Inevitability of knowing less than nothing

Mark M. Wilde
School of Electrical and Computer Engineering, Cornell University

Joint work with Gilad Gour, Sarah Brandsen, Isabelle Geng

Available as arXiv:2208.14424

Main goal of talk

- Give an axiomatic formulation of quantum conditional entropy
- Prove that every function satisfying the two axioms must take on negative values for certain entangled quantum states
- Justifies why any plausible conditional entropy takes on negative values in quantum information (“knowing less than nothing”)

Postulates of classical mechanics

- State of a classical system described by a probability vector $\vec{p}$, with entries satisfying

$$p_x \geq 0 \quad \forall x, \qquad \sum_x p_x = 1$$

- Classical evolution described by a stochastic map/matrix, called a classical channel

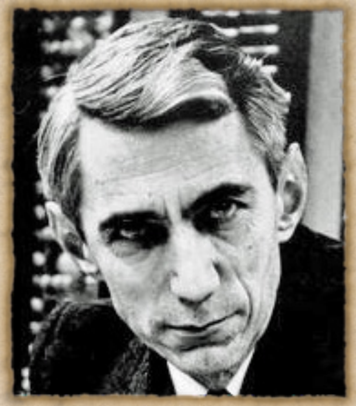
- Probability vectors for composite systems are elements of tensor-product vector spaces

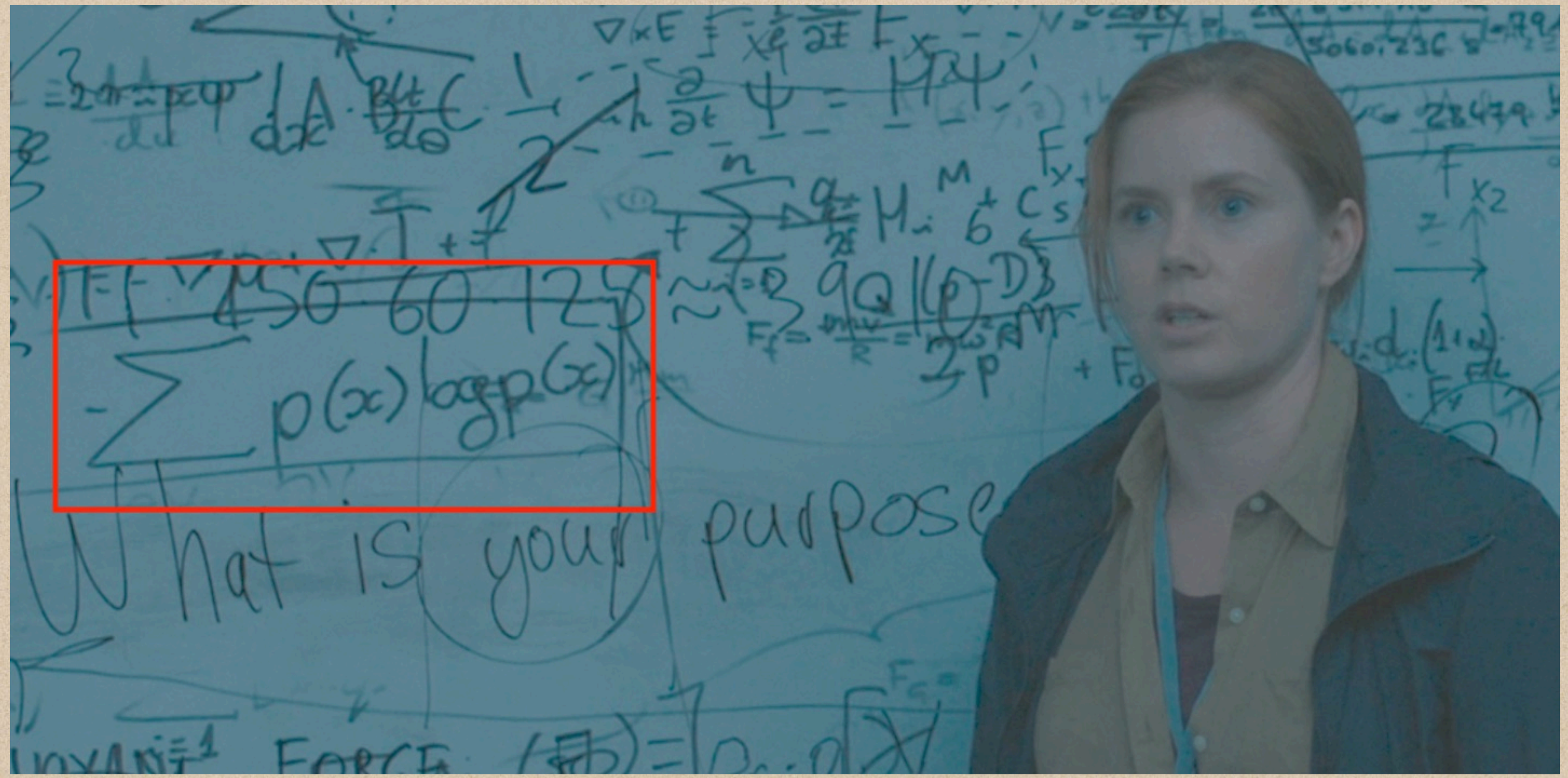
What is entropy?

