Information and Context

J. Acacio de Barros

School of Humanities and Liberal Studies
San Francisco State University

IMBS, UC Irvine, 2019

Why work with information?

- Information is a fundamental concept in Engineering
  - Data science
  - Communication theory
  - Signal processing
  - Compression algorithms

Why work with information?

Information is also an important concept in

- Physics
  - Statistical mechanics
  - In certain interpretations of quantum mechanics
  - Quantum computation
  - Considered by some physicists as part of the underlying fabric of reality:
    - Wheeler's it-from-bit
    - D'Arianno's derivation of QFT
- Philosophy
  - Language
  - Mind (IIT)
- Social sciences?

What is information?

- Information is a difficult to define concept.
- In physics and in engineering:
    - Claude Shannon's definition from theory of communications.
    - Requires a probability space $(\Omega, \mathcal{F}, p)$

Information and context

- Kolmogorov's axioms for $(\Omega, \mathcal{F}, p)$
  - $p : \mathcal{F} \to [0, 1]$
  - $p(\Omega) = 1$
  - $p(A \cup B) = p(A) + p(B)$, $A \cap B = \emptyset$.
- For some random variables, $\neg\exists\,(\Omega, \mathcal{F}, p)$
  - Contextuality
- How do we extend Shannon to those situations?

Outline

Outline

From linguistics

- A text or speech is considered contextual if the parts which are related to it are connected to its meaning.
- Consider the statement:
  - "Cheap dates are great" - True or False?

From linguistics

- A text or speech is considered contextual if the parts which are related to it are connected to its meaning.
- Consider the statement:
    - "Cheap dates are great" - True or False?
    - Context 1: conversation about social engagements
    - Context 2: discussion about the *Phoenix dactylifera* fruits.

Contextuality is about truth values

- A system of propositions is contextual (or exhibit contextuality) if truth values change with context.
- Seems ok in linguistics
    - But what about physics or psychology?
- We need a more measurable definition of contextuality.

Underpinnings for contextuality

- Start with probability space, $(\Omega, \mathcal{F}, p)$
  - Ω is a sample space
  - $\mathcal{F}$ is a σ-algebra over Ω
  - p a function $p : \mathcal{F} \to [0, 1]$
- Kolmogorov's Axioms
  - K1. $p(\Omega) = 1$
  - K2. $p(A \cup B) = p(A) + p(B)$ for $A, B \in \mathcal{F}$ and $A \cap B = \emptyset$.

Representations as RV

- Outcomes of measurements can be modeled with RV.
- Random variable $\mathbf{R} : \Omega \rightarrow O$
  - O is the set of outcomes (e.g. $\{1, 2, 3, 4, 5, 6\}$, $\{-1, 1\}$, $\mathbb{R}$, etc)
  - $\mathbf{R}$ is measurable
- Example: $\Omega = \{(1, 1), (1, 2), \ldots, (5, 6), (6, 6)\}$, $\mathcal{F} = 2^{\Omega}$, $p(\omega_i) = 1/36$

Not all observations can be modeled with RV?

Or not all RV have a joint?

- Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be ± 1-valued RV (e.g. $O = \{-1, 1\}$)
- Let
$$E(\mathbf{XY}) = E(\mathbf{XZ}) = E(\mathbf{YZ}) = -1.$$
- No $(\Omega, \mathcal{F}, p)$ exists.
    - If it existed, we would reach a contradiction: $\mathbf{X} = 1 \rightarrow \mathbf{Y} = -1 \rightarrow \mathbf{Z} = 1 \rightarrow \mathbf{X} = -1$

Contradiction comes context-independency!

$$E(\mathbf{XY}) = E(\mathbf{XZ}) = E(\mathbf{YZ}) = -1.$$

- Contradiction: $\mathbf{X} = 1 \rightarrow \mathbf{Y} = -1 \rightarrow \mathbf{Z} = 1 \rightarrow \mathbf{X} = -1$
- Observations cannot be contradictory: our model is the problem!
  - Assumption: $\mathbf{X}$ in experiment measuring $E(\mathbf{XY})$ is the same as $E(\mathbf{XZ})$
- Consider:

Context 1: X and Y

Context 2: X and Z

Context 3: Y and Z

- If we index variables (e.g. $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Y}_1$, $\mathbf{Y}_3$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$), no contradiction.

