Finite de Finetti theorem for conditional probability distributions describing physical theories

Matthias Christandl[*]
Centre for Quantum Computation, DAMTP, University of Cambridge, Cambridge CB3 0WA, UK

Ben Toner[†]
Centrum voor Wiskunde en Informatica, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands and
Institute for Quantum Information, California Institute of Technology, Pasadena CA 91125, USA

We work in a general framework where the state of a physical system is defined by its behaviour under measurement, and the relation between different systems is constrained by a no-signalling principle. We characterize symmetric states in such theories, showing that their marginals can be approximated by convex combinations of independent and identical conditional probability distributions. This generalizes the classical finite de Finetti theorem of Diaconis and Freedman. Our results have application to the foundations of physics, quantum cryptography, and the study of classical channels. In particular, they apply to correlations obtained from quantum states even when there is no bound on the local dimension, so that known finite quantum de Finetti theorems cannot be used.

The finite de Finetti theorem states that taking a random sample from a set of objects is approximately the same as sampling independent and identically distributed (i.i.d.) random variables, as long as the sample size is small compared to the total number of objects in the set. This is intuitively clear, but the error of the approximation was not quantified until 1980 [1]. We prove the analogous statement when the objects in the set are not classical particles, but particles in a more general, probabilistic, physical theory, which we analyse in the convex sets framework (see [2, 3]).

We now give a brief description of the setting and our results; precise definitions are given later on. A physical system in a probabilistic physical theory is made up of a number of—in our case identical—subsystems, called *particles*. On each particle different measurements from a set $\mathcal{X}$ can be performed and outputs from a set $\mathcal{A}$ are obtained. The state of a particle is specified by a *conditional probability distribution* $P[A|X]$: the probability of obtaining result a when performing measurement x is given by $P[A=a|X=x]$. The set of allowed states Ω is assumed to be convex (so that probabilistic mixtures of states are valid states) and there is a rule for determining which states of n particles are valid. Different rules lead to different theories; they have in common, however, a *no-signalling* property, which ensures that the reduced state on a subset of the particles is always well-defined.

Our main result is that the joint state $P[A^k|X^k] = P[A_1\cdots A_k|X_1\cdots X_k]$ of k particles randomly chosen from n particles—or equivalently, the state of the first k particles of a permutation-invariant state of n particles—can be approximated by a convex combination of identical and independent conditional probability distributions,

$$P[A^k|X^k] \approx \int dm(\lambda) P_\lambda[A|X]^{\otimes k} \qquad (1)$$

and that the error in the approximation is bounded by $|\mathcal{X}|k(k-1)/n$ in the appropriate distance measure, where $|\mathcal{X}|$ is the number of different possible measurements [28]. Our result generalizes the finite de Finetti theorem of Diaconis and Freedman, who proved for classical probability distributions ($|\mathcal{X}|=1$) that the error in the approximation is no more than $k(k-1)/n$ [1][29].

This work is motivated by recent work on finite *quantum* de Finetti theorems, i.e., statements of the form

$$\rho^k \approx \int d\sigma\ \sigma^{\otimes k}, \qquad (2)$$

where ρ^k is the k-particle reduced density matrix of a permutation-invariant density matrix of n d-dimensional particles, with the error at most $4d^2k/n$ in the trace distance [4, 5][30]. Our results do not imply the quantum de Finetti theorem in full generality, but do have consequences for quantum correlations that do not follow from known quantum de Finetti theorems. In particular, our results are relevant when there is no bound on the dimension of the quantum particles (so that the error in Eq. (2) cannot be controlled) but there is a bound on the number of different measurements $|\mathcal{X}|$ (so that the approximation in Eq. (1) is good). We also prove a finite quantum de Finetti theorem for separable ρ^n, in which case there is an approximation of the form in Eq. (2) with error $k(k-1)/n$, independent of the dimension. This highlights the difference between classical correlations and entanglement in quantum states.

Applications.—This work has applications to three areas. The first is to the *foundations of physics*: Just as the classical de Finetti theorem allows a Bayesian of de Finetti's subjective school to connect with the prior probabilities of the objectivist [6], our results establish the same correspondence for measurement results in theories that obey a no-signalling principle.