- "Knowledge gained upon learning the outcome of a random experiment"
- Die Toss: if deterministic, don't learn anything by performing the toss
- If uniformly random, we learn $\log_2 d$ bits
- If successive tosses are independent, expect entropy to be additive
- (stick to finite random variables throughout)

Formulas for entropy

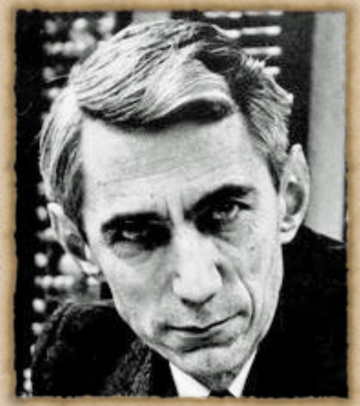
- Shannon entropy $H(X) \equiv \sum_x p(x) \log_2 \left(\frac{1}{p(x)} \right)$





- Renyi entropy $H_\alpha(X) \equiv \frac{1}{1-\alpha} \log_2 \sum_x p^\alpha(x)$, where $\alpha \in (0,1) \cup (1,\infty)$

Axiomatic approach to entropy

- Shannon entropy uniquely defined by some axioms 
- Dropping one of them leads to the Renyi family 
- Why are these natural?
- We can reduce to just two axioms and derive several basic properties of entropy from these

Two basic axioms for entropy

Entropy $\boldsymbol{H}$ is a function that is not equal to the zero function and

1) is an uncertainty measure

2) additive for product distributions: $\boldsymbol{H}(\overrightarrow{p} \otimes \overrightarrow{q}) = \boldsymbol{H}(\overrightarrow{p}) + \boldsymbol{H}(\overrightarrow{q})$

Gour and Tomamichel, arXiv:2006.11164

Mixing operations define uncertainty measures

- What is a mixing operation? A random relabeling of values
- Mathematically: $M\vec{p} \equiv \sum_i q_i P_i \vec{p}$, where $\{q_i\}_i$ is a probability distribution, $\{P_i\}_i$ is a set of permutation matrices
- Mixing operations preserve the uniform distribution: $\vec{u} = M\vec{u}$

Uncertainty measures

Let us define a function f to be an uncertainty measure for a probability distribution $\overrightarrow{p}$ if

1) It does not decrease under the action of a mixing operation: $f(\overrightarrow{p}) \leq f(M\overrightarrow{p})$

2) It is invariant under embeddings: $f(\overrightarrow{p}) = f(\overrightarrow{p} \oplus \overrightarrow{0})$

