Title: Generalised entropies, information causality, and non-local games

Date: May 12, 2011 04:10 PM

URL: http://pirsa.org/11050048

Abstract: We will explore generalisations of the Shannon and von Neumann entropy to other probabilistic theories, and their connection to the principle of information causality. We will also investigate the link between information causality and non-local games, leading to a new quantum bound on computing the inner product non-locally.

# Generalized entropies, information causality, and non-local games

Tony Short

University of Cambridge

(with Stephanie Wehner and Sabri Al-Safi)

# Introduction

- The ability to quantify information and uncertainty using entropies has proved very valuable in understanding quantum and classical theory, and their information-processing capabilities.

- Can the notion of entropy be extended to other probabilistic theories?
  - AJS, S. Wehner, *New J. Phys.* 12 03302 (2010)
  - H. Barnum, J. Barrett, L.O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce and R. Wilke, *New J. Phys.* 12 033024 (2010)
  - G. Kimura, K. Nuida, H. Imai, *Rep. Math. Phys.* 66, 175 (2010)

# General probabilistic theories

- We consider a general probabilistic framework for physical theories based on operational notions (as in many previous works...).

- The framework we use is more general than some, as we do *not* assume
  - That all mathematically possible measurements/transformations are implementable
  - That composite systems can be completely characterised by local measurements. (i.e. Local tomography)

## Overview of framework (not including lots of details)

- Each system (or collection of systems) has a set of allowed **states S**
  - Any mixture can be prepared, and is represented by

$$s_{mix} = ps_1 + (1-p)s_2$$

hence state sets are convex

  - If we independently prepare system A in state $s_A$ and system B in state $s_B$, we generate a well-defined **product state** of the composite AB denoted by $s_A \otimes s_B$
  - States are **separable** if they are equal to a mixture of product states and **entangled** otherwise.

- For each system, there is some set of allowed **measurements E**
  - Each measurement **e** has a finite set of outcomes
  - Each outcome r is associated with an affine **effect** $e_r$: S→ [0,1] such that $e_r(s)$ is the probability of obtaining result r on state s.

$$e_r(s_{mix}) = pe_r(s_1) + (1-p)e_r(s_2)$$

  - The sum over all effects in a measurement is the unit effect *u* satisfying *u(s)*=1 for all states.
  - We assume that we can independently measure two systems to perform a well-defined product measurement $\mathbf{e}_A \otimes \mathbf{e}_B$

- If one measurement gives strictly more information than another we call it a **refinement**

  (e.g. $\{e_1, e_2\} \rightarrow \{e_1, e_{2a}, e_{2b}\}$ where $e_{2a} + e_{2b} = e_2$ and $e_{2a} \neq c\, e_{2b}$).

- We call a measurement fine-grained if it has no refinement.
  - The set of fine-grained measurements is denoted by $E^* \subseteq E$
  - We assume that $E^*$ is non-empty.

- Similarly, if one measurement gives strictly less information than another we call it a **coarse-graining**.
  - We assume that E is closed under coarse graining.

- For clarity, when dealing with entropies of states we will sometimes write $H(s_A)$ as $H(A)$, $H(s_{AB})$ as $H(AB)$, etc.

- For each system, there is some set of allowed **measurements E**
  - Each measurement **e** has a finite set of outcomes
  - Each outcome r is associated with an affine **effect** $e_r$: S→ [0,1] such that $e_r(s)$ is the probability of obtaining result r on state s.

$$e_r(s_{mix}) = pe_r(s_1) + (1-p)e_r(s_2)$$

  - The sum over all effects in a measurement is the unit effect *u* satisfying *u(s)*=1 for all states.
  - We assume that we can independently measure two systems to perform a well-defined product measurement $\mathbf{e}_A \otimes \mathbf{e}_B$

- If one measurement gives strictly more information than another we call it a **refinement**

  (e.g. $\{e_1, e_2\} \rightarrow \{e_1, e_{2a}, e_{2b}\}$ where $e_{2a} + e_{2b} = e_2$ and $e_{2a} \neq c\, e_{2b}$).

- We call a measurement fine-grained if it has no refinement.
  - The set of fine-grained measurements is denoted by $E^* \subseteq E$
  - We assume that $E^*$ is non-empty.

