Cost of postselection in decision theory

Joshua Combes and Christopher Ferrie

Pricing postselection: the cost of indecision in decision theory

Joshua Combes[1, 2, 3, 4] and Christopher Ferrie[1, 5]

[1]*Center for Quantum Information and Control, University of New Mexico, Albuquerque, New Mexico, 87131-0001*
[2]*Centre for Engineered Quantum Systems, School of Mathematics and Physics,*
The University of Queensland, St Lucia, QLD 4072, Australia
[3]*Institute for Quantum Computing, University of Waterloo, Ontario N2L 3G1, Canada*
[4]*Perimeter Institute for Theoretical Physics, 31 Caroline St. N, Waterloo, Ontario, Canada N2L 2Y5*
[5]*Centre for Engineered Quantum Systems, School of Physics,*
The University of Sydney, Sydney, NSW, Australia

Postselection is the process of discarding outcomes from statistical trials that are not the event one desires. Postselection can be useful in many applications where the cost of getting the wrong event is implicitly high. However, unless this cost is specified exactly, one might conclude that discarding all data is optimal. Here we analyze the optimal decision rules and quantum measurements in a decision theoretic setting where a pre-specified cost is assigned to discarding data. Our scheme interpolates between unambiguous state discrimination (when the cost of postselection is zero) and a minimum error measurement (when the cost of postselection is maximal). We also relate our formulation to previous approaches which focus on minimizing the probability of indecision.

I. INTRODUCTION

There has been some confusion over the role of post-selection in quantum information processing protocols. On one hand, postselection is a powerful computational resource [1] and enables technological goals, such as probabilistic photon-photon gates [2]. On the other hand, in some situations postselection can impede quantum information processing.

Probabilistic metrology—also known as metrology with abstention [3] and weak value amplification [4]—is the idea that postselection may improve estimation precision beyond the usual quantum limits. When the performance of probabilistic metrology is evaluated with respect to the standard figure of merit for parameter estimation, mean squared error, postselection is provably suboptimal, even when there are imperfections [5–12]. Counter claims have been made in the literature (see Refs. [13–17]) but the issue is far from settled.

In this article we attempt to reconcile the intuition that postselection can help statistical tasks with the fact that for the standard figures of merit generically it does not. To simplify the analysis and make our assumptions explicit we will use a statistical decision theory approach in the context of quantum state discrimination [18, 19]. To assert that a state discrimination protocol is optimal, we must first specify a *cost* or *loss* function which encapsulates how each decision is penalized. Then we minimize the average loss over decision rules and measurements.

This approach defines a task for which the optimal protocol incurs the least losses for the specified loss function. For example consider a two party discrimination game involving an employer Alice and an employee Bob. Alice gives Bob one of two quantum states Ψ_1 or Ψ_2. Bob is allowed to perform any generalized measurement on the state but then must report which state Alice gave him; he cannot decline to report a state. Bob's bonus, of at most $\mathbb{D}$ dollars, is tied to his performance in this game. If he reports Ψ_i when Ψ_j is true his bonus will

be reduced to $\$(1 - \lambda_{i,j})\mathbb{D}$ where $\lambda_{i,j}$ is called the loss function. Bob wants to devise a strategy to minimise his expected losses. When the cost of reporting the correct answer is "0" and the incorrect answer is "1" or maximal, $\lambda_{i,j}$ is known as the 0-1 loss function. Mimimising the losses from the 0-1 loss function is equivalent to minimizing the probability of misidentifying the states (termed the error probability) [20, 21]. The corresponding optimal measurement strategy, with respect to minimizing losses, is called the Helstrom [20] or minimum error measurement. A postselected strategy will have higher expected losses, that is it is suboptimal with respect to the 0-1 loss function.

Postselected strategies for state discrimination were introduced by Ivanovic [22], Dieks [23], and Peres [24] in what is now known as *unambiguous state discrimination* (USD). In USD one allows for an extra "reject" decision—postselection—then two nonorthogonal states can be distinguished without error, albeit probabilistically. The USD measurement is optimized in the sense that it has minimal probability of reporting the inconclusive result "reject". Prior work on inconclusive state discrimination has focused on exploring and optimizing schemes which interpolate between minimum error probability and minimum inconclusive result probability [25–31]. Typically in USD and its generalizations [32] there is no explicit penalty for reporting "reject". It is unclear if such postselection is optimal with respect to any loss function.

Here we re-formalize the inconclusive state discrimination problem by assigning a cost to discarded outcomes. In particular, we modify the most commonly used cost, the 0-1 loss function, to what we call the 0-1-λ loss function. In the 0-1-λ loss function, λ is the cost of reporting "reject". In our approach, we find that the USD measurement appears when $\lambda \to 0$. In this limit there is an alternative protocol which is equally optimal: always report "reject". Finally we show how our results can be connected to previous approaches where there is a trade-

off between the rejection probablity and the error probablity [25–31]. Our analysis adheres to the desiderata suggested in Ref. [9], and thus is a definitive case where employing postselection can be said to be optimal.

II. STATISTICAL DECISION THEORY

We start by reviewing statistical decision theory and formally introducing the 0-1-λ loss function, which is a special case of Chow's work on hypothesis testing or classification [33, 34]. Consider a set of competing hypotheses $\mathcal{H}_j$ for $j \in \{1, 2, ..., n\}$ with prior probabilities $\Pr(\mathcal{H}_j)$. Given some data $\mathbf{D}$ the posterior probability of the j'th hypothesis is

$$\Pr(\mathcal{H}_j|\mathbf{D}) = \frac{\Pr(\mathbf{D}|\mathcal{H}_j)\Pr(\mathcal{H}_j)}{\Pr(\mathbf{D})}, \tag{1}$$

where

$$\Pr(\mathbf{D}) = \sum_{j=1}^{n} \Pr(\mathbf{D}|\mathcal{H}_j)\Pr(\mathcal{H}_j). \tag{2}$$

What we would like to do is have a *decision rule* $\delta(\mathbf{D})$ that maps the data $\mathbf{D}$ to decision i—that is, report hypothesis i, where in this case $i \in \{0, 1, 2..., n\}$. The decision $i = 0$ allows for the possibility that one may not be able to decide, often referred to as the "don't know" or "abstain" or "reject" option.

In Bayesian decision theory the decision rule must arise from minimizing a loss function, which encapsulates how each decision is penalized. The conditional risk, i.e. the *a posteriori* expected loss, for the decision i conditioned on data $\mathbf{D}$ is

$$\mathcal{R}[i|\mathbf{D}] = \sum_{j=1}^{n} \lambda_{i,j}\Pr(\mathcal{H}_j|\mathbf{D}), \tag{3}$$

where the loss function is denoted by $\lambda_{i,j}$ which corresponds to reporting hypothesis i when hypothesis j is true. The loss function $\lambda_{i,j}$ is a good place to start building intuitions for the role of postselection in detection and estimation theory.

Following Chow, we will require that

$$\lambda_{i,i} < \lambda_{0,j} < \lambda_{i,j} \quad (i \neq j \neq 0), \tag{4}$$

which is interpreted as the loss for making a correct decision $\lambda_{i,i}$ ($i \neq 0$) is less than the cost of reject a decision $\lambda_{0,j}$ which is less than the cost of making a wrong decision $\lambda_{i,j}$. We relax this assumption in Sec. V, such that $\lambda_{0,j} > \lambda_{i,j}$ is possible. A good description of the mathematical and philosophical requirements of a loss function can be found in chapter 2 of Ref. [35].

