

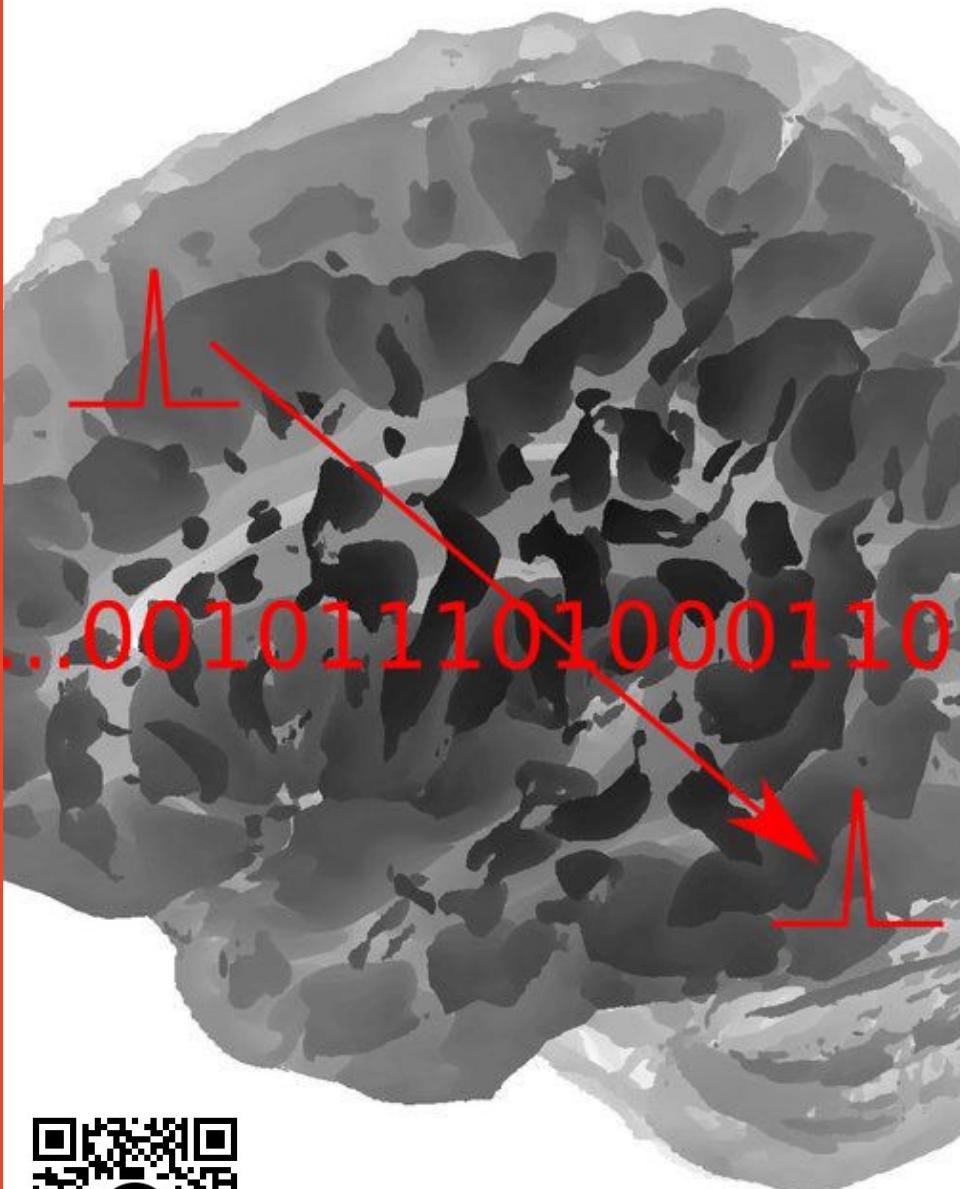
Enabling tools to model information processing in brains