- Similarly, if one measurement gives strictly less information than another we call it a **coarse-graining**.
  - We assume that E is closed under coarse graining.

- For clarity, when dealing with entropies of states we will sometimes write $H(s_A)$ as $H(A)$, $H(s_{AB})$ as $H(AB)$, etc.

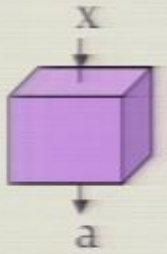
## Examples

- *Classical probability theory*
  - States are finite probability vectors $p_i$
  - Effects are $e_r(s)=\sum_i q_r^i p_i$ for $q_r^i \in [0,1]$
- *Quantum theory*
  - States are finite-dimensional trace 1 positive operators $\rho$
  - Effects are $e_r(s)=\mathrm{tr}(E_r \rho)$ for $0 \leq E_r \leq I$
  - Fine-grained measurements are POVMs with rank 1 elements.
- *Restricted quantum/classical theories*
  - some subset of allowed states / measurements
  - Real quantum theory

# Box world

- A generalised probabilistic theory that has received a lot of attention is **box-world** , previously called `Generalised Non-Signallling Theory' [J. Barrett (2005)].

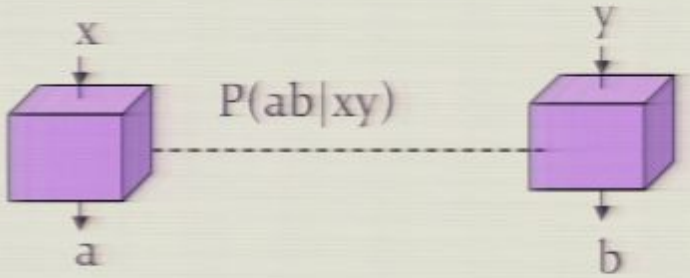
- The state of a single system is given by a conditional probability distribution P(a|x). All such distributions are allowed states.
  - Intuitively, *x* represents an input (a choice of measurement) and *a* an output (the measurement result).
  - Different types of system have different finite input and output sets



- Multipartite states are represented by joint conditional probability distributions $P(a_1\, a_2 \ldots a_n | x_1\, x_2 \ldots\, x_n)$.
  - All distributions satisfying the no-signalling condition are allowed states.

$$\sum_{a_k} P(a_1 \cdots a_n \mid x_1 \cdots x_n) \quad \text{is independent of} \quad x_k$$

  - An important state in box-world is the entangled PR-box state [S. Popescu, D. Rohrlich(1994)] where a,b,x,y are binary

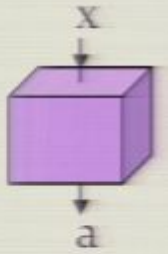



$$P_{PR}(ab \mid xy) = \begin{cases} 1/2 \text{ if } a \oplus b = xy \\ \\ 0 \text{ if } a \oplus b \neq xy \end{cases}$$

# Box world

- A generalised probabilistic theory that has received a lot of attention is **box-world** , previously called `Generalised Non-Signallling Theory' [J. Barrett (2005)].

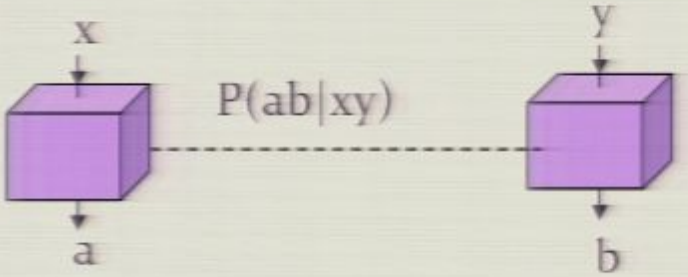
- The state of a single system is given by a conditional probability distribution P(a|x). All such distributions are allowed states.
  - Intuitively, *x* represents an input (a choice of measurement) and *a* an output (the measurement result).
  - Different types of system have different finite input and output sets



- Multipartite states are represented by joint conditional probability distributions $P(a_1\, a_2 \ldots a_n | x_1\, x_2 \ldots\, x_n)$.
  - All distributions satisfying the no-signalling condition are allowed states.