The optimal decision is

$$\delta^*(\mathbf{D}) \equiv \arg\min_i \mathcal{R}[i|\mathbf{D}]. \tag{5}$$
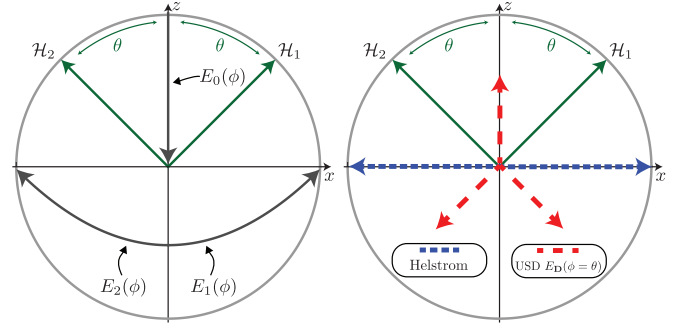

FIG. 1: The Bloch representation of the states and POVM elements involved in the state discrimination protocol. The POVM elements, $E_{\mathbf{D}}(\phi)$ are not mixed states, but subnormalized rank-1 operators, which lie on a circle at a lower level in a cone of positive operators. The grey lines on the left figure are the arc of the POVM elements as ϕ is varied in Eq. (19) from 0 to $\pi/2$. The right figure is illustrates two special cases of the POVM elements $E_{\mathbf{D}}(\phi)$. When $\phi = \pi/2$ there are only two POVM elements and the measurement is the Helstrom measurement. When $\phi = \theta$ we recover the USD measurement.

When we turn our attention to quantum hypothesis testing we will need to determine the optimal measurement to pair with this optimal decision rule. The criterion for optimal we adopt will require us to minimize the average of the posterior risk

$$\mathcal{R}[\delta(\mathbf{D})] = \sum_{\mathbf{D}}\sum_j \lambda_{\delta(\mathbf{D}),j}\Pr(\mathcal{H}_j|\mathbf{D})\Pr(\mathbf{D}), \tag{6a}$$

$$= \sum_{\mathbf{D}}\sum_j \lambda_{\delta(\mathbf{D}),j}\Pr(\mathbf{D}|\mathcal{H}_j)\Pr(\mathcal{H}_j), \tag{6b}$$

over the distribution of data and the measurement. When we assume the optimal decision is being used we denote the total risk as $\mathcal{R}^* = \mathcal{R}[\delta^*(\mathbf{D})]$.

To simplify or analysis we will consider binary hypothesis testing (i.e $\mathcal{H}_1$ vs $\mathcal{H}_2$) and take

$$\begin{aligned}\lambda_{1,1} &= \lambda_{2,2} = 0,\\ \lambda_{1,2} &= \lambda_{2,1} = 1,\\ \lambda_{0,1} &= \lambda_{0,2} = \lambda,\end{aligned} \tag{7}$$

which we call the "0-1-λ" loss function. For the 0-1-λ loss function the conditional risks for decisions i are

$$\begin{aligned}\mathcal{R}[2|\mathbf{D}] &= 1 - \Pr(\mathcal{H}_2|\mathbf{D}),\\ \mathcal{R}[1|\mathbf{D}] &= 1 - \Pr(\mathcal{H}_1|\mathbf{D}),\\ \mathcal{R}[0|\mathbf{D}] &= \lambda,\end{aligned} \tag{8}$$

where we have used $\Pr(\mathcal{H}_1|\mathbf{D}) + \Pr(\mathcal{H}_2|\mathbf{D}) = 1$.

Thus our decision rule $\delta^*(\mathbf{D})$ is

$$\delta^*(\mathbf{D}) = \begin{cases} 2 & \text{if } \mathcal{R}[2|\mathbf{D}] < \mathcal{R}[1|\mathbf{D}] \text{ and } \mathcal{R}[0|\mathbf{D}]\\ 1 & \text{if } \mathcal{R}[1|\mathbf{D}] < \mathcal{R}[2|\mathbf{D}] \text{ and } \mathcal{R}[0|\mathbf{D}]\\ 0 & \text{otherwise} \end{cases} \tag{9}$$

With respect to the posterior probabilities we find

$$\delta^*(\mathbf{D}) = \begin{cases} 2 & \text{if } \Pr(\mathcal{H}_2|\mathbf{D}) \geq 1-\lambda \text{ and } \Pr(\mathcal{H}_1|\mathbf{D}) \\ 1 & \text{if } \Pr(\mathcal{H}_1|\mathbf{D}) \geq 1-\lambda \text{ and } \Pr(\mathcal{H}_2|\mathbf{D}) \\ 0 & \text{otherwise} \end{cases}$$
$$(10)$$

In words, the decision rule is as follows: find the largest posterior probability; if it is greater than or equal to the threshold $1-\lambda$, report it; if it is less than $1-\lambda$ report "reject". Now we connect this decision theoretic framework to quantum hypothesis testing.

III. STATE DISCRIMINATION

In quantum theory the statistics of measurements are described by a positive operator valued measure (POVM) $\{E_\mathbf{D}\}$, the elements of which sum to the identity: $\sum_\mathbf{D} E_\mathbf{D} = \mathbb{I}$. The number of elements of a POVM is the number of outcomes of the measurement. To match this with our previous terminology the outcomes of the measurement are the data $\mathbf{D}$. In order to encompass both USD and Helstrom measurements we must consider a three-outcome POVM $E_\mathbf{D}$ where $\mathbf{D} \in \{0,1,2\}$. Let us make the following symmetry assumptions to make the discussion less cumbersome:

$$\Pr(\mathcal{H}_1) = \Pr(\mathcal{H}_2), \tag{11a}$$
$$\Pr(E_1) = \Pr(E_2), \tag{11b}$$
$$\Pr(E_1|\mathcal{H}_1) = \Pr(E_2|\mathcal{H}_2), \tag{11c}$$
$$\Pr(E_1|\mathcal{H}_2) = \Pr(E_2|\mathcal{H}_1), \tag{11d}$$
$$\Pr(E_0|\mathcal{H}_1) = \Pr(E_0|\mathcal{H}_2). \tag{11e}$$

These symmetries are implied, for example, by the states and operators in Fig. 1.

