) = \Pi(\cdot)\Pi$  be the map that projects onto the subspace  $\Pi$ . Suppose  $\mathcal{E}_S$  is path-preserving, i.e.,  $\mathcal{E}_S(|0\rangle\langle 0|) = |0\rangle\langle 0|$  and  $\mathcal{E}_S(|1\rangle\langle 1|) = |1\rangle\langle 1|$  and furthermore suppose  $\mathcal{E}_S$  commutes with  $\mathcal{G}$ . Then

$$\min_{W \in XY} H_{\max}(W)_{\bar{\rho}} = \log(1 + \sqrt{1 - \mathcal{V}^2}), \quad (35)$$

where  $H_{\max}(W)$  is evaluated for the state  $\bar{\rho}_{Q'}^{D_0}$ .

*QBS example.*—Finally, we treat the quantum beam splitter shown in Fig. 5. (Note that  $S = Q$  in the MZI.) This setup involves first a quantum channel  $\mathcal{E}$  that describes the interaction of  $S$  with an environment  $F$  between times  $t_1$  and  $t_2$ , followed by another channel associated with the QBS that interacts  $S$  with the polarization  $P$ , followed by a post-selected detection at  $D_0$ . Together these three steps form a quantum operation  $\mathcal{A}$  that maps  $S \rightarrow FP$ , and hence this falls under our retrodictive framework.

To prove (17) we apply (5) to the state in (31) while choosing  $E_1 = FP$  and  $E_2$  to be trivial, giving

$$H_{\min}(Z|FP)_{\bar{\rho}} + \min_{W \in XY} H_{\max}(W)_{\bar{\rho}} \geq 1. \quad (36)$$

We then use relations analogous to those in (33) and (34), where the former relation now involves conditioning also on the polarisation  $P$ . Finally, we note that Lemma 2 applies to the QBS case.

### ACKNOWLEDGEMENTS

We thank B. Englert and S. Tanzilli for helpful correspondence, and acknowledge helpful discussions with M.

Woods, M. Tomamichel, C. J. Kwong, and L. C. Kwek. We acknowledge funding from the Ministry of Education (MOE) and National Research Foundation Singapore, as well as MOE Tier 3 Grant “Random numbers from quantum processes” (MOE2012-T3-1-009).

- 
- [1] R. P. Feynman, *Feynman Lectures on Physics* (Addison Wesley, Longman, 1970).
- [2] B.-G. Englert, Phys. Rev. Lett. **77**, 2154 (1996).
- [3] G. Jaeger, A. Shimony, and L. Vaidman, Phys. Rev. A **51**, 54 (1995).
- [4] W. K. Wootters and W. H. Zurek, Phys. Rev. D **19**, 473 (1979).
- [5] D. M. Greenberger and A. Yasin, Physics Letters A **128**, 391 (1988), ISSN 0375-9601.
- [6] B.-G. Englert, D. Kaszlikowski, L. C. Kwek, and W. H. Chee, International Journal of Quantum Information **06**, 129 (2008).
- [7] N.-L. Liu, L. Li, S. Yu, and Z.-B. Chen, Phys. Rev. A **79**, 052108 (2009).
- [8] J.-H. Huang, S. Wölk, S.-Y. Zhu, and M. S. Zubairy, Phys. Rev. A **87**, 022107 (2013).
- [9] T. Qureshi, Progress of Theoretical and Experimental Physics **2013** (2013).
- [10] L. Li, N.-L. Liu, and S. Yu, Phys. Rev. A **85**, 054101 (2012).
- [11] B.-G. Englert and J. A. Bergou, Optics Communications **179**, 337 (2000), ISSN 0030-4018.
- [12] K. Banaszek, P. Horodecki, M. Karpiński, and C. Radzewicz, Nat Commun **4** (2013).
- [13] A.-A. Jia, J.-H. Huang, W. Feng, T.-C. Zhang, and S.-Y. Zhu, Chinese Physics B **23**, 30307 (2014).
- [14] B.-G. Englert, M. O. Scully, and H. Walther, Nature **375**, 367 (1995).
- [15] P. Storey, S. Tan, M. Collett, and D. Walls, Nature **367**, 626 (1994).
- [16] H. Wiseman and F. Harrison, Nature **377**, 584 (1995).
- [17] N. Bohr, Nature **121**, 580 (1928).
- [18] W. Heisenberg, Zeitschrift für Physik **43**, 172 (1927).
- [19] E. Kennard, Z. Phys **44**, 326 (1927).
- [20] H. P. Robertson, Phys. Rev. **34**, 163 (1929).
- [21] D. Deutsch, Physical Review Letters **50**, 631 (1983).
- [22] I. Białynicki-Birula and J. Mycielski, Communications in Mathematical Physics **44**, 129 (1975).
- [23] H. Maassen and J. B. M. Uffink, Phys. Rev. Lett. **60**, 1103 (1988).
- [24] S. Wehner and A. Winter, New J. Phys. **12**, 025009 (2010).
- [25] J. M. Renes and J.-C. Boileau, Phys. Rev. Lett. **103**, 020402 (2009).
- [26] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, Nature Physics **6**, 659 (2010).
- [27] P. J. Coles, L. Yu, V. Gheorghiu, and R. B. Griffiths, Phys. Rev. A **83**, 062338 (2011).
- [28] P. J. Coles, R. Colbeck, L. Yu, and M. Zwolek, Phys. Rev. Lett. **108**, 210405 (2012).
- [29] M. Tomamichel and R. Renner, Phys. Rev. Lett. **106**, 110506 (2011).
- [30] S. Durr and G. Rempe, American Journal of Physics **68**, 1021 (2000).
- [31] P. Busch and C. Shilladay, Physics Reports **435**, 1 (2006), ISSN 0370-1573.
- [32] R. König, R. Renner, and C. Schaffner, IEEE Trans. Inf. Theory **55**, 4337 (2009).
- [33] R. Ionicioiu and D. R. Terno, Phys. Rev. Lett. **107**, 230406 (2011).
- [34] F. Kaiser, T. Coudreau, P. Milman, D. B. Ostrowsky, and S. Tanzilli, Science **338**, 637 (2012).
- [35] A. Peruzzo, P. Shadbolt, N. Brunner, S. Popescu, and J. L. O’Brien, Science **338**, 634 (2012).
- [36] J.-S. Tang, Y.-L. Li, C.-F. Li, and G.-C. Guo, Phys. Rev. A **88**, 014103 (2013).
- [37] V. Jacques, E. Wu, F. Grosshans, F. Treussart, P. Grangier, A. Aspect, and J.-F. Roch, Phys. Rev. Lett. **100**, 220402 (2008).
- [38] J. D. Franson, Phys. Rev. Lett. **62**, 2205 (1989).
- [39] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner, Nature Communications **3**, 634 (2012).
- [40] G. Björk, J. Söderholm, A. Trifonov, T. Tsegaye, and A. Karlsson, Phys. Rev. A **60**, 1874 (1999).
- [41] G. M. Bosyk, M. Portesi, F. Holik, and A. Plastino, Physica Scripta **87**, 065002 (2013).
- [42] F. Buscemi, M. J. W. Hall, M. Ozawa, and M. M. Wilde, Phys. Rev. Lett. **112**, 050401 (2014).
- [43] A. K. Ekert, J. G. Rarity, P. R. Tapster, and G. Massimo Palma, Phys. Rev. Lett. **69**, 1293 (1992).
- [44] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000), 5th ed.
- [45] C. W. Helstrom, *Quantum detection and estimation theory* (Academic Press, New York, USA, 1976), ISBN 0123400503.
- [46] C. King and M. Ruskai, IEEE Trans. Inf. Theory **47**, 192 (2001).
- [47] S. Tanzilli, private communication.
- [48] U. Herzog and J. A. Bergou, Phys. Rev. A **70**, 022302 (2004).
- [49] See the Supplementary Information.

# Supplementary Information

## Contents

<b>I. Introduction</b>	10
<b>II. Relating max-entropy to guessing probability</b>	10
<b>III. Hybrid of predictive and retrodictive scenarios</b>	13
A. Introduction	13
B. General treatment of hybrid scenario	13
C. Proof of Lemma S7 (generalized version of Lemma 2)	16
<b>IV. Enhanced visibility</b>	17
A. Quantum erasure	18
1. Results of Ref. [11]	18
2. Our treatment	18
B. Polarization-enhanced visibility and discussion of Ref. [12]	19
1. Result of Ref. [12]	19
2. Our treatment	20
3. Proof of Lemma S9	21
<b>V. Testing coherence in a quantum beam splitter</b>	22
A. Quantities sensitive to coherence	22
B. Discussion of Ref. [34]	23
C. Derivation of distinguishability formulas	25
D. Measuring $\mathcal{D}_i^P$	25

## I. INTRODUCTION

In this Supplementary Information, we elaborate on the technical details justifying our claims. Furthermore, to emphasize the universality of our framework, we provide additional results showing that other WPDRs appearing in the literature can be phrased within our framework.

In what follows, Sec. II proves the relation between the max-entropy and the guessing probability given in the Methods section. In Sec. III we extend our framework to scenarios where both the preparation (at the interferometer input) and the measurement outcome (at the interferometer output) may provide information about the quanton’s path (inside the interferometer). Such situations are a “hybrid” of the predictive and retrodictive scenarios discussed in the main text. This extension allows us to reinterpret a result in Ref. [10] for asymmetric beam splitters as an entropic uncertainty relation and hence show that it falls under our framework. Subsection IIIC proves Lemma 2 from the main text, or in fact, proves a generalized version of this lemma that holds for these “hybrid” scenarios. In Sec. IV we consider WPDRs involving enhanced visibility. In particular we show that the results of Ref. [11] for quantum erasure and Ref. [12] for polarization dynamics can both be viewed as entropic uncertainty relations, where the visibility term is enhanced by conditioning on additional information. Finally in Sec. V we elaborate on our novel WPDR for the quantum beam splitter, demonstrating that it captures the coherence in the QBS and discussing how the polarization-enhanced distinguishability can be measured.

## II. RELATING MAX-ENTROPY TO GUESSING PROBABILITY

Here we relate the max-entropy to the guessing probability. The main lemma that we prove explicitly solves for the max-entropy of a classical-quantum (cq) state where the classical register is binary, which to our knowledge is a new result. An arbitrary cq state where the classical register is binary, which is the only case relevant to our analysis, can be written as  $\rho_{XB} = |0\rangle\langle 0| \otimes \sigma_0 + |1\rangle\langle 1| \otimes \sigma_1$ , where  $\sigma_0$  and  $\sigma_1$  are subnormalized states that satisfy  $\text{Tr}\sigma_0 + \text{Tr}\sigma_1 = 1$ . In this case, the optimal guessing probability takes the form

$$p_{\text{guess}}(X|B) := \max_{M_0, M_1} \text{Tr}(M_0\sigma_0) + \text{Tr}(M_1\sigma_1), \quad (\text{S37})$$

where the maximization is taken over all POVMs on subsystem  $B$ , namely operators  $M_0, M_1 \geq 0$  such that  $M_0 + M_1 = \mathbf{1}$ . Since this is exactly the state discrimination problem solved by Helstrom [45] we have

$$p_{\text{guess}}(X|B) = \frac{1}{2} + \frac{1}{2}\|\sigma_0 - \sigma_1\|_1. \quad (\text{S38})$$

Hence the guessing probability is related to the trace distance between the conditional states.

