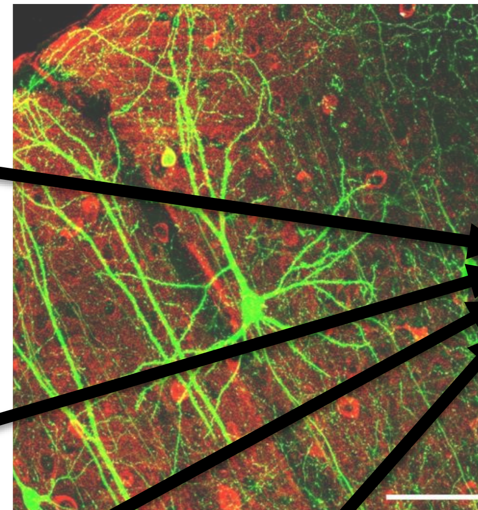
Detecting causal information flows

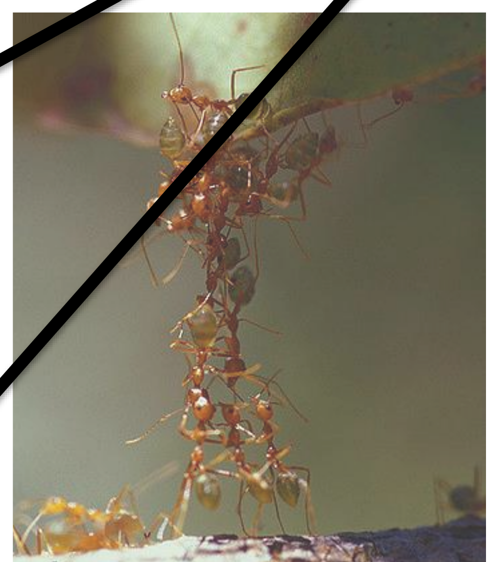
Nudge causality

Rick Quax,
Assistant Professor in Computational Science Lab;
Research Fellow in Institute of Advanced Studies

Research focus: *emergent dynamics*

Complex Systems
- Neuronal network
- Ant colony
- Immune system
- Financial market
- Cell regulation
- Crowd, flock
- Social network
- Ecosystem
- Coral colony
- Bacterial colony

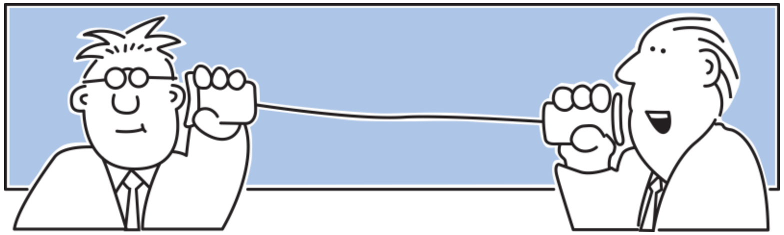


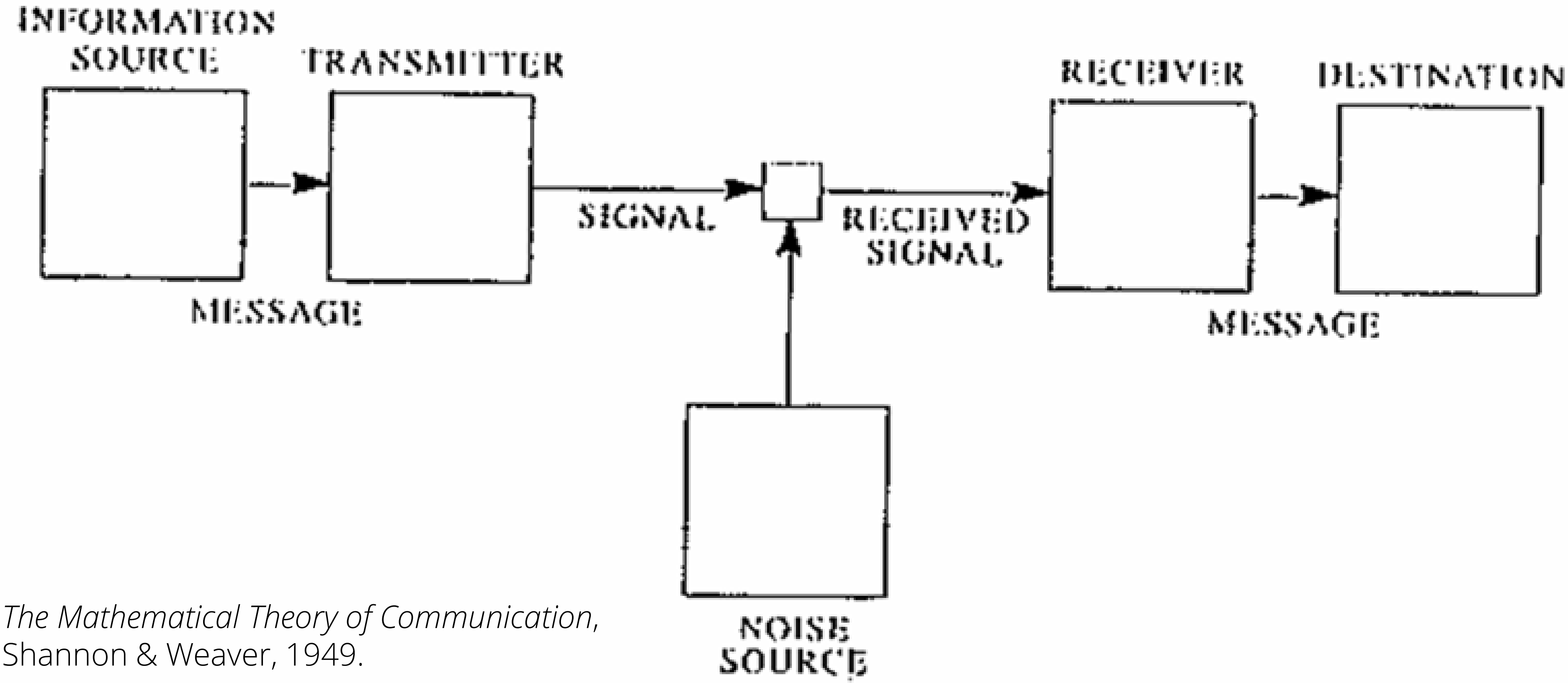


Emergent phenomena
- Phase transitions
- Self-organized criticality
- Tipping points
- Sensitivity/robustness
- Adaptation/learning
- Pattern formation
- Multistability
- Evolution

Computational Science

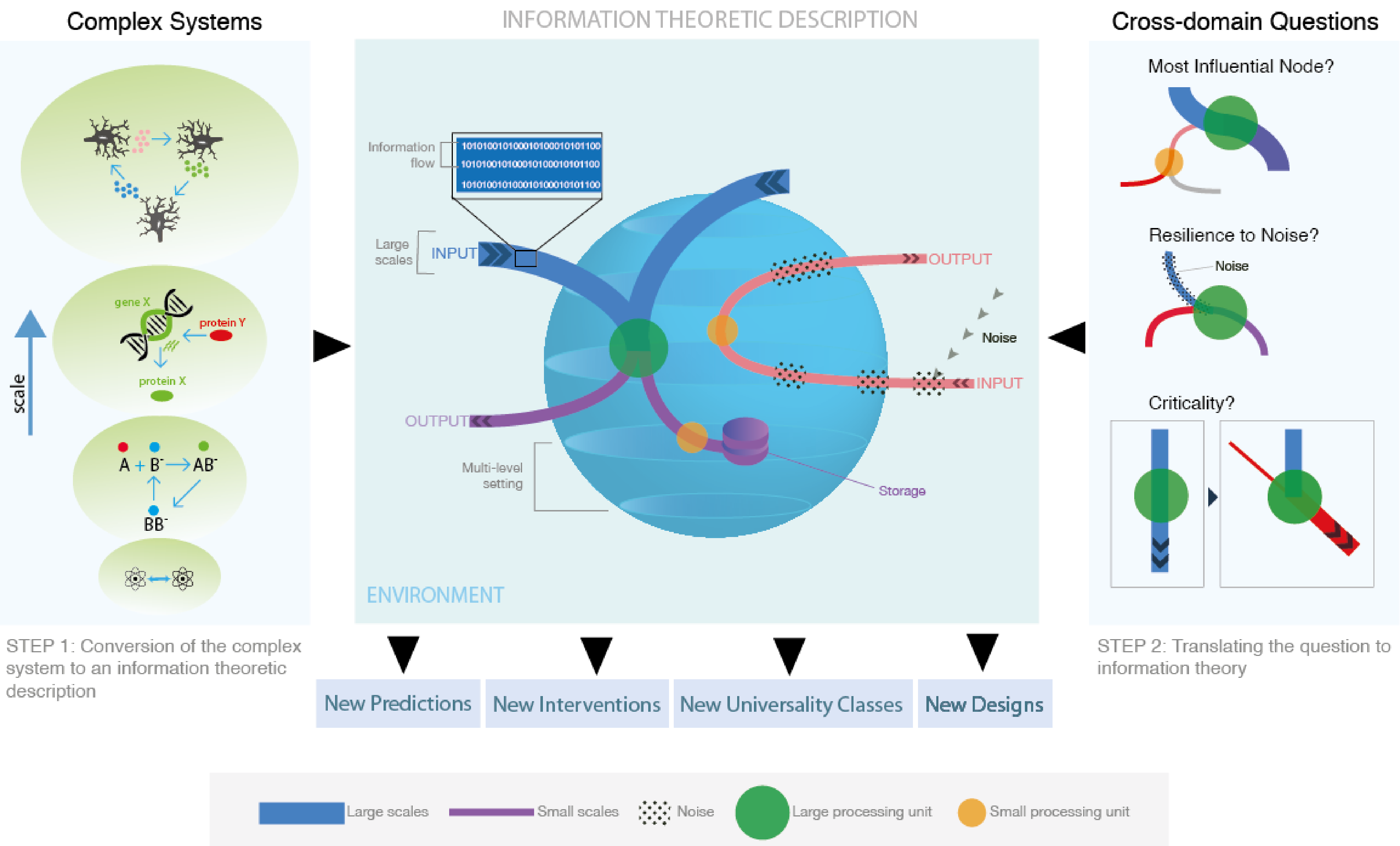
Shannon's communication channel





The Mathematical Theory of Communication,
Shannon & Weaver, 1949.

Everything is information (?)