$$\sum_{a_k} P(a_1 \cdots a_n \mid x_1 \cdots x_n) \quad \text{is independent of} \quad x_k$$

  - An important state in box-world is the entangled PR-box state [S. Popescu, D. Rohrlich(1994)] where a,b,x,y are binary




$$P_{PR}(ab \mid xy) = \begin{cases} 1/2 \text{ if } a \oplus b = xy \\ \\ 0 \text{ if } a \oplus b \neq xy \end{cases}$$

- In box world all mathematically well-defined measurements and transformations are allowed.
  - The allowed measurements E include `putting an input *x* in the box, and obtaining result *a*', but also mixtures of these things.
  - On bipartite systems all measurements can be performed only using the box inputs and outputs. E.g.


  - However there exist non-trivial tri-partite measurements, that cannot be performed in this way. [AJS, J.Barrett (2010)]

# Generalised entropy

- Our aim is to define an entropy that is meaningful for any operational theory, yet reduces to the Shannon and Von-Neumann entropy in classical and quantum theory.

- The more properties of the usual entropy that our definition maintains, the better.
  - To preserve the valuable intuitions we have developed
  - To allow the possibility of lifting proofs from the quantum to the general case

- Many definitions are possible, and we will consider another natural alternative later (the decomposition entropy)

- However, a good definition is :

$$H(s) \equiv \inf_{\mathbf{e} \in E^*} H_c(\mathbf{e}(s))$$

  - $H_c$ is the classical Shannon entropy $H_c(\mathbf{e}(s)) \equiv -\sum_{r \in R_e} e_r(s) \log_2 e_r(s)$

  - Intuitively, H(s) is the minimal outcome uncertainty for any fine grained measurement on the system.

## Properties of H(s)

***1) Reduction:*** Crucially, H(s) reduces to the von-Neumann entropy in quantum theory, and the Shannon entropy in classical theory

- In the quantum case, the optimal fine-grained measurement is a projective measurement in the eigenbasis of ρ.

***2) Positivity and Finiteness:*** let d be the minimal number of outcomes of a fine grained measurement. Then it is easy to see that

$$0 \leq H(s) \leq \log d$$

### 3) Concavity:

$$H(s_{mix}) \geq pH(s_1) + (1-p)H(s_2)$$

*Proof:* Suppose the infimum is attained for $\mathbf{e} \in E^*$, then

$$\begin{aligned} H(s_{mix}) &= H_c(\mathbf{e}(s_{mix})) \\ &\geq pH_c(\mathbf{e}(s_1)) + (1-p)H_c(\mathbf{e}(s_2)) \\ &\geq pH(s_1) + (1-p)H(s_2) \end{aligned}$$

- Two additional properties hold with *weak* additional assumptions. In particular, they hold in classical theory, quantum theory and box-world

4) ***(Limited) Subadditivity:*** Suppose that a product of two fine grained measurements is also fine-grained. That is

$$\mathbf{e} \in E_A^*, \mathbf{f} \in E_B^* \Rightarrow \mathbf{e} \otimes \mathbf{f} \in E_{AB}^*$$

then

$$H(A) + H(B) \geq H(AB)$$

5) ***(Limited) Continuity:*** Suppose that for a given system, restricting $E^*$ to measurements with a bounded number of outcomes does not change the entropy. Then the entropy is continuous on states with the natural distance measure (distinguishability using E)

## (Limited) Coding theorem

- Ideally , we would like an operational understanding of our entropy in terms of data compression
- With additional (relatively strong) assumptions, which hold in quantum and classical theory, we can prove such a result - that the entropy gives an achievable `compression' rate.
- In particular, we assume that all relevant measurements are
  - *Repeatable* - Repeating the measurement yields the same result
  - *Weakly disturbing* - If a measurement result is almost certain to occur, obtaining that result doesn't change the state much.