The formula for the max-entropy was given in Eq. (6b) from the main text and was expressed in terms of the fidelity, which is defined as

$$F(M, N) := \|\sqrt{M}\sqrt{N}\|_1, \quad (\text{S39})$$

for two positive semi-definite operators  $M$  and  $N$ . In our case of a cq state with a binary register, the formula given in Eq. (6b) simplifies to [32]

$$H_{\text{max}}(X|B) = 2 \log \max_{\rho} (F(\sigma_0, \rho) + F(\sigma_1, \rho)), \quad (\text{S40})$$

where the maximization is taken over all normalized states on  $B$ . The following lemma derives the optimal value of this optimization problem.

**Lemma S3.** Let  $M, N \geq 0$  be positive semi-definite operators and let  $\mathcal{S}$  be the set of positive semi-definite operators with unit trace. Then

$$\max_{\rho \in \mathcal{S}} (F(M, \rho) + F(N, \rho)) = \sqrt{\text{Tr}M + \text{Tr}N + 2F(M, N)}. \quad (\text{S41})$$

*Proof.* First, we show that the right-hand side constitutes a valid upper bound and then we give an explicit choice of  $\rho$  that achieves it.

For arbitrary unitaries  $U_0$  and  $U_1$  let  $X^\dagger = U_0\sqrt{M} + U_1\sqrt{N}$  and  $Y = \sqrt{\rho}$ . The Cauchy-Schwarz inequality,  $|\text{Tr}(X^\dagger Y)|^2 \leq \text{Tr}(X^\dagger X) \cdot \text{Tr}(Y^\dagger Y)$ , implies that

$$\left| \text{Tr} \left( (U_0\sqrt{M} + U_1\sqrt{N})\sqrt{\rho} \right) \right|^2 \leq \text{Tr}(X^\dagger X) \cdot \text{Tr}\rho = \text{Tr}(X^\dagger X). \quad (\text{S42})$$

Since

$$X^\dagger X = U_0 M U_0^\dagger + U_1 N U_1^\dagger + U_0 \sqrt{M} \sqrt{N} U_1^\dagger + U_1 \sqrt{N} \sqrt{M} U_0^\dagger \quad (\text{S43})$$

we have

$$\text{Tr}(X^\dagger X) = \text{Tr}M + \text{Tr}N + \text{Tr}(U_0\sqrt{M}\sqrt{N}U_1^\dagger + U_1\sqrt{N}\sqrt{M}U_0^\dagger) \quad (\text{S44})$$

$$\leq \text{Tr}M + \text{Tr}N + \|U_0\sqrt{M}\sqrt{N}U_1^\dagger + U_1\sqrt{N}\sqrt{M}U_0^\dagger\|_1 \quad (\text{S45})$$

$$\leq \text{Tr}M + \text{Tr}N + 2\|\sqrt{M}\sqrt{N}\|_1 = \text{Tr}M + \text{Tr}N + 2F(M, N), \quad (\text{S46})$$

where we have used the fact that for Hermitian matrices  $\text{Tr} T \leq \|T\|_1$  followed by the triangle inequality for the 1-norm. Note that this bound is valid for all unitaries  $U_0$  and  $U_1$ .

Let  $L$  be a linear operator and let  $L = U_0 S U_1$  be its singular value decomposition. Clearly,  $\|L\|_1 = \text{Tr}(VL)$  for  $V = U_1^\dagger U_0^\dagger$ . Therefore, for every pair of positive semi-definite operators  $A$  and  $B$  there exists a unitary  $V$  such that  $F(A, B) = \|\sqrt{A}\sqrt{B}\|_1 = \text{Tr}(V\sqrt{A}\sqrt{B})$ . Let us choose unitaries  $V_0$  and  $V_1$  such that

$$F(M, \rho) = \text{Tr}(V_0\sqrt{M}\sqrt{\rho}) \quad \text{and} \quad F(N, \rho) = \text{Tr}(V_1\sqrt{N}\sqrt{\rho}). \quad (\text{S47})$$

Adding these two terms together gives

$$F(M, \rho) + F(N, \rho) = \text{Tr}(V_0\sqrt{M}\sqrt{\rho}) + \text{Tr}(V_1\sqrt{N}\sqrt{\rho}) = \text{Tr} \left( (V_0\sqrt{M} + V_1\sqrt{N})\sqrt{\rho} \right). \quad (\text{S48})$$

Since for this particular choice of unitaries the quantity on the right-hand side is real and positive we can apply (S42) to obtain

$$F(M, \rho) + F(N, \rho) = \left| \text{Tr} \left( (V_0\sqrt{M} + V_1\sqrt{N})\sqrt{\rho} \right) \right| \leq \sqrt{\text{Tr}M + \text{Tr}N + 2F(M, N)}. \quad (\text{S49})$$

Now, we simply need to provide a state  $\rho$  that saturates this inequality. Taking advantage of the singular value decomposition of  $\sqrt{M}\sqrt{N} = U_0 S U_1$  (where  $S$  is a diagonal matrix of real, non-negative numbers and  $\text{Tr} S = \|\sqrt{M}\sqrt{N}\|_1$ ) we define  $V = U_1^\dagger U_0^\dagger$  and

$$K = M + N + \sqrt{M}V^\dagger\sqrt{N} + \sqrt{N}V\sqrt{M}. \quad (\text{S50})$$

Note that  $K \geq 0$  since  $K = L^\dagger L$  for  $L = \sqrt{M} + V^\dagger\sqrt{N}$ . It is easy to verify that

$$\text{Tr}(\sqrt{M}V^\dagger\sqrt{N}) = \text{Tr}(\sqrt{N}V\sqrt{M}) = \text{Tr} S = \|\sqrt{M}\sqrt{N}\|_1 = F(M, N), \quad (\text{S51})$$

which implies that

$$\text{Tr} K = \text{Tr} M + \text{Tr} N + 2F(M, N) \quad (\text{S52})$$

To calculate  $F(M, K) = \|\sqrt{M}\sqrt{K}\|_1 = \text{Tr}\sqrt{\sqrt{M}K\sqrt{M}}$  note that

$$\sqrt{M}K\sqrt{M} = M^2 + \sqrt{M}N\sqrt{M} + MV^\dagger\sqrt{N}\sqrt{M} + \sqrt{M}\sqrt{N}VM \quad (\text{S53})$$

$$= M^2 + U_0 S^2 U_0^\dagger + MU_0 S U_0^\dagger + U_0 S U_0^\dagger M \quad (\text{S54})$$

$$= (M + U_0 S U_0^\dagger)^2. \quad (\text{S55})$$

Therefore,  $F(M, K) = \text{Tr} M + F(M, N)$  and similarly  $F(N, K) = \text{Tr} N + F(M, N)$ . Since

$$F(\alpha A, B) = \|\sqrt{\alpha A}\sqrt{B}\|_1 = \sqrt{\alpha}\|\sqrt{A}\sqrt{B}\|_1 = \sqrt{\alpha}F(A, B) \quad (\text{S56})$$

we can define  $\rho = K/\text{Tr} K$  which satisfies

$$F(M, \rho) + F(N, \rho) = \frac{F(M, K) + F(N, K)}{\sqrt{\text{Tr} K}} = \sqrt{\text{Tr} M + \text{Tr} N + 2F(M, N)} \quad (\text{S57})$$

and saturates the bound (S49).  $\square$

Now taking the above lemma and setting  $M = \sigma_0$  and  $N = \sigma_1$  allows us to solve the maximization in (S40). We obtain the following result.

**Lemma S4.** For any cq state  $\rho_{XB} = |0\rangle\langle 0| \otimes \sigma_0 + |1\rangle\langle 1| \otimes \sigma_1$  where  $X$  is binary,

$$H_{\max}(X|B) = \log(1 + 2F(\sigma_0, \sigma_1)). \quad (\text{S58})$$

Finally we relate the fidelity in (S58) to the trace distance in the guessing probability (S38) with the following lemma.

**Lemma S5.** Let  $M, N \geq 0$  be two positive semi-definite operators. Then we have

$$\|M - N\|_1^2 + 4\|\sqrt{M}\sqrt{N}\|_1^2 \leq (\text{Tr} M + \text{Tr} N)^2. \quad (\text{S59})$$

*Proof.* Let  $|\Omega\rangle = \sum_k |k\rangle|k\rangle$  and consider  $|\psi_M\rangle = (\sqrt{M} \otimes \mathbb{1})|\Omega\rangle$  and  $|\psi_N\rangle = (\sqrt{N} \otimes U)|\Omega\rangle$ , where  $U$  is a unitary. It is easy to verify that

$$\langle\psi_M|\psi_N\rangle = \langle\Omega|\sqrt{M}\sqrt{N} \otimes U|\Omega\rangle = \text{Tr}(\sqrt{M}\sqrt{N}U^T), \quad (\text{S60})$$

where  $T$  denotes the transpose in the standard basis. By choosing  $U^T = U_2^\dagger U_1^\dagger$ , where  $U_1$  and  $U_2$  come from the singular value decomposition of  $\sqrt{M}\sqrt{N} = U_1 S U_2$  we obtain  $\langle\psi_M|\psi_N\rangle = \text{Tr} S = \|\sqrt{M}\sqrt{N}\|_1$ . Let  $\text{Tr}_2$  denote partial trace over the second subsystem. It is easy to check that  $\text{Tr}_2|\psi_M\rangle\langle\psi_M| = M$  and  $\text{Tr}_2|\psi_N\rangle\langle\psi_N| = N$ . Since the trace norm is non-increasing under the partial trace we have

$$\|M - N\|_1 \leq \left\| |\psi_M\rangle\langle\psi_M| - |\psi_N\rangle\langle\psi_N| \right\|_1. \quad (\text{S61})$$

The rank of the Hermitian matrix  $H = |\psi_M\rangle\langle\psi_M| - |\psi_N\rangle\langle\psi_N|$  is at most 2 and let us denote the non-zero eigenvalues by  $\lambda_1$  and  $\lambda_2$ . It is easy to verify that

$$\lambda_1 + \lambda_2 = \text{Tr} H = \langle\psi_M|\psi_M\rangle - \langle\psi_N|\psi_N\rangle, \quad (\text{S62})$$

$$\lambda_1^2 + \lambda_2^2 = \text{Tr} H^2 = (\langle\psi_M|\psi_M\rangle)^2 + (\langle\psi_N|\psi_N\rangle)^2 - 2|\langle\psi_M|\psi_N\rangle|^2. \quad (\text{S63})$$

Since  $\lambda_1\lambda_2 = \frac{1}{2}[(\lambda_1 + \lambda_2)^2 - (\lambda_1^2 + \lambda_2^2)] = |\langle\psi_M|\psi_N\rangle|^2 - \langle\psi_M|\psi_M\rangle\langle\psi_N|\psi_N\rangle$ , the Cauchy-Schwarz inequality ensures that  $\lambda_1\lambda_2 \leq 0$ . As  $H$  is Hermitian, we have  $\|H\|_1 = |\lambda_1| + |\lambda_2|$  and since the eigenvalues have opposite signs we can write it as

$$\|H\|_1 = |\lambda_1 - \lambda_2| = \sqrt{(\lambda_1 - \lambda_2)^2}. \quad (\text{S64})$$

Expanding the square gives

$$(\lambda_1 - \lambda_2)^2 = \lambda_1^2 + \lambda_2^2 - 2\lambda_1\lambda_2 \quad (\text{S65})$$

$$= (\langle\psi_M|\psi_M\rangle)^2 + (\langle\psi_N|\psi_N\rangle)^2 - 2|\langle\psi_M|\psi_N\rangle|^2 + 2\langle\psi_M|\psi_M\rangle\langle\psi_N|\psi_N\rangle - 2|\langle\psi_M|\psi_N\rangle|^2 \quad (\text{S66})$$

$$= (\langle\psi_M|\psi_M\rangle + \langle\psi_N|\psi_N\rangle)^2 - 4|\langle\psi_M|\psi_N\rangle|^2, \quad (\text{S67})$$

which combined with (S64) implies

$$\left\| |\psi_M\rangle\langle\psi_M| - |\psi_N\rangle\langle\psi_N| \right\|_1 = \sqrt{(\langle\psi_M|\psi_M\rangle + \langle\psi_N|\psi_N\rangle)^2 - 4|\langle\psi_M|\psi_N\rangle|^2} \quad (\text{S68})$$

$$= \sqrt{(\text{Tr}M + \text{Tr}N)^2 - 4\|\sqrt{M}\sqrt{N}\|_1^2}. \quad (\text{S69})$$

□

Combining Lemmas S4 and S5 gives the following result.