The second is in *quantum cryptography*, specifically the security of Quantum Key Distribution (QKD). QKD schemes [7, 8] allow the generation of cryptographic keys that are provably secure under the assumption that quantum mechanics provides a correct description of physical reality. This presents a major advantage over classical schemes, where security relies on assumptions about the limited computational power of an eavesdropper. The role of a quantum de Finetti theorem in proving the security of a key distribution scheme is to establish that the most general attack an eavesdropper can make is only a little more powerful than a *collective* attack, where the adversary is assumed to apply the same transformation to each subsystem [9, 10]. Recently, Barrett, Hardy, and Kent have devised a scheme to produce a key bit from quantum correlations and their protocol is secure even when the eavesdropper is only restricted by a no-signalling principle and can do operations not allowed by quantum theory [11] (see also [12, 13, 14]). Although it is reasonable to assume that quantum theory is correct, such a strong security proof has advantages as it still applies when the communicating parties do not have full control over their measurement apparatus and, in particular, when they cannot control the dimension of the quantum states from which they obtain the correlations. Therefore, quantum de Finetti theorems do not directly apply, since they are necessarily dimension-dependent [5], but our results do not have this limitation.

The third application is to the study of classical channels. Up to now we have described de Finetti theorems as bounding the distance between states in a physical theory, and have not addressed operations or channels on those states. Fuchs, Schack and Scudo have used the Jamiolkowski isomorphism to transfer the infinite quantum de Finetti theorem ($n = \infty$, $k < \infty$) [15, 16, 17] to quantum channels [18]. Since a conditional probability distribution can be viewed as a classical channel with probability distributions as input and output, our results also provide a de Finetti theorem for classical channels.

Outline.—Our first task is to define an appropriate distance measure on states of k particles in probabilistic theories, in order to quantify the error in Eq. (1). The distance between states should bound the probability of distinguishing them by measurement, and so we need to be clear about what measurement strategies are allowed. One possibility, which we explore in [19], is to restrict to strategies where each of the k particles is measured individually. But when the conditional probability distributions arise from making informationally complete local measurements on entangled quantum states, the resulting distance measure fails to bound the trace distance between the quantum states. Therefore, in this Letter we work in the convex sets framework, a very general setting in which all noncontextual measurements are allowed, including all joint quantum-mechanical measurements. We describe this framework in the next section. We then state and prove our results.

Convex sets framework.—Let Ω be the set of states of a particle. We assume that Ω is convex, compact, and has affine dimension n. In probability theory, for example, Ω is the simplex of probability distributions $(\omega_1, \ldots, \omega_{n+1})$, $\omega_i \geq 0, \sum_i \omega_i = 1$, while in quantum theory, Ω is (isomorphic to) the set of positive operators ω with trace one on a Hilbert space $\mathcal{H} \cong \mathbb{C}^d$. We are particularly interested in the case where Ω is specified by a set of conditional probability distributions $\{P_\lambda[A|X]\}$, whose elements are indexed by a label λ. This is partly because quantum states can be described in this way. For instance, the state ρ of a qubit, a spin-$\frac{1}{2}$ system, is uniquely determined by the probabilities of obtaining spin up or down when it is measured along the x, y, or z axes of the Bloch sphere. Thus a qubit can be described by a conditional probability distribution $P[A|X]$ with $\mathcal{A} = \{\uparrow, \downarrow\}$ and $\mathcal{X} = \{x, y, z\}$. Not all conditional probability distributions can be obtained by making local measurements on quantum states. This led Barrett to define generalized theories [20], where the state space Ω is the set of all conditional probability distributions $\{P_\lambda[A|X]\}$, denoted $\Box$. When $|\mathcal{X}| = 1$, this reduces to classical probability theory. In fact, every Ω can be mapped to a convex subset of $\Box$ for some number of *fiducial* measurements and outcomes (e.g., one measurement and $\dim A(\Omega)$ outcomes [2, Lemma 1]).