4) ***(Limited) Subadditivity:*** Suppose that a product of two fine grained measurements is also fine-grained. That is

$$\mathbf{e} \in E_A^*, \mathbf{f} \in E_B^* \Rightarrow \mathbf{e} \otimes \mathbf{f} \in E_{AB}^*$$

then

$$H(A) + H(B) \geq H(AB)$$

5) ***(Limited) Continuity:*** Suppose that for a given system, restricting $E^*$ to measurements with a bounded number of outcomes does not change the entropy. Then the entropy is continuous on states with the natural distance measure (distinguishability using E)

## (Limited) Coding theorem

- Ideally , we would like an operational understanding of our entropy in terms of data compression
- With additional (relatively strong) assumptions, which hold in quantum and classical theory, we can prove such a result - that the entropy gives an achievable `compression' rate.
- In particular, we assume that all relevant measurements are
  - *Repeatable* - Repeating the measurement yields the same result
  - *Weakly disturbing* - If a measurement result is almost certain to occur, obtaining that result doesn't change the state much.

## Other entropic quantities

- Given H(s), we can also define other entropic quantities analogously to the quantum case

- **Conditional entropy:** $H(A|B) \equiv H(AB) - H(B)$

- **Mutual Information:** $I(A:B) = H(A) + H(B) - H(AB)$

- However, these do not maintain their intuitive properties as nicely as the entropy itself. For example, the conditional entropy is not subadditive in box-world.

## Decomposition entropy

- There are many alternative definitions of the entropy which we could have chosen. One example is the decomposition entropy $H_D(s)$, which measures the mixedness of a state

- Denote the extreme points of S by $S^*$. Call these **pure states**, with all other states being **mixed.** The decomposition entropy of s is the minimal Shannon entropy of the probability distribution over pure states, for all pure state decompositions of s

$$H_D(s) \equiv \inf_{\left\{(p_i, s_i) \middle| s = \sum_i p_i s_i\right\}} H_c(p_i)$$

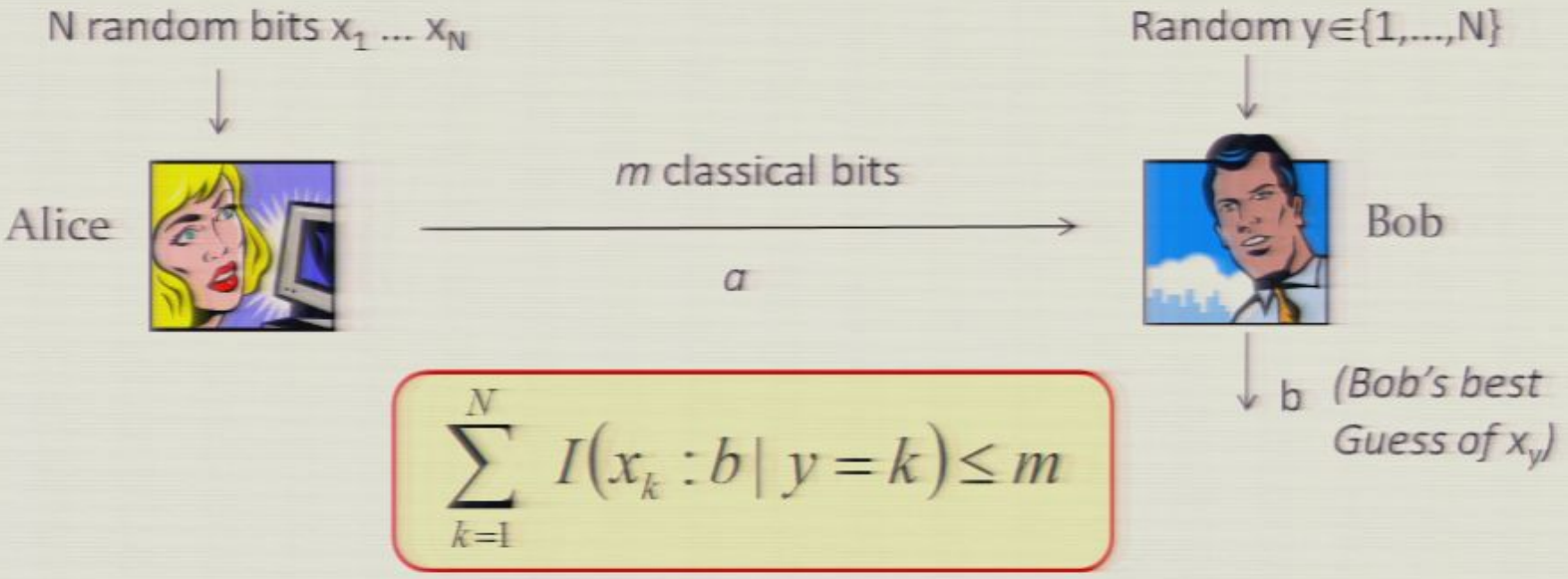
- Like H(s), it can be proved that $H_D(s)$ reduces to the von-Neumann and Shannon entropy in classical and quantum theory respectively.
  - The fact that these two conceptually different entropies are the same in quantum theory is intriguing (Barnum et al.)