**Lemma S6.** For any cq state  $\rho_{XB}$  where  $X$  is binary,

$$H_{\max}(X|B) \leq \log \left( 1 + \sqrt{1 - [2p_{\text{guess}}(X|B) - 1]^2} \right). \quad (\text{S70})$$

*Proof.* Apply Lemma S5 to the subnormalized states  $\sigma_0$  and  $\sigma_1$  appearing in Lemma S4 to give

$$\|\sigma_0 - \sigma_1\|_1^2 + 4\|\sqrt{\sigma_0}\sqrt{\sigma_1}\|_1^2 \leq 1. \quad (\text{S71})$$

Since  $\|\sqrt{\sigma_0}\sqrt{\sigma_1}\|_1 = F(\sigma_0, \sigma_1)$  combining this inequality with (S58) and (S38) gives the desired inequality. □

### III. HYBRID OF PREDICTIVE AND RETRODICTIVE SCENARIOS

#### A. Introduction

In the main text we discussed how to derive WPDRs from preparation and measurement uncertainty relations, which respectively deal with predicting the future and retrodicting the past. In this section we show that our framework can also be applied to scenarios that involves a hybrid (or mixture) of prediction and retrodiction.

Let us first mention a motivating example from the literature for when this hybrid situation can arise. Ref. [10] considered a simple yet insightful scenario involving a MZI where both beamsplitters  $\text{BS}_1$  and  $\text{BS}_2$  (see, e.g., Fig. S6) may be asymmetric. Since  $\text{BS}_1$  is asymmetric, the experimenter has prior knowledge about which path the photon will take. Since  $\text{BS}_2$  is asymmetric, the experimenter can use the final measurement outcome of which detector clicked to help retrodict which path the photon took. Ref. [10] formulated a WPDR for this scenario, and by the end of this section it will be clear that this falls under our framework. But let us develop the general idea first.

#### B. General treatment of hybrid scenario

To treat the hybrid case we will consider our retrodictive framework (discussed in the Methods) and add in the possibility of pre-experiment information about which path the quanton will take. Recall that, in the retrodictive case, we introduced a qubit register  $Q'$  that is maximally entangled to the quanton  $S$  at time  $t_1$ , the time just after the quanton enters the interferometer. More precisely  $Q'$  is maximally entangled to the interfering subspace  $Q$  of  $S$ . The purpose of  $Q'$  is to store a record of the quanton's properties at time  $t_1$ , so that when the quanton evolves and changes over time, we can still go back to  $Q'$  to ask about the quanton's properties at the earlier time.

The fact that we chose a maximally entangled state is connected to the fact that there is no prior knowledge about the path the quanton will take. But now we are relaxing that assumption, so we will consider a *partially* entangled

state. It is useful to think of this partially entangled state as arising from taking the physical state  $\rho_S^{(1)}$  at time  $t_1$ , and then applying an isometry  $V_c$  that copies the which-path information and stores it in  $Q'$ . This isometry expands the Hilbert space, mapping  $S$  to  $Q'S$  as follows:

$$V_c = \sum_{j=0}^1 |j\rangle_{Q'} \otimes |j\rangle_S, \quad \bar{\rho}_{Q'S}^{(1)} := V_c \rho_Q^{(1)} V_c^\dagger. \quad (\text{S72})$$

There is a minor technical detail in (S72) that is irrelevant to the MZI but becomes relevant, e.g., in the Franson interferometer. Namely, in (S72), instead of using the initially-prepared state  $\rho_S^{(1)}$ , which may have support outside of the interfering subspace  $Q$ , we use the projected and renormalized state  $\rho_Q^{(1)}$ , defined as

$$\rho_Q^{(1)} := N_1 \cdot (\Pi \rho_S^{(1)} \Pi), \quad \text{with } N_1 := 1/\text{Tr}(\Pi \rho_S^{(1)}), \quad \text{and } \Pi = |0\rangle\langle 0| + |1\rangle\langle 1|. \quad (\text{S73})$$

As discussed in the Methods section, the physical motivation behind this projection is that it corresponds to the experimenter post-selecting on the interfering portion of the data. Note that if  $\rho_Q^{(1)}$  is a pure state from the  $XY$  plane of the Bloch sphere (i.e., of the form  $(|0\rangle + e^{i\phi}|1\rangle)/\sqrt{2}$ ), then  $\bar{\rho}_{Q'S}^{(1)}$  is maximally entangled and then we are just back to the retrodictive case discussed in the Methods. The generality in the present treatment comes from the fact that  $\rho_Q^{(1)}$  is arbitrary.

As in the Methods, we treat the dynamics after time  $t_1$  very generally by saying that some *quantum operation* [44] (completely positive trace non-increasing map) denoted  $\mathcal{A}$  acts on system  $S$ , mapping  $S$  to the joint system  $E_1 E_2$ . The output system does not contain  $S$  because the quanton is eventually detected by a detector, at which point we are no longer interested in discussing the quanton's spatial degree of freedom quantum mechanically; we only care where it was detected. The map  $\mathcal{A}$  corresponds to a particular detection event; for concreteness let us say that detector  $D_0$  clicking is the associated detection event. The probability for this event is the trace of the state after the action of  $\mathcal{A}$ , and upon renormalizing we arrive at the final state

$$\bar{\rho}_{Q'E_1 E_2}^{D_0} := \frac{(\mathcal{I} \otimes \mathcal{A})(\bar{\rho}_{Q'S}^{(1)})}{\text{Tr}[(\mathcal{I} \otimes \mathcal{A})(\bar{\rho}_{Q'S}^{(1)})]}. \quad (\text{S74})$$

To derive WPDRs for the hybrid scenario, we apply the our main uncertainty relation, Eq. (5) from the main text, to the state  $\bar{\rho}_{Q'E_1 E_2}^{D_0}$ .

Consider the following important special case, where the map  $\mathcal{A}$  involves two sequential steps. First there is interaction between  $S$  and an environment  $F$  inside the interferometer between times  $t_1$  and  $t_2$ , which corresponds to feeding  $S$  through a channel  $\mathcal{E}$ , mapping  $S$  to  $SF$ , obtaining the state  $\bar{\rho}_{Q'SF}^{(2)} = (\mathcal{I} \otimes \mathcal{E})(\bar{\rho}_{Q'S}^{(1)})$ . Second, system  $S$  is detected at the interferometer output, say at detector  $D_0$ , which we can model as a map  $\mathcal{B}(\cdot) = \text{Tr}_S[C_0(\cdot)]$ , where  $C_0$  is the POVM element (acting on  $S$  at time  $t_2$ ) associated with detector  $D_0$  clicking. Hence we choose  $\mathcal{A} = \mathcal{B} \circ \mathcal{E}$ , giving the (renormalized) state:

$$\bar{\rho}_{Q'F}^{D_0} := \frac{\text{Tr}_S(C_0 \bar{\rho}_{Q'SF}^{(2)})}{\text{Tr}(C_0 \bar{\rho}_{Q'SF}^{(2)})}. \quad (\text{S75})$$

Now applying our main uncertainty relation to the state  $\bar{\rho}_{Q'F}^{D_0}$  and choosing  $E_2$  to be trivial and  $E_1 = F$  gives

$$H_{\min}(Z_{Q'}|F)_{\bar{\rho}^{D_0}} + \min_{W \in XY} H_{\max}(W_{Q'})_{\bar{\rho}^{D_0}} \geq 1, \quad (\text{S76})$$

where the subscript  $\bar{\rho}^{D_0}$  means evaluating the entropy on the state in (S75) and the subscript  $Q'$  is used to emphasize that the observables  $Z$  and  $W$  refer to system  $Q'$ .

At this point we must remark on the physical meaning of a relation such as (S76). In the special where the state  $\bar{\rho}_{Q'S}^{(1)}$  was maximally entangled, as considered in the main text, we noted that (S76) can be interpreted as a joint measurement relation. This is because a maximally entangled state is special in that it maps observables on register  $Q'$  to the transpose observables on the system of interest. However this interpretation is lost once we relax the form of  $\bar{\rho}_{Q'S}^{(1)}$ , so we can no longer interpret (S76) as a joint measurement relation, in the general case. Rather, one can think of (S76) as a hybrid between a preparation and measurement uncertainty relation.

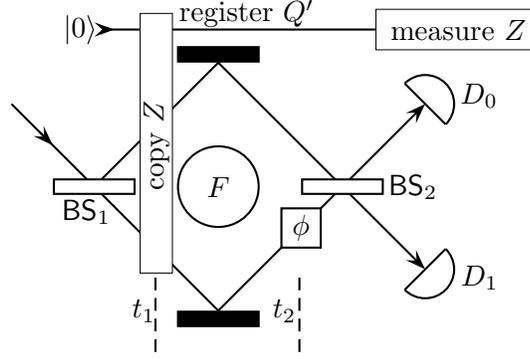


FIG. S6: In the case, e.g., where both  $\text{BS}_1$  and  $\text{BS}_2$  are asymmetric beam splitters, the distinguishability (denoted  $\mathcal{D}_{Q'}$  in the text) can be measured as follows. The experimenter can copy the which-path information at time  $t_1$  to a register  $Q'$  (e.g.,  $Q'$  may be the photon's polarisation) via a controlled-not gate. Then the experimenter can measure the observable on  $Q'$  that encodes the which-path information and estimate the probability of correctly guessing this measurement's outcome.