Two basic axioms for entropy

Entropy $\boldsymbol{H}$ is a function that is not equal to the zero function and

1) is an uncertainty measure

2) additive for product distributions: $\boldsymbol{H}(\overrightarrow{p} \otimes \overrightarrow{q}) = \boldsymbol{H}(\overrightarrow{p}) + \boldsymbol{H}(\overrightarrow{q})$

Gour and Tomamichel, arXiv:2006.11164

Consequences of entropy axioms

- $\boldsymbol{H}$ is non-negative for all probability distributions and equal to zero for degenerate distributions
- $\boldsymbol{H}$ is maximal for uniform distribution $\vec{u}_d$ of size d (among all distributions of size d)
- if we normalize $\boldsymbol{H}$ such that $\boldsymbol{H}(\vec{u}_2) = 1$, then $\boldsymbol{H}(\vec{u}_d) = \log_2 d$

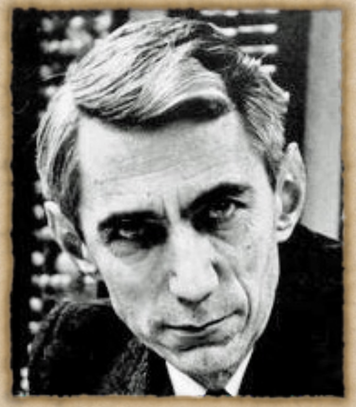
Gour and Tomamichel, arXiv:2006.11164

What is conditional entropy?

- "Knowledge gained upon learning outcome of X given that value of Y has already been observed" (X and Y are two random variables)
- If X and Y independent, no difference between entropy & conditional entropy
- If dependent, knowledge of Y informs about X, and so conditional entropy of X given Y is not greater than entropy of X
- Expect also to be additive for independent trials

Formulas for conditional entropy

- Conditional Shannon entropy



$$H(X|Y) \equiv H(XY) - H(Y) = \sum_y p(y) \left[\sum_x p(x|y) \log_2 \left(\frac{1}{p(x|y)} \right) \right]$$

- Conditional Renyi entropy

$$H_\alpha(X|Y) = \frac{1}{1-\alpha} \log_2 \sum_{x,y} p^\alpha(x,y) p^{1-\alpha}(y)$$

Two basic axioms for conditional entropy

Conditional entropy $\boldsymbol{H}$ is a function that is not equal to the zero function and

1) is a conditional uncertainty measure

2) additive for product distributions:

$$\boldsymbol{H}(X_1X_2 \,|\, Y_1Y_2)_{\vec{p}_{X_1Y_1} \otimes \vec{q}_{X_2Y_2}} = \boldsymbol{H}(X_1 \,|\, Y_1)_{\vec{p}_{X_1Y_1}} + \boldsymbol{H}(X_2 \,|\, Y_2)_{\vec{q}_{X_2Y_2}}$$

Maximal conditional uncertainty

- A mixing operation preserves the uniform distribution, and this implies that the uniform distribution has maximal uncertainty

- What is a bivariate distribution of maximal conditional uncertainty?
  First guess: $\vec{u}_{XY} = \vec{u}_X \otimes \vec{u}_Y$

Maximal conditional uncertainty (ctd.)

- However, many others: $\vec{u}_X \otimes \vec{q}_Y$, where $\vec{q}_Y$ is an arbitrary distribution
- Conditional uncertainty: how well one can guess X when Y is available
- If Y is independent of X, then it is of no use in trying to guess X and uniform distribution for X is most difficult to guess
- This justifies $\{\vec{u}_X \otimes \vec{q}_Y : \vec{q}_Y \in \mathscr{P}_Y\}$ as a maximal conditional uncertainty set