No need for perfect correlations

- For three **X**, **Y**, and **Z** that are ± 1-valued, $-1 \leq \mathbf{XY} + \mathbf{XZ} + \mathbf{YZ}$.
- Therefore

$$-1 \leq E(\mathbf{XY}) + E(\mathbf{XZ}) + E(\mathbf{YZ}) \leq 3.$$

- Necessary and sufficient conditions for the non-contextuality.
- For $E(\mathbf{X}) = E(\mathbf{Y}) = E(\mathbf{Z}) = 0$, $E(\mathbf{XY}) = E(\mathbf{XZ}) = E(\mathbf{YZ}) = \epsilon$, and $E(\mathbf{XYZ}) = \beta$, non-contextual if within the bounds of a polytope on (ϵ, β) with vertices:
    - $(1, 0)$, $(0, -1)$, $(0, 1)$, and $(-1/3, 0)$.

Outline

Properties in physics

- Newtonian physics:
  - state of a system is a point in phase space (e.g. $(\mathbf{p}, \mathbf{r})$ for a single particle)
  - properties are defined by subsets of the phase space
  - properties are non-contextual
- Quantum physics:
  - states are vectors (or operators) in a Hilbert space
  - properties are given by Hermitian operators
  - properties may be contextual

Crash course in quantum physics

All you need to know about quantum mechanics in one single slide!

- Binary properties are projection operators
  - for a **w** and a projector $\hat{P}$, $\hat{P}\mathbf{w} = \mathbf{w}$ means **w** is an eigenvector with eigenvalue 1
    - we say **w** has property P associated to $\hat{P}$.
  - For a **v** orthogonal to **w**, $\hat{P}\mathbf{v} = 0$; **v** is eigenvector with eigenvalue 0
    - **v** does not have property P.
  - Linear combinations $a\mathbf{w} + b\mathbf{v}$ are not an eigenvector of $\hat{P}$
    - We cannot tell it has property P (unless we measure it).
    - Once we measure P, state "collapses" to either **w** or **v**.

Example of quantum contextual properties

Four dimensional vector space $\mathbb{R}^4$.

$$
\begin{aligned}
\hat{P}_{0,0,0,1} + \hat{P}_{0,0,1,0} + \hat{P}_{1,1,0,0} + \hat{P}_{1,-1,0,0} &= 1, \\
\hat{P}_{0,0,0,1} + \hat{P}_{0,1,0,0} + \hat{P}_{1,0,1,0} + \hat{P}_{1,0,-1,0} &= 1, \\
\hat{P}_{1,-1,1,-1} + \hat{P}_{1,-1,-1,1} + \hat{P}_{1,1,0,0} + \hat{P}_{0,0,1,1} &= 1, \\
\hat{P}_{1,-1,1,-1} + \hat{P}_{1,1,1,1} + \hat{P}_{1,0,-1,0} + \hat{P}_{0,1,0,-1} &= 1, \\
\hat{P}_{0,0,1,0} + \hat{P}_{0,1,0,0} + \hat{P}_{1,0,0,1} + \hat{P}_{1,0,0,-1} &= 1, \\
\hat{P}_{1,-1,-1,1} + \hat{P}_{1,1,1,1} + \hat{P}_{1,0,0,-1} + \hat{P}_{0,1,-1,0} &= 1, \\
\hat{P}_{1,1,-1,1} + \hat{P}_{1,1,1,-1} + \hat{P}_{1,-1,0,0} + \hat{P}_{0,0,1,1} &= 1, \\
\hat{P}_{1,1,-1,1} + \hat{P}_{-1,1,1,1} + \hat{P}_{1,0,1,0} + \hat{P}_{0,1,0,-1} &= 1, \\
\hat{P}_{1,1,1,-1} + \hat{P}_{-1,1,1,1} + \hat{P}_{1,0,0,1} + \hat{P}_{0,1,-1,0} &= 1.
\end{aligned}
$$

Another contextual example in $\mathbb{C}^4$

- Alice: A, A'
- Bob: B, B'
- For non-contextuality

$$-2 \leq E\left(\mathbf{AB}\right) + E\left(\mathbf{AB'}\right) + E\left(\mathbf{A'B}\right) - E\left(\mathbf{A'B'}\right) \leq 2$$

$$-2 \leq E\left(\mathbf{AB}\right) + E\left(\mathbf{AB'}\right) - E\left(\mathbf{A'B}\right) + E\left(\mathbf{A'B'}\right) \leq 2$$

$$-2 \leq E\left(\mathbf{AB}\right) - E\left(\mathbf{AB'}\right) + E\left(\mathbf{A'B}\right) + E\left(\mathbf{A'B'}\right) \leq 2$$

$$-2 \leq -E\left(\mathbf{AB}\right) + E\left(\mathbf{AB'}\right) + E\left(\mathbf{A'B}\right) + E\left(\mathbf{A'B'}\right) \leq 2$$

- Quantum theory predicts
  $E\left(\mathbf{AB}\right) + E\left(\mathbf{AB'}\right) + E\left(\mathbf{A'B}\right) - E\left(\mathbf{A'B'}\right) < -2$
- Experimentally verified.