- However, examples from box world illustrate that $H_D(s)$ is neither **concave** or **subadditive** in general.

- $H_D(s)$ also has a number of other counterintuitive properties in box world. E.g.
  - The decomposition entropy of a maximally random binary-input/output box is 1.
  - However, the decomposition entropy of two random boxes is also 1 (as they are an equal mixture of PR and anti-PR).

# Information causality

- A useful arena in which to try out our generalised entropy is **information causality** [Pawlowski et al, Nature 461, 1101 (2009)]




$$\sum_{k=1}^{N} I(x_k : b \mid y = k) \leq m$$

- Information causality is respected by quantum and classical theory, but violated in box-world.

- In particular, for the case in which N=2 and m=1, using a single PR-box state, Bob can perfectly discover whichever of Alice's bits is requested, achieving

$$2 = \sum_k I_c\left(x_k : b \mid y = k\right) \nleq m = 1$$

- The proof of information causality depends on manipulating the quantum mutual information.
  - By following the same steps using our generalised mutual information, we can see where the proof fails for box world.

- The failure actually happens for the state after Alice has sent the message *a*, which is separable


$$P(x_0 x_1 ac \mid z) = \begin{cases} \frac{1}{8} & \text{if } c = x_z \oplus a \\ 0 & \text{otherwise} \end{cases}$$

- The next step of the proof uses the data processing inequality to deduce that

$$I(x_0 : x_1 aZ) \geq I(x_0 : aZ)$$

- However, this relation does not hold in box world. Indeed, for the state given it is easy to compute that

$$0 = I(x_0 : x_1 a Z) \ngeq I(x_0 : aZ) = 1$$

- Applying our usual intuitions about the mutual information, this would suggest that `forgetting' $x_1$ gives more information about $x_0$

- This shows that our general entropy does not satisfy strong subadditivity in box world

$$H(ABC) + H(C) \nleq H(AB) + H(AC)$$

- Interestingly, given any theory which includes classical systems (and some natural transformations), information causality will hold if there exists any entropy function $H_T(s)$ for states in that theory satisfying

  i) **Classical reduction:** $H_T(s)=H_c(s)$ when s is a classical system

  ii) **Data processing:** For any joint system AB, and any transformation on A

$$\Delta H_T(AB) \geq \Delta H_T(A)$$

*(this is equivalent to $I(A:B) \geq I(T[A]:B)$, and is satisfied by the Shannon and von Neumann entropy)*