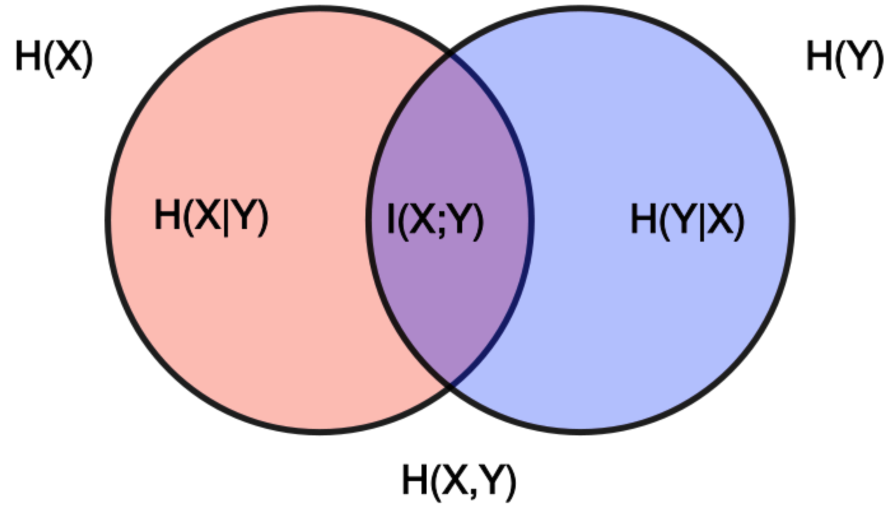
Basics of information theory

$$H(X) = -\sum_{X=x} p(x) \log p(x)$$

"Entropy"

$$I(X:Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)},$$
$$= H(X) - H(X|Y).$$

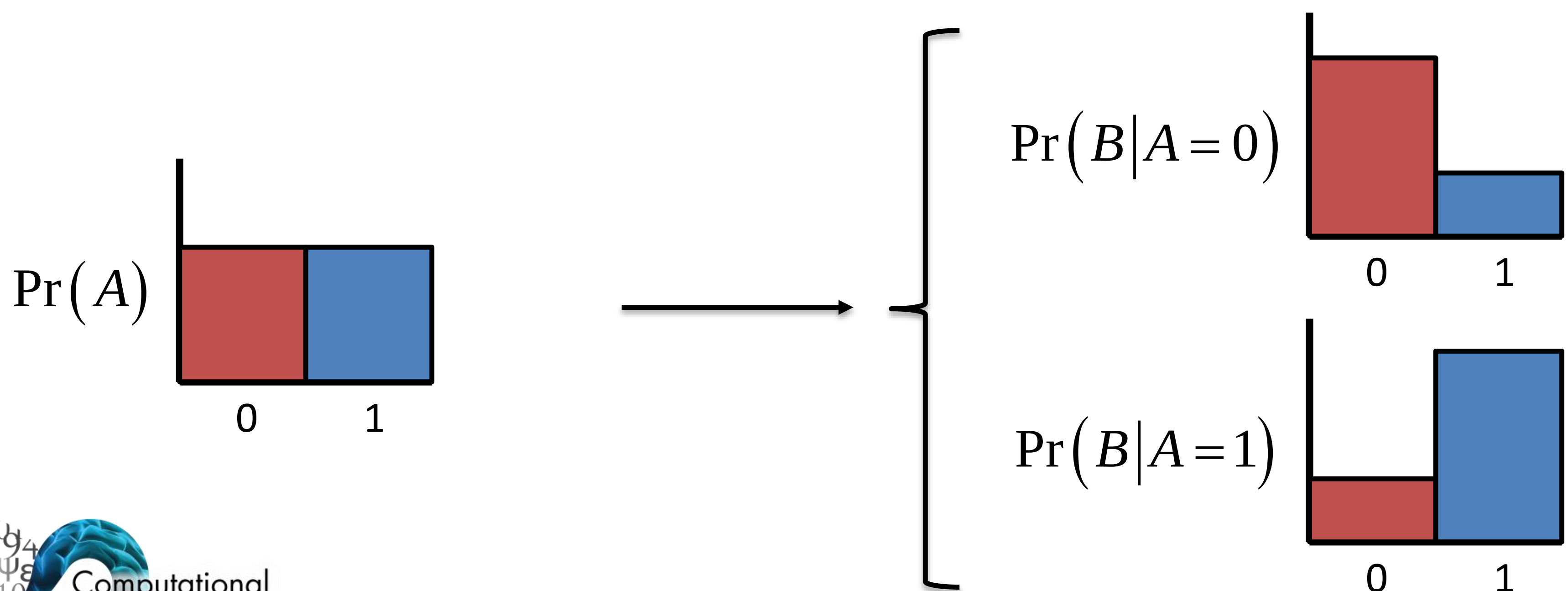
"Mutual information"



From probabilistic causal relationship to information flow

BASIC IDEA

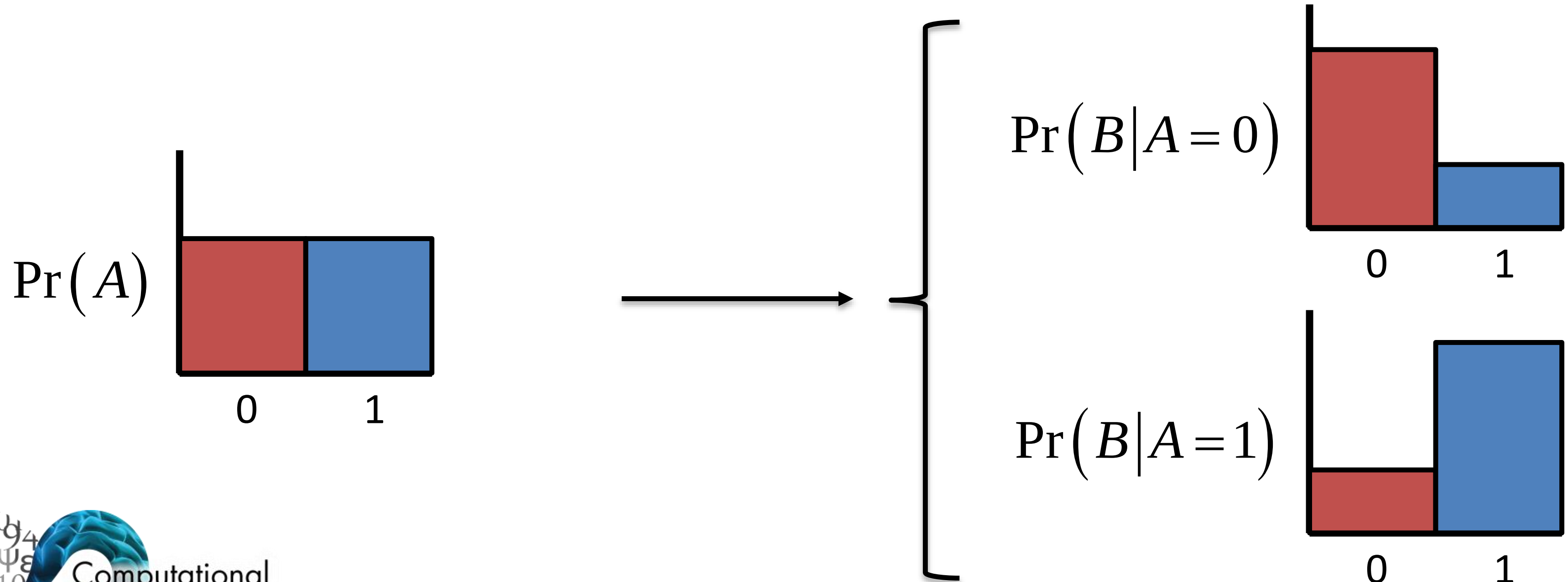
Probabilistic causal relation



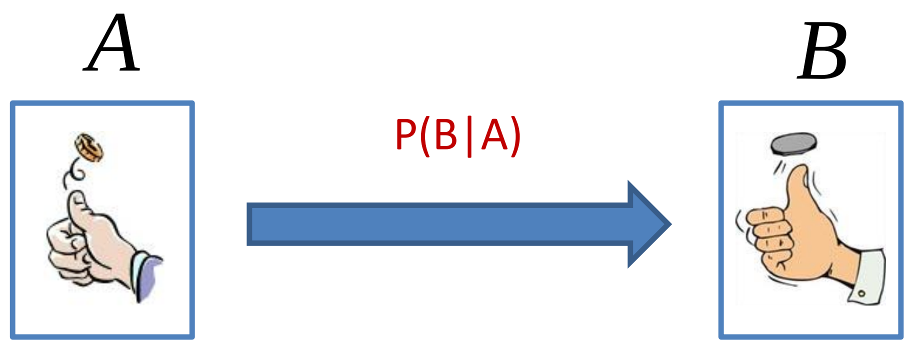
Probabilistic causal relation

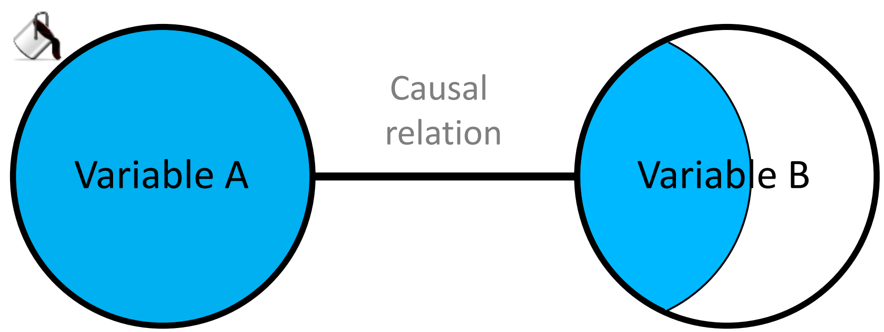
Kullback-Leibler divergence → Mutual information

$$\mathrm{E}_A\left[D_{\mathrm{KL}}\left(\Pr(B \mid A=a) \,\|\, \Pr(B)\right)\right] = I(A:B)$$



Causality → information flow



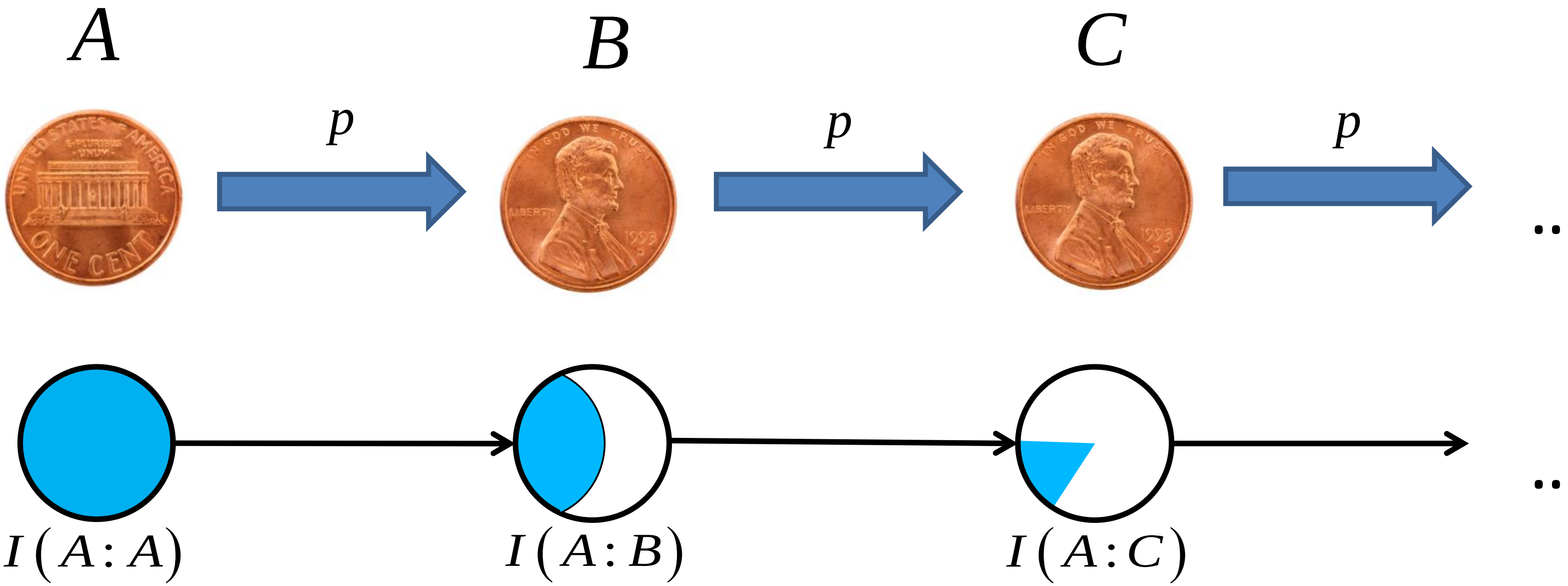


$I(A:A)$ $\qquad$ $I(A:B)$

“Entropy” $\qquad$ “Mutual information”