Why is contextuality important?

- Presents challenges to interpretations of theory and concept of property
- It is purely quantum
  - Cannot be reproduced classically (not without cost)
- It is a resource in quantum computation

What about psychology?

- Order effect (explicit contextuality)
  - Clinton is honest and trustworthy? (non-comparative: 50%; comparative 57%)
  - Gore is honest and trustworthy? (non-comparative: 68%; comparative 60%)
- Cervantes and Dzhafarov (hidden contextuality)
  - Gerda, Troll; Beautiful, Unattractive
  - Gerda, Troll; Kind, Evil
  - Snow Queen, Old Finn Woman; Beautiful, Unattractive
  - Snow Queen, Old Finn Woman; Kind, Evil
- But it is not like quantum contextuality!

Outline

How to define and measure information?

Shannon's mathematical theory of communication

- Information is related to how surprising a source s is.
  - Quantified by $-\log p(s)$
  - E.g. for four equally surprising outcomes:
    - $p(s_i) = 1/4$
    - $I = -\log_2 p(s_1) = 2$
    - Interpret as two bits of information.
- For a source represented by r.v. $\mathbf{X}$ with output $\{x_1, x_2, \ldots, x_N\}$, average amount of information is

$$H = -\sum_{i=1}^{N} p(x_i) \log_2 p(x_i) \,. \tag{1}$$

- H is known as Shannon's *entropy*.

Classical Entropy

- In

$$H = -\sum_{i=1}^{N} p(x_i) \log_2 p(x_i), \tag{2}$$

  base 2 gives unit of information
  - if 2, measure in bits.
- Shannon's coding theorem:
  - it is impossible to compress data such that the average number of bits per symbol is less than H without data loss
  - it is possible to create a code whose rate is arbitrarily close to H
- Example:
  - ASCII 8 bits
  - 26 english letters require 4.7 binary bits in simple coding (no probability)
  - Morse code allows for 4.14 bits (average)

Quantum sources

- A property O of a system S is a Hermitian operator $\hat{O}$ on a Hilbert space $\mathcal{H}$.
    - Hermitian operators are called *observables*.
    - State is a vector in $\mathcal{H}$ or by a density operator (a positive semidefinite observable with trace one).
        - vectors are pure states
        - density operator are mixed states.
    - E.g. the normalized vector $\mathbf{w} \in \mathcal{H}$ is a pure state; the density operator is $\hat{\rho}_w = \mathbf{w}\underline{\omega}$, where $\underline{\omega}$ is the dual to $\mathbf{w}$.
    - Density operators are more general.

Expectations from $\hat{\rho}$

- Three dimensional Hilbert space, basis $\mathbf{e}_i$, $i = 1, 2, 3$.
- In vector formalism, the expectation $\left\langle \hat{P}_i \right\rangle = \left| \hat{P}_i \mathbf{w} \right|^2$.
- For $\hat{\rho} = \mathbf{w}\underline{\omega}$, $\left\langle \hat{P}_i \right\rangle = \left| \hat{P}_i \mathbf{w} \right|^2 = \mathrm{Tr}\left(\hat{\rho} \hat{P}_i \right)$
- If we write $\hat{\rho}_M = c_1 \hat{\rho}_1 + c_2 \hat{\rho}_2$, the linearity of the trace gives us that $\left\langle \hat{P}_i \right\rangle = \mathrm{Tr}\left(\hat{\rho}_M \hat{P}_i \right) = c_1 \mathrm{Tr}\left(\hat{\rho}_1 \hat{P}_i \right) + c_2 \mathrm{Tr}\left(\hat{\rho}_2 \hat{P}_i \right)$, or

$$\left\langle \hat{P}_i \right\rangle = c_1 \left\langle \hat{P}_i \right\rangle_1 + c_2 \left\langle \hat{P}_i \right\rangle_2$$

- For any observable $\hat{O}$ it follows that $\left\langle \hat{O} \right\rangle = \mathrm{Tr}\left(\hat{\rho} \hat{O} \right)$.