Conditional mixing operations

- A channel $M_{XY\to XY'}$ is a conditional mixing operation if for every distribution $\overrightarrow{q}_Y$, there exists a distribution $\vec{r}_{Y'}$ such that $M_{XY\to XY'}(\overrightarrow{u}_X \otimes \overrightarrow{q}_Y) = \overrightarrow{u}_X \otimes \vec{r}_{Y'}$
- That is, conditional mixing operations preserve the set of bivariate distributions of maximal conditional uncertainty

Conditional uncertainty measure

A function f is a conditional uncertainty measure for a bivariate probability distribution $\overrightarrow{p}_{XY}$ if

1) It does not decrease under the action of a conditional mixing operation: $f(\overrightarrow{p}) \leq f(M\overrightarrow{p})$

2) It is invariant under a local embedding of X:

$f(\overrightarrow{p}) = f((U_{X\to X'} \otimes I_Y)\overrightarrow{p})$, where $U_{X\to X'} = \begin{bmatrix} I \\ 0 \end{bmatrix}$ is a local embedding

Two basic axioms for conditional entropy

Conditional entropy $\boldsymbol{H}$ is a function that is not equal to the zero function and

1) is a conditional uncertainty measure

2) additive for product distributions:

$$\boldsymbol{H}(X_1X_2 \mid Y_1Y_2)_{\vec{p}_{X_1Y_1} \otimes \vec{q}_{X_2Y_2}} = \boldsymbol{H}(X_1 \mid Y_1)_{\vec{p}_{X_1Y_1}} + \boldsymbol{H}(X_2 \mid Y_2)_{\vec{q}_{X_2Y_2}}$$

Consequences of axioms

- $\boldsymbol{H}$ is non-negative for all bivariate probability distributions
- $\boldsymbol{H}$ reduces to an entropy for product distributions: $\boldsymbol{H}(X|Y)_{\vec{p}_X \otimes \vec{q}_Y} = \boldsymbol{H}(\vec{p}_X)$
- $\boldsymbol{H}$ is maximal for uniform distribution of size d (among all distributions of size d)
- if we normalize $\boldsymbol{H}$ such that $\boldsymbol{H}(\vec{u}_2) = 1$, then $\boldsymbol{H}(\vec{u}_d) = \log_2 d$

Let us now enter the quantum world….

Postulates of quantum mechanics

- State of a quantum system described by a density operator ρ:

$$\rho \geq 0, \quad \mathrm{Tr}[\rho] = 1$$

- Evolution of a quantum system described by a completely positive and trace-preserving map, called a quantum channel

- Density operators for composite systems act on tensor-product Hilbert spaces

What is quantum entropy?

- Inspired by the classical case, let us take an axiomatic approach

Axiomatic approach to quantum entropy

Two axioms:

1) $\boldsymbol{H}$ is an uncertainty measure

2) Additivity: $\boldsymbol{H}(\rho \otimes \sigma) = \boldsymbol{H}(\rho) + \boldsymbol{H}(\sigma)$

Quantum mixing operation

- $\mathscr{M}$ is a quantum mixing operation if it is a channel that preserves the uniform state: $\mathscr{M}(\boldsymbol{u}_d) = \boldsymbol{u}_d$

Uncertainty measure

A function f is an uncertainty measure for a quantum state ρ if

1) It does not decrease under action of a quantum mixing operation: $f(\rho) \leq f(\mathcal{M}(\rho))$

2) It is invariant under embeddings: $f(\rho) = f(\rho \oplus \mathbf{0})$

Axiomatic approach to quantum entropy

Two axioms:

1) $\boldsymbol{H}$ is an uncertainty measure

2) Additivity: $\boldsymbol{H}(\rho \otimes \sigma) = \boldsymbol{H}(\rho) + \boldsymbol{H}(\sigma)$

Consequences of axioms

- $\boldsymbol{H}$ is non-negative for all quantum states and equal to zero for pure states
- $\boldsymbol{H}$ is maximal for uniform state $\boldsymbol{u}_d$ of dimension d (among all distributions of size d)
- if we normalize $\boldsymbol{H}$ such that $\boldsymbol{H}(\boldsymbol{u}_2) = 1$, then $\boldsymbol{H}(\boldsymbol{u}_d) = \log_2 d$