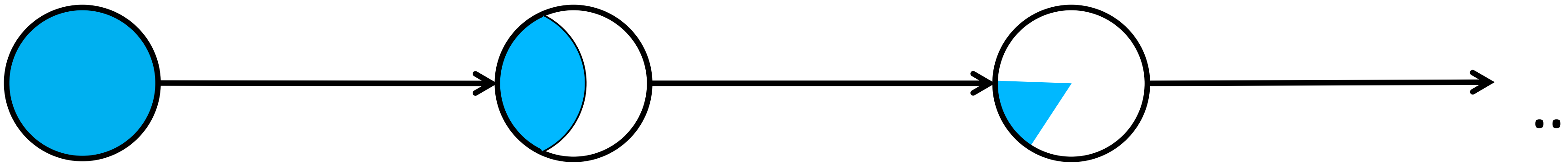
TOWARDS NETWORKS

Chain of interactions



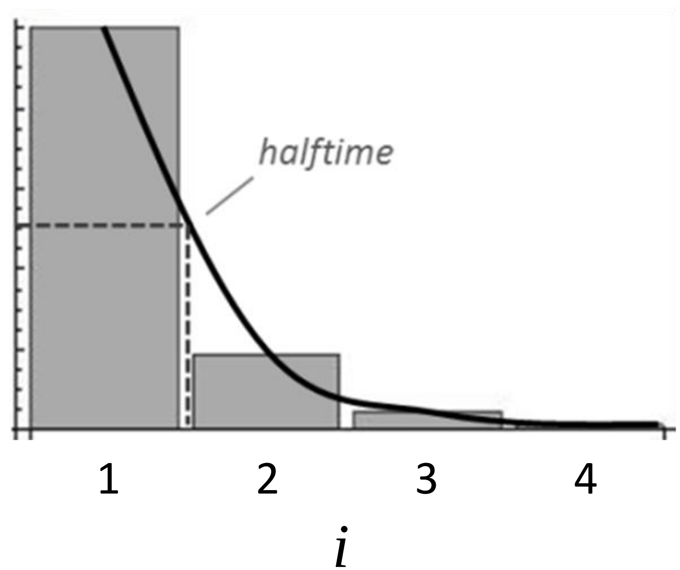
How far can information travel?

Information dissipation length

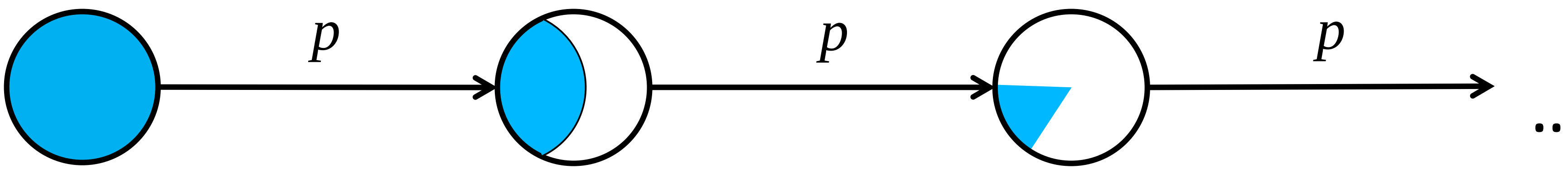


$$I(X_1 : X_1) \longrightarrow I(X_1 : X_2) \longrightarrow I(X_1 : X_3) \longrightarrow \ldots$$

$$I(X_1 : X_i)$$



Information dissipation length



(Apply l'Hôpital's rule twice)

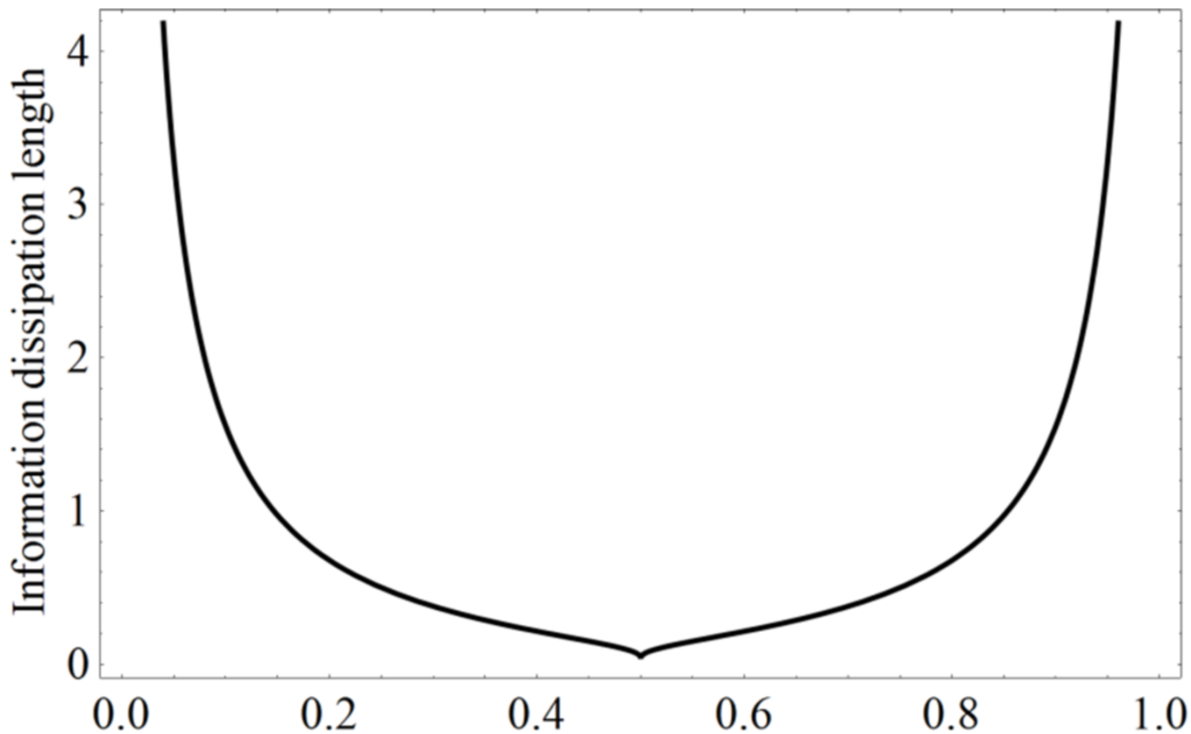
$$f \equiv \lim_{i\to\infty} \frac{I(X_1 : X_{i+1})}{I(X_1 : X_i)} = \ldots = (2p-1)^2$$

$$I(X_1 : X_1) = 1 \longrightarrow I(X_1 : X_2) = f \longrightarrow I(X_1 : X_3) = f^2 \longrightarrow \quad \ldots$$

When is $f^n = \frac{1}{2}$?

Information dissipation length

$$\log_f \frac{1}{2}$$



p

Influence length $\rightarrow$ instability

Interface

The diminishing role of hubs in dynamical processes on complex networks

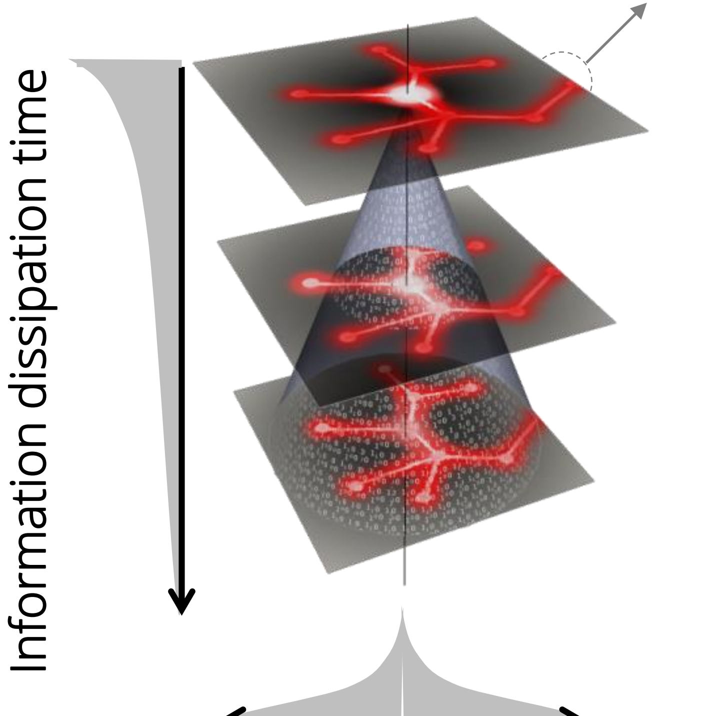
Rick Quax[1†], Andrea Apolloni[2,†] and Peter M. A. Sloot[1,3,4]

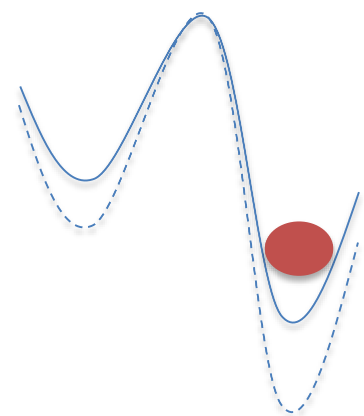
Towards *networks*

- Locally tree-like (i.e., no short loops)
- Any degree distribution

Generalized energy function

$$p(s_i^{t+1} = x \mid s_j^t, \ldots) \propto \exp \sum_j -E(x, s_j^t)$$

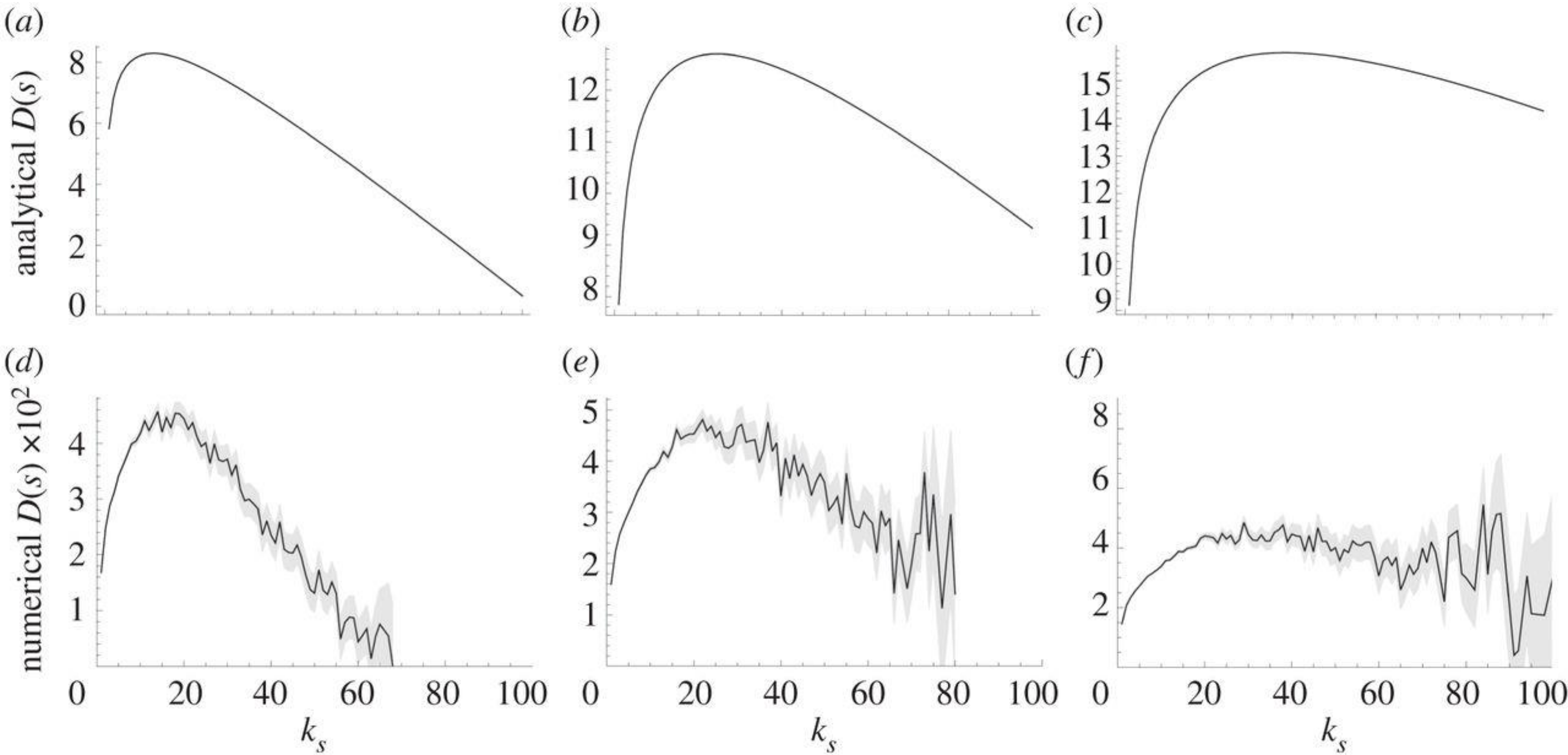


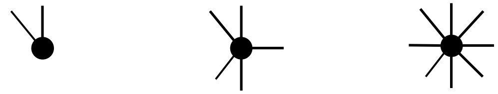


halftime of decay:

$$\Delta_i : I\left(x_i^t : X^{t+\Delta_i}\right) = \frac{1}{2} H\left(x_i^t\right)$$