Utilizing some of these these symmetries the total risk in Eq. (6b) becomes

$$\mathcal{R} = \tfrac{1}{2}[(\lambda_{\delta(0),1} + \lambda_{\delta(0),2})\Pr(E_0|\mathcal{H}_1)+$$
$$(\lambda_{\delta(1),1} + \lambda_{\delta(2),2})\Pr(E_1|\mathcal{H}_1)+$$
$$(\lambda_{\delta(2),1} + \lambda_{\delta(1),2})\Pr(E_2|\mathcal{H}_1)]. \tag{12}$$

Next we use the optimal decision rule, Eq. (9) or Eq. (10), and more of the symmetries to massage this expression. Further, we assume that $\lambda < 1/2$; as for $\lambda \geq 1/2$ one can always randomly choose to report $\mathcal{H}_1$ or $\mathcal{H}_2$ and reduce the expected risk (in Sec. V we will relax this assumption). Equation (11e) implies $\Pr(\mathcal{H}_1|E_0) = \Pr(\mathcal{H}_2|E_0) = 1/2$, thus the lowest conditional risk i.e. Eq. (8) implies that the optimal decision for $\mathbf{D} = 0$ is $\delta^*(0) = 0$ always. Also $\lambda_{\delta^*(1),1} = \lambda_{\delta^*(2),2}$ and $\lambda_{\delta^*(2),1} = \lambda_{\delta^*(1),2}$ are implied by symmetry as well. Using these relations we obtain

$$\mathcal{R}^* = \lambda_{0,1}\Pr(E_0|\mathcal{H}_1) + \lambda_{\delta^*(1),1}\Pr(E_1|\mathcal{H}_1)+$$
$$\lambda_{\delta^*(2),1}\Pr(E_2|\mathcal{H}_1). \tag{13}$$

Recall from Eq. (7) that $\lambda_{0,1} = \lambda$. Using this and Bayes rule we obtain

$$\mathcal{R}^* = 2[\lambda\Pr(\mathcal{H}_1|E_0)\Pr(E_0) + \lambda_{\delta^*(1),1}\Pr(\mathcal{H}_1|E_1)\Pr(E_1)$$
$$+ \lambda_{\delta^*(2),1}\Pr(\mathcal{H}_1|E_2)\Pr(E_2)]. \tag{14}$$

Then using $\Pr(E_0) = 1 - \Pr(E_1) - \Pr(E_2) = 1 - 2\Pr(E_1)$ we have

$$\mathcal{R}^* = 2\{\tfrac{1}{2}\lambda[1 - 2\Pr(E_1)]+ \tag{15}$$
$$[\lambda_{\delta^*(1),1}\Pr(\mathcal{H}_1|E_1) + \lambda_{\delta^*(1),2}\Pr(\mathcal{H}_2|E_1)]\Pr(E_1)\},$$

where we have used $\Pr(\mathcal{H}_1|E_2) = \Pr(\mathcal{H}_2|E_1)$ and Eq. (11 b). The term $T = [\lambda_{\delta^*(1),1}\Pr(\mathcal{H}_1|E_1) + \lambda_{\delta^*(2),1}\Pr(\mathcal{H}_1|E_2)]$ still depends on the optimal decision rule so we must explictly use it. It is important to note that we can't assume $\delta^*(1) = 1$, this means we must consider two cases ($\delta^*(1) = 2$ is obviously ruled out by symmetry): (1) $\delta^*(1) = 0$: this implies $T = \lambda[\Pr(\mathcal{H}_1|E_1) + \Pr(\mathcal{H}_2|E_1)] = \lambda$; or (2) $\delta(1) = 1$: this implies $T = \Pr(\mathcal{H}_2|E_1)$. Using the optimal decision rule, the risk becomes

$$\mathcal{R}^* = \begin{cases} \lambda\Pr(E_0|\mathcal{H}_2) + \Pr(E_1|\mathcal{H}_2) & \text{if } \Pr(\mathcal{H}_2|E_1) \leq \lambda \\ \lambda & \text{otherwise} \end{cases}.$$
$$(16)$$

Equivalently this can be written as

$$\mathcal{R}^* = \lambda + \min\{0, \Pr(E_1|\mathcal{H}_2) - \lambda[1 - \Pr(E_0|\mathcal{H}_2)]\} \tag{17}$$

The above risk is true for the 0-1-λ loss function and any two hypotheses and measurements satisfying the symmetry conditions. The first term represents the part of the expected risk when a rejection is made. The second term is not yet optimized over the possible measurements.

As a specific example, here we will consider the problem of discriminating the following two quantum states:

$$\mathcal{H}_1: \quad |\Psi_1\rangle = \cos\tfrac{\theta}{2}|0\rangle + \sin\tfrac{\theta}{2}|1\rangle, \tag{18a}$$
$$\mathcal{H}_2: \quad |\Psi_2\rangle = \cos\tfrac{\theta}{2}|0\rangle - \sin\tfrac{\theta}{2}|1\rangle, \tag{18b}$$

where $0 \leq \theta \leq \pi/2$, $|\langle\Psi_2|\Psi_1\rangle| = \cos\theta$ and the prior probabilities are $\Pr(\mathcal{H}_1) = \Pr(\mathcal{H}_2) = 1/2$.

The symmetry we imposed in Eq. (11), imply the measurement is in fact a generalized measurement with POVM elements

$$E_2(\phi) = \frac{1}{2\cos^2\tfrac{\phi}{2}} \begin{pmatrix} \sin^2\tfrac{\phi}{2} & -\sin\tfrac{\phi}{2}\cos\tfrac{\phi}{2} \\ -\sin\tfrac{\phi}{2}\cos\tfrac{\phi}{2} & \cos^2\tfrac{\phi}{2} \end{pmatrix},$$

$$E_1(\phi) = \frac{1}{2\cos^2\tfrac{\phi}{2}} \begin{pmatrix} \sin^2\tfrac{\phi}{2} & \sin\tfrac{\phi}{2}\cos\tfrac{\phi}{2} \\ \sin\tfrac{\phi}{2}\cos\tfrac{\phi}{2} & \cos^2\tfrac{\phi}{2} \end{pmatrix}, \tag{19}$$

$$E_0(\phi) = \begin{pmatrix} 1 - \tan^2\tfrac{\phi}{2} & 0 \\ 0 & 0 \end{pmatrix},$$

