Quantifying information synergy

Using intermediate variables

Quax R, Har-Shemesh O, Sloot P. *Quantifying synergistic information using intermediate stochastic variables*. Entropy. 2017 Feb;19(2):85.

Rick Quax: Computational Science, University of Amsterdam, The Netherlands.

Counterexample?

$$X_1, X_2 \mapsto S_1 \qquad S_1 \equiv (2 - X_1 + X_2) \bmod 3$$

$$X_1, X_2 \mapsto S_1, S_2 \qquad S_2 = (X_1 + X_2) \bmod 3$$

Rows: X_1; columns: X_2

2	0	1
1	2	0
0	1	2

$S_1(X_1, X_2)$

Columns: X_2

0	1	2
1	2	0
2	0	1

$S_2(X_1, X_2)$

$$I(X_1 : S_i) = I(X_2 : S_i) = 0 \qquad I(X_1 : S_1, S_2) = \log_2 3$$

Counterexample?

$$X_1, X_2 \mapsto S_1 \qquad S_1 \equiv (2 - X_1 + X_2) \bmod 3$$

$$X_1, X_2 \mapsto S_1, S_2 \qquad S_2 = (X_1 + X_2) \bmod 3$$

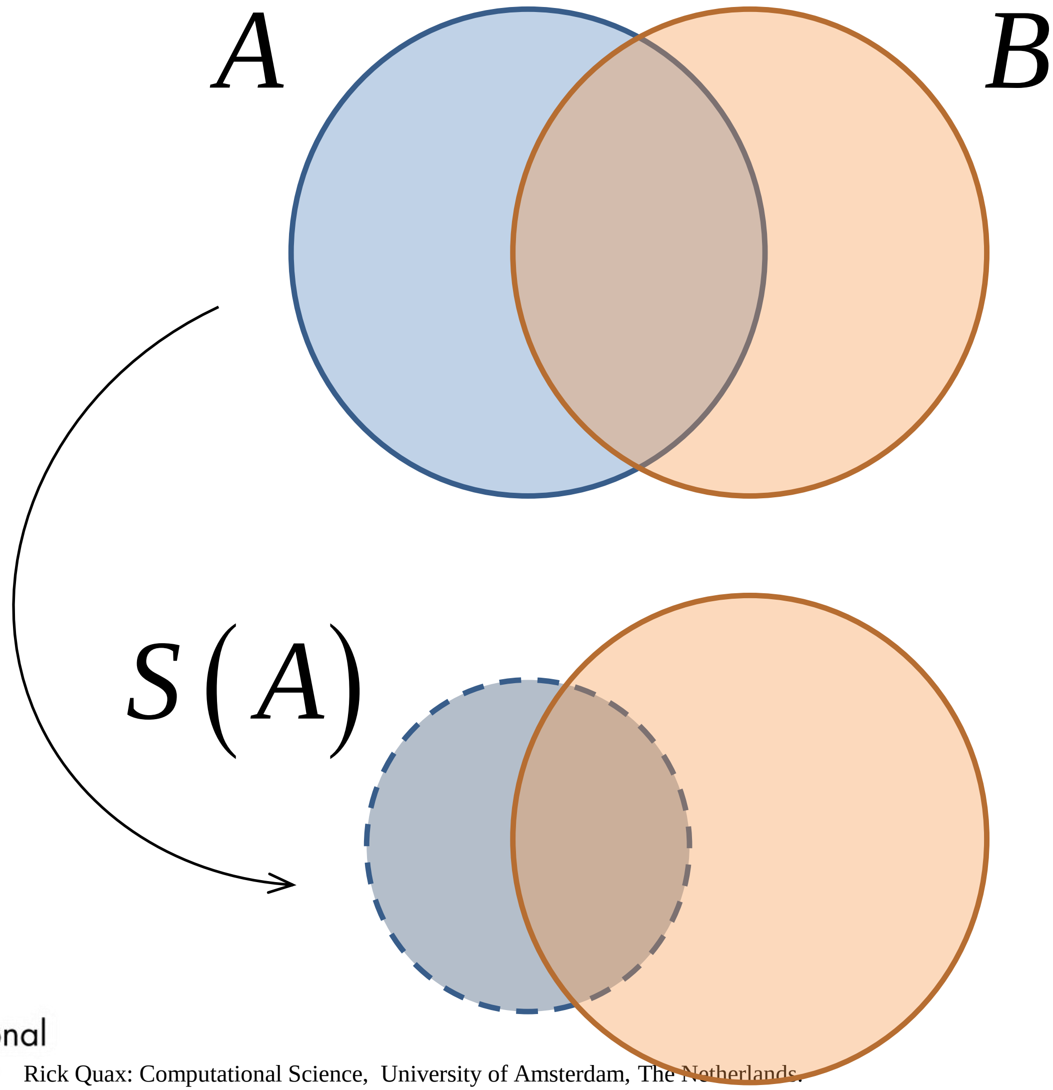
$$S_1(X_1, X_2) = \begin{array}{c|c} & \overbrace{\qquad\qquad\qquad}^{X_2} \\ X_1 & \begin{bmatrix} 2 & 0 & 1 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix} \end{array} \qquad S_2(X_1, X_2) = \begin{array}{c} \overbrace{\qquad\qquad\qquad}^{X_2} \\ \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{bmatrix} \end{array}$$

$$I_{\text{syn}}(X_1, X_2 : S_i) = \log_2 3 \qquad I(S_1 : S_2) = 0 \qquad I_{\text{syn}}(X_1, X_2 : S_1, S_2) = 0$$