Regardless, (S76) can be written in the traditional WPDR form. First we note that the isometry  $V_c$  in (S72) maps the which-path observable  $Z$  of  $Q$  onto the corresponding  $Z$  observable of  $Q'$ . So the observable  $Z_{Q'}$  is effectively the which-path observable, and we can define path distinguishability by

$$\mathcal{D}_{Q'} := 2p_{\text{guess}}(Z_{Q'}|F)_{\bar{p}^{D_0}} - 1, \quad \text{or in other words, } H_{\min}(Z_{Q'}|F)_{\bar{p}^{D_0}} = 1 - \log(1 + \mathcal{D}_{Q'}). \quad (\text{S77})$$

Note that  $\mathcal{D}_{Q'}$  is a generalization of the input distinguishability  $\mathcal{D}_i$  defined in the main text. Experimentally measuring  $\mathcal{D}_{Q'}$  involves the procedure shown in Fig. S6 for the special case of the MZI. That is, the experimenter perfectly copies the which-path information to a register and then measures the register. Note that a register is necessary in this case since other procedures to measure distinguishability discussed in the main text - such as removing the second beam splitter or randomly inserting a blocker in one arm - would not capture the inherent which-path asymmetry associated with both beam splitters.

Moving onto the which-phase observable  $W$ , we prove a general and powerful lemma that the second term in (S76) corresponds precisely to the fringe visibility  $\mathcal{V}$ . Let us remind the reader that the latter is given by

$$\mathcal{V} = \frac{p_{\max}^{D_0} - p_{\min}^{D_0}}{p_{\max}^{D_0} + p_{\min}^{D_0}} \quad (\text{S78})$$

where  $p^{D_0}$  is the probability for the quanton to be detected at detector  $D_0$ ,  $p_{\max}^{D_0} = \max_{\phi} p^{D_0}$ , and  $p_{\min}^{D_0} = \min_{\phi} p^{D_0}$ . To establish this connection, we need to restrict the form of the channel  $\mathcal{E}_S = \text{Tr}_F \circ \mathcal{E}$  acting on the quanton such that  $\mathcal{E}_S$  is path preserving. We emphasize that the following lemma is a generalization of Lemma 2 from the main text and hence its proof implies Lemma 2. See the next subsection for the proof.

**Lemma S7.** Consider any binary interferometer. Let  $\mathcal{E}_S = \text{Tr}_F \circ \mathcal{E}$  denote the quantum channel describing the quanton's interaction with the environment  $F$  inside the interferometer between times  $t_1$  and  $t_2$ , and let  $\mathcal{G}(\cdot) = \Pi(\cdot)\Pi$  denote the map that projects onto the subspace  $\Pi = |0\rangle\langle 0| + |1\rangle\langle 1|$ . Suppose that  $\mathcal{E}_S$  is path-preserving, i.e.,  $\mathcal{E}_S(|0\rangle\langle 0|) = |0\rangle\langle 0|$  and  $\mathcal{E}_S(|1\rangle\langle 1|) = |1\rangle\langle 1|$  and furthermore suppose that  $\mathcal{E}_S$  commutes with  $\mathcal{G}$ , i.e.,  $\mathcal{G} \circ \mathcal{E}_S = \mathcal{E}_S \circ \mathcal{G}$ . Then

$$\min_{W \in XY} H_{\max}(W_{Q'})_{\bar{p}^{D_0}} = \log(1 + \sqrt{1 - \mathcal{V}^2}), \quad (\text{S79})$$

where  $H_{\max}(W_{Q'})$  is evaluated for the state  $\bar{p}_{Q'}^{D_0} = \text{Tr}_F(\bar{p}_{Q',F}^{D_0})$  from (S75).

Now making the assumption in Lemma S7 regarding the form of the interaction  $\mathcal{E}_S$  allows us to rewrite (S76), using (S77) and (S79), as follows:

$$\mathcal{D}_{Q'}^2 + \mathcal{V}^2 \leq 1. \quad (\text{S80})$$

This WPDR holds for generic binary interferometers.

Ref. [10] derived a WPDR similar to (S80) but for the special case of a MZI where both  $\text{BS}_1$  and  $\text{BS}_2$  are possibly asymmetric. They restricted to interactions with the environment  $F$  that had the form of a controlled-unitary, which are path-preserving and hence are included in the class of dynamics that we assumed to derive (S80). Of course (S80)

applies very generally to binary interferometers, but it can be applied to the MZI (note that  $Q = S$  in the MZI case) where both  $\text{BS}_1$  and  $\text{BS}_2$  are asymmetric. Thus, we find that the result in Ref. [10] can be understood as an entropic uncertainty relation, namely a special case of (S76) [from which we derived (S80)]. To be more precise, Ref. [10] also generalized their relation to allow for non-optimal strategies for measuring  $F$ ; we do not treat this generalization here. We believe the most important conceptual advance of Ref. [10] was to prove a WPDR that applies to a scenario that - in our language - is a hybrid of preparation and measurement uncertainty. What we have emphasized in this section is that our framework naturally extends to this hybrid scenario.

### C. Proof of Lemma S7 (generalized version of Lemma 2)

To prove Lemma S7 we will make use of the following lemma.

**Lemma S8.** Let  $\mathcal{Q}$  be a qubit quantum channel, i.e., whose input and output are operators on a 2-dimensional Hilbert space, and suppose that  $\mathcal{Q}(|j\rangle\langle j|) = |j\rangle\langle j|$  for  $j = 0, 1$ . Likewise let  $\mathcal{R}_\phi$  be a qubit quantum channel whose action is given by

$$\mathcal{R}_\phi(\cdot) = U_\phi(\cdot)U_\phi^\dagger, \quad \text{with } U_\phi = |0\rangle\langle 0| + e^{i\phi}|1\rangle\langle 1|. \quad (\text{S81})$$

Then, for any  $\phi$ ,  $\mathcal{Q}$  and  $\mathcal{R}_\phi$  commute, i.e.,  $\mathcal{R}_\phi \circ \mathcal{Q} = \mathcal{Q} \circ \mathcal{R}_\phi$ .

*Proof.* Since  $\mathcal{Q}$  is unital and furthermore preserves the  $Z$ -basis, its action on the Bloch sphere can only involve a rotation  $\mathcal{R}_\theta$  about the  $Z$ -axis composed with a shrinking  $\mathcal{S}$  of the Bloch sphere, and this shrinking must be cylindrically symmetric about the  $Z$ -axis (see, e.g., Ref. [46]). The rotation  $\mathcal{R}_\theta$  obviously commutes with the rotation  $\mathcal{R}_\phi$ , and likewise  $\mathcal{R}_\phi$  commutes with the shrinking  $\mathcal{S}$  due to the cylindrical symmetry of  $\mathcal{S}$ .  $\square$

Now we prove Lemma S7, which relates the fringe visibility to the max-entropy of the which-phase observable in our “hybrid” framework. Lemma S7 generalizes Lemma 2 from the main text, which is the corresponding result for our retrodictive framework.

*Proof.* In the formula for  $\mathcal{V}$  in (S78), the notation

$$p^{D_0} = \text{Tr}[C_0 \mathcal{R}_\phi(\rho_Q^{(2)})] \quad (\text{S82})$$

refers to the probability for detector  $D_0$  to click when the quanton’s state at time  $t_2$  (the time just before the phase-shift  $\phi$  is applied) is  $\rho_Q^{(2)}$ . This state was defined in Eq. (25) of the Methods section as

$$\rho_Q^{(2)} = N_2 \cdot (\Pi \rho_S^{(2)} \Pi), \quad \text{with } N_2 := 1/\text{Tr}(\Pi \rho_S^{(2)}), \quad (\text{S83})$$

which notes that the experimenter post-selects on the interfering subspace, associated with projector  $\Pi$ . Defining  $\tilde{C}_0 := \Pi C_0 \Pi$  and taking the maximum of (S82) over  $\phi$  gives

$$p_{\max}^{D_0} = N_2 \max_{\phi} \text{Tr}[\tilde{C}_0 \mathcal{R}_\phi(\Pi \rho_S^{(2)} \Pi)] \quad (\text{S84})$$

$$= N_2 \max_{\phi} \text{Tr}[\tilde{C}_0 \mathcal{R}_\phi(\Pi \mathcal{E}_S(\rho_S^{(1)}) \Pi)] \quad (\text{S85})$$

$$= N_2 \max_{\phi} \text{Tr}[\tilde{C}_0 \mathcal{R}_\phi(\mathcal{E}_S(\Pi \rho_S^{(1)} \Pi))] \quad (\text{S86})$$

$$= (N_2/N_1) \max_{\phi} \text{Tr}[\tilde{C}_0 \mathcal{R}_\phi(\mathcal{E}_S(\rho_Q^{(1)}))] \quad (\text{S87})$$

$$= (N_2/N_1) f(\hat{\phi}), \quad \text{where } f(\phi) := \text{Tr}[\tilde{C}_0 \mathcal{R}_\phi(\mathcal{E}_S(\rho_Q^{(1)}))], \quad (\text{S88})$$

and we use  $\hat{\phi}$  to denote that phase that maximises  $f(\phi)$ . Now, by thinking of  $f(\phi)$  as the inner product between two vectors in the Bloch sphere, one can see that the phase  $\phi$  that minimizes  $f(\phi)$  is 180 degrees added to the phase that maximizes it. So we have

$$p_{\min}^{D_0} = (N_2/N_1) \min_{\phi} f(\phi) = (N_2/N_1) f(\hat{\phi} + \pi). \quad (\text{S89})$$

Hence from (S78) we compute the fringe visibility to be

$$\mathcal{V} = \frac{f(\widehat{\phi}) - f(\widehat{\phi} + \pi)}{f(\widehat{\phi}) + f(\widehat{\phi} + \pi)}. \quad (\text{S90})$$

Now consider the left-hand side of (S79), which we write as

$$\min_{W \in XY} H_{\max}(W_{Q'})_{\overline{\rho}^{D_0}} = \log \left( 1 + \sqrt{1 - \mathcal{V}_{Q'}^2} \right), \quad (\text{S91})$$

which defines the visibility-like quantity  $\mathcal{V}_{Q'}$ , and ultimately we wish to show that  $\mathcal{V}_{Q'} = \mathcal{V}$ . The formula for the unconditional max-entropy is  $H_{\max}(\{p_j\}) = 2 \log \sum_j \sqrt{p_j}$ , which implies that

$$H_{\max}(W_{Q'})_{\overline{\rho}^{D_0}} = \log \left( 1 + \sqrt{1 - (2 \Pr(w_+)_{\overline{\rho}^{D_0}} - 1)^2} \right) \quad (\text{S92})$$

where  $\Pr(w_+)_{\overline{\rho}^{D_0}} := \langle w_+ | \overline{\rho}_{Q'}^{D_0} | w_+ \rangle$ . Comparing (S91) with (S92), we see that

$$\mathcal{V}_{Q'} = 2 \max_{W \in XY} \Pr(w_+)_{\overline{\rho}^{D_0}} - 1. \quad (\text{S93})$$

Using the formula for the state  $\overline{\rho}_{Q'}^{D_0}$  in (S75), we have

$$\Pr(w_+)_{\overline{\rho}^{D_0}} = \frac{\langle w_+ | \text{Tr}_S(C_0 \overline{\rho}_{Q'}^{(2)}) | w_+ \rangle}{\text{Tr}(C_0 \overline{\rho}_S^{(2)})} = \frac{\text{Tr}_{Q'}[ (|w_+\rangle\langle w_+| \otimes C_0) \cdot (\mathcal{I} \otimes \mathcal{E}_S)(V_c \rho_Q^{(1)} V_c^\dagger) ]}{\text{Tr}(C_0 \overline{\rho}_S^{(2)})}. \quad (\text{S94})$$