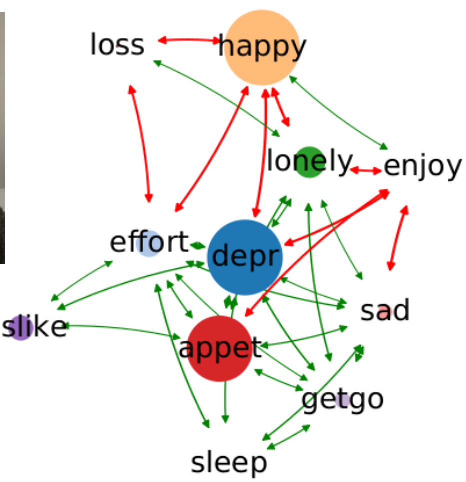
Not the *influentials* but the *man in the street* drives behavior



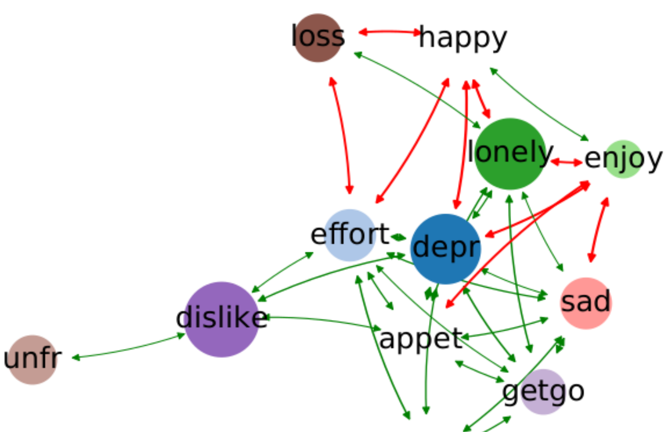


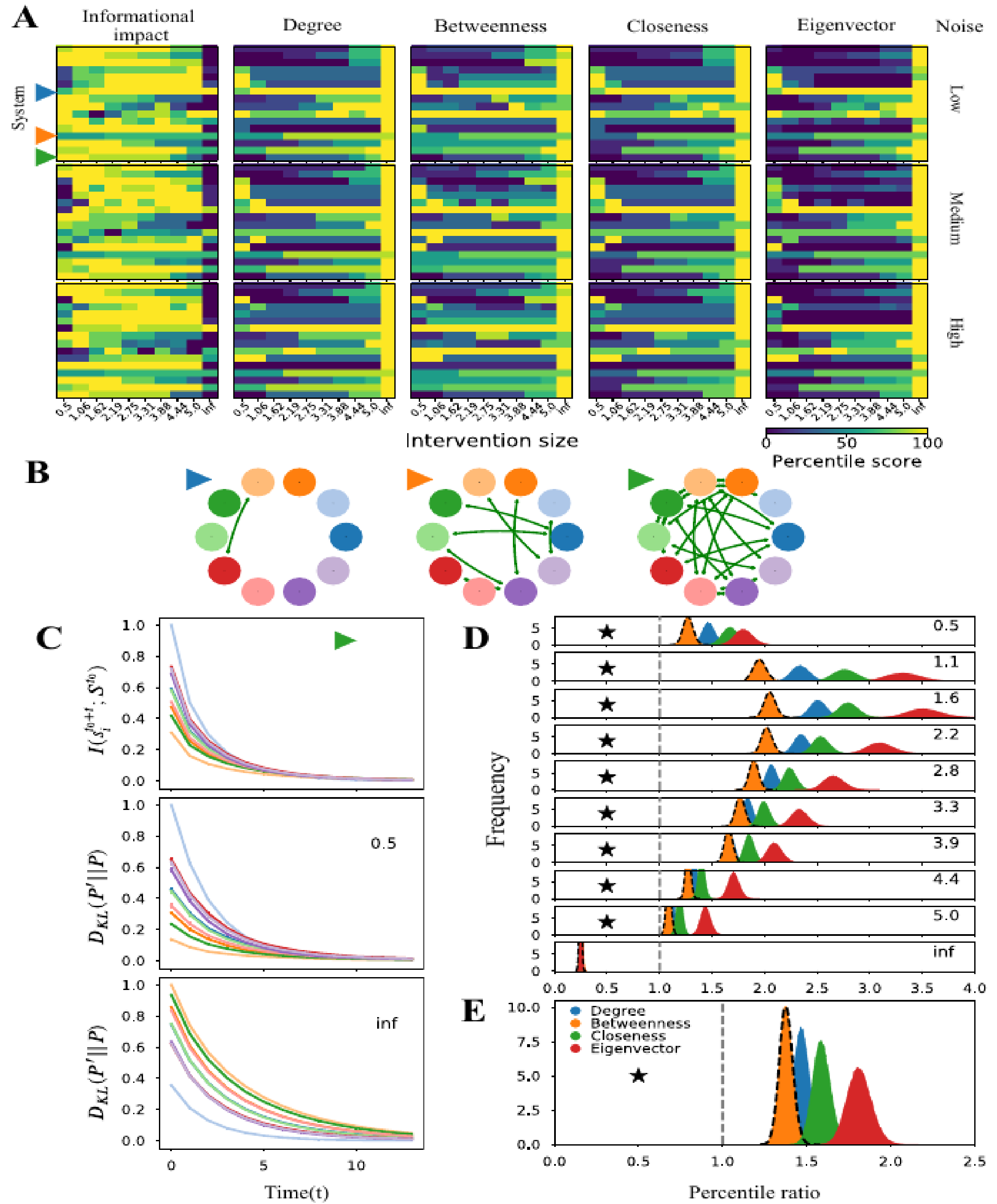
Driver nodes?

Betweenness



Eigenvector





To be published

, Uni

From shared information to causal information flow?

TOWARDS CAUSAL DISCOVERY

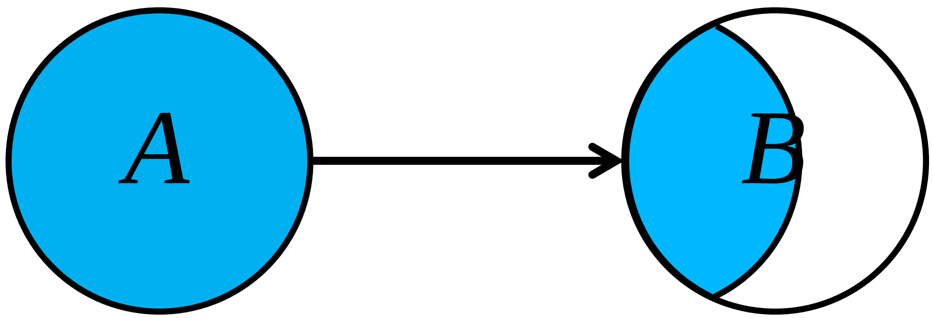
work in progress

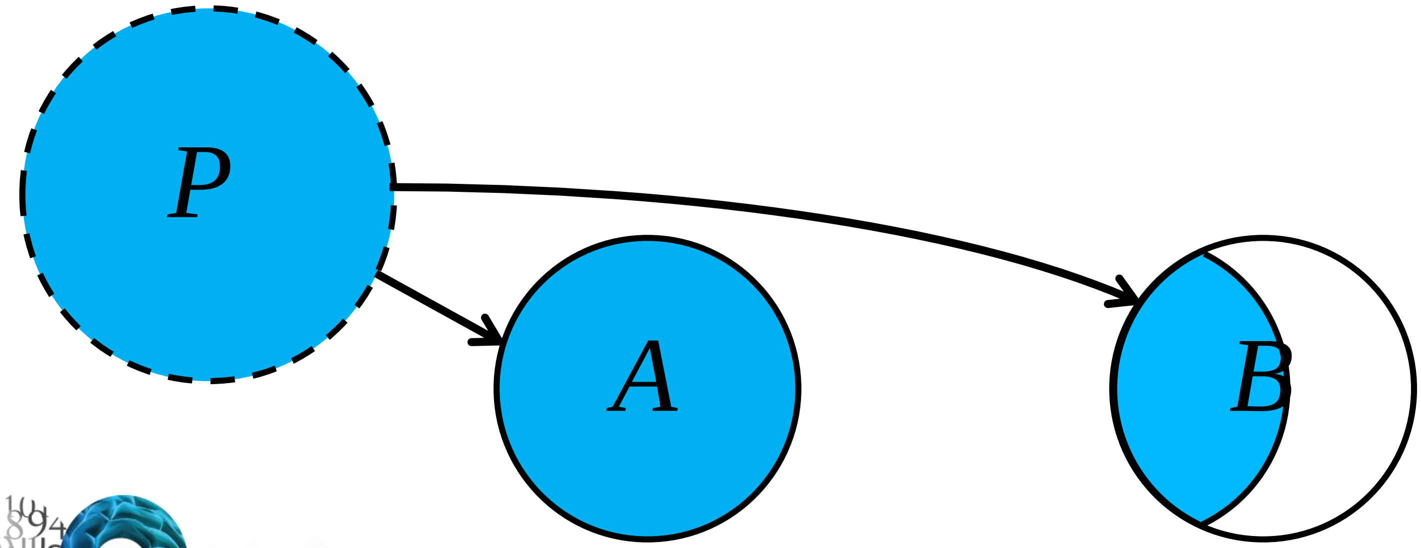
Context

- Type of interactions:
- Work presented:
- Time-series:
- Depends on a 'do':
- Type of values

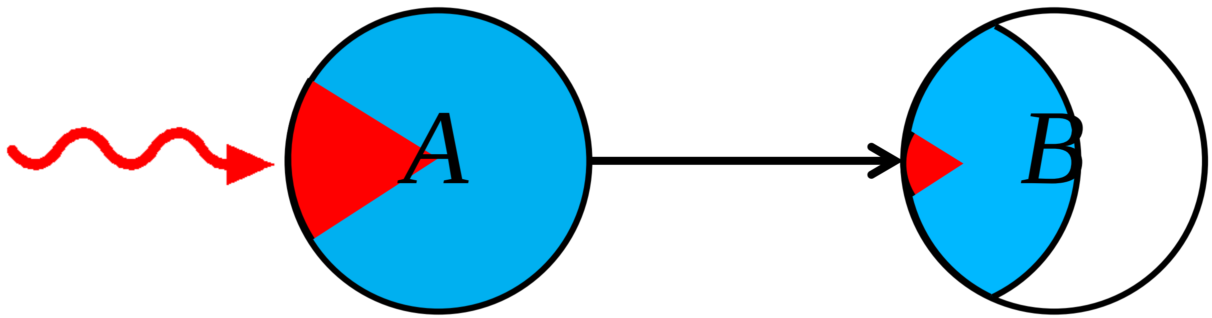
- Any (non-linear)
- New and unknown (!)
- Local *and* global stationarity ('long') / cross-section ('many') + nudge variable
- Yes
- Discrete