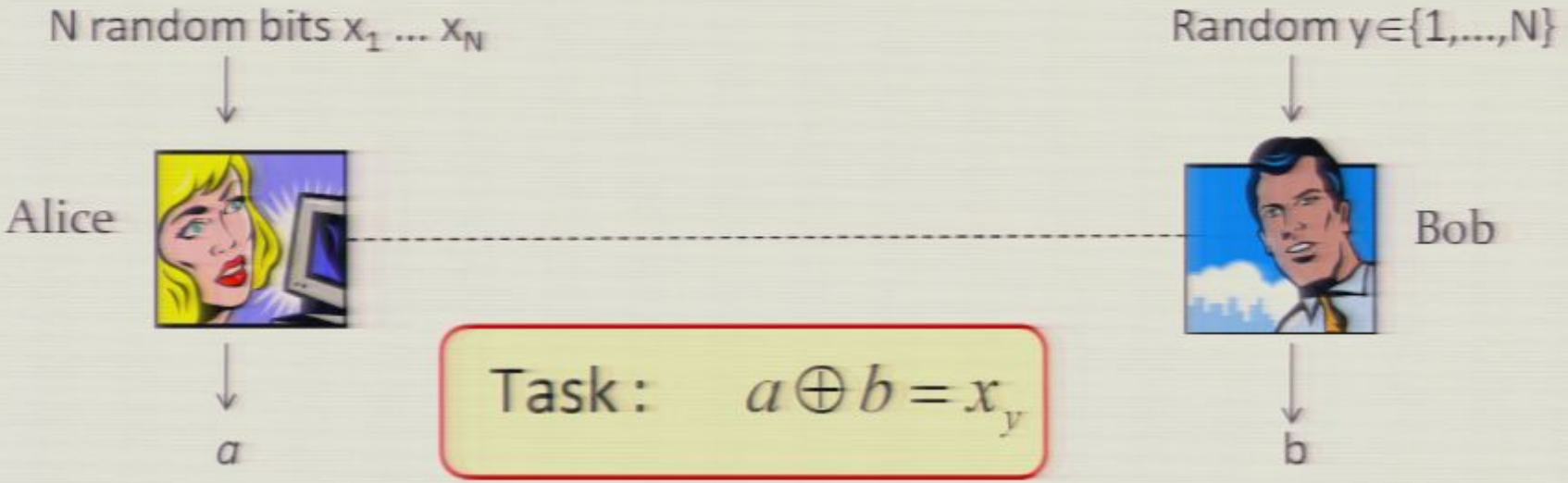
- The fact that quantum theory admits an entropy which shares many of the powerful properties of the Shannon entropy is surprising, and may be very special in the set of theories.
  - Information Causality seems to be a consequence of this.

- Can we find other interesting tasks for which there is a classical entropic bound, and see if they hold in quantum theory but not in general?

- Entropies are strange non-linear functions
  - Surprisingly however, Information causality can be used to derive part of the boundary of the set of quantum correlations [Pawlowski et al, and Allcock et al (2009)]. How?

# Information causality as a non-local game

- The proof of Tsirelson's bound from information causality involves only 1 bit of communication, which is added to Bob's guess (mod 2). Hence we can think of it as a non-local game.



- For non-local games, the normal figure of merit is the probability of success $P_{success}$. Quantum theory can do better than classical in this case (e.g. For N=2 we get the same probabilities as CHSH)

- Define $P_y$ as the probability of success when Bob is given y, and the corresponding bias $B_y = 2P_y - 1$

- When proving Tsirelson's bound from Information causality
  - A quadratic bound on the entropy is used to derive a probabilistic bound on this game given by

$$\sum_{y=1}^{N} B_y^2 \leq 2 \ln 2$$

  - Can we derive a similar bound directly from quantum theory?

- For non-local games, the normal figure of merit is the probability of success $P_{success}$. Quantum theory can do better than classical in this case (e.g. For N=2 we get the same probabilities as CHSH)

- Define $P_y$ as the probability of success when Bob is given y, and the corresponding bias $B_y = 2P_y - 1$

- When proving Tsirelson's bound from Information causality
  - A quadratic bound on the entropy is used to derive a probabilistic bound on this game given by

$$\sum_{y=1}^{N} B_y^2 \leq 2 \ln 2$$

  - Can we derive a similar bound directly from quantum theory?

- For any quantum strategy

$$B_y = \frac{1}{2^N} \sum_x \langle \psi | (-1)^{\hat{a}_x + \hat{b}_y + x_y} | \psi \rangle$$

- Using similar techniques to those in the non-local computation paper [Linden et al (2007)] we define

$$|\alpha\rangle = \frac{1}{\sqrt{2^N}} \sum_x (-1)^{\hat{a}_x} |\psi\rangle \otimes |x\rangle \qquad |\beta_y\rangle = \frac{1}{\sqrt{2^N}} \sum_x (-1)^{\hat{b}_y + x_y} |\psi\rangle \otimes |x\rangle$$

and note that

$$\sum_y B_y^2 = \langle \alpha | \left( \sum_y |\beta_y\rangle\langle\beta_y| \right) | \alpha \rangle \leq 1$$

- The quantum bound

$$\sum_y B_y^2 \leq 1$$

is easily saturated classically, by answering one question perfectly.