The most general measurement that can be performed on a quantum system is a positive operator-valued measure (POVM), whose elements are termed *effects*. An effect r can be written as $r(\omega) = \mathsf{Tr}\,(R\omega)$ for some nonnegative operator R with $R \leq \mathbf{1}$, where $r(\omega)$ is the probability of obtaining the outcome associated with effect r when the state is ω. Effects in a generalized theory will be functions mapping states to probabilities, and these functions should be affine so that they are compatible with preparing convex combinations. The vector space of affine functions $a : \Omega \to \mathbb{R}$, denoted $A(\Omega)$, is isomorphic to $\mathbb{R}^{n+1}$. The cone of nonnegative affine functions on Ω is denoted $A_+(\Omega)$. The *order unit* of $A(\Omega)$ is the element $e \in A(\Omega)$ satisfying $e(\omega) = 1$ for all $\omega \in \Omega$. An *effect* is an element $a \in A(\Omega)$ satisfying $0 \leq a(\omega) \leq 1$ for all $\omega \in \Omega$. The set of all effects is denoted $[0, e]$. There is a natural embedding of Ω into $A(\Omega)^*$, the dual space of $A(\Omega)$, given by $\omega \mapsto \hat{\omega}$, where $\hat{\omega}(a) = a(\omega)$ for all $a \in A(\Omega)$. Furthermore, if $\hat{\omega} \in A(\Omega)^*$ satisfies $\hat{\omega}(a) \geq 0$ for all $a \in A_+(\Omega)$ and $\hat{\omega}(e) = 1$, then $\hat{\omega}$ is the image of some state $\omega \in \Omega$ [21, Section 2.6]. We identify $\hat{\omega}$ with ω in what follows. It is easy to check that $\| \cdot \| = \sup_{a \in [0,e]} |a(\cdot)|$ is a norm on $A(\Omega)^*$. For more details about the convex sets framework, see [2, 3].

A natural distance measure on the set of states, which generalises the variational distance between classical probability distributions and the trace distance between quantum states, is given by

$$\|\omega - \omega'\| = \sup_{a\in[0,e]} |a(\omega) - a(\omega')|. \tag{3}$$

In quantum theory, systems are combined by taking the *tensor product* of the Hilbert spaces for each system. In the convex sets framework, the space $A(\Omega)^\star$ is not a Hilbert space but a Banach space and, although we still combine systems via a tensor product, the tensor product space is no longer unique. The action of the tensor product on states is the usual one: $\omega \otimes \omega'$ is defined to be the *product state* where system Ω is in state ω, system Ω' is in state ω', and the two systems are independent. Taking the closure under convex combinations yields:

Definition 1. *The* minimal tensor product *of* Ω *and* Ω'*, denoted by* $\Omega \otimes_{min} \Omega'$ *consists of convex combinations of product states* $\omega \otimes \omega'$*,* $\omega \in \Omega$ *and* $\omega' \in \Omega'$*.*

States in $\Omega \otimes_{\min} \Omega'$ are said to be *separable*. The action on effects is also the standard one: if a is a valid effect for system Ω and a' a valid effect for system Ω', then $a \otimes a'$ is the effect defined on product states via $a \otimes a'(\omega \otimes \omega') = a(\omega)a'(\omega')$. If such effects (and convex combinations thereof) are to be allowed, the state space must only contain states in the *maximal tensor product*, defined via duality as:

Definition 2. *The* maximal tensor product *of* Ω *and* Ω'*, denoted by* $\Omega \otimes_{max} \Omega'$ *consists of all bilinear functions* $\mu : A(\Omega) \times A(\Omega') \to \mathbb{R}$ *that satisfy* $\mu(a \otimes b) \geq 0$ *for* $a, b \geq 0$*, and* $\mu(e \otimes e') = 1$*.*

Thus $\mu \in \Omega \otimes_{\max} \Omega'$ can be written as a linear combination of product states, possibly with negative weights. In classical probability theory, the minimal and the maximal tensor product coincide. In general, a tensor product $\Omega \otimes \Omega'$ is a convex set with $\Omega \otimes_{\min} \Omega' \subseteq \Omega \otimes \Omega' \subseteq \Omega \otimes_{\max} \Omega'$. In quantum theory, $\Omega \otimes \Omega'$ is the set of trace one positive operators on the (unique) Hilbert space tensor product of $\mathcal{H}$ and $\mathcal{H}'$. Note that $\Omega \otimes \Omega'$ lies strictly between the maximal and minimal tensor products in the quantum case. The set of separable quantum states is $\Omega \otimes_{\min} \Omega'$.