von Neumann entropy

- The quantum system itself is a source, equivalent to **X** in classical communications theory.
- For a state $\hat{\rho}$, its von Neumann entropy [**?**] is defined as

$$S = -\mathrm{Tr}\left(\hat{\rho}\log\hat{\rho}\right). \tag{3}$$

- E.g. for three binary properties, represented by basis $\mathbf{e}_i$, $i = 1, \ldots, N$.
  - Observables $\hat{P}_i = \mathbf{e}_i \underline{\epsilon}_i$ have binary outcomes 0 or 1.
  - If the state is $\hat{\rho} = \sum_i c_i \mathbf{w}_i \underline{\omega}_i$, it follows that $S = -\sum c_i \log c_i$.

H is not the same as S!

- Despite similarities, Shannon is not the same as von Neumann
  - If $\hat{\rho}$ was not a proper mixture of non-orthogonal projectors, Shannon would not follow.
- However, there exists a quantum coding theorem.
- For a general orthomodular lattice, natural measure of informational content is von Neumann's entropy
  - Shannon's entropy emerges as the measure for classical-like situations.

Outline

Two sources

- Consider **X** and **Y**, valued $+1$ or -1 with zero expectations.
- Shannon's entropy yields for each, separately,

$$H(\mathbf{X}) = -p_x \log p_x - p_{\overline{x}} \log p_{\overline{x}} \qquad (4)$$
$$= -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} =$$
$$H(\mathbf{Y}) = 1.$$

- **X** and **Y** seem to have have, combined, two bits of information.

Two sources

- Consider **X** and **Y**, valued $+1$ or -1 with zero expectations.
- Shannon's entropy yields for each, separately,

$$\begin{aligned} H(\mathbf{X}) &= -p_x \log p_x - p_{\overline{x}} \log p_{\overline{x}} \\ &= -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = \\ H(\mathbf{Y}) &= 1. \end{aligned} \tag{4}$$

- **X** and **Y** seem to have have, combined, two bits of information.
- True only if uncorrelated.
  - E.g. $E(\mathbf{XY}) = 1$ or $E(\mathbf{XY}) = -1$ would reduce information.

How to compute information for two sources?

- Consider **X** and **Y** as a pair: outcomes are all pairs, namely xy, $x\overline{y}$, $\overline{x}y$, and $\overline{xy}$.
- Setting $E(\mathbf{XY}) = \alpha$, it follows that

$$p_{xy} + p_{x\overline{y}} + p_{\overline{x}y} + p_{\overline{xy}} = 1, \tag{5}$$

$$p_{xy} + p_{x\overline{y}} - p_{\overline{x}y} - p_{\overline{xy}} = 0, \tag{6}$$

$$p_{xy} - p_{x\overline{y}} + p_{\overline{x}y} - p_{\overline{xy}} = 0, \tag{7}$$

$$p_{xy} - p_{x\overline{y}} - p_{\overline{x}y} + p_{\overline{xy}} = \alpha. \tag{8}$$

For two sources, here is entropy

- Solution is

$$p_{xy} = p_{\overline{x}\overline{y}} = \frac{1}{4}(1+\alpha), \tag{9}$$

$$p_{\overline{x}y} = p_{\overline{x}y} = \frac{1}{4}(1-\alpha), \tag{10}$$

- Entropy:

$$H(\mathbf{X}, \mathbf{Y}) = -\frac{1}{4}(1+\alpha)\log\left(\frac{1}{4}(1+\alpha)\right) - \frac{1}{4}(1-\alpha)\log\left(\frac{1}{4}(1-\alpha)\right).$$

Information for two sources

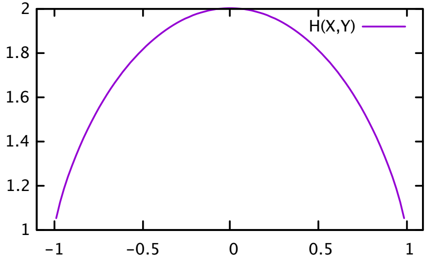

Figure: Joint entropy of **X** and **Y** as a function of the correlation α.