Now let  $|w_+\rangle = (|0\rangle + e^{i\phi}|1\rangle)/\sqrt{2} = U_\phi|+\rangle$ , and let us maximise (S94) over all  $W$  in the  $XY$  plane, which corresponds to maximising over  $\phi$ . Noting that the denominator on the right-hand side of (S94) is independent of  $W$ , we have

$$\max_{W \in XY} \Pr(w_+)_{\overline{\rho}^{D_0}} = \frac{1}{\text{Tr}(C_0 \overline{\rho}_S^{(2)})} \max_{W \in XY} \text{Tr}_S[C_0 \cdot \mathcal{E}_S(\text{Tr}_{Q'}((|w_+\rangle\langle w_+| \otimes \mathbb{1}_S) V_c \rho_Q^{(1)} V_c^\dagger))] \quad (\text{S95})$$

$$= \frac{1}{\text{Tr}(C_0 \overline{\rho}_S^{(2)})} \max_{\phi} \text{Tr}_S[C_0 \cdot \mathcal{E}_S(\text{Tr}_{Q'}((U_\phi|+\rangle\langle +| U_\phi^\dagger \otimes \mathbb{1}_S) V_c \rho_Q^{(1)} V_c^\dagger))] \quad (\text{S96})$$

$$= \frac{1}{2 \text{Tr}(C_0 \overline{\rho}_S^{(2)})} \max_{\phi} \text{Tr}_S[C_0 \cdot \mathcal{E}_S(U_\phi^\dagger \rho_Q^{(1)} U_\phi)] \quad (\text{S97})$$

$$= \frac{1}{2 \text{Tr}(C_0 \overline{\rho}_S^{(2)})} \max_{\phi} \text{Tr}_S[C_0 \cdot \mathcal{R}_\phi(\mathcal{E}_S(\rho_Q^{(1)}))] \quad (\text{S98})$$

$$= \frac{1}{2 \text{Tr}(C_0 \overline{\rho}_S^{(2)})} f(\widehat{\phi}), \quad (\text{S99})$$

where (S98) invoked Lemma S8. Next, using the fact that  $\overline{\rho}_S^{(1)} = \mathcal{R}_\phi(\overline{\rho}_S^{(1)})$  is diagonal in the standard basis, and furthermore that  $\rho_Q^{(1)} + \mathcal{R}_\pi(\rho_Q^{(1)}) = 2\overline{\rho}_S^{(1)}$ , we have

$$2 \text{Tr}(C_0 \overline{\rho}_S^{(2)}) = 2 \text{Tr}(\tilde{C}_0 \mathcal{E}_S(\overline{\rho}_S^{(1)})) = 2 \text{Tr}[\tilde{C}_0 \mathcal{E}_S(\mathcal{R}_{\widehat{\phi}}(\overline{\rho}_S^{(1)}))] = \text{Tr}[\tilde{C}_0 \mathcal{E}_S(\mathcal{R}_{\widehat{\phi}}(\rho_Q^{(1)}) + \mathcal{R}_{\widehat{\phi}+\pi}(\rho_Q^{(1)}))] = f(\widehat{\phi}) + f(\widehat{\phi} + \pi). \quad (\text{S100})$$

Combining (S93), (S99), and (S100) gives

$$\mathcal{V}_{Q'} = \frac{2f(\widehat{\phi})}{f(\widehat{\phi}) + f(\widehat{\phi} + \pi)} - 1, \quad (\text{S101})$$

which is equivalent to the formula in (S90), and hence completes the proof.  $\square$

#### IV. ENHANCED VISIBILITY

In this section, we consider two examples from the literature, Ref. [11] and [12], of WPDRs where the visibility is enhanced by utilizing a portion of the environment. This corresponds to system  $E_2$  being non-trivial in our main WPDR, Eq. (5) from the main text. We show how both literature results fit into our framework.

## A. Quantum erasure

### 1. Results of Ref. [11]

Ref. [11] derived some WPDRs in which the visibility is enhanced by conditioning on a measurement on the environment. This scenario is called quantum erasure since it aims to erase the which-path information stored in the environment. Here we show that this scenario can be treated in our framework, and hence, that the main results of [11] can be viewed as entropic uncertainty relations.

Ref. [11] considered interferometers that can be modeled as a qubit, which are slightly less general than our notion of binary interferometers, where the interfering subspace is a qubit living inside a larger space. While it should be clear that the treatment can be extended to binary interferometers, for simplicity we will present the treatment as in Ref. [11], as follows.

Suppose the qubit system of interest  $Q$  is initially in state  $\rho_Q^{(1)}$  at time  $t_1$  (see, e.g., Fig. 1 from the main text). Ref. [11] allowed the system  $Q$  to interact with an environment  $F$  resulting in a bipartite state  $\rho_{QF}^{(2)} = \mathcal{E}_{\text{int}}(\rho_Q^{(1)})$  at time  $t_2$ , and then an observable  $\Gamma$  on system  $F$  is measured. We can represent  $\Gamma$  as a set of orthogonal projectors  $\{\Gamma_k\}$  with  $\sum_k \Gamma_k = \mathbb{1}_F$ . (We do not lose generality by assuming the  $\Gamma_k$  are projectors instead of arbitrary positive operators, since system  $F$  is arbitrary and any POVM can be thought of as a projective measurement on an enlarged Hilbert space.) Obtaining outcome  $k$  leaves system  $Q$  in the conditional state

$$\rho_{Q,k}^{(2)} = \frac{1}{g_k} \text{Tr}_F[(\mathbb{1}_Q \otimes \Gamma_k) \rho_{QF}^{(2)}], \quad \text{with } g_k := \text{Tr}[(\mathbb{1}_Q \otimes \Gamma_k) \rho_{QF}^{(2)}]. \quad (\text{S102})$$

One can define the path predictability and fringe visibility associated with this conditional state as

$$\mathcal{P}_k := 2p_{\text{guess}}(Z)_k - 1 \quad (\text{S103})$$

$$\mathcal{V}_k := \frac{p_{\text{max},k}^{D_0} - p_{\text{min},k}^{D_0}}{p_{\text{max},k}^{D_0} + p_{\text{min},k}^{D_0}} \quad (\text{S104})$$

where the subscript  $k$  just means evaluating the quantity for the state  $\rho_{Q,k}^{(2)}$ . Ref. [11] now defined the average predictability and visibility (i.e., averaged over all measurement outcomes) as

$$\mathcal{P}(\Gamma) := \sum_k g_k \mathcal{P}_k, \quad \mathcal{V}(\Gamma) := \sum_k g_k \mathcal{V}_k. \quad (\text{S105})$$

Ref. [11] noted that maximizing  $\mathcal{P}(\Gamma)$  over all  $\Gamma$  gives the distinguishability, while they defined a quantity called ‘‘coherence’’ as the supremum over all  $\Gamma$  of  $\mathcal{V}(\Gamma)$ , as follows

$$\mathcal{D} = \max_{\Gamma} \mathcal{P}(\Gamma), \quad \mathcal{C} := \sup_{\Gamma} \mathcal{V}(\Gamma). \quad (\text{S106})$$

They noted the hierarchies  $\mathcal{P} \leq \mathcal{P}(\Gamma) \leq \mathcal{D}$  and  $\mathcal{V} \leq \mathcal{V}(\Gamma) \leq \mathcal{C}$ . The two main results that were highlighted in [11] were the WPDRs

$$\mathcal{P}(\Gamma)^2 + \mathcal{V}(\Gamma)^2 \leq 1, \quad (\text{S107})$$

$$\mathcal{P}^2 + \mathcal{C}^2 \leq 1, \quad (\text{S108})$$

where (S107) holds for any choice of  $\Gamma$ . Actually, Ref. [11] noted that (S107) implies (S108) by taking the supremum such that  $\mathcal{V}(\Gamma)$  approaches  $\mathcal{C}$ . So let us focus on proving (S107).

### 2. Our treatment

The overall dynamics described above can be separated into three steps:

1. The system  $Q$  interacts with an environment  $F$ , via CPTP map  $\mathcal{E}_{\text{int}}$ .
2. System  $F$  is measured and the outcome is stored in a register  $R$ , via CPTP map  $\mathcal{E}_{\text{meas}}$ .

3. The experimenter uses this measurement result to enhance the visibility on system  $Q$  (i.e., to sort the data point into a sub-ensemble and determine the optimal phase shift for that sub-ensemble). This is modelled as a CPTP map  $\mathcal{E}_{\text{enh}}$  that couples  $R$  to  $Q$ .

The overall CPTP map  $\mathcal{E}$  is a composition of these three maps:

$$\mathcal{E} = \mathcal{E}_{\text{enh}} \circ \mathcal{E}_{\text{meas}} \circ \mathcal{E}_{\text{int}}. \quad (\text{S109})$$

As noted above, the interaction with  $F$  results in the state  $\rho_{QF}^{(2)} := \mathcal{E}_{\text{int}}(\rho_Q^{(1)})$ . Next,  $\mathcal{E}_{\text{meas}}$  performs the projective measurement  $\Gamma = \{\Gamma_k\}$  on system  $F$  and stores the outcome in two (redundant) registers  $R$  and  $R'$ :

$$\rho_{QRR'}^{(3)} := \mathcal{E}_{\text{meas}}(\rho_{QF}^{(2)}) = \sum_k g_k \rho_{Q,k}^{(2)} \otimes |k\rangle\langle k|_R \otimes |k\rangle\langle k|_{R'} \quad (\text{S110})$$

where state  $|k\rangle$  corresponds to obtaining outcome  $k$  from measuring  $\Gamma$ , and the set  $\{|k\rangle\}$  forms an orthonormal basis on the register Hilbert space. The point of having two registers is that one register will act as system  $E_1$  from the main text - to be used to enhance the distinguishability - while the other will act as system  $E_2$  from the main text - to be used to enhance the visibility.

For each measurement outcome  $k$ , we wish to obtain the full visibility that is available, so we allow the experimenter to choose the optimal basis  $W_k$  in the  $XY$  plane of the Bloch sphere for each  $k$ . We can think of this as allowing the experimenter, given the outcome  $k$ , to rotate the system  $Q$  via a unitary  $U_k^Z$  that is diagonal in the  $Z$  basis. Suppose this unitary is tailored to rotate the optimal basis  $W_k$  to the  $X$  basis, i.e.,  $X = U_k^Z W_k (U_k^Z)^\dagger$  for each  $k$ . Accounting for all possible values of  $k$ , the overall unitary is a controlled unitary  $U_{\text{enh}} := \sum_k U_k^Z \otimes |k\rangle\langle k|_R$  where  $R$  acts as the control system. Hence the action of the map that enhances the visibility is:

$$\rho_{QRR'}^{(4)} := \mathcal{E}_{\text{enh}}(\rho_{QRR'}^{(3)}) = U_{\text{enh}} \rho_{QRR'}^{(3)} U_{\text{enh}}^\dagger = \sum_k g_k \tilde{\rho}_{Q,k}^{(2)} \otimes |k\rangle\langle k|_R \otimes |k\rangle\langle k|_{R'}, \quad \text{with } \tilde{\rho}_{Q,k}^{(2)} := U_k^Z \rho_{Q,k}^{(2)} (U_k^Z)^\dagger. \quad (\text{S111})$$

Finally, we apply our main WPDR, Eq. (5) from the main text, to the state  $\rho_{QRR'}^{(4)}$ . Specifically we choose  $E_1 = R$  and  $E_2 = R' = R$  noting that  $R'$  and  $R$  are identical copies, giving

$$H_{\min}(Z|R)_{\rho^{(4)}} + H_{\max}(X|R)_{\rho^{(4)}} \geq 1, \quad (\text{S112})$$

where the subscript  $\rho^{(4)}$  emphasizes that the entropy terms are evaluated for the state  $\rho_{QRR'}^{(4)}$ , for which  $X$  is the basis that achieves the minimization in  $\min_{W \in XY} H_{\max}(W|R)$ .