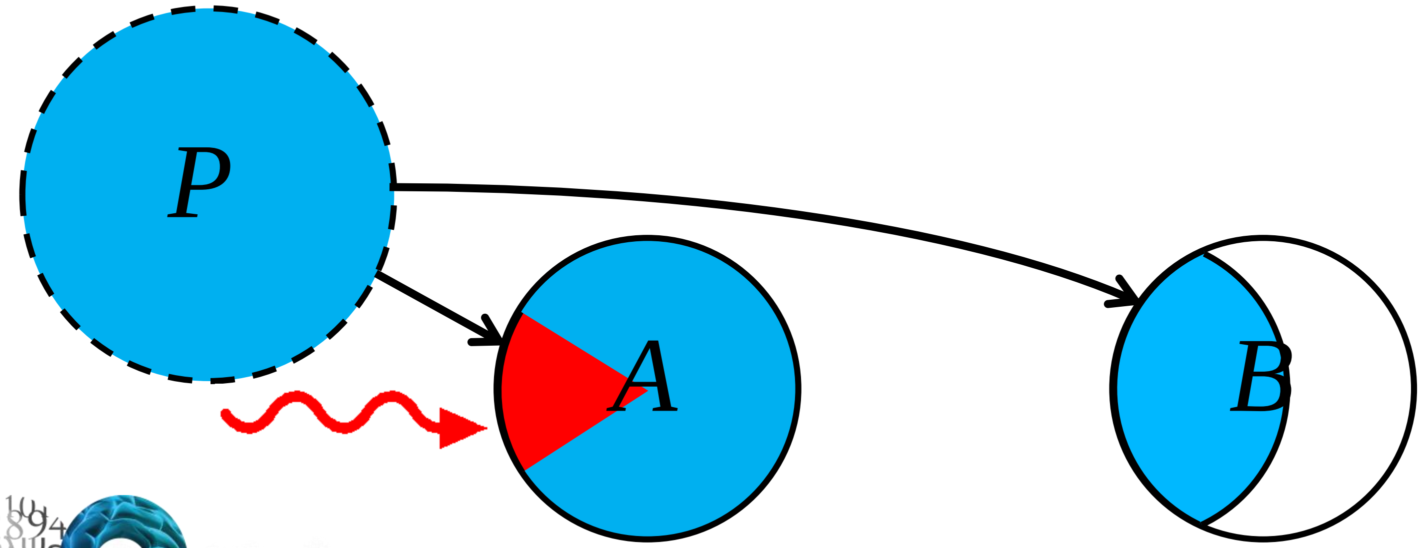
Information flow: causal?



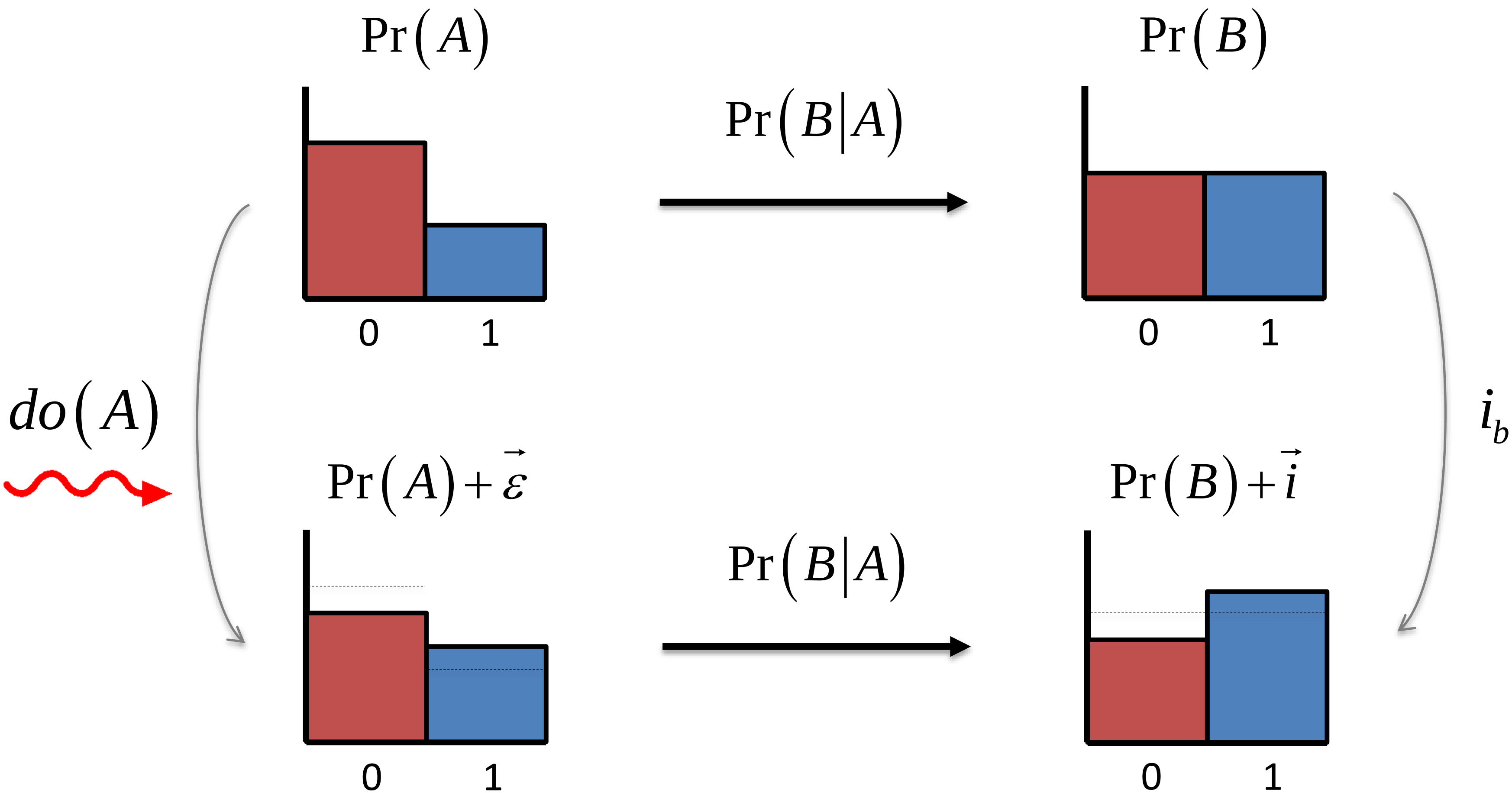


'Do' something!





‘Do’: *Nudging*



Intuition: stay as close as possible to original system dynamics

Nudging a *causal* relation

start: $$I(A:B)=\sum_a \Pr(a)\sum_b \Pr(b|a)\log\frac{\Pr(b|a)}{\Pr(b)}.$$

$do(A)$: $$\Pr'(a)=\Pr(a)+\varepsilon_a$$

$$\Pr'(b)=\sum_a \Pr(b|a)\Pr'(a)$$ (if fully causal)

$$=\Pr(b)+\sum_a \varepsilon_a \Pr(b|a)$$ (if fully causal)

$$\equiv \Pr(b)+i_b$$

Nudging a *causal* relation

$$I(A':B') = \sum_a \Pr'(a) \sum_b \Pr(b|a) \log \frac{\Pr(b|a)}{\Pr'(b)}.$$

$$I(A':B') = \sum_a \Pr'(a) \sum_b \Pr(b|a) \left[\log \Pr(b|a) - \log\left(\Pr(b) + i_b\right) \right].$$

$$I(A':B') = \sum_a \Pr'(a) \sum_b \Pr(b|a) \left[\log \Pr(b|a) - \log \Pr(b) - \log\left(1 + \frac{i_b}{\Pr(b)}\right) \right].$$

$$I(A':B') = \sum_a \Pr'(a) \sum_b \Pr(b|a) \left[\log \Pr(b|a) - \log \Pr(b) - \frac{i_b}{\Pr(b)} \right].$$

$\vdots$

$$\left| \frac{i_b}{\Pr(b)} \right| \ll 1$$

‘Causal information’?

$$do(A)$$

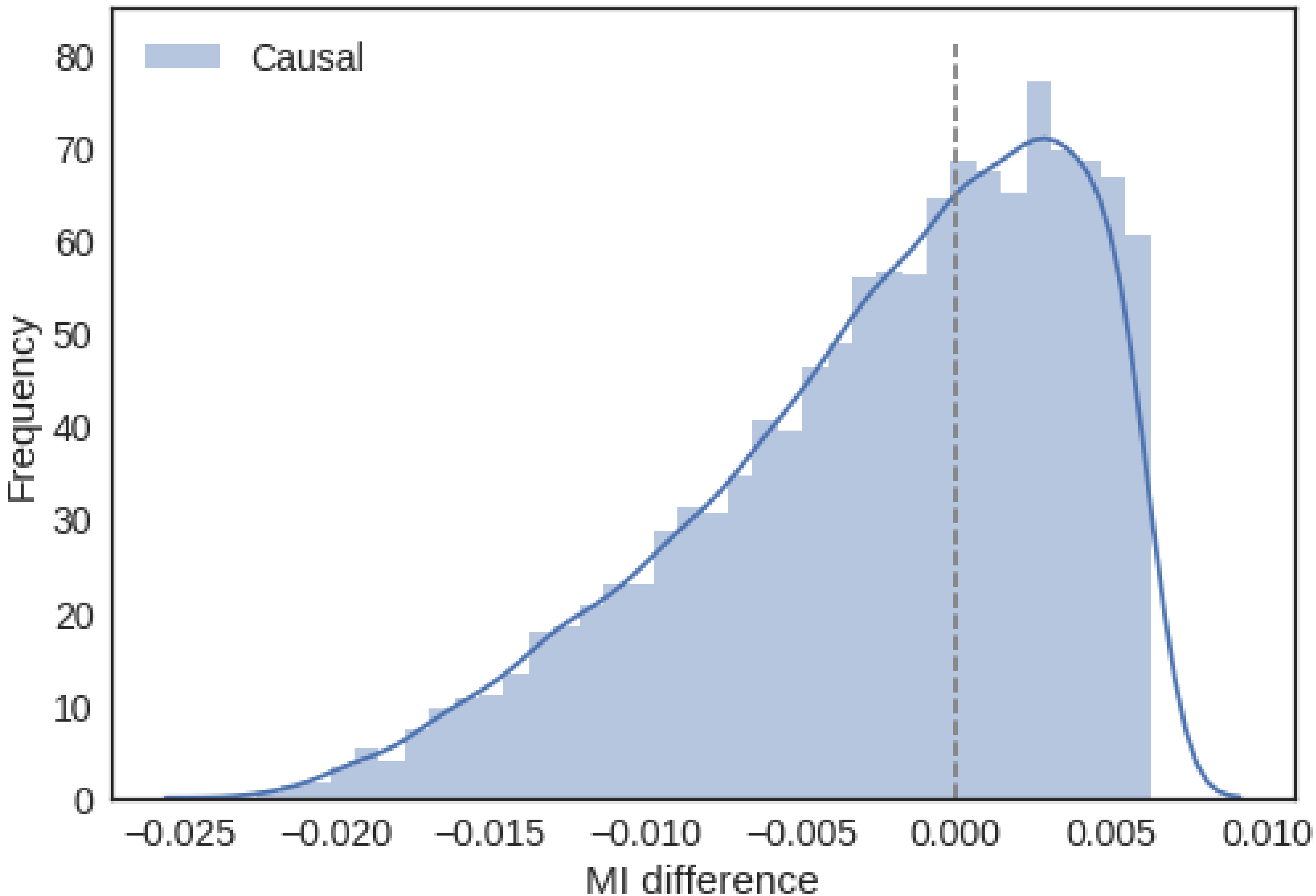
$$\downarrow$$

$$I(A':B') - I(A:B) = \underbrace{\sum_a \varepsilon_a \cdot \sigma(a)}_{\text{Correlation}} - \overbrace{\sum_b \frac{\dot{i}_b^{\,2}}{\Pr(b)}}^{\text{Causation}}.$$

where $\sigma(a) \equiv \sum_b \Pr(b|a) \log \frac{\Pr(b|a)}{\Pr(b)}$

(“specific surprise”)

‘Causal information’?