For a state $\mu \in \Omega \otimes \Omega'$, we say that $\mu_\Omega \in \Omega$, defined by $a(\mu_\Omega) = a \otimes e'(\mu)$ for all effects a, is the *partial trace* of μ with respect to Ω'. An effect on the tensor product is an element $a \in A(\Omega \otimes \Omega')$ satisfying $0 \leq a \leq e \otimes e'$. The larger the set of joint states, the smaller the set of allowed effects. This means that the distance measure that we defined in Eq. (3), when applied to states of more than one particle, depends on which tensor product we use. It is true, however, that $\|\omega - \omega'\| \leq \|\omega - \omega'\|_{\min}$, the distance measure for the minimal tensor product, since in that case the set of effects is largest. Also note that a physical theory may place additional restrictions on which effects are allowed but, even then, $\|\omega - \omega'\|$ provides an upper bound on the probability of distinguishing ω and ω'.

In *generalized no-signalling theory* (GNST), particles are combined according to the maximal tensor product. It can be shown that the state space of n particles is precisely the set of all *no-signalling* conditional probability distributions [20, 22], including Popescu-Rohrlich boxes [23]. A bipartite distribution $P[A_1A_2|X_1X_2]$ is said to be *no-signalling* if the marginal distribution $P[A_1|X_1X_2]$ is independent of X_2 and likewise $P[A_2|X_1X_2]$ is independent of X_1. A multipartite distribution is no-signalling if all the bipartite distributions obtained by grouping the particles into two sets are no-signalling. The no-signalling principle ensures that the marginal distribution $P[A^m|X^m] \in \square^{\otimes_{\max} m}$ of a state $P[A^n|X^n] \in \square^{\otimes_{\max} n}$ is well-defined.

Results.—Suppose we have n particles in state $P[A^n|X^n] \in \Omega^{\otimes n}$. If we interchange the particles according to a permutation $\pi \in S_n$, the resulting state is

$$\begin{aligned}&\pi P[A^n = a_1 \cdots a_n | X^n = x_1 \cdots x_n] \\ &= P[A^n = a_{\pi^{-1}(1)} \cdots a_{\pi^{-1}(n)} | X^n = x_{\pi^{-1}(1)} \cdots x_{\pi^{-1}(n)}].\end{aligned}$$

We say that a conditional probability distribution $P[A^n|X^n]$ is *symmetric* if it is invariant under all permutations $\pi \in S_n$. If $|\mathcal{X}| = 1$, this definition reduces to the usual definition of a symmetric probability distribution. We can now state our main result:

Theorem 3. *Let* Ω *be a convex subset of* $\square$*. Suppose that* $P[A^n|X^n] \in \Omega^{\otimes n}$ *is symmetric. Then there is a measure* $m(\lambda)$ *on single-particle conditional probability distributions* $P_\lambda[A|X] \in \square$ *such that*

$$\begin{aligned}&\left\|P[A^k|X^k] - \int dm(\lambda) P_\lambda[A|X]^{\otimes k}\right\| \\ &\qquad \leq \min\left(\frac{2k|\mathcal{X}||\mathcal{A}|^{|\mathcal{X}|}}{n}, \frac{|\mathcal{X}|k(k-1)}{n}\right).\end{aligned} \tag{4}$$

This establishes that the state of a random subset of k out of n particles is well approximated by a convex combination of independent and identical conditional probability distributions. To prove Theorem 3, we first show that if $P[A^n|X^n]$ is symmetric and m is chosen to be sufficiently small, then $P[A^m|X^m]$ is separable (Lemma 4). We then establish a de Finetti theorem for separable states, Theorem 5, which will complete the proof of our main result, Theorem 3. We continue with Lemma 4.

Lemma 4. *Let* $n \geq |\mathcal{X}|$ *and set* $m = \lceil n/|\mathcal{X}| \rceil$*. Suppose that* $P[A^n|X^n] \in \Omega^{\otimes n}$ *is symmetric. Then* $P[A^m|X^m] \in \square^{\otimes_{min} m}$*.*

Proof. In order not to obscure the main argument, we prove the statement for integral $m = n/|\mathcal{X}|$ [31]. Our technique can be traced to Werner [24, 25, 26]. We imagine the m particles to be separated in space and note that $P[A^m|X^m]$ is separable if and only if it can be simulated by a local hidden variable model. Such a simulation is described in Fig. 1. We now provide the formal proof. We