We now show how our quantum erasure relation in (S112) implies (S107), which in turn implies (S108) as noted previously. First note that

$$\mathcal{P}(\Gamma) = \sum_k g_k \mathcal{P}_k = \sum_k g_k [2p_{\text{guess}}(Z)_k - 1] = 2 \left[ \sum_k g_k p_{\text{guess}}(Z)_k \right] - 1 = 2p_{\text{guess}}(Z|R)_{\rho^{(4)}} - 1,$$

which implies that

$$H_{\min}(Z|R)_{\rho^{(4)}} = -\log p_{\text{guess}}(Z|R)_{\rho^{(4)}} = 1 - \log[1 + \mathcal{P}(\Gamma)]. \quad (\text{S113})$$

Next, using the relation (S70) and noting that  $\mathcal{V}(\Gamma) = 2p_{\text{guess}}(X|R)_{\rho^{(4)}} - 1$ , we have that

$$H_{\max}(X|R)_{\rho^{(4)}} \leq \log(1 + \sqrt{1 - \mathcal{V}(\Gamma)^2}). \quad (\text{S114})$$

Substituting (S113) and (S114) into (S112) gives (S107).

We remark that the uncertainty relation (S112) corresponds to the ‘‘preparation uncertainty’’ scenario discussed in the main text, addressing the question of the predictability of the measurement at the interferometer *output*. Hence the results of Ref. [11] are of the preparation uncertainty variety.

## B. Polarization-enhanced visibility and discussion of Ref. [12]

### 1. Result of Ref. [12]

The aim of this section is to show that the main result of Ref. [12] can be viewed as, or is a direct consequence of, the uncertainty relation for the min- and max-entropies, and hence is covered by our framework. Their result is

a WPDR for a MZI where a fairly general interaction occurs inside the interferometer between the photon's spatial degree of freedom  $Q$ , its polarization  $P$ , and an environment  $F$ .

In their WPDR they replaced the usual fringe visibility with a quantity they called generalized visibility with the motivation that it provides a stronger bound on the path distinguishability (via the WPDR). We will denote their generalized visibility as  $\mathcal{V}_B$ , where the subscript  $B$  refers to the first author of Banaszek et al. [12]. (Similarly we will use  $\mathcal{D}_B$  for their path distinguishability.) Their generalized visibility  $\mathcal{V}_B$  was written in an abstract form; however, its operational meaning was not stated. Here we give an operational meaning to  $\mathcal{V}_B$ , showing that it is directly proportional to the guessing probability for the which-phase observable, given the optimal measurement on the photon's polarisation  $P$  (see below for the precise statement). Hence the visibility is *enhanced* by gaining further information from measuring the polarisation.

We remark that enhancing the visibility in this way can be done either in our predictive or retrodictive framework. In the previous section we discussed Ref. [11], which considered the predictive scenario. On the other hand, it turns out that Ref. [12] considered the retrodictive (or joint measurement) scenario, where one uses the final measurement outcomes to retrodict the photon's properties at an earlier time. Thus, taken together, Refs. [11] and [12] nicely illustrate how visibility can be enhanced in the two different scenarios.

Ref. [12] allowed for an interaction to occur within the interferometer coupling  $Q$  to  $P$  and  $F$ . We can model this as an isometry  $V$  that maps states on  $Q$  at time  $t_1$  to states on the larger system  $QPFT$  at time  $t_2$ , where  $T$  acts as a purifying system for the overall state. Ref. [12] assumed that  $V$  is path-preserving. From  $V$  we may define two complementary quantum channels

$$\mathcal{F}(\cdot) = \text{Tr}_{QPT}[V(\cdot)V^\dagger], \quad \mathcal{F}_c(\cdot) = \text{Tr}_F[V(\cdot)V^\dagger], \quad (\text{S115})$$

which respectively map  $Q$  at time  $t_1$  to  $F$  (or  $QPT$ ) at time  $t_2$ . Now, consider the (normalized) conditional states on the environment at time  $t_2$  respectively associated with the  $|0\rangle$  and  $|1\rangle$  input states on  $Q$ :

$$\rho_F^0 = \mathcal{F}(|0\rangle\langle 0|), \quad \rho_F^1 = \mathcal{F}(|1\rangle\langle 1|). \quad (\text{S116})$$

The main result of Ref. [12] is a WPDR of the form

$$\mathcal{D}_B^2 + \mathcal{V}_B^2 \leq 1, \quad (\text{S117})$$

where

$$\mathcal{D}_B := \frac{1}{2} \left\| \rho_F^0 - \rho_F^1 \right\|_1, \quad (\text{S118})$$

$$\mathcal{V}_B := \left\| \sqrt{\rho_F^0} \sqrt{\rho_F^1} \right\|_1. \quad (\text{S119})$$

To clarify, Ref. [12] wrote  $\mathcal{V}_B$  in several ways, one of which was  $\mathcal{V}_B = \max_{|\psi^0\rangle, |\psi^1\rangle} |\langle \psi^0 | \psi^1 \rangle|$  where the maximisation is over all purifications  $|\psi^0\rangle$  and  $|\psi^1\rangle$  of  $\rho_F^0$  and  $\rho_F^1$  respectively. But this form of  $\mathcal{V}_B$  is equivalent to that in (S119) due to Uhlmann's theorem (see, e.g., [44]).

## 2. Our treatment

We will reinterpret (S117) within the context of the our complementary guessing game discussed in the main text. In particular let us consider the retrodictive scenario, discussed in the main text, where a sender Bob inputs states randomly (with equal probability for each state in the ensemble) into the interferometer chosen from one of two ensembles,  $Z_i = \{|0\rangle, |1\rangle\}$  or  $W_i = \{|w_\pm\rangle\}$ . Alice tries to guess which state Bob sent (given that Bob announces which ensemble he is considering,  $Z_i$  or  $W_i$ ). With this scenario in mind, we can rewrite the path distinguishability as

$$\mathcal{D}_B = 2p_{\text{guess}}(Z_i|F)_{\mathcal{F}} - 1 \quad (\text{S120})$$

where the subscript  $\mathcal{F}$  refers to the relevant quantum channel. Mathematically speaking, (S120) follows from (S38).

Similarly we give an operational interpretation for  $\mathcal{V}_B$  with the following lemma.

**Lemma S9.** Let the interaction  $V$  be path-preserving, then

$$\mathcal{V}_B = \max_{W \in XY} [2p_{\text{guess}}(W_i|QPT)_{\mathcal{F}_c} - 1] \quad (\text{S121})$$

where  $p_{\text{guess}}(W_i|QPT)_{\mathcal{F}_c}$  denotes the probability of guessing  $W_i = \{|w_\pm\rangle\}$  correctly given the optimal measurement on the joint system  $QPT$ , i.e., the output of the channel  $\mathcal{F}_c$ .

See the next subsection for the proof. This says that  $\mathcal{V}_B$  can be interpreted like an *input* visibility, measuring how well Alice can distinguish between the different  $W_i$  states that Bob sends, by doing a measurement on the output system  $QPT$ . In this sense it quantifies how much phase information or wave-like information gets transmitted from the input  $Q$  to the output  $QPT$ . This clear operational interpretation sheds light on the meaning on  $\mathcal{V}_B$ , which was previously only given in an abstract form in [12].

Now we show how (S117) can be recast in terms of the min- and max-entropic uncertainty relation. In the retrodictive scenario (as discussed in the Methods) we consider the Choi-Jamiołkowski state obtained from applying  $V$  to half of a maximally-entangled state  $|\Phi\rangle$  on  $Q'Q$ ,

$$|\Lambda\rangle_{Q'QFPT} = (\mathbb{1} \otimes V)|\Phi\rangle. \quad (\text{S122})$$

Now we apply the uncertainty relation for the min- and max-entropies to the state  $|\Lambda\rangle_{Q'QFPT}$ , giving

$$H_{\min}(Z_{Q'}|F)_{|\Lambda\rangle} + \min_{W \in XY} H_{\max}(W_{Q'}|QPT)_{|\Lambda\rangle} \geq 1. \quad (\text{S123})$$

As discussed in the Methods section, the observables  $Z_{Q'}$  and  $W_{Q'}$  on  $Q'$  get mapped to the input ensembles  $Z_i$  and  $W_i$  on  $Q$  (more precisely, one takes the transpose, but this does not matter since we are minimizing  $W$  over the  $XY$  plane). Hence (S123) can be rewritten as

$$H_{\min}(Z_i|F)_{\mathcal{F}} + \min_{W \in XY} H_{\max}(W_i|QPT)_{\mathcal{F}_c} \geq 1, \quad (\text{S124})$$

where it should be clear that (S124) refers to the systems  $F$  and  $QPT$  *after* the interaction  $V$ , which we emphasise with the subscripts  $\mathcal{F}$  and  $\mathcal{F}_c$  indicating the relevant quantum channels that map the input to the output. Note that this uncertainty relation is a special case of our main result given in the main manuscript, which can be seen by choosing:

$$E_1 = F, \quad E_2 = QPT. \quad (\text{S125})$$

Also, we remind the reader that (S124) can be interpreted as a joint measurement uncertainty relation in the following sense. The input ensembles  $Z_i$  and  $W_i$  are used to calibrate the measurement apparatus, i.e., to assess how well it can measure  $Z$  and  $W$ . Equation (S124) says the output of the apparatus,  $QFPT$ , cannot provide full information about both the  $Z$  and  $W$  observables on the input.

To show that (S124) implies (S117) we use the relations

$$H_{\min}(Z_i|F)_{\mathcal{F}} = 1 - \log(1 + \mathcal{D}_B), \quad (\text{S126})$$

$$\min_{W \in XY} H_{\max}(W_i|QPT)_{\mathcal{F}_c} \leq \log(1 + \sqrt{1 - \mathcal{V}_B^2}), \quad (\text{S127})$$

where (S126) follows from the relation between the min-entropy and the guessing probability, and (S127) follows from combining Lemmas S6 and S9. Inserting these into (S124) gives (S117). This shows that the main result of [12] is an entropic uncertainty relation in disguise, following directly from (S124).