such that $E_2(\phi) + E_1(\phi) + E_0(\phi) = \mathbb{I}$. When $\phi = \pi/2$ we get $E_0 = 0, E_1 = |+\rangle\langle+|, E_2 = |-\rangle\langle-|$ (where $|\pm\rangle$
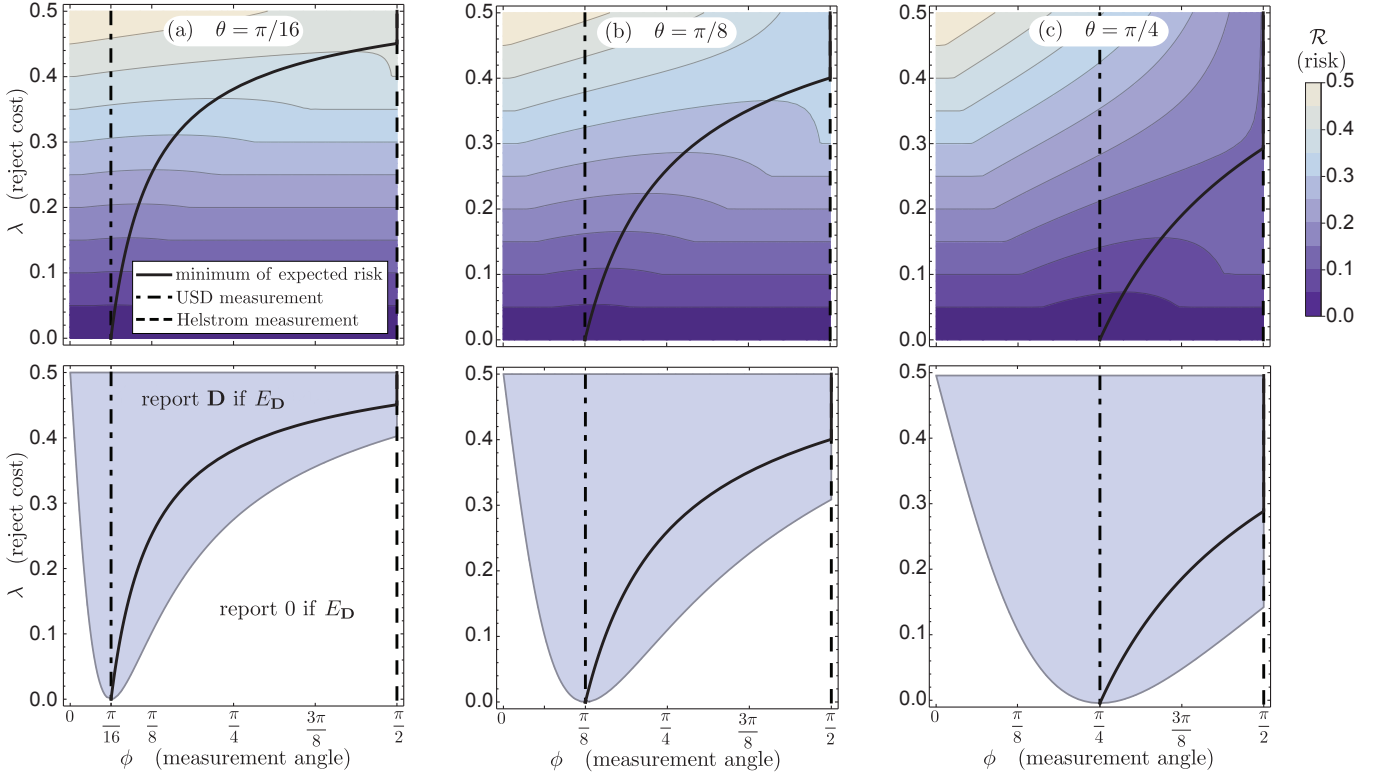
FIG. 2: Expected risk $\mathcal{R}$ (row 1) and decision rule (row 2) for the 0-1-λ loss function. In all figures the abscissa is ϕ (the measurement angle) and the ordinate is λ (the cost of reporting "reject"). The dark black line is the minimum risk ($\mathcal{R}^*[\phi^*]$) for a given λ and thus specifies the optimal measurement angle. The shaded regions in the second row are simply the region for which the expected risk is less than λ; in this region one always reports i if one obtained outcome E_i.

are the eigenstates of the Pauli X operator), which is the Helstrom measurement for all θ. When $\phi = \theta$ we obtain the USD measurement for all θ. In Fig. 1 the grey lines are the arc traced by Eq. (19) as a function of ϕ. Note that for $\phi > \pi/2$ the POVM element E_0 is not a positive operator, thus we do not allow these values of ϕ.

To apply the above decision theoretic formalism we need to compute the probabilities given in Eq. (17). All of these probabilities can be computed using the usual rule:

$$\Pr(E_{\mathbf{D}}|\mathcal{H}_i,\phi) = \langle\Psi_i| E_{\mathbf{D}}(\phi) |\Psi_i\rangle, \quad (20)$$

see footnote [36] for some examples. Notice how all of the probabilities depend on the measurement angle ϕ, this means the expected risk will also be a function of ϕ.

Given the POVM elements in Eq. (19) the expected risk is

$$\mathcal{R}^*[\phi] = \lambda + \quad (21)$$
$$\min\left[0, \frac{(2\lambda-1)(\cos\theta\cos\phi - 1) - \sin\theta\sin\phi}{2(1+\cos\phi)}\right],$$

Intuitively this says the risk is at most λ and sometimes less. This risk is plotted in Fig. 2 as a function of λ and ϕ for particular values of θ. To find the optimal angle we fix λ and ask which ϕ minimizes $\mathcal{R}^*[\phi]$. This can

be done analytically. The trival case is when $\mathcal{R}^*[\phi] = \lambda$ an thus no optimization over ϕ is possible. The optimal measurement found by solving

$$\frac{\partial}{\partial\phi}\left[\lambda + \frac{(2\lambda-1)(\cos\theta\cos\phi - 1) - \sin\theta\sin\phi}{2(1+\cos\phi)}\right] = 0, \quad (22)$$

for ϕ. The constraint on the positivity of the measurement operators, i.e. $\phi \leq \pi/2$, results in following peicewise defintion of optimal measurement angle

$$\phi^* = \begin{cases} 2\cot^{-1}\left[(1-2\lambda)\cot\frac{\theta}{2}\right] & \text{if } \lambda < \frac{1}{2}\left(1-\tan\frac{\theta}{2}\right) \\ \frac{\pi}{2} & \text{if } \lambda \geq \frac{1}{2}\left(1-\tan\frac{\theta}{2}\right) \end{cases}. \quad (23)$$

This optimal angle is plotted as the solid black lines in Fig. 2. The decision functions plotted in the second row of Fig. 2 are particularly simple: in the shaded regions report $\mathbf{D}$ if $E_{\mathbf{D}}$ is observed and report "reject" or 0 if $E_{\mathbf{D}}$ is observed in the non shaded regions.

From Fig. 2 it is clear that, as a function of λ the optimal measurement angle interpolates between the USD and the Helstrom measurement. This can be made explicit as follows. The second branch of Eq. (23), i.e.
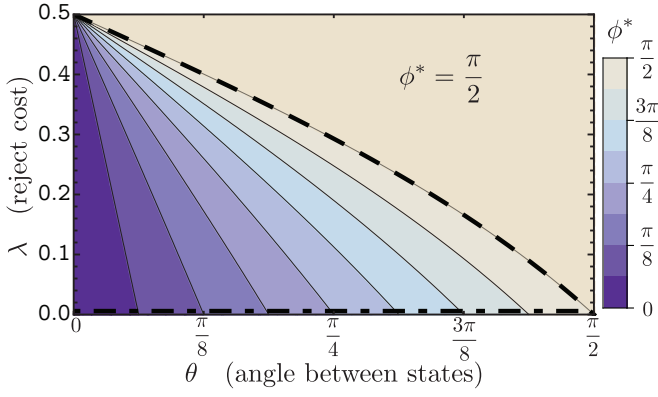
FIG. 3: The angle ϕ^* of the optimal measurement minimizing the risk for the 0-1-λ loss, i.e. Eq. (23), as a function of λ and θ. The dot dashed line at $\lambda = 0$ corresponds to the USD measurement when $\phi^* = \theta$. Above the dashed line the Helstrom measurement is optimal. The optimal angle has been discretized for ploting

.

when $\phi^* = \pi/2$, is the Helstrom measurement. To recover the USD measurement we plug $\lambda = 0$ into Eq. (23) gives $\phi^* = \theta$, so $\lambda = 0$ implies the USD measurement. However, $\lambda = 0$ is also a degenerate case where no cost is assigned to reporting "reject". Thus, the risk is *also* minimized by reporting "reject" for *any* outcome of *any* measurement or, equivelently, not bothering to make the measurement and simply reporting "reject". Recall that what we are calling *the* USD measurement is the one which minimizes the probability of obtaining the "reject" outcome in the usual paradigm. Here, as expected, the USD measurement is approached for $\lambda \to 0$. This is also when the probablity for reporting "reject" is maximized, see Fig. 5 of Sec. IV.