Formulas for quantum entropy

- von Neumann entropy $H(\rho) \equiv -\,\mathrm{Tr}[\rho \log_2 \rho]$

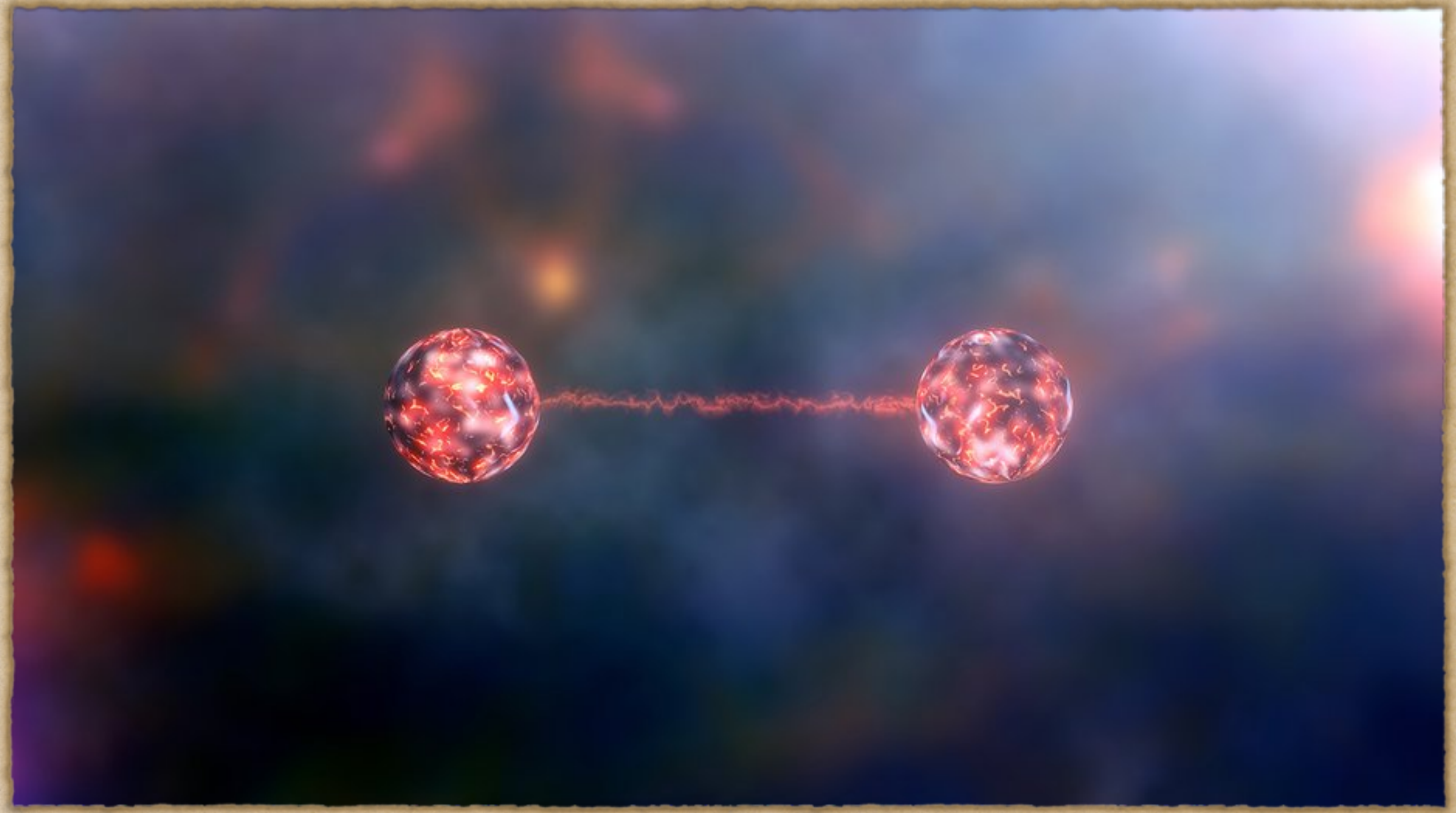
- Renyi entropy $H_\alpha(\rho) \equiv \dfrac{1}{1-\alpha} \log_2 \mathrm{Tr}[\rho^\alpha]$, where $\alpha \in (0,1) \cup (1,\infty)$