### 3. Proof of Lemma S9

We first note that the formula in (S121) can be rewritten as

$$\mathcal{V}_B = \max_{W \in XY} [2p_{\text{guess}}(W_{Q'}|QPT)_{|\Lambda\rangle} - 1] \quad (\text{S128})$$

where  $p_{\text{guess}}(W_{Q'}|QPT)_{|\Lambda\rangle}$  denotes the probability of correctly guessing the observable  $W$  on system  $Q'$  given the optimal measurement on  $QPT$ , for the state  $|\Lambda\rangle$  in (S122). Now to prove (S128), we apply (S129) in the following lemma, noting that the state  $|\Lambda\rangle$  in (S122) has the special form assumed in this lemma since  $V$  is path-preserving.

**Lemma S10.** Let  $|\Psi\rangle_{ABC}$  be a tripartite pure state such that the reduced state on  $A$  and  $B$  is classical-quantum:  $\rho_{AB} = |0\rangle\langle 0| \otimes \sigma_0 + |1\rangle\langle 1| \otimes \sigma_1$ , where  $\sigma_0$  and  $\sigma_1$  are subnormalized such that  $\text{Tr}\sigma_0 + \text{Tr}\sigma_1 = 1$ . Let  $W$  correspond to the outcome of a projective measurement in the  $XY$  plane performed on  $A$ . Then it holds that

$$2\|\sqrt{\sigma_0}\sqrt{\sigma_1}\|_1 = 2p_{\text{guess}}(W|C)_{|\psi\rangle} - 1. \quad (\text{S129})$$

*Proof.* Since the quantity  $p_{\text{guess}}(W|C)$  is invariant under unitaries on  $C$  we can choose  $|\Psi\rangle_{ABC}$  to be an arbitrary purification of  $\rho_{AB}$ . We find it convenient to split up  $C$  into  $C_1$  and  $C_2$  which purify  $B$  and  $A$ , respectively:

$$|\Psi\rangle_{ABC_1C_2} = |0\rangle_A |\psi_{\sigma_0}\rangle_{BC_1} |0\rangle_{C_2} + |1\rangle_A |\psi_{\sigma_1}\rangle_{BC_1} |1\rangle_{C_2}. \quad (\text{S130})$$

Without loss of generality the two orthonormal basis states associated with  $W$  can be written as

$$|w_{\pm}\rangle = \frac{1}{\sqrt{2}} \left( |0\rangle \pm e^{i\theta} |1\rangle \right), \quad (\text{S131})$$

where  $\theta \in [0, 2\pi]$  is a parameter. Performing the measurement of  $W$  leads to the following cq state

$$\begin{aligned} \rho_{WBC_1C_2} &= \frac{1}{2} |w_+\rangle \langle w_+|_W \otimes \left( |\psi_{\sigma_0}\rangle \langle \psi_{\sigma_0}|_{BC_1} \otimes |0\rangle \langle 0|_{C_2} + e^{i\theta} |\psi_{\sigma_0}\rangle \langle \psi_{\sigma_1}|_{BC_1} \otimes |0\rangle \langle 1|_{C_2} \right. \\ &\quad \left. + e^{-i\theta} |\psi_{\sigma_1}\rangle \langle \psi_{\sigma_0}|_{BC_1} \otimes |1\rangle \langle 0|_{C_2} + |\psi_{\sigma_1}\rangle \langle \psi_{\sigma_1}|_{BC_1} \otimes |1\rangle \langle 1|_{C_2} \right) \\ &\quad + \frac{1}{2} |w_-\rangle \langle w_-|_W \otimes \left( |\psi_{\sigma_0}\rangle \langle \psi_{\sigma_0}|_{BC_1} \otimes |0\rangle \langle 0|_{C_2} - e^{i\theta} |\psi_{\sigma_0}\rangle \langle \psi_{\sigma_1}|_{BC_1} \otimes |0\rangle \langle 1|_{C_2} \right. \\ &\quad \left. - e^{-i\theta} |\psi_{\sigma_1}\rangle \langle \psi_{\sigma_0}|_{BC_1} \otimes |1\rangle \langle 0|_{C_2} + |\psi_{\sigma_1}\rangle \langle \psi_{\sigma_1}|_{BC_1} \otimes |1\rangle \langle 1|_{C_2} \right). \end{aligned} \quad (\text{S132})$$

It is easy to verify that  $\text{Tr}_B |\psi_{\sigma_u}\rangle \langle \psi_{\sigma_v}|_{BC} = (\sqrt{\sigma_u} \sqrt{\sigma_v})_C$  for  $u, v \in \{0, 1\}$ . Therefore, we get

$$\begin{aligned} \rho_{WC_1C_2} &= \frac{1}{2} |w_+\rangle \langle w_+| \otimes \left( \sigma_0 \otimes |0\rangle \langle 0| + e^{i\theta} \sqrt{\sigma_0} \sqrt{\sigma_1} \otimes |0\rangle \langle 1| + e^{-i\theta} \sqrt{\sigma_1} \sqrt{\sigma_0} \otimes |1\rangle \langle 0| + \sigma_1 \otimes |1\rangle \langle 1| \right) \\ &\quad + \frac{1}{2} |w_-\rangle \langle w_-| \otimes \left( \sigma_0 \otimes |0\rangle \langle 0| - e^{i\theta} \sqrt{\sigma_0} \sqrt{\sigma_1} \otimes |0\rangle \langle 1| - e^{-i\theta} \sqrt{\sigma_1} \sqrt{\sigma_0} \otimes |1\rangle \langle 0| + \sigma_1 \otimes |1\rangle \langle 1| \right). \end{aligned} \quad (\text{S133})$$

Using the relation between the guessing probability and the trace distance from (S38), it is straightforward to find:

$$\begin{aligned} 2p_{\text{guess}}(W|C)_{|\psi\rangle} - 1 &= \left\| e^{i\theta} \sqrt{\sigma_0} \sqrt{\sigma_1} \otimes |0\rangle \langle 1| + e^{-i\theta} \sqrt{\sigma_1} \sqrt{\sigma_0} \otimes |1\rangle \langle 0| \right\|_1 \\ &= 2 \left\| \sqrt{\sigma_0} \sqrt{\sigma_1} \right\|_1 \end{aligned} \quad (\text{S134})$$

It is worth noting that the phase  $\theta$  does not appear in (S134), i.e., all measurements in the  $XY$  plane lead to the same guessing probability.  $\square$

## V. TESTING COHERENCE IN A QUANTUM BEAM SPLITTER

### A. Quantities sensitive to coherence

Here we further elaborate on the treatment of the quantum beam splitter (QBS) depicted in Fig. 5 of the main text. The QBS is modelled as a controlled-unitary, whose form is

$$U_{PQ} = |H\rangle \langle H|_P \otimes \mathbb{1}_S + |V\rangle \langle V|_P \otimes U(R), \quad \text{with } U(R) = \begin{pmatrix} \sqrt{R} & \sqrt{1-R} \\ \sqrt{1-R} & -\sqrt{R} \end{pmatrix}. \quad (\text{S135})$$

Here the photon's polarisation  $P$  in the  $\{|H\rangle, |V\rangle\}$  basis determines whether one applies the transformation associated with a (classical) beam splitter with reflection coefficient  $R$ . If  $P$  is fed in as a superposition of the  $\{|H\rangle, |V\rangle\}$  basis states then the beam splitter is said to be "quantum".

Our main goal in what follows is to show that our novel WPDR stated in the main text:

$$(\mathcal{D}_i^P)^2 + \mathcal{V}^2 \leq 1 \quad (\text{S136})$$

captures the coherence of the beam splitter, whereas a weaker WPDR:

$$\mathcal{D}_i^2 + \mathcal{V}^2 \leq 1 \quad (\text{S137})$$

does not. Here the different distinguishabilities are

$$\mathcal{D}_i = 2p_{\text{guess}}(Z_i)_{D_0} - 1, \quad (\text{S138})$$

$$\mathcal{D}_i^P = 2p_{\text{guess}}(Z_i|P)_{D_0} - 1, \quad (\text{S139})$$

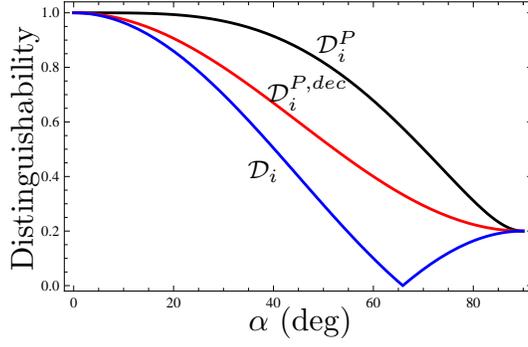


FIG. S7: Plot of  $\mathcal{D}_i$ ,  $\mathcal{D}_i^{P,dec}$ , and  $\mathcal{D}_i^P$  as a function of  $\alpha$ , for  $R = 0.4$ .

where  $P$  in (S139) refers to the final polarization after the QBS. For simplicity, we will neglect any interaction with an external environment  $F$  in what follows, and hence conditioning these distinguishabilities on  $F$  is not necessary. (Since we are interested in demonstrating that (S136) can capture coherence, it suffices to demonstrate it for a special case where  $F$  plays no role.)

For comparison, we will also define “decohered” versions of  $\mathcal{D}_i$  and  $\mathcal{D}_i^P$ ,  $\mathcal{D}_i^{dec}$  and  $\mathcal{D}_i^{P,dec}$  respectively, where the latter correspond to feeding in a decohered version of the polarization state  $\rho_P^{(2)}$ , i.e., feeding in the corresponding classical mixture of  $|H\rangle$  and  $|V\rangle$  rather than a coherent superposition. Precisely this means replacing  $\rho_P^{(2)}$  with  $\mathcal{Z}(\rho_P^{(2)})$  where  $\mathcal{Z}(\cdot) = |H\rangle\langle H|(\cdot)|H\rangle\langle H| + |V\rangle\langle V|(\cdot)|V\rangle\langle V|$  is the quantum channel that decoheres the polarization state.

The first noteworthy point is that  $\mathcal{D}_i = \mathcal{D}_i^{dec}$ , hence measuring  $\mathcal{D}_i$  does not reveal the coherence in the QBS. This is because  $\mathcal{D}_i$  is not conditioned on  $P$ , so we evaluate it on the reduced state obtained from tracing over  $P$ . But tracing over  $P$  removes any dependence on the off-diagonal elements of  $\rho_P^{(2)}$  in  $\{|H\rangle, |V\rangle\}$  basis, since the unitary  $U_{PQ}$  is controlled by the  $\{|H\rangle, |V\rangle\}$  basis.

On the other hand we show that, in general,  $\mathcal{D}_i^P \neq \mathcal{D}_i^{P,dec}$ , so  $\mathcal{D}_i^P$  has the potential to reveal coherence. We also remark that the following hierarchy holds in general:

$$\mathcal{D}_i \leq \mathcal{D}_i^{P,dec} \leq \mathcal{D}_i^P. \quad (\text{S140})$$

The first inequality holds because  $\mathcal{D}_i = \mathcal{D}_i^{dec}$  and conditioning on  $P$  can never decrease the guessing probability. The second inequality holds because the decoherence operation  $\mathcal{Z}(\cdot)$  commutes with  $U_{PQ}$  and hence can be viewed as restricting the class of measurements over which one optimizes to evaluate the guessing probability.