To complete the example we plot in Fig. 3 the optimal measurement angle ϕ^* as a function of λ and the angle between the states θ and the z axis. The USD protocol corresponds the line at $\lambda = 0$ while the Helstrom measurement is performed for when $\phi^* = \pi/2$. The area where $\phi^* = \pi/2$ is approximately half of the parameter space, i.e. $\lambda \gtrsim \frac{1}{2}(1 - \theta/2) + O(\theta^3)$, thus even when the loss function encourages postselection it is not guaranteed to be optimal.

Other studies of inconclusive state discrimination [27, 28, 30, 31] concern themselves with the probabilities of error and reporting the "reject" result. This avoids the question of what to do *given* the outcome of some measurement. Here we have phrased the problem as a decision theoretic one where the loss is incurred on the decisions and once that loss is specified, a definitive answer can be given. In real applications, it would be unlikely that an agent's decisions are constrained to be deterministic functions of measurement operators. Indeed, our results imply that loosening that constraint can only decrease the agent's risk if they can not measure at the optimal angle for a given λ.

IV. RELATIONSHIP BETWEEN RISK AND ERROR AND REJECT PROBABILITIES

So far we have focused on the decision function and the loss function. In this section we connect our approach to the previous approaches which focus on tradeoffs between reject and error probabilities [31], and rejection thresholds [30].

For equal prior probabilities the optimal decision rule when measuring at the optimal angle, is particularly simple: report $\mathbf{D}$ if $E_{\mathbf{D}}$. Let probability of making the correct decision be C, the probability of error be E, the probability of rejection be R and the probability that a piece of data is accepted be A. These probabilities can be written explicitly as follows:

$$\Pr(C|\theta, \lambda) = \sum_{i \in \{1,2\}} \Pr(\mathcal{H}_i) \Pr[E_i(\phi^*)|\Psi_i], \tag{24a}$$

$$\Pr(E|\theta, \lambda) = \sum_{i,j \in \{1,2\}, i \neq j} \Pr(\mathcal{H}_i) \Pr[E_j(\phi^*)|\Psi_i], \tag{24b}$$

$$\Pr(R|\theta, \lambda) = \sum_{i \in \{1,2\}} \Pr(\mathcal{H}_i) \Pr[E_0(\phi^*)|\Psi_i], \tag{24c}$$

$$\Pr(A|\theta, \lambda) = \Pr(C|\theta, \lambda) + \Pr(E|\theta, \lambda), \tag{24d}$$

These probabilities obey $\Pr(E) + \Pr(C) + \Pr(R) = 1$ which implies $\Pr(A) + \Pr(R) = 1$.

In Fig. 4 we plot these probabilities as a function of the angle θ between the states. A strategy without postselection adheres to the lines of Fig. 4 when $\lambda = 0$. Deviating from this behavior indicates postselection. Notice that as $\theta \to 0$ $\Pr(R) \to 1$ for all λ except $\lambda = 0.5$. While, in Fig. 5 we plot the error probability and reject probability as a function of the rejection threshold. Postselection occurs whenever $\Pr(A) < 1$. Notice that as λ approaches 0, the probablilty of rejection gets closer to 1 for all values of θ.

In 1970 Chow [34] showed a particularly simple relationship between the error probabilities and the minimum risk under the optimal decision rule

$$\mathcal{R}^*[\phi^*] = \Pr(E|\theta, \lambda) + \lambda \Pr(R|\theta, \lambda), \tag{25a}$$

$$= \int_0^\lambda d\lambda' \Pr(R|\theta, \lambda', \phi). \tag{25b}$$

Both of these expressions can be visualized graphically, see Fig. 6. Prior to our work the expression given in Eq. (25) (a) is one of the ways the loss function has been explained, see e.g. [31]. It is important that the optimal decision rule and measurement angle is used otherwise the risk will generally be different to the above risk.

It turns out that $\Pr(E)$ can be derived from $\Pr(R)$ for a particular rejection threshold. Chow [34] has shown that the Stieltjes integral of λ with respect to $\Pr(R|\theta, \lambda)$ is precisely the error probability

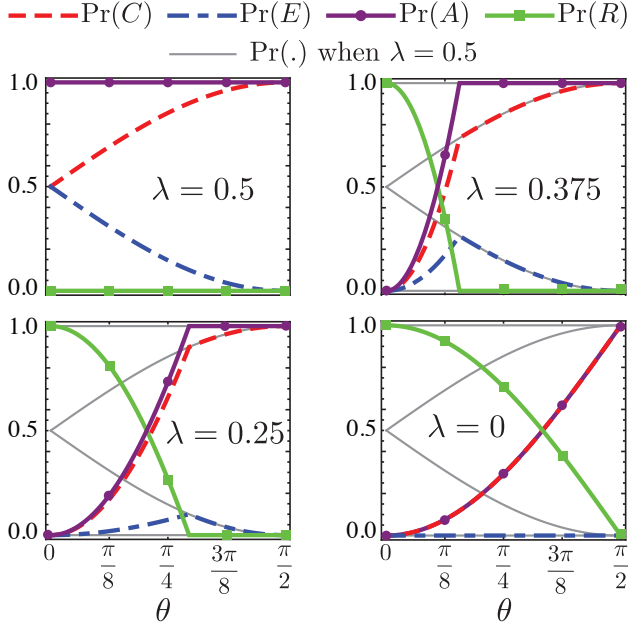$$\Pr(E|\theta, \lambda) = - \int_0^\lambda \lambda' \, d\Pr(R|\theta, \lambda'). \tag{26}$$

FIG. 4: The probabilities in Eq. (24) as a function of the angle between the states θ. When $\lambda = 0.5$ it is easy to show that $\Pr(C|\theta) = 1 - \Pr(E|\theta) = (1/2)(1 + \sin\theta)$, $\Pr(A|\theta) = 1$, and $\Pr(R) = 0$ as plotted in the top left plot. These lines are the gray lines on the other figures. Generically as $\theta \to 0$ probability for reporting "don't know" indeed approaches one except when $\lambda = 0.5$. When the equality in the second branch of Eq. (23) is satisfied we see the measurement switches from one with an inconclusive outcome to the Helstrom measurement i.e. $\Pr(A) = 1$ and $\Pr(R) = 0$ and $\Pr(C|\theta) = 1 - \Pr(E|\theta) = (1/2)(1 + \sin\theta)$.
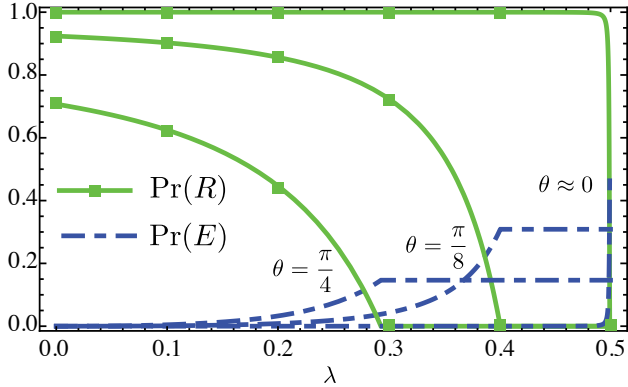
FIG. 5: The rejection and error probabilities as a function of λ. When $\lambda = 0$ the measurement strategy is precisely the USD measurement and the rejection probability attains its maximum $\Pr(R) = \cos\theta$. Now consider the values of λ for which $\Pr(R) = 0$. For example when $\theta = \pi/8$, $\Pr(R) = 0$ when $\lambda \in [0.4, 0.5]$. As λ is decreased the probability of reject increases and probability of error decreases with diminishing returns.
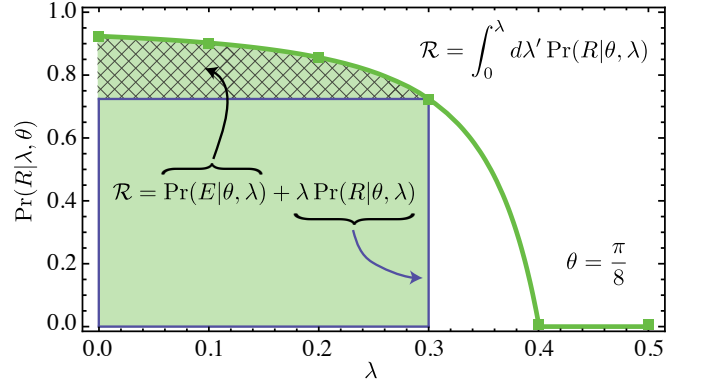