Nudging a *non-causal* relation

start: $$I(A:B) = \sum_a \Pr(a) \sum_b \Pr(b|a) \log \frac{\Pr(b|a)}{\Pr(b)}.$$

$do(A)$:

$$\Pr'(a) = \Pr(a) + \varepsilon_a,$$

$$\Pr'(b) = \Pr'(b),$$ (if fully non-causal)

$$\Pr'(b|a) = \Pr(b|a) + \Delta_{b|a}.$$ (if fully non-causal)

Definition of non-causal info.

$$do\left(A\right)$$

$$\downarrow$$

$$I\left(A':B'\right)-I\left(A:B\right)=$$

$$\sum_a p_a \sum_b \Delta_{b|a}\cdot\sigma_{b|a}+\sum_a p_a \sum_b \frac{\Delta_{b|a}^2}{p_{b|a}}+\sum_a \varepsilon_a \sum_b \Delta_{b|a}\cdot\sigma_{b|a}+\sum_a \varepsilon_a\cdot\sigma_a+\sum_a \varepsilon_a \sum_b \frac{\Delta_{b|a}^2}{p_{b|a}}.$$

where $\sigma\left(a\right)\equiv\sum_b \Pr\left(b|a\right)\log\frac{\Pr\left(b|a\right)}{\Pr\left(b\right)}$

Definition of non-causal info.

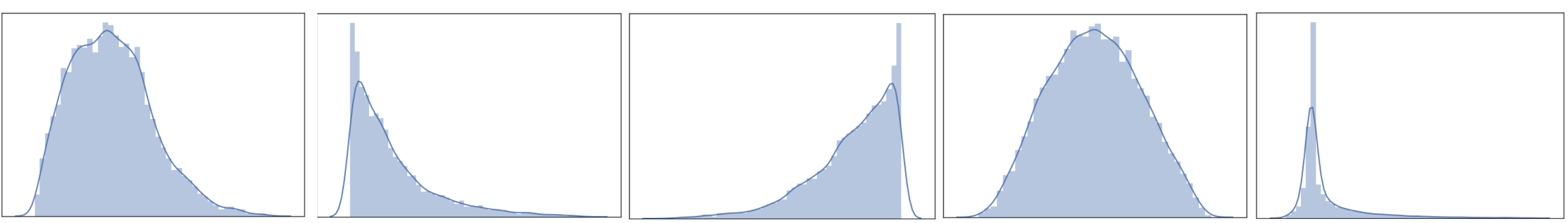
$$do(A)$$

$$\Big\downarrow$$

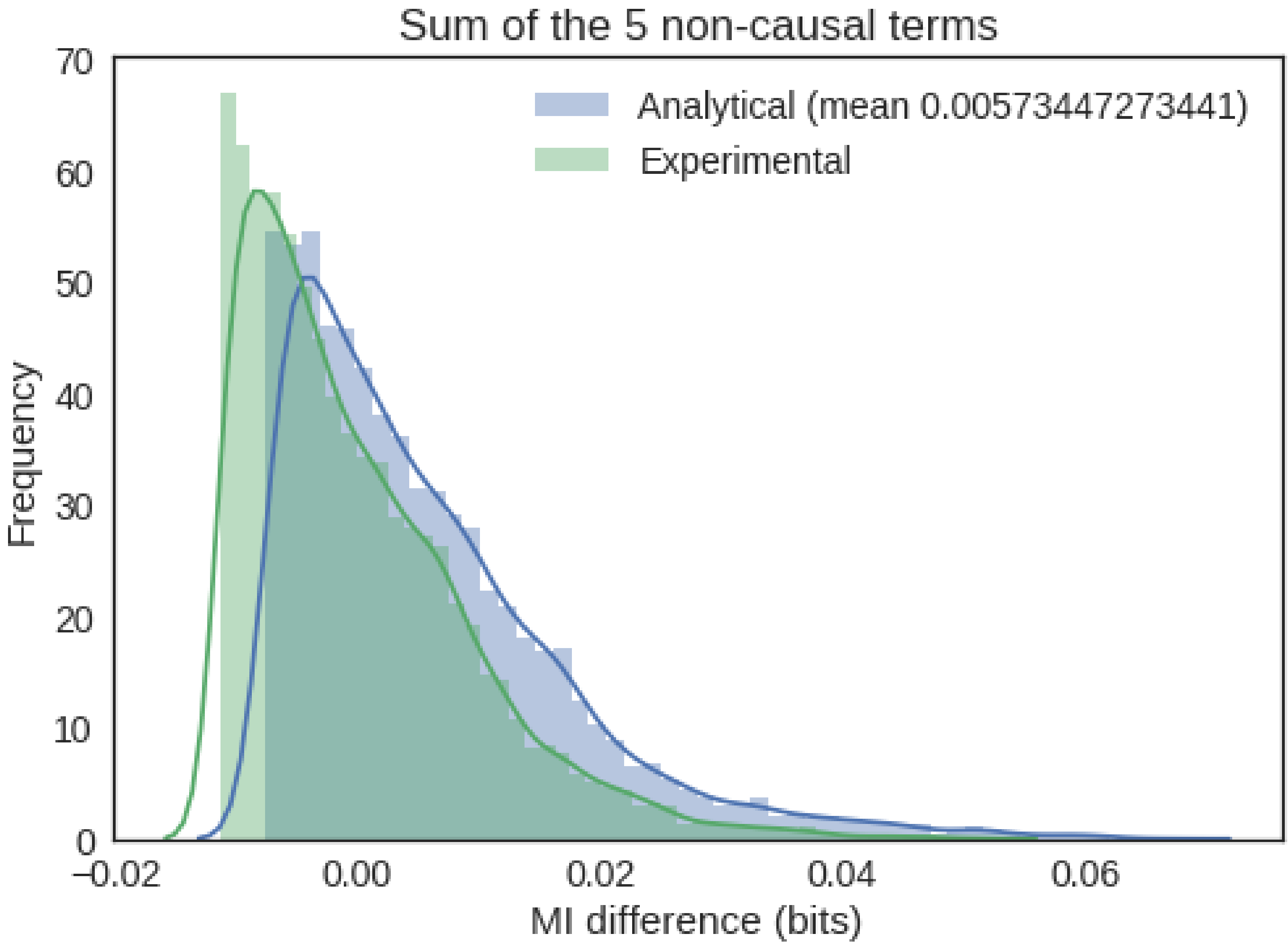
$$I(A':B') - I(A:B) =$$

$$\sum_a p_a \sum_b \Delta_{b|a} \cdot \sigma_{b|a} + \sum_a p_a \sum_b \frac{\Delta_{b|a}^2}{p_{b|a}} + \sum_a \varepsilon_a \sum_b \Delta_{b|a} \cdot \sigma_{b|a} + \sum_a \varepsilon_a \cdot \sigma_a + \sum_a \varepsilon_a \sum_b \frac{\Delta_{b|a}^2}{p_{b|a}}.$$

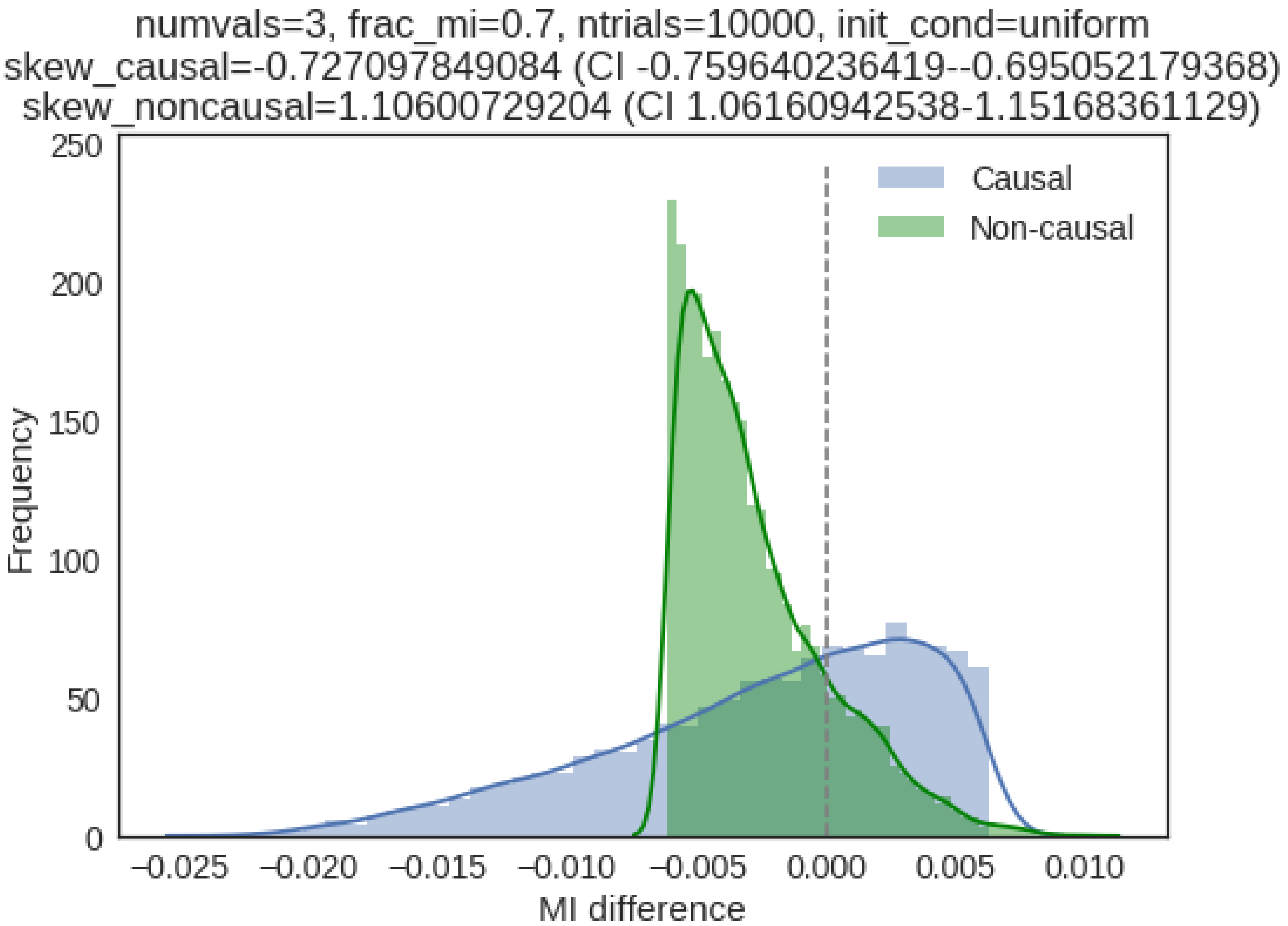
1.0 1.0 0.3 0.3 0.1



Definition of non-causal info.