An alternative starting point for defining

SYNERGISTIC INFORMATION

Intuition

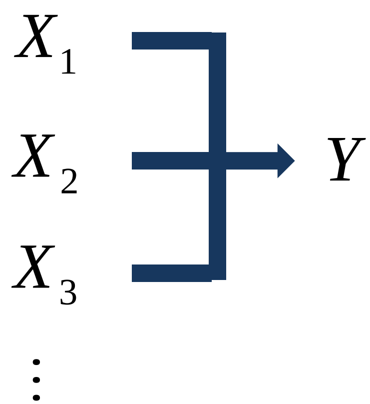


Setting

$$X \equiv \{X_1, X_2, \ldots\}$$

How much synergy in Y about X

$$p(X)$$



$$p(Y|X)$$

Proposed definition: “synergistic”

Reminder: $X \equiv \{X_1, X_2, \ldots\}.$

S_j is fully synergistic about X iff

$$I(S_j : X) > 0,$$

$$I(S_j : X_i) = 0.$$

Proposed definition: “synergistic”

Reminder: $X \equiv \{X_1, X_2, \ldots\}.$

S_j is fully synergistic about X iff

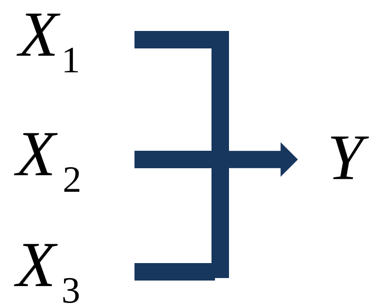
$$I\left(S_j : X\right) > 0,$$

$$I\left(S_j : X_i\right) = 0.$$

$$S \equiv \left\{S_j\right\}_j$$

First intuition

$p(X)$



$p(Y|X)$

$$\text{"synergy"} \overset{?}{=} I(Y:S)$$

Counterexample

$$S_1, S_2 \in S$$

$$X = \{X_1, X_2\}$$

$$X_i \in \{0,1,2\}$$

$$\Pr(X) = 1/9$$

$$S_1 = \begin{cases} (X_1 + X_2) \bmod 3 = 0: & 1 \\ \text{otherwise}: & 0 \end{cases}$$

$$S_2 = \begin{cases} X_1 = (X_2 + 2) \bmod 3: & 1 \\ \text{otherwise}: & 0 \end{cases}$$

$$I(S_1 : X_1) = 0$$

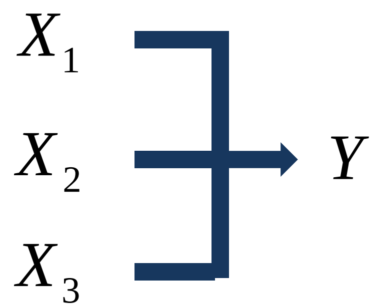
$$I(S_2 : X_1) = 0$$

$$I(S_1, S_2 : X_1) \approx 1.22$$

$Y=X_1$ would be synergistic?!

Second intuition

$p(X)$



$p(Y|X)$

$$\text{"synergy"} \overset{?}{=} \sum_j I(Y:S_j)$$

Counterexample

$$X \equiv \{X_1, X_2, X_3\}$$

$$X_i \in \{0,1\}$$

$$\Pr(X)$$

$$S_1 = X_1 \oplus X_2$$

$$S_2 = X_2 \oplus X_3$$

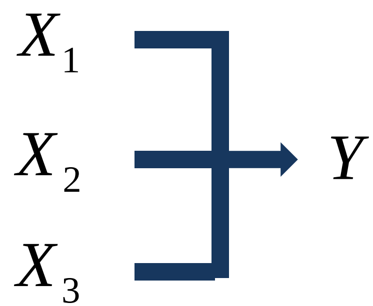
$$S_3 = X_1 \oplus X_3$$

$$S_4 = X_1 \oplus X_2 \oplus X_3$$

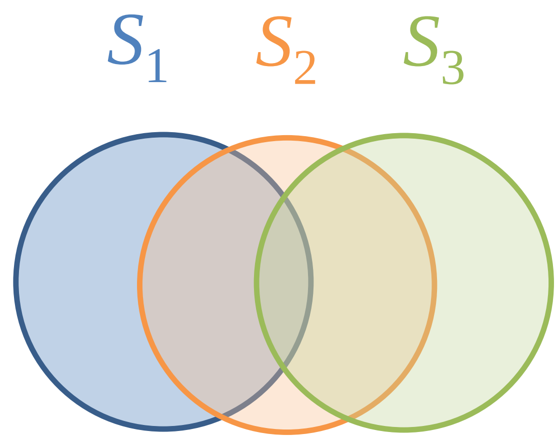
Choosing Y={S_1,S_2} would result in 3 bits of synergy

Why fails? → S_i overlap!