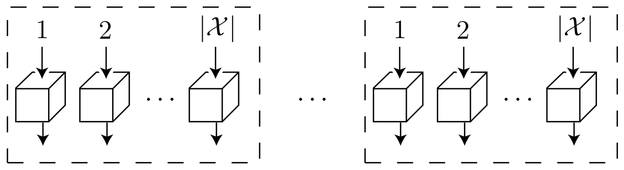

FIG. 1: Since $n = m|\mathcal{X}|$, we can divide the particles into m groups of $|\mathcal{X}|$ particles. In each of these groups we measure one particle according to each measurement in X *in advance* and record a list of all the results. In the simulation, if particle i is supposed to be measured according to a measurement $x \in X$, we just look through the ith group until we come to the particle on which measurement x was performed in advance, and output the result we find.

construct a separable conditional distribution $Q[A^m|X^m]$ and then show that it is equal to $P[A^m|X^m]$. We assume that $\mathcal{X} = \{1, 2, \ldots, |\mathcal{X}|\}$, define a vector $y^n = (y_j)_{j=1,\ldots,n}$ with coordinates $y_j = (j \mod |\mathcal{X}|) + 1$, and define the separable state

$$Q[A^m|X^m] = \sum_{b^n} q_{b^n} Q_{b^n,1}[A_1|X_1] \cdots Q_{b^n,m}[A_m|X_m],$$

where $b^n \in A^n$ is distributed according to $q_{b^n} = P[A^n = b^n|X^n = y^n]$ and the single-particle conditional distributions are deterministic and defined by $Q_{b^n,i}[A_i = a_i|X_i = x_i] = [a_i = b_{(i-1)|\mathcal{X}|+x_i}]$, where $[t] = 1$ if t is true and 0 otherwise. Let $\mathcal{L} = \{1, 2, \ldots, n\}$, $\mathcal{L}_1 = \{(i-1)|\mathcal{X}|+x_i : i = 1, 2, \ldots, m\}$ and $\mathcal{L}_2 = \mathcal{L} \backslash \mathcal{L}_1$. Further let $A^{\mathcal{L}} = A^n$, $A^{\mathcal{L}_1} = (A_{x_1}, A_{|\mathcal{X}|+x_2}, \ldots, A_{(m-1)|\mathcal{X}|+x_m})$ and $A^{\mathcal{L}_2} = A^{\mathcal{L}} \backslash A^{\mathcal{L}_1}$ and define $b^{\mathcal{L}}, b^{\mathcal{L}_1}$ and $b^{\mathcal{L}_2}$ similarly. We find

$$\begin{aligned}
&Q[A^m = a^m|X^m = x^m] \\
&\quad = \sum_{b^n} P[A^n = b^n|X^n = y^n] \\
&\qquad\qquad \times [a_1 = b_{x_1}] \cdots [a_m = b_{(m-1)|\mathcal{X}|+x_m}] \\
&\quad = \sum_{b^{\mathcal{L}_2}} P[A^{\mathcal{L}_1} = a^m, A^{\mathcal{L}_2} = b^{\mathcal{L}_2}|X^{\mathcal{L}_1} = x^m, X^{\mathcal{L}_2} = y^{\mathcal{L}_2}] \\
&\quad = P[A^{\mathcal{L}_1} = a^m|X^{\mathcal{L}_1} = x^m] = P[A^m = a^m|X^m = x^m],
\end{aligned}$$

where we started with the definition of $Q[A^m|X^m]$, split the summation over $\mathcal{L}_1$ and $\mathcal{L}_2$, dropped the conditioning over $X^{\mathcal{L}_2} = y^{\mathcal{L}_2}$ because of the no-signalling property of P, used the definition of a marginal state, and, lastly, the permutation-invariance of P. □

Our next statement is a de Finetti theorem for convex set theories with the minimum tensor product.