  - Hence with this figure of merit quantum theory is no better than classical. Yet in box-world the sum can equal N

- Note that this also gives a bound on the probability of success

$$P_{success}^{quantum} \leq \frac{1}{2}\left(1+\sqrt{\left\langle B_y^2 \right\rangle}\right) \leq \frac{1}{2}\left(1+\frac{1}{\sqrt{N}}\right)$$

which is saturated when N is a power of 2

- For any quantum strategy

$$B_y = \frac{1}{2^N} \sum_x \langle \psi | (-1)^{\hat{a}_x + \hat{b}_y + x_y} | \psi \rangle$$

- Using similar techniques to those in the non-local computation paper [Linden et al (2007)] we define

$$|\alpha\rangle = \frac{1}{\sqrt{2^N}} \sum_x (-1)^{\hat{a}_x} |\psi\rangle \otimes |x\rangle \qquad |\beta_y\rangle = \frac{1}{\sqrt{2^N}} \sum_x (-1)^{\hat{b}_y + x_y} |\psi\rangle \otimes |x\rangle$$

and note that

$$\sum_y B_y^2 = \langle \alpha | \left( \sum_y |\beta_y\rangle\langle\beta_y| \right) | \alpha \rangle \leq 1$$

- The quantum bound

$$\sum_y B_y^2 \leq 1$$

is easily saturated classically, by answering one question perfectly.

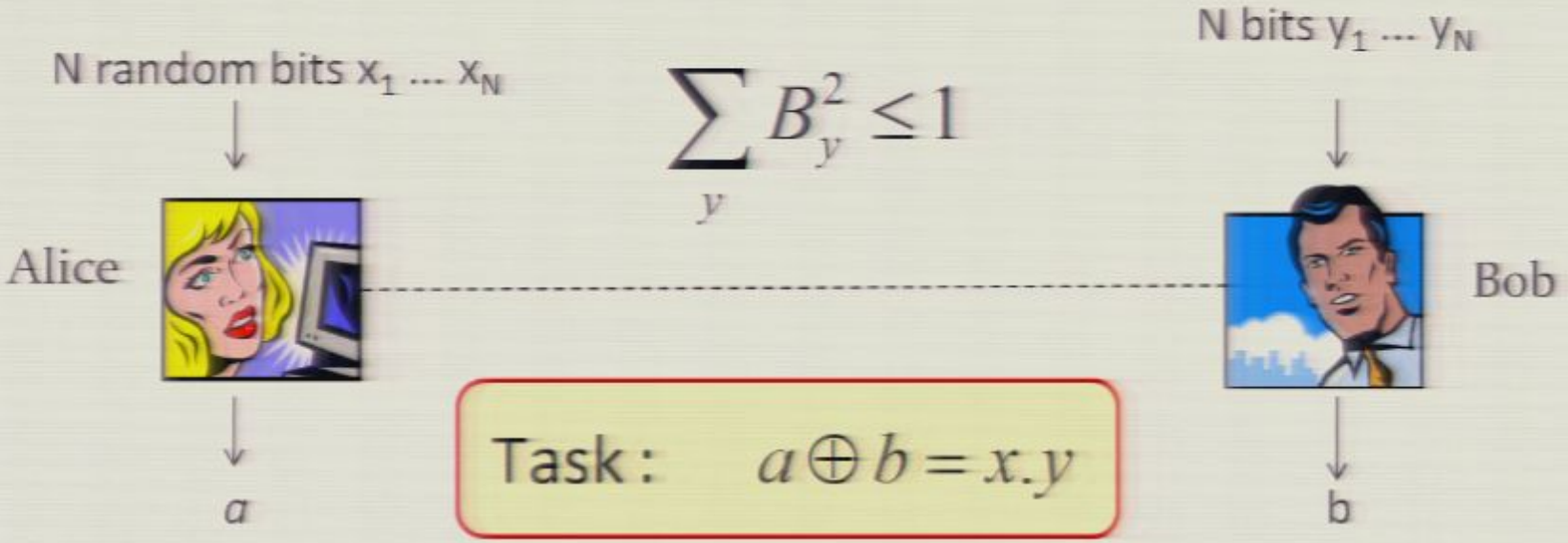
  - Hence with this figure of merit quantum theory is no better than classical. Yet in box-world the sum can equal N