$p(X)$

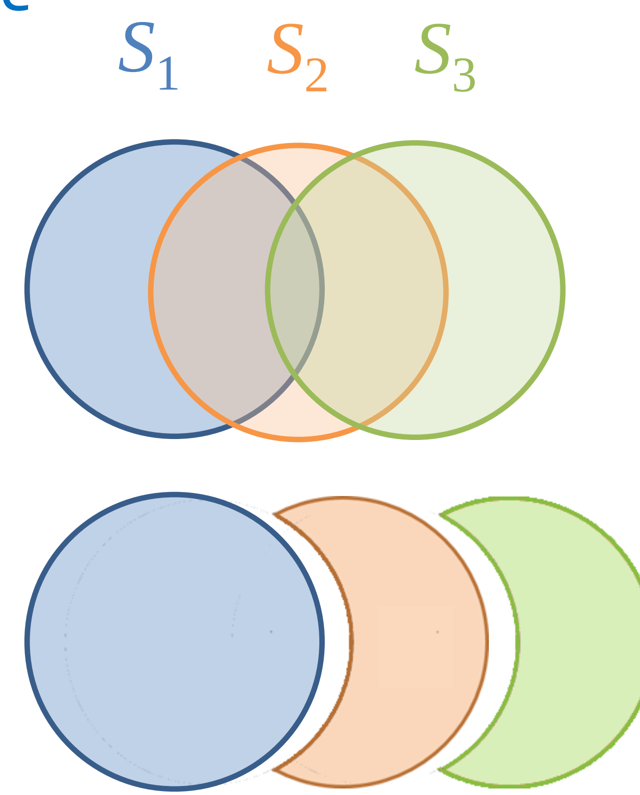


$p(Y|X)$



$$\text{"synergy"} \stackrel{?}{=} \sum_j I(Y:S_j)$$

Final result



$$S$$

$$S^* \equiv \left(S_i^\perp : S_i^\perp \in D\left(S_i ; S_1, \ldots, S_{i-1} \right) \right)_i ,$$

For a given ordering of S_i

$$I_{\text{syn}}\left(X \rightarrow Y \right) \equiv \max_{\text{ordering for } S^*} \sum_{S_i \in S^*} I\left(Y : S_i \right).$$

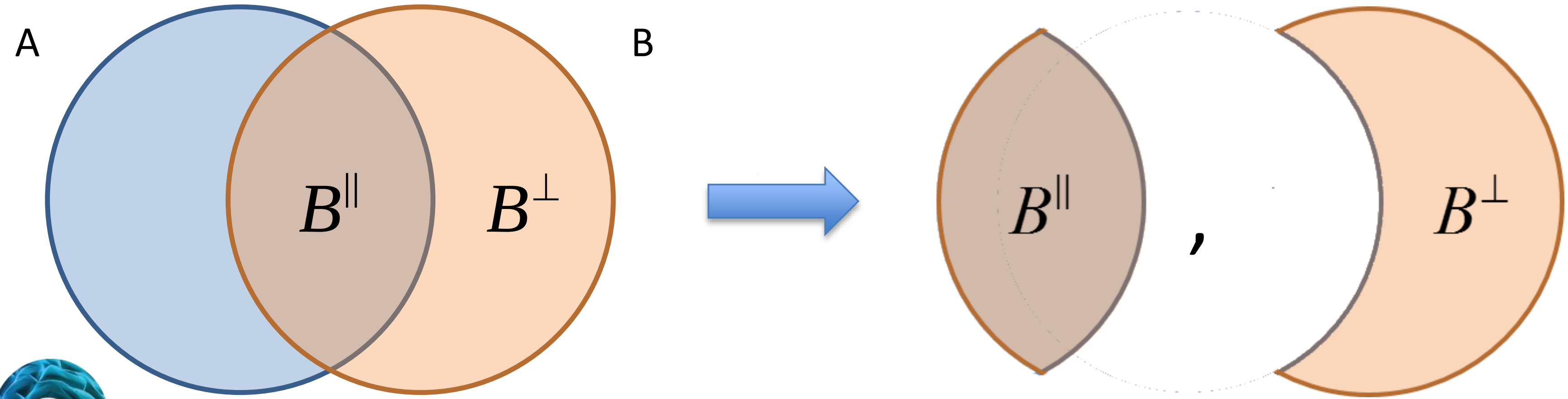
Orthogonal decomposition

$$D : B \mapsto B^{\perp}, B^{\parallel} \text{ such that}$$

$$I\left(B^{\perp}, B^{\parallel} : B\right) = H(B),$$

$$I\left(B^{\perp} : A\right) = 0,$$

$$I\left(B^{\parallel} : A\right) = I(B : A).$$



Proved consequences

No ‘overcounting’

$$I_{\text{syn}}\left(X \to Y\right) \leq H\left(S\right)$$

No ‘undercounting’

$$\forall X, \exists Y : I_{\text{syn}}\left(X \to Y\right) = H\left(S\right)$$

$$I\left(Y : S_3\right) \text{ not counted } \Rightarrow \text{ already counted}$$