Theorem 5. *Let Ω be a convex set with E extreme points (E may be infinite). Suppose $\omega^n \in \Omega^{\otimes_{min} n}$ is symmetric. Then there is a measure $m(\tau)$ on states $\tau \in \Omega$ such that*

$$\big\|\omega^k - \int dm(\tau)\, \tau^{\otimes k}\big\|_{min} \leq \min\left(\frac{2kE}{n}, \frac{k(k-1)}{n}\right). \quad (5)$$

Proof. Let $\tau_1, \ldots, \tau_E$ be the extreme points of Ω. Any symmetric separable state is a convex combination of states of the form $\omega^n = \frac{1}{n!}\sum_\pi \tau_{i_{\pi^{-1}(1)}} \otimes \cdots \otimes \tau_{i_{\pi^{-1}(n)}}$, where $1 \leq i_1, \ldots, i_n \leq E$. Define $\tau := \frac{1}{n}\sum_{j=1}^n \tau_{i_j}$. We expand

$$\tau^{\otimes k} = \sum_{j_1=1}^n \cdots \sum_{j_k=1}^n M_n(i_{j_1}, \ldots, i_{j_k}) \tau_{i_{j_1}} \otimes \cdots \otimes \tau_{i_{j_k}}, \quad (6)$$

where $M_n(i_{j_1}, \ldots, i_{j_k}) = 1/n^k$ is the multinomial distribution. To compare this expression with ω^k, write

$$\omega^k = \sum_{j_1=1}^n \cdots \sum_{j_k=1}^n H_n(i_{j_1}, \ldots, i_{j_k}) \tau_{i_{j_1}} \otimes \cdots \otimes \tau_{i_{j_k}}, \quad (7)$$

where $H_n(i_{j_1}, \ldots, i_{j_k})$ is the hypergeometric distribution for an urn with n balls (see [1]). Then

$$\begin{aligned}
\big\|\omega^k - \tau^{\otimes k}\big\|_{\min} &= \big\| \sum_{j_1,\ldots,j_k} \big(H_n(i_{j_1}, \ldots, i_{j_k}) \\
&\qquad M_n(i_{j_1}, \ldots, i_{j_k})\big) \tau_{i_{j_1}} \otimes \cdots \otimes \tau_{i_{j_k}} \big\|_{\min} \\
&\leq \sum_{j_1,\ldots,j_k} \big|H_n(i_{j_1}, \ldots, i_{j_k}) - M_n(i_{j_1}, \ldots, i_{j_k})\big| \\
&\leq \min\left(\frac{2kE}{n}, \frac{k(k-1)}{n}\right), \qquad (8)
\end{aligned}$$

where we used the triangle inequality and Diaconis and Freedman's result on estimating the hypergeometric distribution with a multinomial distribution [1]. □

Proof of Theorem 3. Set $m = \lceil n/|\mathcal{X}| \rceil$ and apply Lemma 4. Then apply Theorem 5 to $P[A^m|X^m]$, noting that □ has $|\mathcal{A}|^{|\mathcal{X}|}$ extreme points (the deterministic functions $\mathcal{X} \mapsto \mathcal{A}$) and that $\| \cdot \| \leq \| \cdot \|_{\min}$. □

Our final result is an application to quantum theory. Let Ω be the set of density operators on $\mathbb{C}^d$ and note that a separable density operator on n systems is an element of $\Omega^{\otimes_{\min} m}$. Since $||\rho - \sigma||_1 = \mathsf{Tr}|\rho - \sigma| = 2\sup_{0 \leq R \leq \mathbf{1}} |\mathsf{Tr}(\rho - \sigma)R| = 2\|\rho - \sigma\|_{\min}$, we obtain:

Corollary 6. *If ρ is a separable permutation-invariant density operator on $(\mathbb{C}^d)^{\otimes n}$, then there is a measure $m(\sigma)$ on states σ on $\mathbb{C}^d$ such that*

$$\big\|\rho^k - \int dm(\sigma)\, \sigma^{\otimes k}\big\|_1 \leq 2\frac{k(k-1)}{n}. \quad (9)$$

We conclude with an open question. Is Theorem 3 true if the conditional probability distributions $P_\lambda[A|X]$ are restricted to be elements of Ω (rather than □)? If so, we would have a finite de Finetti theorem for all theories in the convex sets framework.

Acknowledgements.—This work was carried out at the same time as related work by J. Barrett and M. Leifer [27]. We thank them for discussions of their work, and especially for explaining how to define the analogue of trace distance. We also thank R. Colbeck and R. Renner for discussions, G. Mitchison for valuable comments on the manuscript, and the organizers of the FQXi workshop *Operational probabilistic theories as foils to quantum theory*, where part of this work was done. MC would like to thank the IQI at Caltech and the CWI in Amsterdam for their hospitality and acknowledge the Research Fellowships of the UK's EPSRC and Magdalene College Cambridge. This work was supported in part by the NSF under Grants PHY-0456720 and CCF-0524828, the EU's FP6-FET Integrated Projects SCALA (CT-015714) and QAP (CT-015848), NWO VICI project 639-023-302, and the Dutch BSIK/BRICKS project.