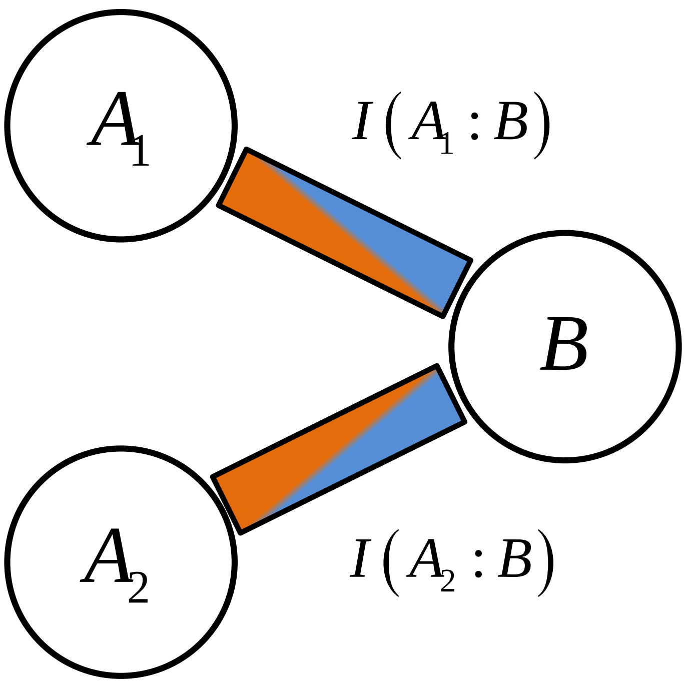


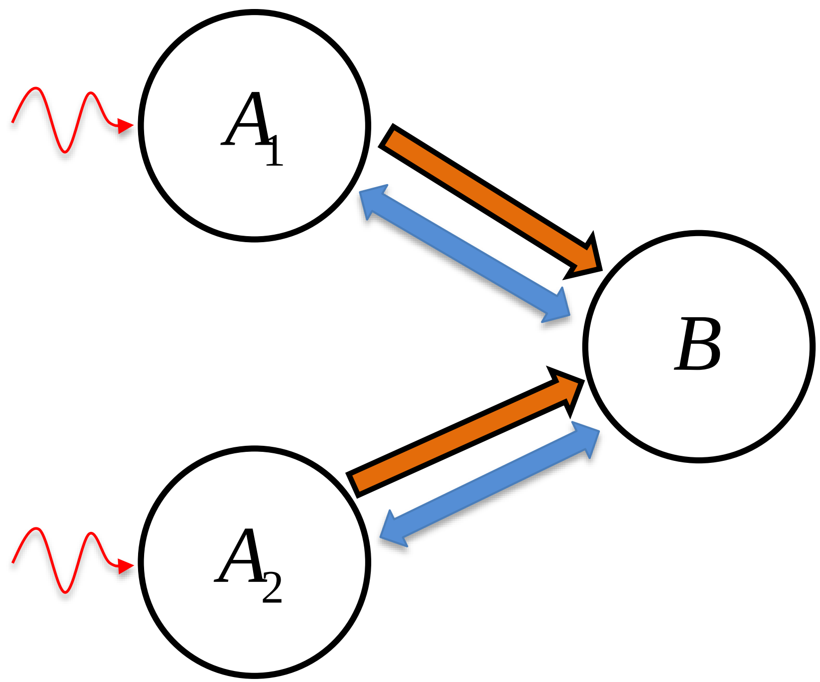
Causal versus non-causal MI



Information flows

- **Before**: Mixed MI
- do(A_i): Flow + corr.



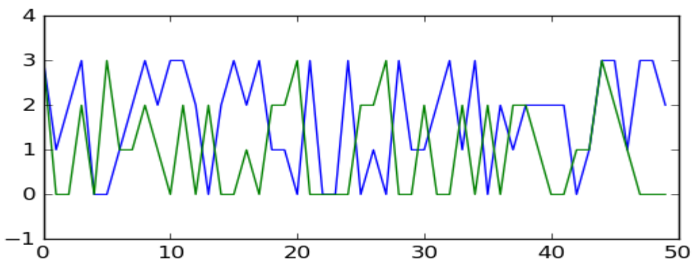


idea: $I_{\mathrm{corr}}(A_1:B)+I_{\mathrm{causal}}(A_1:B)\equiv I(A_1:B)$

Simulated ime-series data

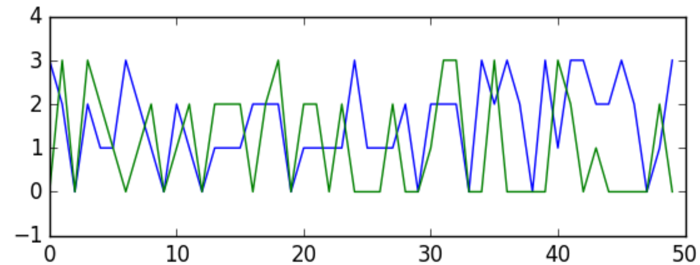
$$\Pr(A)\Pr(B)$$

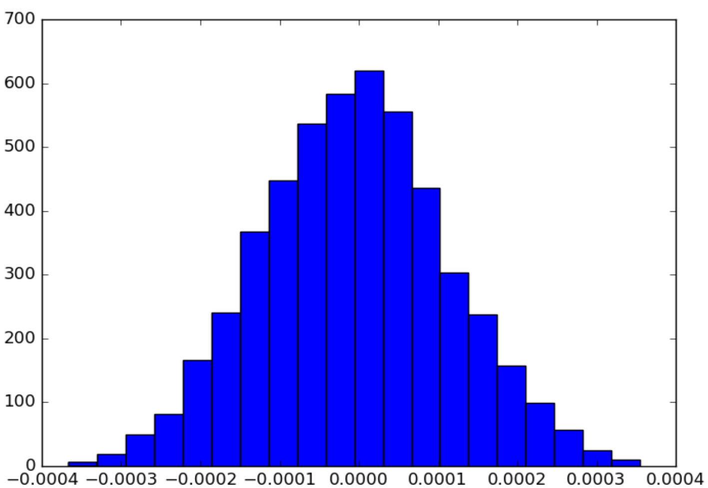
$$\Pr(A)\Pr(B|A)$$



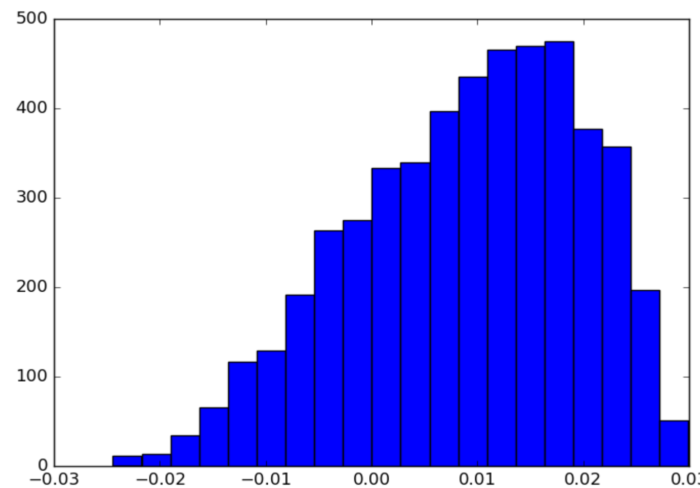
1k

10k points



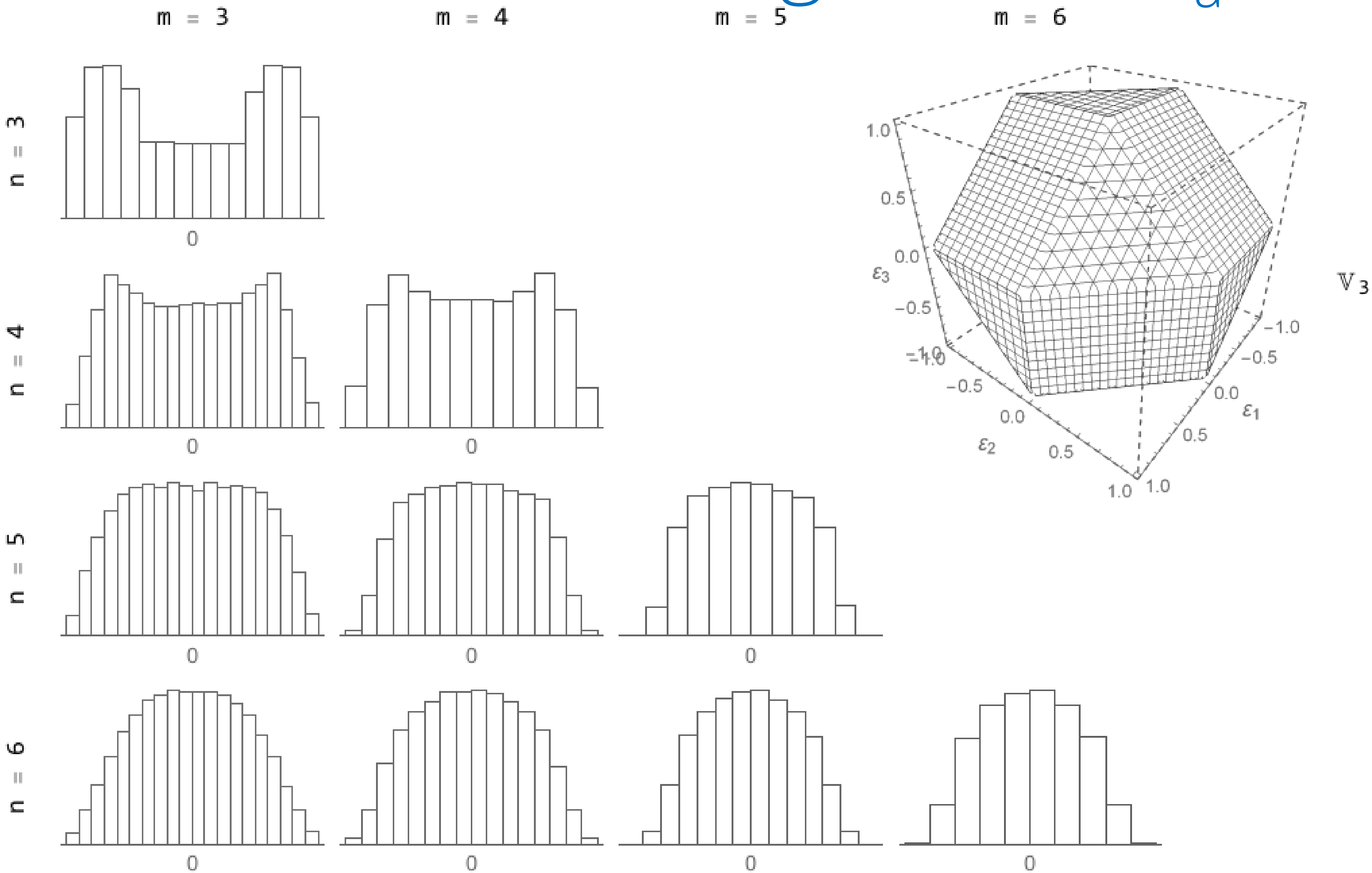


$$\Delta I$$



$$\Delta I$$

Elements of nudge vector ε_a



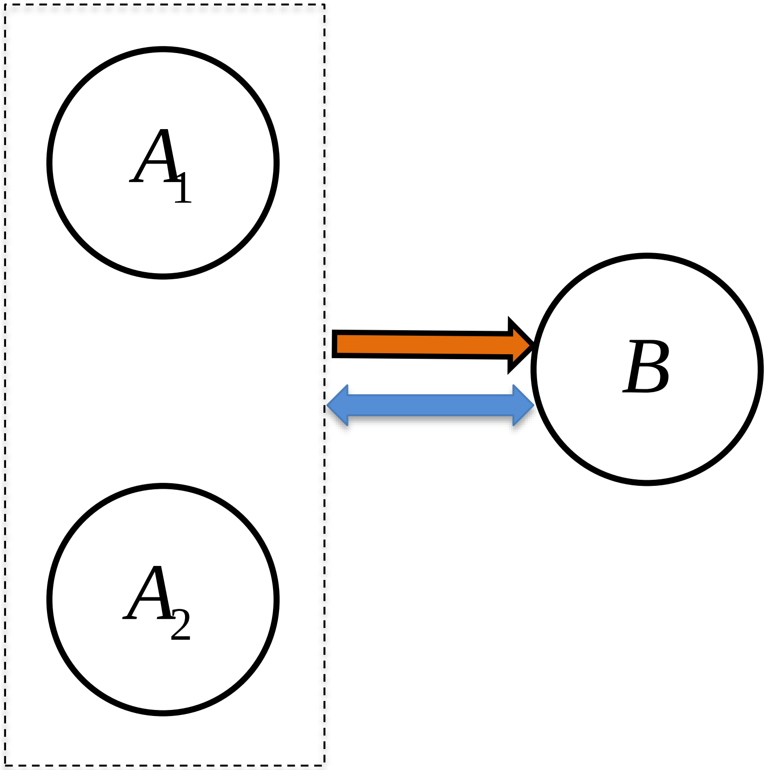
Is it enough to control a single variable only?