- Note that this also gives a bound on the probability of success

$$P_{success}^{quantum} \leq \frac{1}{2}\left(1+\sqrt{\left\langle B_y^2 \right\rangle}\right) \leq \frac{1}{2}\left(1+\frac{1}{\sqrt{N}}\right)$$

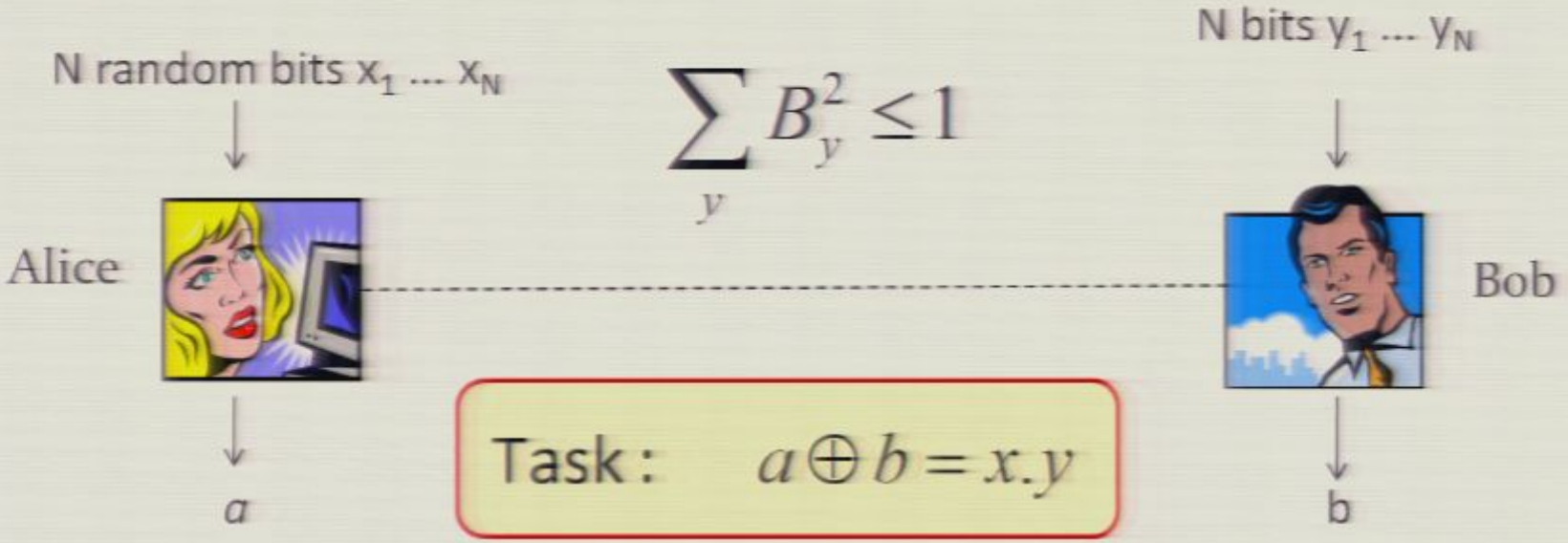
which is saturated when N is a power of 2

- We can use essentially the same proof to get a quantum bound for the inner product game (with Bob's input having any distribution)




- When Bob's bit string is restricted to contain a single 1, this implies the information causality result. When N=1, it yields Tsirelson's bound, and the stronger quadratic version [Uffink 2002]

- We can use essentially the same proof to get a quantum bound for the inner product game (with Bob's input having any distribution)




- When Bob's bit string is restricted to contain a single 1, this implies the information causality result. When N=1, it yields Tsirelson's bound, and the stronger quadratic version [Uffink 2002]

# Open questions

- Quantum theory has an entropy with many of the intuitive properties of the Shannon entropy. Can we find other theories like this?
- Are there other interesting informational principles which hold in quantum theory but not in general?
- Is there a connection to statistical physics?
- Can we find quadratic bounds for other non-local games?