* Electronic address: matthias.christandl@qubit.org
† Electronic address: Ben.Toner@cwi.nl

[1] P. Diaconis and D. Freedman, Ann. Probab. **8**, 745 (1980).

[2] H. Barnum, J. Barrett, M. Leifer, and A. Wilce (2006), quant-ph/0611295.

[3] H. Barnum, J. Barrett, M. Leifer, and A. Wilce (2007), arXiv:0707.0620.

[4] R. König and R. Renner, J. Math. Phys. **46**, 122108 (2005), quant-ph/0410229.

[5] M. Christandl, R. König, G. Mitchison, and R. Renner, Comm. Math. Phys. **273**, 473 (2007), quant-ph/0602130.

[6] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory* (Wiley, Chichester, 1994).

[7] C. H. Bennett and G. Brassard, in *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing* (IEEE, 1984), pp. 175–179.

[8] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

[9] R. Renner, Ph.D. thesis, Swiss Federal Institute of Technology, Zurich (2005), quant-ph/0512258.

[10] R. Renner, Nature Physics **3**, 645 (2007), quant-ph/0703069.

[11] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005), quant-ph/0405101.

[12] A. Acín, N. Gisin, and Ll. Masanes, Phys. Rev. Lett. **97**, 120405 (2006), quant-ph/0510094.

[13] V. Scarani, N. Gisin, N. Brunner, Ll. Masanes, S. Pino, and A. Acín, Phys. Rev. A **74**, 042339 (2006), quant-ph/0606197.

[14] Ll. Masanes and A. Winter, quant-ph/0606049v1.

[15] E. Størmer, J. Funct. Anal. **3**, 48 (1969).

[16] R. L. Hudson and G. R. Moody, Z. Wahrschein. verw. Geb. (Probab. Theory Related Fields) **33**, 343 (1976).

[17] C. M. Caves, C. A. Fuchs, and R. Schack, J. Math. Phys. **43**, 4537 (2002), quant-ph/0104088.

[18] C. A. Fuchs, R. Schack, and P. F. Scudo, Phys. Rev. A **69**, 062305 (2004), quant-ph/0307198.

[19] M. Christandl and B. Toner (2007), in preparation.

[20] J. Barrett, Phys. Rev. A **75**, 032304 (2007), quant-ph/0508211.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004), available online at http://www.stanford.edu/˜boyd/cvxbook/.

[22] C. H. Randall and D. J. Foulis, in *Interpretations and Foundations of Quantum Mechanics*, edited by H. Neumann (Bibliographisches Institut, Wissenschaftsverlag, Manheim, 1981).

[23] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).

[24] R. F. Werner, Lett. Math. Phys. **17**, 359 (1989).

[25] B. M. Terhal, A. C. Doherty, and D. Schwab, Phys. Rev. Lett. **90**, 157903 (2003), quant-ph/0210053.

[26] B. F. Toner (2006), quant-ph/0601172.

[27] J. Barrett and M. Leifer (2007), in preparation.

[28] Individual particles are labeled by subscripts and $m(\lambda)$ is a measure on the set of conditional probability distributions on a single particle.

[29] Diaconis and Freedman also obtain a second bound $k|\mathcal{A}|/n$. The analogous bound within our framework is $k|\mathcal{A}|^{|\mathcal{X}|}/n$. Restricting to *adaptive* measurements on individual particles, we are able to improve this bound to $k|\mathcal{X}|^2|\mathcal{A}|(1+4\sqrt{\frac{2+\log|\mathcal{X}|}{k}})/n$ [19].

[30] A density operator ρ^n is permutation-invariant if $\rho^n = \pi\rho^n\pi^{-1}$ for all permutations $\pi \in S_n$.

[31] This immediately implies the result for $\lfloor n/|\mathcal{X}|\rfloor$. The extension to the case $\lceil n/|\mathcal{X}|\rceil$ is more technical and can be found in [19].