Further proved consequences

$$I_{\text{syn}}\left(X \to Y\right) \geq 0.$$

$$I_{\text{syn}}\left(X \to Y\right) \leq I\left(X : Y\right).$$

$$I_{\text{syn}}\left(\left(X_i\right)_i \to \left(Y_i\right)_j\right) = I_{\text{syn}}\left(\left(X_{i'}\right)_{i'} \to \left(Y_{i'}\right)_{i'}\right).$$

$$I_{\text{syn}}\left(X_1 \to Y\right) = 0.$$

$$I_{\text{syn}}\left(X \to X_1\right) = 0.$$

$$I_{\text{syn}}\left(X \to X\right) = H\left(\sigma^{\perp}_{\left(S_i\right)_i}\left(X\right)\right) = H\left(\sigma\left(X\right)\right) = \max.$$

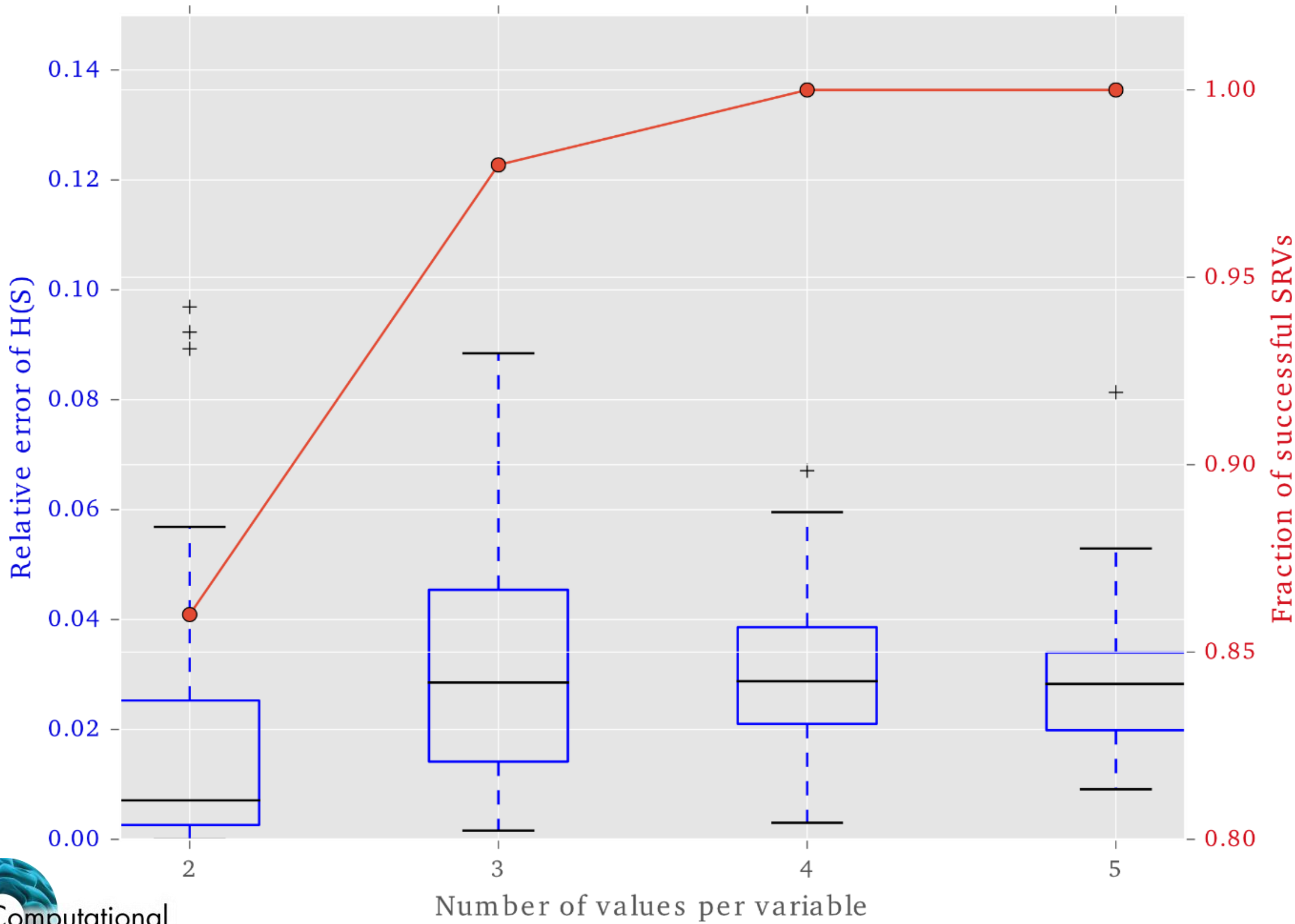
$$I(X : Y) \leq H\left(X_1, \ldots, X_N\right) - \max_i H\left(X_i\right), \text{ where } Y \in \sigma\left(X\right).$$