Three variables

- Consider ± 1-valued random variables $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, zero expectation, correlations $E(\mathbf{XY}) = E(\mathbf{XZ}) = E(\mathbf{YZ}) = \epsilon$, $E(\mathbf{XYZ}) = \beta$.

$$p_{xyz} + p_{xy\bar{z}} + p_{x\bar{y}z} + p_{\bar{x}yz} + p_{\overline{xy}z} + p_{\bar{x}y\bar{z}} + p_{x\overline{yz}} + p_{\overline{xyz}} = 1, \tag{11}$$

$$p_{xyz} + p_{xy\bar{z}} + p_{x\bar{y}z} - p_{\bar{x}yz} - p_{\overline{xy}z} - p_{\bar{x}y\bar{z}} + p_{x\overline{yz}} - p_{\overline{xyz}} = 0, \tag{12}$$

$$p_{xyz} + p_{xy\bar{z}} - p_{x\bar{y}z} + p_{\bar{x}yz} - p_{\overline{xy}z} + p_{\bar{x}y\bar{z}} - p_{x\overline{yz}} - p_{\overline{xyz}} = 0, \tag{13}$$

$$p_{xyz} - p_{xy\bar{z}} + p_{x\bar{y}z} + p_{\bar{x}yz} + p_{\overline{xy}z} - p_{\bar{x}y\bar{z}} - p_{x\overline{yz}} - p_{\overline{xyz}} = 0, \tag{14}$$

$$p_{xyz} + p_{xy\bar{z}} - p_{x\bar{y}z} - p_{\bar{x}yz} + p_{\overline{xy}z} - p_{\bar{x}y\bar{z}} - p_{x\overline{yz}} + p_{\overline{xyz}} = \epsilon, \tag{15}$$

$$p_{xyz} - p_{xy\bar{z}} + p_{x\bar{y}z} - p_{\bar{x}yz} - p_{\overline{xy}z} + p_{\bar{x}y\bar{z}} - p_{x\overline{yz}} + p_{\overline{xyz}} = \epsilon, \tag{16}$$

$$p_{xyz} - p_{xy\bar{z}} - p_{x\bar{y}z} + p_{\bar{x}yz} - p_{\overline{xy}z} - p_{\bar{x}y\bar{z}} + p_{x\overline{yz}} + p_{\overline{xyz}} = \epsilon, \tag{17}$$

$$p_{xyz} - p_{xy\bar{z}} - p_{x\bar{y}z} - p_{\bar{x}yz} + p_{\overline{xy}z} + p_{\bar{x}y\bar{z}} + p_{x\overline{yz}} - p_{\overline{xyz}} = \beta. \tag{18}$$

Solution

- Solution is

$$p_{xyz} = \frac{1}{8}\left(1+\beta+3\epsilon\right), \;\; p_{xy\overline{z}} = p_{x\overline{y}z} = p_{\overline{x}yz} = \frac{1}{8}\left(1-\beta-3\epsilon\right) \tag{19}$$

$$p_{\overline{xy}z} = p_{\overline{x}y\overline{z}} = p_{x\overline{yz}} = \frac{1}{8}\left(1+\beta-3\epsilon\right), \;\; p_{\overline{xyz}} = \frac{1}{8}\left(1-\beta+3\epsilon\right). \tag{20}$$

- Some p's may be negative for certain values of ϵ and β.
  - Non-negative solutions correspond to non-contextual cases
  - Negative solutions are contextual.

Non-contextual entropy

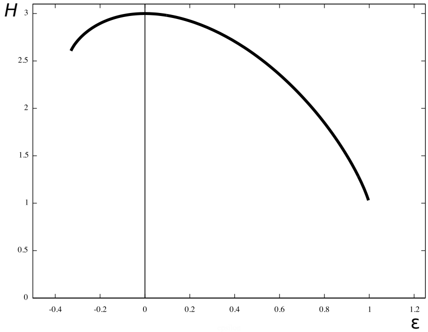

Figure: Entropy $H(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ for $E(\mathbf{XY}) = E(\mathbf{XZ}) = E(\mathbf{YZ}) = \epsilon$ and $E(\mathbf{XYZ}) = 0$ as a function of ϵ. The maximum of 3 bits occurs when $\epsilon = 0$.

What about contextual sources?

- Outside of $-1/3 \leq \epsilon \leq 1$, Shannon's entropy is not defined.
- But **X**, **Y**, and **Z** have informational content outside of the probability polytope.
- How do we measure information outside of the polytope?

What about contextual sources?

- Outside of $-1/3 \leq \epsilon \leq 1$, Shannon's entropy is not defined.
- But **X**, **Y**, and **Z** have informational content outside of the probability polytope.
- How do we measure information outside of the polytope?