SYNERGY

Synergy

	X1	X2	Y
25%	0	0	0
25%	0	1	1
25%	1	0	1
25%	1	1	0

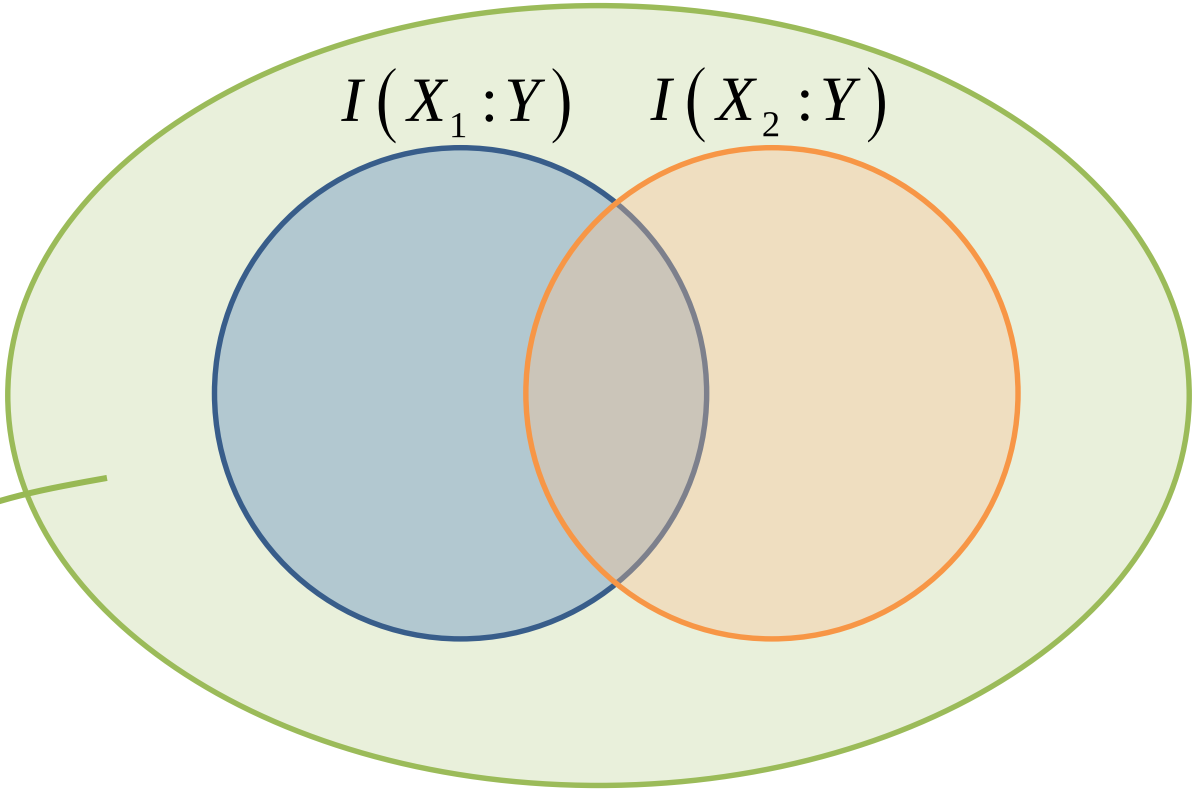
$$I\left(Y:X_1\right)=0,$$
$$I\left(Y:X_2\right)=0,$$
$$I\left(Y:X_1,X_2\right)=1.$$



Rick Quax, Omri Har-Shemesh, Peter M. A. Sloot: Quantifying synergistic information using intermediate stochastic variables. *Entropy*, 2017

Synergistic information

$$I\left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} : Y\right) =$$

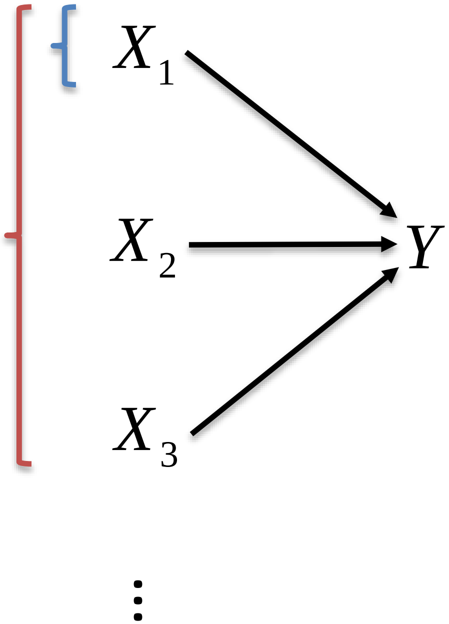


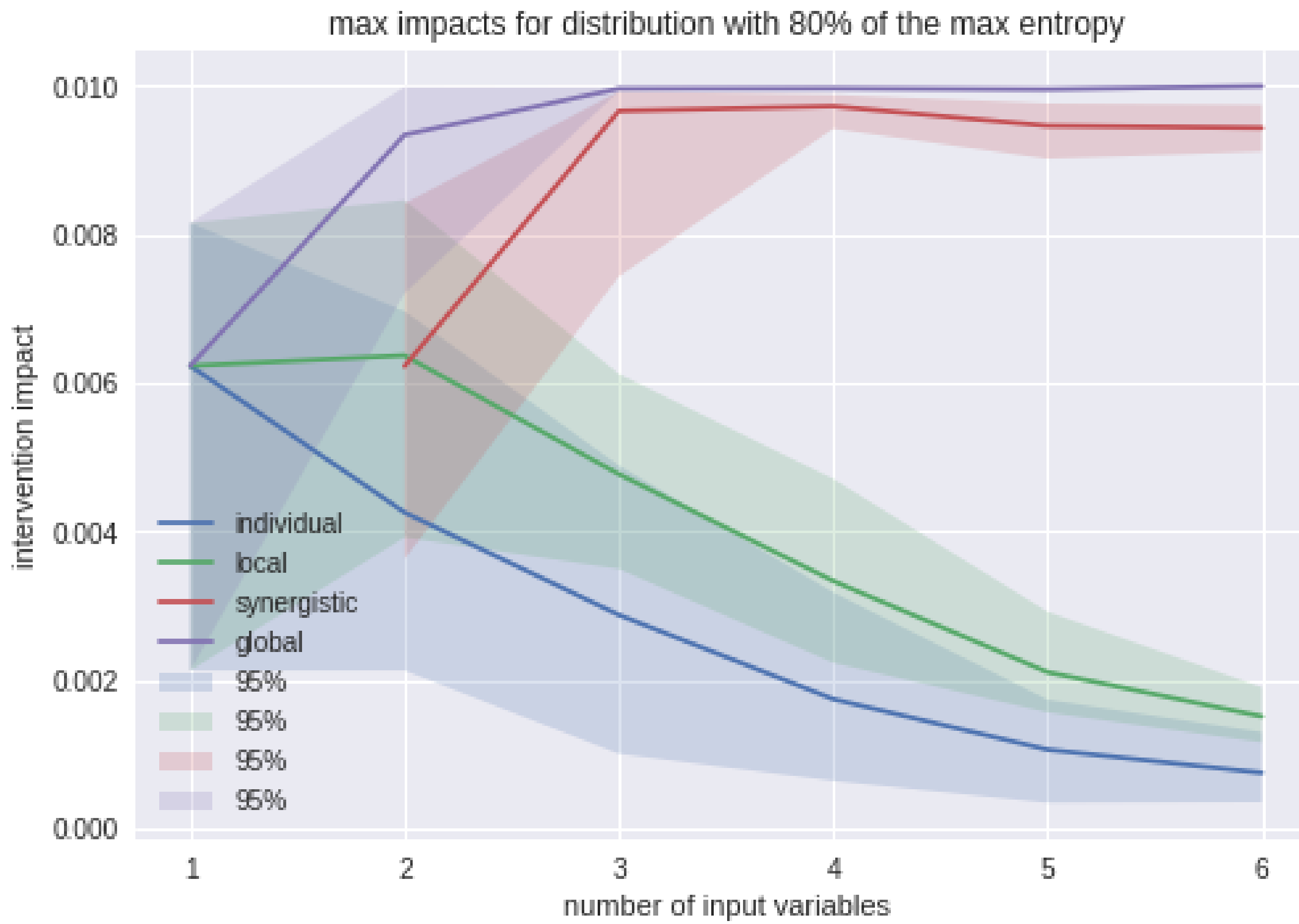
$$I_{\text{syn}}\left(\{X_i\}_i \to Y\right)?$$

Quax, R.; Har-Shemesh, O.; Sloot, P.M.A. **Quantifying Synergistic Information Using Intermediate Stochastic Variables.** Entropy 2017, 19, 85.
http://www.mdpi.com/1099-4300/19/2/85

One-to-one causality?

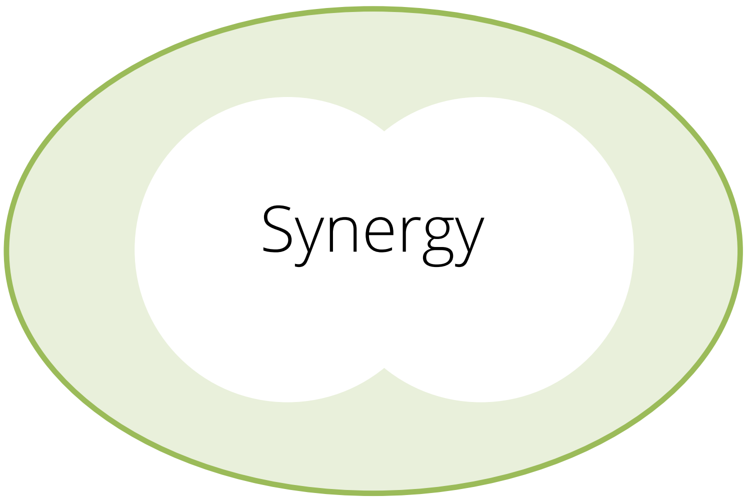

Derk-Jan Riesthuis





(In preparation.)

So in multi-input causation…



= Important

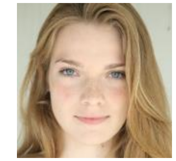
Concluding remarks

- Mutual information measures causal effect of a nudge in 1-to-1 causation (if choosing KL-Div. as 'impact' measure)
- Nudging the system reveals 'causal MI'
- *Local* as opposed to *global* like Transfer Entropy or D-separation-like techniques
- Requirements:
  - slow nudging variable
  - Discrete
  - Lots of data
- Synergy is important but often neglected

Software: https://bitbucket.org/rquax/jointpdf
https://github.com/cvanelteren/information_impact

Current projects

- Diabetes type 2 (ZonMw)


- Criminal networks (RIEC)


- SocialHealth (Radboud)


- Alzheimer as a system (Radboud)

- Immune system failure (ITMO)

- Disease networks, interactome (EU)
- Information processing in neural networks
- **Information theory meets causality**