Python package: `jointpdf`

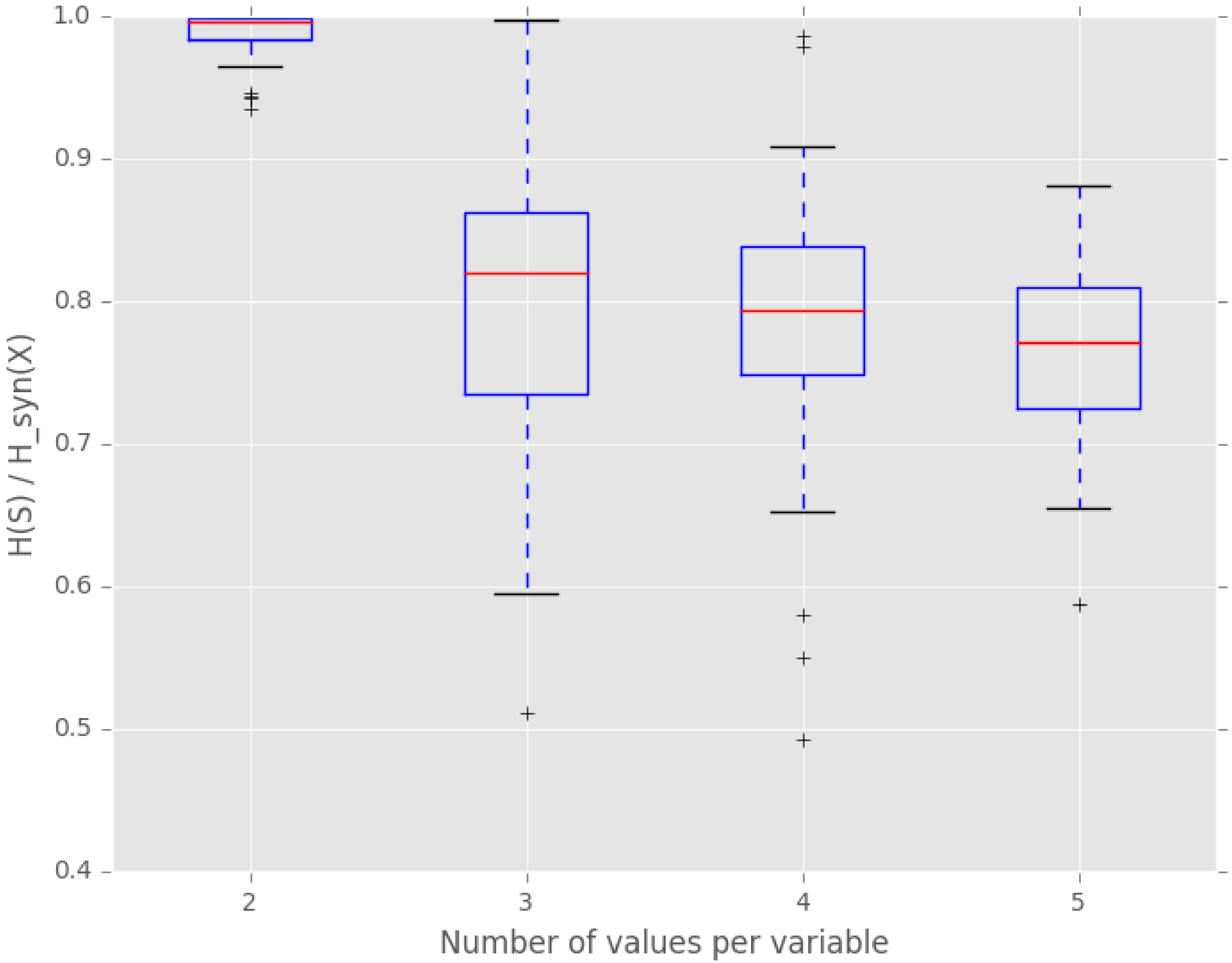
https://bitbucket.org/rquax/jointpdf

```
from jointpdf import JointProbabilityMatrix

# randomly generate a joint probability mass function p(A,B,C)
# of 3 discrete stochastic variables, each having 4 possible values
p_ABC = JointProbabilityMatrix(3,4)

# compute e.g. mutual information between I(A:B)
p_ABC.mutual_information([0], [1])

# compute the information synergy that C contains about A and B (takes a while)
p_ABC.synergistic_information([2], [0,1])
```