- Going back to $\epsilon = -1$.
  - If we know **X** we know **Y** and **Z**, so 1 bit?
  - No. **X** in the context of **Y** cannot be the same as in the context of **Z**, so information content is more than one bit.
  - Variables are $\mathbf{X_Y}$, $\mathbf{X_Z}$, $\mathbf{Y_X}$, $\mathbf{Y_Z}$, $\mathbf{Z_X}$, and $\mathbf{Z_Y}$

Outline

Negative Probabilities

- "Negative probabilities" were introduced by Dirac.
- Name *negative probabilities* may be misleading:
  - some probabilities of atomic events may be negative, but not all
  - observable probabilities may not be negative.
- Perhaps better term is "signed probability".

Extending Shannon

- Let $(\Omega, \mathcal{F}, p^*)$ be a negative probability space.
- The extended entropy S_{NP} for p^* is

$$S_{NP} = -\sum_{\omega_i \in \Omega} \left| p^*(\omega_i) \right| \log_2 \left| p^*(\omega_i) \right|. \qquad (21)$$

- The absolute value comes from using the expression $A = \sum_{\omega_i \in \Omega} \left| p^*(\omega_i) \right|$ as a measure of how contextual a probabilistic system is [**?**].
  - If $A = 1$, no contextuality, since $A = \sum_{\omega_i \in \Omega} p^*(\omega_i)$, which implies that p^*'s are non-negative.
  - If $A > 1$, then the system is contextual.
- S_{NP} is Shannon when p^* is non-negative.

S_{NP} for symmetric case

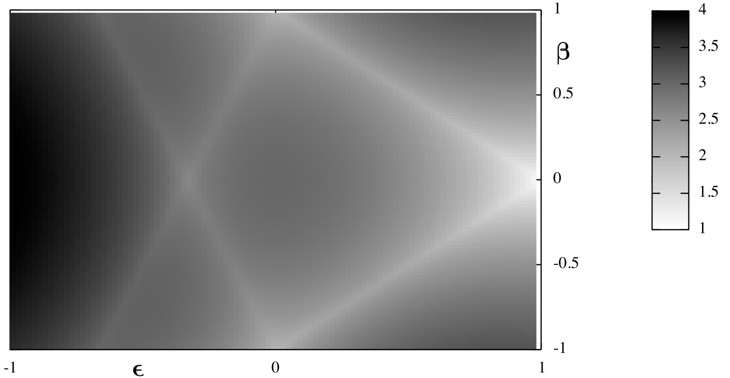

Figure: Surface plot of S_{NP} as a function of ϵ (horizontal axis) and β (vertical axis). Lighter regions correspond to less entropy, whereas darker regions to more entropy.

Cross section of S_{NP}

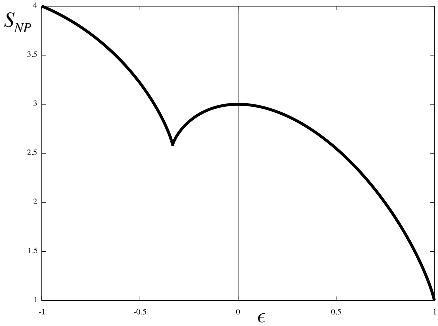

Figure: Cross section of S_{NP} as a function of ϵ and for $\beta = 0$. We see a local maxima at $\epsilon = 0$ and a global maxima at $\epsilon = -1$, where $S_{NP} = 4$.

Outline

Final remarks

- Contextual informational in key in
  - physics
  - engineering
- Shannon and von Neumann provide measures
  - von Neumann deals with contextual information, but requires $\mathcal{H}$
- S_{NP} generalizes Shannon for some non-quantum contextual states
  - consistent with von Neumann's quantum inequalities

Acknowledgments

"I not only use all the brains I have, but all I can borrow"
- *Woodrow Wilson*

- Patrick Suppes (Stanford); Gary Oas (Stanford); Pawel Kurzynski (Adam Mickiewicz); Federico Holik (La Plata); Carlos Montemayor (SFSU); John Perry (Stanford); Stephan Hartmann (LMU); Francisco Doria (UFRJ); Gregory Chaitin (UFRJ)

Acknowledgments

"I not only use all the brains I have, but all I can borrow"
- Woodrow Wilson

- Patrick Suppes (Stanford); Gary Oas (Stanford); Pawel Kurzynski (Adam Mickiewicz); Federico Holik (La Plata); Carlos Montemayor (SFSU); John Perry (Stanford); Stephan Hartmann (LMU); Francisco Doria (UFRJ); Gregory Chaitin (UFRJ)

Thank you!!