Prelude to quantum conditional entropy

- Conditional entropy is defined for a bipartite state
- Before getting to it, let us discuss the phenomenon of quantum entanglement and features of it that distinguish it from the classical case of bivariate distributions

What is entanglement?

- Strong correlation that two parties can share
- Key phenomenon that distinguishes the classical and quantum theories of information
- Useful for teleportation and quantum key distribution



Quantum entanglement 8788
R Horodecki, P Horodecki, M Horodecki, K Horodecki
Reviews of modern physics 81 (2), 865, 2009

Mathematical definition of entanglement

- A state of systems A and B is entangled if it cannot be written as

$$\sum_x p_X(x)\sigma_A^x \otimes \tau_B^x$$

  where p_X is a prob. distribution and $\{\sigma_A^x\}_x$ and $\{\tau_B^x\}_x$ are sets of states

- Separable states can be prepared by local operations and classical communication (i.e., a classical procedure)

"I would not call that one but rather the characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought.

– Erwin Schroedinger, 1935

"Another way of expressing the peculiar situation is: the best possible knowledge of a whole does not necessarily include the best possible knowledge of all its parts, even though they may be entirely separated and therefore virtually capable of being 'best possibly known,' i.e. of possessing, each of them, a representative of its own."

– Erwin Schroedinger, 1935

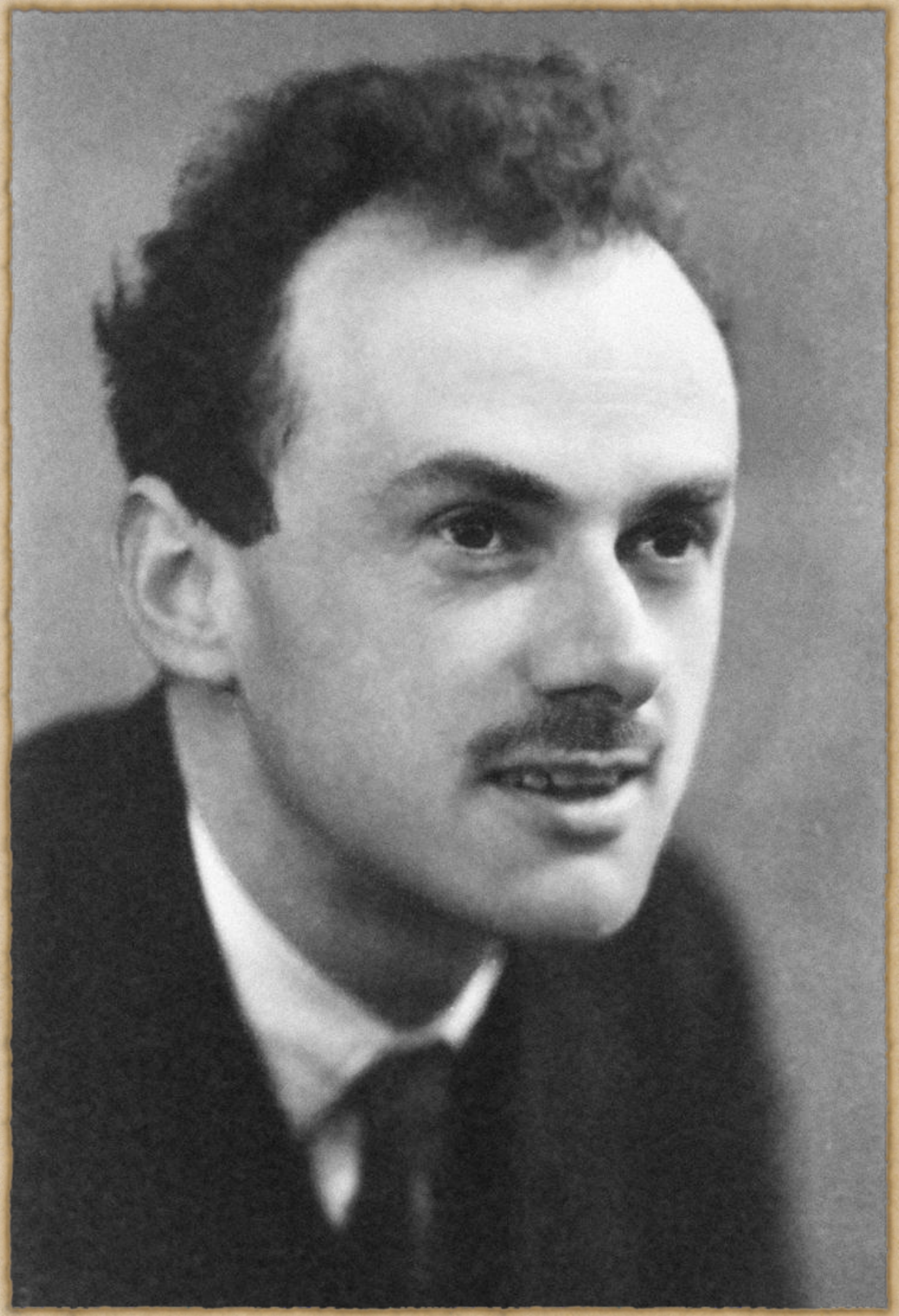
Basic notation

- Dirac notation widely used in quantum information:

$$|0\rangle \equiv \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad |1\rangle \equiv \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- Above are "kets." Dual vectors are "bras":

$$\langle 0| \equiv [1 \quad 0], \qquad \langle 1| \equiv [0 \quad 1]$$



Paul A. M. Dirac

Basic form of entanglement

- The most basic form is the ebit / Bell state / EPR pair:

$$|\Phi^+\rangle\langle\Phi^+|_{AB}$$

where

$$|\Phi^+\rangle_{AB} := \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B)$$

- Closest classical analog of entanglement is a shared secret key, due to concept of monogamy of entanglement

Bell state

- The Bell state is what we call a pure state, which we definitely know and thus should have zero entropy
- Reduced state of Bob's system is a uniformly random mixture of $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$ and thus should have entropy equal to one bit
- Thus, the knowledge of the whole is greater than the knowledge of the parts, and so the conditional entropy goes negative….

Widely used formulas for conditional entropy

- Given a bipartite state ρ_{AB}:
- von Neumann conditional entropy $H(A|B)_\rho \equiv H(\rho_{AB}) - H(\rho_B)$

- Petz-Renyi conditional entropy

$$H_\alpha(A|B)_\rho \equiv \frac{1}{1-\alpha} \log_2 \mathrm{Tr}[\rho_{AB}^\alpha (I_A \otimes \rho_B^{1-\alpha})] \text{ where } \alpha \in (0,1) \cup (1,\infty)$$


Two basic axioms for conditional entropy

Conditional entropy $\boldsymbol{H}$ is a function that is not equal to the zero function and

1) is a conditional uncertainty measure

2) additive for product states:

$$\boldsymbol{H}(A_1A_2|B_1B_2)_{\rho_{A_1B_1}\otimes\sigma_{A_2B_2}} = \boldsymbol{H}(A_1|B_1)_{\rho_{A_1B_1}} + \boldsymbol{H}(A_2|B_2)_{\sigma_{A_2B_2}}$$