Numerical experiments

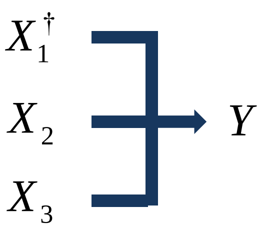


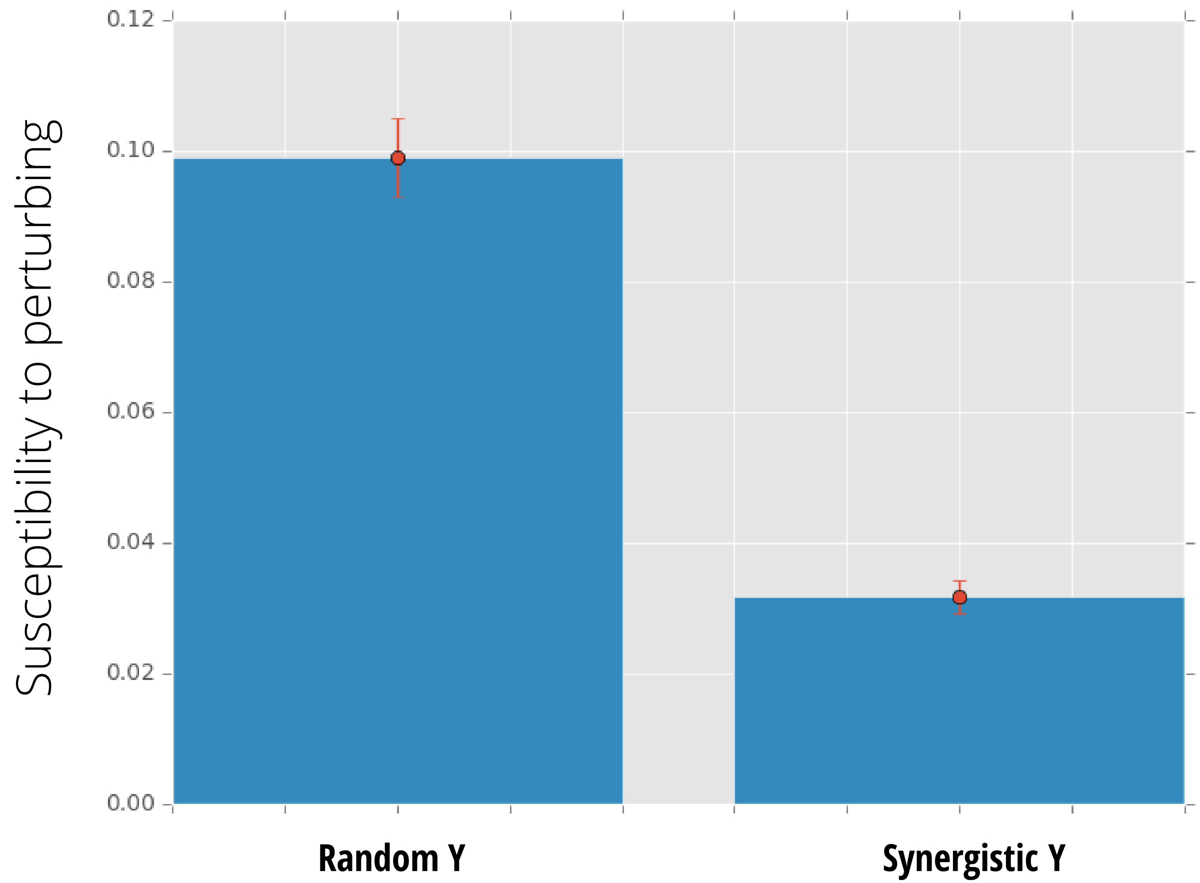
Numerical experiments



Operational meaning of synergy