FIG. 6: The relationship between risk and probability for rejection. The rejection probability is plotted as a function of the rejection threshold λ when $\theta = \pi/8$. Consider a rejection threshold of $\lambda = 0.3$, given this threshold and the angle between the states the expected risk can be computed from Eq. (22) to be $\mathcal{R} \approx 0.26$. Equation 25 (b) shows this equivalent to the (shaded) area under the curve up to the rejection threshold. The area under the curve can be decomposed into a rectangle with height $\Pr(R|\theta, \lambda) \approx 0.724$ and width $\lambda = 0.3$ so $\lambda \Pr(R|\theta, \lambda) \approx 0.2172$ the integral given Eq. (27) results in $\Pr(E|\theta, \lambda) \approx 0.0428$ and thus $\mathcal{R}^* = \Pr(E|\theta, \lambda) + \lambda \Pr(R|\theta, \lambda)$.

As noted by Chow, this expression is suggestive of an error probability-reject probability tradeoff relation, see Fig. 7. If $\Pr(R|\theta, \lambda)$ is differentiable with respect to λ then the Stieltjes integral reduces to the Riemann integral

$$\Pr(E|\theta, \lambda) = -\int_0^\lambda \lambda' \left[\frac{d}{d\lambda'} \Pr(R|\theta, \lambda') \right] d\lambda'. \quad (27)$$

From Eq. (26) and Eq. (27) it is clear that the slope of the error-reject tradeoff curve in Fig. 7 is exactly value of the rejection threshold. Consequently the tradeoff is most effective initially and is less rewarding as the desired errror decreases. In Fig. 7 we also see that specifying a particular rejection threshold, e.g. $\Pr(R) = Q$ as in [30], implies a value for λ and $\Pr(E)$ (once θ is fixed).

V. THE 0-λ_E-λ_R LOSS FUNCTION

Here we generalize the 0-1-λ loss function to the 0-λ_E-λ_R loss function, where λ_E is the cost of reporting the incorrect decision and λ_R is the cost of reporting reject –i.e.,

$$\begin{aligned} \lambda_{1,1} &= \lambda_{2,2} = 0, \\ \lambda_{1,2} &= \lambda_{2,1} = \lambda_E, \\ \lambda_{0,1} &= \lambda_{0,2} = \lambda_R. \end{aligned} \quad (28)$$

For the 0-λ_E-λ_R loss function in Eq. (28) the conditional risks for decisions i are

$$\begin{aligned} \mathcal{R}[2|\mathbf{D}] &= \lambda_E[1 - \Pr(\mathcal{H}_2|\mathbf{D})], \\ \mathcal{R}[1|\mathbf{D}] &= \lambda_E[1 - \Pr(\mathcal{H}_1|\mathbf{D})], \\ \mathcal{R}[0|\mathbf{D}] &= \lambda_R, \end{aligned} \quad (29)$$
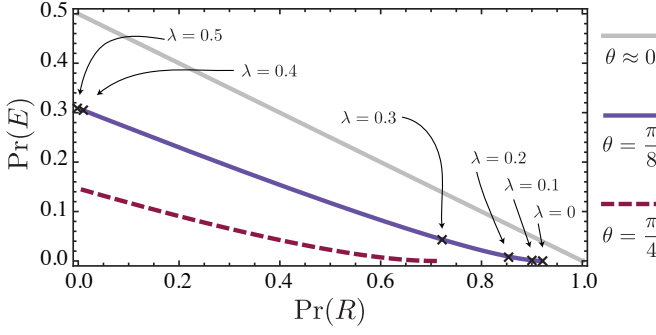
FIG. 7: Error-reject tradeoff curve. In fact the derivative of $\Pr(E)$ with respect to $\Pr(R)$ is λ. These curves are implicit functions of λ. The trade off is not linear in the rejection threshold λ. This is evident on the line corresponding to $\theta = \pi/8$ where six crosses corresponding to $\lambda \in [0, 0.1, 0.2, 0.3, 0.4, 0.5]$ are plotted.
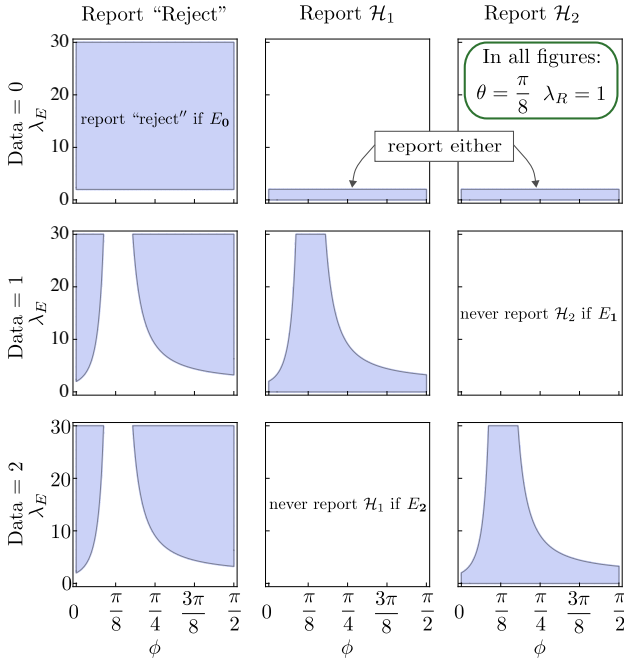

FIG. 8: Decision regions for the 0-λ_E-λ_R loss function. In all figures the angle between the states is $\theta = \pi/8$ and the reject loss was chosen to be $\lambda_R = 1$. The shaded regions should be interpreted as report the column heading. In row one, the reporting of a hypothesis given the inconclusinve outcome is a result of Eq. (30). Evidently, as λ_E becomes large the decision rule becomes more like unambigous state discrimination.