For simplicity, let us consider a one-parameter family of input states  $\rho_P^{(2)} = |\alpha\rangle\langle\alpha|$  with  $|\alpha\rangle := \cos\alpha|H\rangle + \sin\alpha|V\rangle$ , which also happens to be the family considered in Ref. [34] (see the next subsection). Thus, we have an open interferometer when  $\alpha = 0$  and a closed one when  $\alpha = 90$  deg. For such states the visibility becomes

$$\mathcal{V} = 2\sqrt{R(1-R)}\sin^2(\alpha). \quad (\text{S141})$$

Likewise, solving for the different distinguishabilities (see below for derivation) gives:

$$\mathcal{D}_i = |1 - 2(\sin^2\alpha)(1-R)|, \quad (\text{S142})$$

$$\mathcal{D}_i^{P,dec} = 1 - (\sin^2\alpha) \cdot (1 - |2R - 1|), \quad (\text{S143})$$

$$\mathcal{D}_i^P = \sqrt{1 - 4R(1-R)(\sin^4\alpha)}. \quad (\text{S144})$$

Clearly these formulas indicate that, in general,  $\mathcal{D}_i^P \neq \mathcal{D}_i^{P,dec}$ , hence showing that  $\mathcal{D}_i^P$  reveals coherence. We can see a clear distinction between these distinguishabilities in Fig. S7, which considers the case of  $R = 0.4$ .

## B. Discussion of Ref. [34]

We remark that  $\mathcal{D}_i^P$  could be measured using polarization-resolving detection at the interferometer output, as in Fig. 5 of the main text, assuming one chooses the optimal polarization basis to measure. (See below where we explicitly

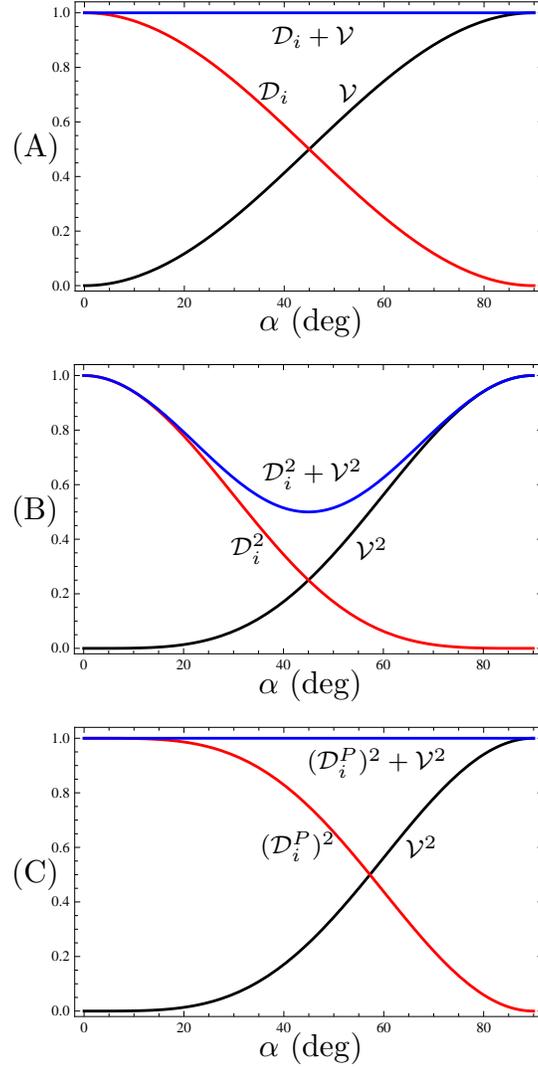


FIG. S8: For  $R = 0.5$ , we plot (A)  $\mathcal{V}$ ,  $\mathcal{D}_i$ , and  $\mathcal{V} + \mathcal{D}_i$ ; (B)  $\mathcal{V}^2$ ,  $\mathcal{D}_i^2$ , and  $\mathcal{V}^2 + \mathcal{D}_i^2$ ; and (C)  $\mathcal{V}^2$ ,  $(\mathcal{D}_i^P)^2$ , and  $\mathcal{V}^2 + (\mathcal{D}_i^P)^2$ . Notice that  $\mathcal{V}^2 + (\mathcal{D}_i^P)^2 = 1$ , showing that our novel WPDR is perfectly tight.

solve for the optimal polarization basis to measure.) We note that the setup in Ref. [34] has polarization-resolving detectors, and hence could measure  $\mathcal{D}_i^P$ . However, the procedure outlined in [34] for measuring distinguishability corresponds to measuring our  $\mathcal{D}_i$ . Figure S8 shows our theoretical predictions for the situation in [34], corresponding to  $R = 0.5$ . At first sight our predictions appear to disagree with [34] in the sense that our Fig. S8B, which plots  $\mathcal{V}^2$ ,  $\mathcal{D}_i^2$ , and  $\mathcal{V}^2 + \mathcal{D}_i^2$ , looks very different from the corresponding plot of these quantities in Fig. 4 of [34]. An explanation for the disagreement is that [34] may have actually plotted  $\mathcal{V}$ ,  $\mathcal{D}_i$ , and  $\mathcal{V} + \mathcal{D}_i$  in their Fig. 4. Indeed, their Fig. 4 looks similar to our predictions for  $\mathcal{V}$  and  $\mathcal{D}_i$  in Fig. S8A, and we have  $\mathcal{V} + \mathcal{D}_i = 1$  which is consistent with their Fig. 4. The authors of [34] have confirmed that their Fig. 4 plotted visibility and distinguishability as opposed to their squares [47]. We emphasize that this minor issue with their plot does not affect the conclusions of Ref. [34].

Since we predict that  $\mathcal{V}^2 + \mathcal{D}_i^2$  can be strictly less than 1, then testing our novel relation  $\mathcal{V}^2 + (\mathcal{D}_i^P)^2 \leq 1$  can give a more stringent test of wave-particle duality. Indeed we show that for the setup in [34], this relation is as strong as possible, i.e., it is satisfied with equality  $\mathcal{V}^2 + (\mathcal{D}_i^P)^2 = 1$ . This is depicted in Fig. S8C.

We remark that Ref. [34] verified the coherence of their QBS by violating a Bell inequality, which involves verifying correlations in multiple bases. Our work here shows that one can verify the coherence of a QBS without switching bases in the experiment. Simply measuring the quantities appearing in our WPDR (S136), in particular  $\mathcal{D}_i^P$ , can verify the coherence of the QBS.

### C. Derivation of distinguishability formulas

The derivation of Eqs. (S142)-(S144) proceeds as follows. As discussed in the Methods section of the main text, our retrodictive framework starts with the quanton  $Q$  being maximally entangled to a register  $Q'$ , in the state  $|\Phi\rangle_{QQ'} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ . Meanwhile the polarization  $P$  enters the QBS in the state  $|\alpha\rangle := c|H\rangle + s|V\rangle$ , where  $c = \cos \alpha$  and  $s = \sin \alpha$ . The action of the QBS, given by  $U_{PQ}$  in (S135), results in the state:

$$(U_{PQ} \otimes \mathbb{1}_{Q'})|\alpha\rangle_P \otimes |\Phi\rangle_{QQ'} = c|H\rangle_P \otimes \frac{|0\rangle_Q|0\rangle_{Q'} + |1\rangle_Q|1\rangle_{Q'}}{\sqrt{2}} + s|V\rangle_P \otimes \frac{U(R)|0\rangle_Q|0\rangle_{Q'} + U(R)|1\rangle_Q|1\rangle_{Q'}}{\sqrt{2}}. \quad (\text{S145})$$

Conditioning on detector  $D_0$  clicking corresponds to measuring  $Q$  in the standard basis and post-selecting on the outcome 0. In addition, measuring  $Q'$  to obtain  $Z$  gives

$$\bar{\rho}_{ZZ}^{D_0} = |0\rangle\langle 0|_Z \otimes \sigma_P^0 + |1\rangle\langle 1|_Z \otimes \sigma_P^1, \quad (\text{S146})$$

$$\text{where } \sigma_P^0 = \begin{pmatrix} c^2 & sc\sqrt{R} \\ sc\sqrt{R} & s^2R \end{pmatrix} \quad \text{and} \quad \sigma_P^1 = \begin{pmatrix} 0 & 0 \\ 0 & s^2(1-R) \end{pmatrix}. \quad (\text{S147})$$

Now we compute the distinguishability by using the fact that it is the trace distance between the conditional states. For the polarization-enhanced distinguishability, this gives:

$$\mathcal{D}_i^P = \|\sigma_P^0 - \sigma_P^1\|_1 = \sqrt{c^4 + s^4(2R-1)^2 + 2(sc)^2}. \quad (\text{S148})$$

Decohering  $\sigma_P^0$  and  $\sigma_P^1$  before calculating the trace norm leads to the decohered distinguishability:

$$\mathcal{D}_i^{P,dec} = c^2 + s^2 \cdot |2R-1|. \quad (\text{S149})$$

Finally, to calculate the non-enhanced distinguishability we need to trace out the polarization register to get

$$\mathcal{D}_i = |\text{Tr}\sigma_P^0 - \text{Tr}\sigma_P^1| = |c^2 + s^2(2R-1)|. \quad (\text{S150})$$

### D. Measuring $\mathcal{D}_i^P$

Measuring the polarization-enhanced distinguishability  $\mathcal{D}_i^P$  requires polarization-resolving detection at the output of the interferometer, as in Fig. 5 of the main text. As defined,  $\mathcal{D}_i^P$  corresponds to measuring the optimal polarization basis at the output, i.e., optimally helpful for guessing which path the photon took. We now solve for this optimal polarization basis. We remark that varying the polarization measurement basis could be accomplished by varying the angle of a half-wave plate inserted just prior to the PBS's in Fig. 5 of the main text.

Finding the optimal measurement is simply a two-state discrimination problem on a qubit, and the solution is well-known [48]. Consider the (unnormalized) conditional states on system  $P$  associated with  $Z = 0$  and  $Z = 1$ , respectively denoted  $\sigma_P^0$  and  $\sigma_P^1$  and given by Eq. (S147). (Note that both states are conditioned on detector  $D_0$  clicking.) From [48], the optimal polarization basis to measure is given by the eigenvectors of the following Hermitian operator:

$$O_P = \sigma_P^0 - \sigma_P^1. \quad (\text{S151})$$

From (S147), we compute that this corresponds to the following polarization observable (represented as a matrix in the  $\{|H\rangle, |V\rangle\}$  basis):

$$\hat{O}_P = \frac{1}{\mathcal{D}_i^P} \begin{pmatrix} 1 - 2(\sin^2 \alpha)R & (\sin 2\alpha)\sqrt{R} \\ (\sin 2\alpha)\sqrt{R} & -[1 - 2(\sin^2 \alpha)R] \end{pmatrix}, \quad (\text{S152})$$

where we have normalized the observable (hence the hat symbol) such that it squares to the identity. For example, choosing  $R = 0.5$  and  $\alpha = 45$  deg (corresponding to an equal superposition of BS<sub>2</sub> being ‘‘absent’’ and ‘‘present’’) gives

$$\hat{O}_P = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & \sqrt{2} \\ \sqrt{2} & -1 \end{pmatrix}. \quad (\text{S153})$$