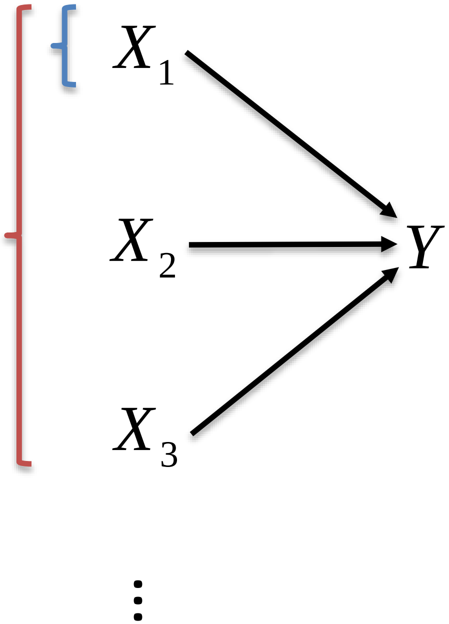
$$X_1 \to X_1^\dagger$$

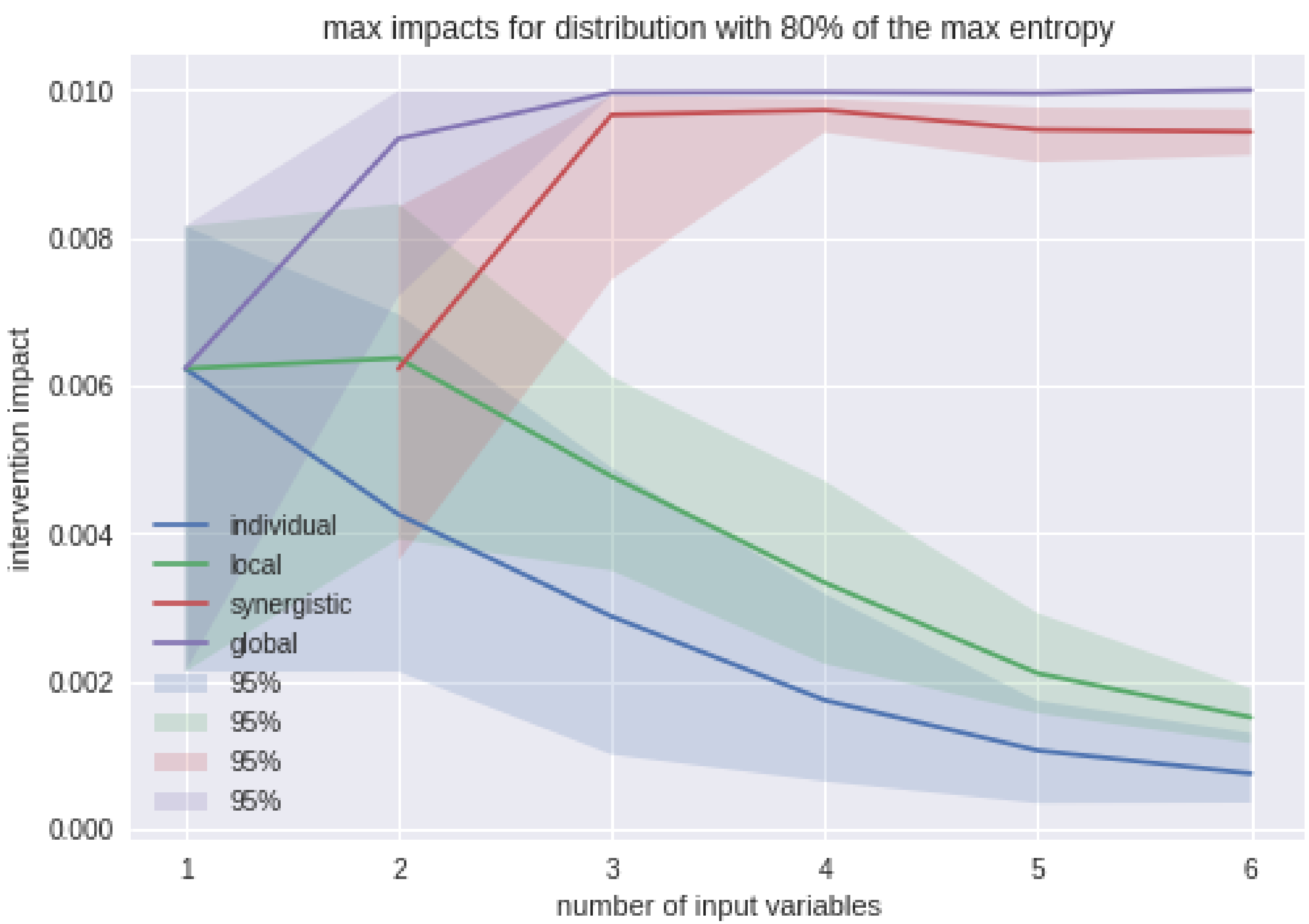




One-to-one causality?