The following analysis assumes the same states [Eq. (18a)], prior probablities [$\Pr(\mathcal{H}_i) = 1/2$], and measurements [Eq. (19)], as before. Of particular interest is the case when the measurement outcome $E_0(\phi)$ is obtained, i.e. $\mathbf{D} = 0$, then the conditional risks are

$$
\begin{aligned}
\mathcal{R}[2|0] &= \lambda_E/2, \\
\mathcal{R}[1|0] &= \lambda_E/2, \\
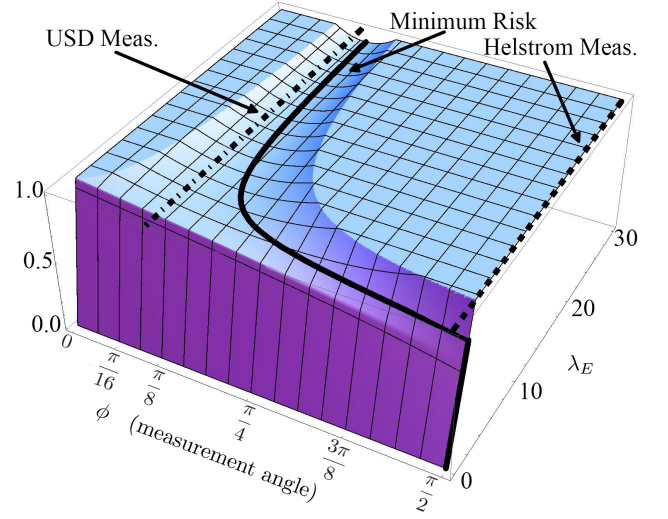\mathcal{R}[0|0] &= \lambda_R.
\end{aligned}
\tag{30}
$$

FIG. 9: Risk as a function of measurement angle ϕ and the cost of reporting the wrong decision λ_E for the 0-λ_E-λ_R loss function. Here $\theta = \pi/8$ and the reject loss was chosen to be $\lambda_R = 1$. For $\lambda_E < 2.5$ we see the optimal measurement is the Helstrom measurement and as $\lambda_E \to \infty$ the optimal measurement approaches the USD measurement.

Thus if $\lambda_R > \lambda_E/2$ we should never reject, instead we should report either hypothesis, as illustrated in row 1 of Fig. 8. In Fig. 8 we have chosen $\lambda_R = 1$ so that for all $\lambda_E \leq 2$ we must report either hypothesis to minimize our risk. In particular if we perform the a measurement with an inconclusive outcome $\phi < \pi/2$ and obtain the inconclusive outcome E_0 we should randomly choose between reporting $\mathcal{H}_1$ and $\mathcal{H}_2$. For $\lambda_R < \lambda_E/2$ we find

$$
\delta(\mathbf{D}) = \begin{cases} 2 & \text{if } \Pr(\mathcal{H}_2|\mathbf{D}) \geq 1 - \frac{\lambda_R}{\lambda_E} \text{ and } \Pr(\mathcal{H}_1|\mathbf{D}) \\ 1 & \text{if } \Pr(\mathcal{H}_1|\mathbf{D}) \geq 1 - \frac{\lambda_R}{\lambda_E} \text{ and } \Pr(\mathcal{H}_2|\mathbf{D}) \\ 0 & \text{otherwise} \end{cases} .
\tag{31}
$$

In words, the decision rule is as follows: find the largest posterior probability; if it is greater than or equal to the threshold $1 - \frac{\lambda_R}{\lambda_E}$, report it; if it is less than $1 - \frac{\lambda_R}{\lambda_E}$, report "reject".

Like the 0-1-λ loss function, the 0-λ_E-λ_R loss function also interpolates between the Helstrom measurement and unambiguous state discrimination, as illustrated in Fig. 9. Notice, for both loss functions, we did not need to "normalize" the loss function or add additional contraints such as $\Pr(R) = 0$ or $\Pr(E) = 0$, unlike other approaches [31].

VI. DISCUSSION

In the ongoing debate about postselection for information theoretic tasks in quantum theory, we have given a plausible example where postselection is a feature of the

optimal solution. We say plausible because the loss function on the decisions was not tailored to favor full-blown postselection—the solution was not obvious.

In Sec. III we have shown that USD measurements only arise in the limit when the cost assigned to discarding data is exactly zero, which corresponds to the line $\lambda = 0$ for all θ in Fig. 3. In contrast, the Helstrom measurement appears to be the natural measurement for approximately half of the paramter space $\lambda \gtrsim \frac{1}{2}(1 - \theta/2)$. For the remainder of the parameter space, i.e. $\lambda \lesssim \frac{1}{2}(1-\theta/2)$, strategies involving postselection (that are not USD) are optimal. In Sec. IV we unified three seemingly separate approaches, namely the decision theoretic approach (i.e. our 0-1-λ loss function), the rejection threshold approach [30], and the probability tradeoff approach [31]. Section V highlighted that the decision function can not simply be ignored—in some situations it is better to report an answer even if the inconclusive outcome was obtained.

It is natural to ask what the implications of our analysis are. In practical situations it could be desirable to reduce errors by rejecting some data, but excessive rejection is required to reduce error to zero. And, at the point where the error is zero one can equivalently reject without bothering to perform any experiment, as the cost of rejection is also zero. Generally this implies when a loss function is specified as conditional on some event being successful that this is equivalent to assigning *cost* to a rejection option. Again, if the cost of rejection is zero why should you bother to perform the experiment at all? We have suggested a sensible approach is to embed a postselection protocol into a class of protocols which assign loss for discarding data, this makes clear the price of postselection.

For example, consider offline magic state distillation for quantum computation [37]. The success probability is relevant for quantifying efficiency (or expected yield in Sec. VI. of [38]) of the magic state distillation routine. When the success probability for the scheme is too small then the overall distillation routine is inefficient, even if it performs very well when it does succeed. This is generically true in offline state preparation. If costs are low, we are happy to wait for some time for a state to be prepared. But the cost are not zero, as we actually want to make a state and perform a useful task.

The virtue of the decision theoretic approach is that all the assumptions, constraints and figures of merit are made explicit at the outset—the rest is derived. Thus, within this framework it is quite natural to include new constraints and features. For example, if experimental noise or inaccuracies or constraints are of concern, one must include those at the highest level—that is, they must be specified in the initial states, POVM, or loss function. Questions of robustness or imperfections, which plague other approaches, are simply a category mistake to ask here.

A number of open questions remain. The first class of questions are about extensions to the specific ideas developed in this manuscript. A simple modification is when Alice makes collective measurements on N copies of $|\Psi_1\rangle$ or $|\Psi_2\rangle$. In this case the states look more orthogonal because $|\langle\Psi_1|\Psi_2\rangle|^{2N} \leq |\langle\Psi_1|\Psi_2\rangle|^2$. Based on our results in Fig. 3 we conjecture that the optimal joint measurement for the 0-1-λ loss function will look closer to a Helstrom measurement than the USD measurement. The obvious question is: does a bound on the N copy risk exist? Ideally the solution would be something like the quantum Chernoff bound [39] which bounds the minimum error probability asymptotically in N (i.e. the risk of the 0/1 loss function).

The second class of questions are about the role of postselection in quantum information tasks. Although we have conjured an exotic loss function for which the optimal strategy includes postselection, it is not tied explicitly to an existing operational task. Nevertheless we suggest that our decision theoretic approach should be taken for any practical state discrimination (or estimation) problem which allows for the possibility of postselection. Extending our approach to parameter estimation seems to be the next great challenge. The results in this manuscript add weight to our suggested loss function [9]: report "reject" and incur loss λ for mean squared error (MSE) above some threshold and incur the MSE loss below that threshold.