Conditional mixing operations

- A channel $\mathcal{M}_{AB \to AB'}$ is a conditional mixing operation if for every state σ_B, there exists a state $\omega_{B'}$ such that $\mathcal{M}_{AB \to AB'}(\boldsymbol{u}_A \otimes \sigma_B) = \boldsymbol{u}_A \otimes \omega_{B'}$
- That is, conditional mixing operations preserve the set of states of maximal conditional uncertainty

Conditional uncertainty measure

- Let us define a function f to be a conditional uncertainty measure for a bipartite state ρ_{AB} if

- 1) It does not decrease under the action of a conditional mixing operation: $f(\rho_{AB}) \leq f(\mathscr{M}(\rho_{AB}))$

- 2) It is invariant under a local embedding of A: $f(\rho_{AB}) = f((U_{A\to A'} \otimes \mathrm{id}_B)(\rho_{AB}))$, where $U_{A\to A'}(\omega_A) = \omega_A \oplus \mathbf{0}$ is a local embedding

Two basic axioms for conditional entropy

Conditional entropy $\boldsymbol{H}$ is a function that is not equal to the zero function and

1) is a conditional uncertainty measure

2) additive for product states:

$$\boldsymbol{H}(A_1A_2|B_1B_2)_{\rho_{A_1B_1}\otimes\sigma_{A_2B_2}} = \boldsymbol{H}(A_1|B_1)_{\rho_{A_1B_1}} + \boldsymbol{H}(A_2|B_2)_{\sigma_{A_2B_2}}$$

Consequences of axioms

- $\boldsymbol{H}$ is non-negative for all separable, unentangled states
- $\boldsymbol{H}$ is can be negative for some entangled states!
- $\boldsymbol{H}$ reduces to an entropy for product states:
  $\boldsymbol{H}(A\,|\,B)_{\rho_A \otimes \sigma_B} = \boldsymbol{H}(\rho_A)$
- $\boldsymbol{H}$ is maximal for $\{\boldsymbol{u}_A \otimes \sigma_B : \sigma_B \in \mathscr{D}_B\}$

Proof of negative conditional entropy

- Construct channel from $\tilde{A}, A, B$ to A', the last of which has dimension $|A|^2$
- Channel first discards system $\tilde{A}$
- Then performs measurement $\{\Phi_{AB}, I_{AB} - \Phi_{AB}\}$ & prepares pure state $|1\rangle\langle 1|_{A'}$ if 1st outcome obtained & orthogonal state $\dfrac{I_{A'} - |1\rangle\langle 1|_{A'}}{|A|^2 - 1}$ else
- This is a conditional mixing operation

Proof of negative conditional entropy (ctd.)

- Then

$$
\begin{aligned}
\boldsymbol{H}(A\,|\,B)_{\Phi_{AB}} + \log_2|A| &= \boldsymbol{H}(A\,|\,B)_{\Phi_{AB}} + \boldsymbol{H}(\tilde{A})_{\boldsymbol{u}_{\tilde{A}}} \\
&= \boldsymbol{H}(A\tilde{A}\,|\,B)_{\Phi_{AB}\otimes \boldsymbol{u}_{\tilde{A}}} \\
&\leq \boldsymbol{H}(A')_{\mathcal{N}_{AB\tilde{A}\to A'}(\Phi_{AB}\otimes \boldsymbol{u}_{\tilde{A}})} \\
&= \boldsymbol{H}(A')_{|1\rangle\langle 1|} \\
&= 0
\end{aligned}
$$

Summary

- Formulated an axiomatic approach to quantum conditional entropy, based on two sensible and simple axioms
- Used these axioms in a simple proof to conclude that quantum conditional entropy is negative for certain entangled states
- Open question: In our work, we proved that the conditional min-entropy is a lower bound on any plausible conditional entropy. We would like to prove that the conditional max-entropy is an upper bound