Derk-Jan Riesthuis



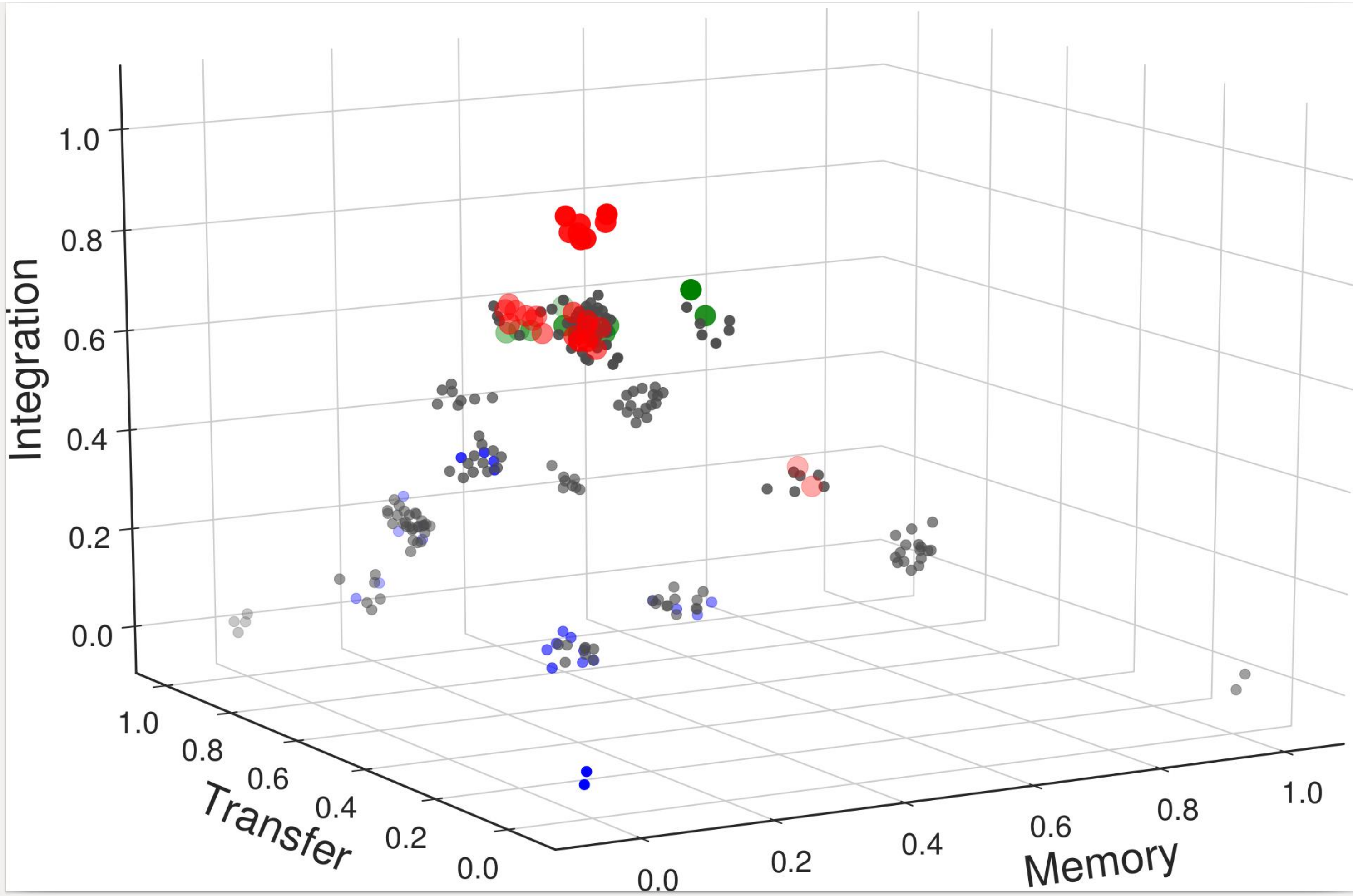


(In preparation.)

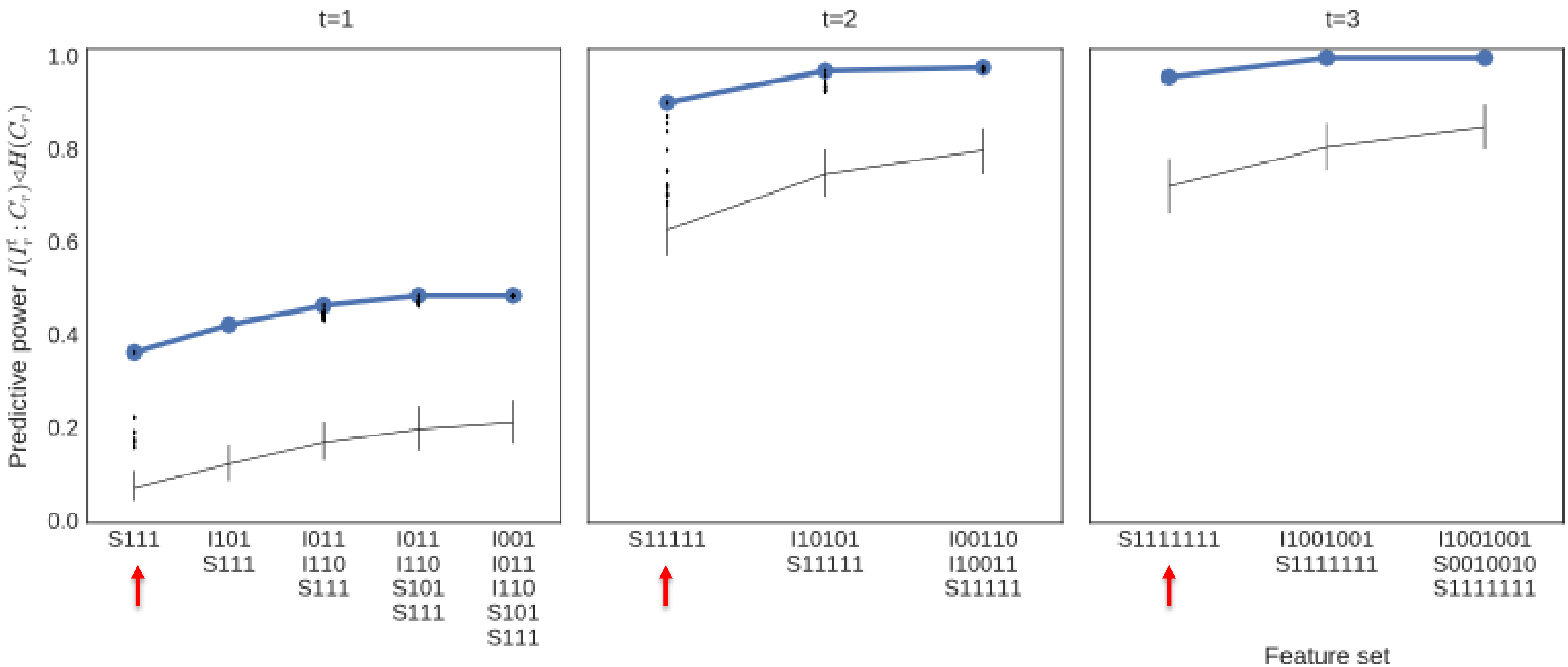
Information → system behavior

= static = cyclic = chaotic = complex

$I_{\text{syn}}\left(\{X_i\}_i \to Y\right)$



Predictability of Wolfram class



The crucial observation is the "S11..11" feature ("synergy"):

- It is always the single most predictive feature
- The predictability of Wolfram class goes very quickly to 1 (=perfect) with time

Conclusion

- We use two simple ingredients:
  - Definition of 'synergistic' variable *S*
  - Correlate *Y* with intermediate *S* instead of directly *X*
- $I(X:Y) \neq I_{\text{indiv}} + I_{\text{syn}}$ (=incompatible with PID)
- Limitation: computationally very expensive

Software: https://bitbucket.org/rquax/jointpdf

Idea for “individual” information?