[1] S. Aaronson, *Quantum computing, postselection, and probabilistic polynomial-time*, Proc. R. Soc. A **461**, no. 2063, 3473 (2005).

[2] J. L. O'Brien, G. J. Pryde, A. G. White, T. C. Ralph, and D. Branning, *Demonstration of an all-optical quantum controlled-NOT gate*, Nature **426**, 264 (2003).

[3] B. Gendra, E. Ronco-Bonvehi, J. Calsamiglia, R. Muñoz-Tapia, and E. Bagan, *Quantum metrology assisted by abstention*, Phys. Rev. Lett. **110**, 100501 (2013).

[4] P. B. Dixon, D. J. Starling, A. N. Jordan, and J. C. Howell, *Ultrasensitive Beam Deflection Measurement via Interferometric Weak Value Amplification*, Phys. Rev. Lett. 102, 173601 (2009).

[5] G. C. Knee, Andrew, S. C. Benjamin, and E. M. Gauger, *Quantum sensors based on weak-value amplification cannot overcome decoherence*, Phys. Rev. A **87**, 012115 (2013).

[6] S. Tanaka and N. Yamamoto, *Information amplification via postselection: A parameter-estimation perspective*, Phys. Rev. A **88**, 042116 (2013).

[7] C. Ferrie and J. Combes, *Weak Value Amplification is Suboptimal for Estimation and Detection*, Phys. Rev. Lett. **112**, 040406 (2014); in particular, see the Supplementary Material.

[8] G. C. Knee and E. M. Gauger, *When Amplification with Weak Values Fails to Suppress Technical Noise*, Phys. Rev. X **4**, 011032 (2014) .

[9] J. Combes, C. Ferrie, Z. Jiang, C. M. Caves, *Quantum limits on postselected, probabilistic quantum metrology*, Phys. Rev. A 89, 052117 (2014).

[10] L. Zhang, A. Datta, I. A. Walmsley, *Precision metrology using weak measurements*, arXiv:1310.5302.

[11] S. Pandey, Z. Jiang, J. Combes, C. M. Caves, *Quantum limits on probabilistic amplifiers*, Phys. Rev. A **88**, 033852 (2013).

[12] G. C. Knee, J. Combes, C. Ferrie, E. M. Gauger, *Weak-value amplification: state of play*, arXiv:1410.6252

[13] A. N. Jordan, J. Martínez-Rincón, and J. C. Howell, *Technical Advantages for Weak-Value Amplification: When Less Is More*, Phys. Rev. X 4, 011031 (2014).

[14] S. Pang, J. Dressel, T. A. Brun, *Entanglement-assisted weak value amplification*, Phys. Rev. Lett. **113**, 030401 (2014).

[15] J. Calsamiglia, B. Gendra, R. Munoz-Tapia, E. Bagan, *Probabilistic metrology defeats ultimate deterministic bound*, arXiv:1407.6910

[16] S. Pang, T. A. Brun, *Improving the precision of weak measurements by postselection*, arXiv:1409.2567

[17] Y. Susa, S. Tanaka, *Statistical Hypothesis Testing by Weak-Value Amplification: Proposal and Evaluation*, arXiv:1502.06334.

[18] J.A. Bergou, U. Herzog, M. Hillery, *Discrimination of Quantum States*, Lect. Notes Phys. **649**, 417 (2004).

[19] Stephen M. Barnett and Sarah Croke, *Quantum state discrimination*, Advances in Optics and Photonics **1**, Issue 2, 238 (2009).

[20] C. W Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).

[21] C. A. Fuchs, Ph.D. thesis, The University of New Mexico, Albuquerque, NM, 1996.

[22] I. D. Ivanovic, *How to differentiate between non-orthogonal states*, Phys. Lett. A **123**, 257 (1987).

[23] D. Dieks, *Overlap and distinguishability of quantum states*, Phys. Lett. A **126**, 303 (1988).

[24] A. Peres, *How to differentiate between non-orthogonal states*, Phys. Lett. A 128, 19 (1988).

[25] A. Chefles and S. M. Barnett, *Strategies for discriminating between non-orthogonal quantum states*, J. Mod. Opt. **45**, 1295 (1998).

[26] C.-W. Zhang, C.-F. Li, and G.-C. Guo, *General strategies for discrimination of quantum states*, Phys. Lett. A **261**, 25 (1999).

[27] M. A. P. Touzel, R. B. A. Adamson, and A. M. Steinberg, *Optimal bounded-error strategies for projective measurements in nonorthogonal-state discrimination*, Phys. Rev. A **76**, 062314 (2007).

[28] A. Hayashi, T. Hashimoto, and M. Horibe, *State discrimination with error margin and its locality*, Phys. Rev. A **78**, 012333 (2008).

[29] H. Sugimoto, T. Hashimoto, M. Horibe, and A. Hayashi, Phys. Rev. A **80**, 052322 (2009).

[30] E. Bagan, R. Muñoz-Tapia, G. A. Olivares-Rentería, and J. A. Bergou, *Optimal discrimination of quantum states with a fixed rate of inconclusive outcomes*, Phys. Rev. A 86, 040303(R) (2012).

[31] J. Dressel, T. A. Brun, and A. N. Korotkov, *Violating the modified Helstrom bound with nonprojective measurements*, Phys. Rev. A 91, 040301(R) (2014).

[32] S. Croke, E. Andersson, S. M. Barnett, C. R. Gilson, and J. Jeffers, *Maximum confidence quantum measurements*, Phys. Rev. Lett. **96**, 070401 (2006).

[33] C. K. Chow, *An optimum character recognition system using decision functions*, IEEE Trans. on Electronic Computers **EC-6**, 247 (1957).

[34] C. K. Chow, *On optimum recognition error and reject tradeoff*, IEEE Trans. on Info. Theory **16**, 41 (1970).

[35] J. O. Berger, *Statistical decision theory and Bayesian analysis*, (Springer Science , 1985).

[36] Some examples are

$$\mathrm{Pr}(E_\mathbf{2}|\mathcal{H}_1,\phi) = \frac{1 - \cos(\theta - \phi)}{2[1 + \cos(\phi)]}, \ \mathrm{Pr}(E_\mathbf{1}|\mathcal{H}_1,\phi) = \frac{1 - \cos(\theta + \phi)}{2[1 + \cos(\phi)]},$$
$$\text{and } \mathrm{Pr}(E_\mathbf{0}|\mathcal{H}_1,\phi) = \cos^2(\theta/2)\cos(\phi)\sec^2(\phi/2),$$

which infact specify all the probablities due to the symmetry assumptions in Eq. (11).

[37] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, Phys. Rev. A 71, 022316 (2005).

[38] E. T. Campbell, H. Anwar, and D. E. Browne, *Magic-State Distillation in All Prime Dimensions Using Quantum Reed-Muller Codes*, Phys. Rev. X 2, 041021 (2012).

[39] K.M.R. Audenaert, J. Calsamiglia, Ll. Masanes, R. Munoz-Tapia, A. Acin, E. Bagan, F. Verstraete, *The Quantum Chernoff Bound*, Phys. Rev. Lett. **98**, 160501 (2007).