A/Prof. Joseph Lizier

Centre for Complex Systems
School of Computer Science
Faculty of Engineering

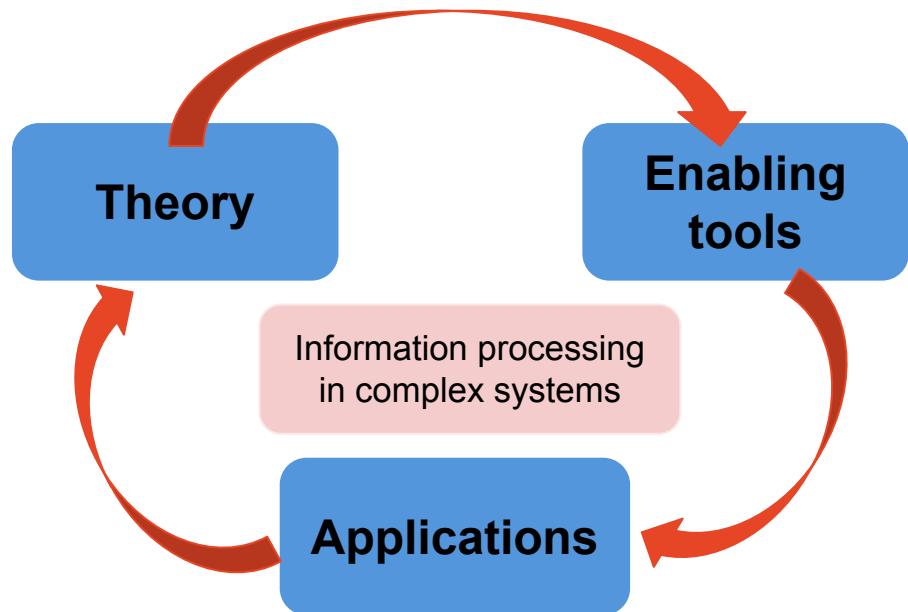
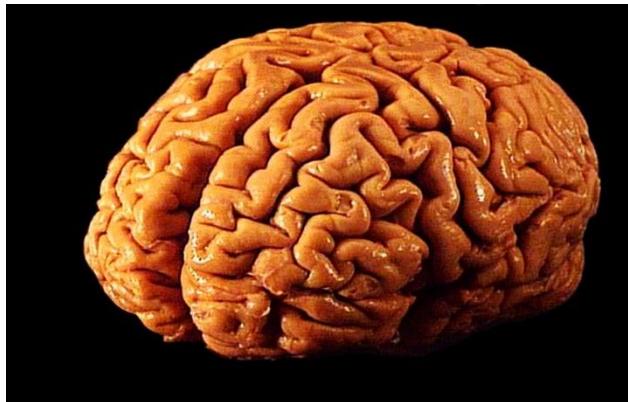


THE UNIVERSITY OF
SYDNEY



Download slides

Neuroscience and me



USyd



Mac
Shine

Richard
Spinney



Oliver
Cliff



Conor
Finn

David
Shorten

Leonardo
Novelli

Mike
Li

Michael
Wibral

Viola
Priesemann



Ben
Fulcher



The University of Sydney

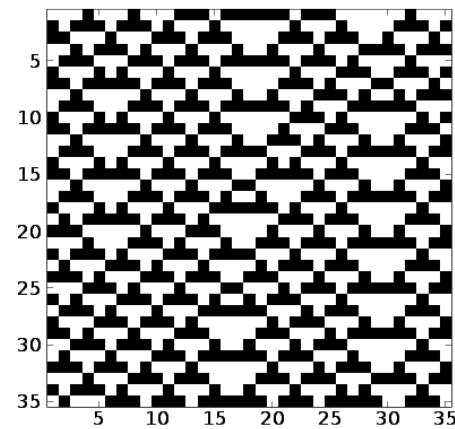
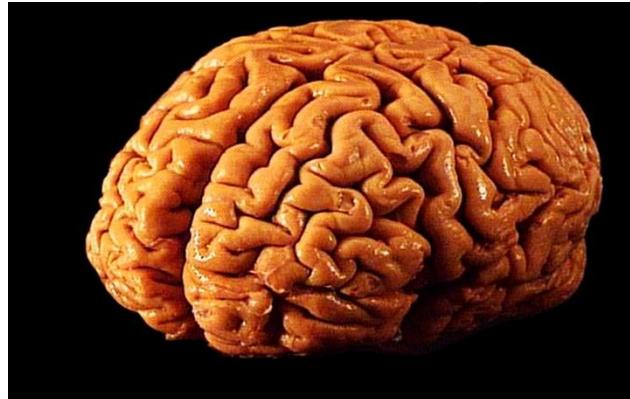


Patricia
Wollstadt



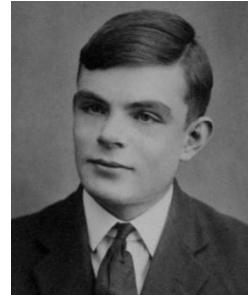
Pedro
Mediano

Complex systems

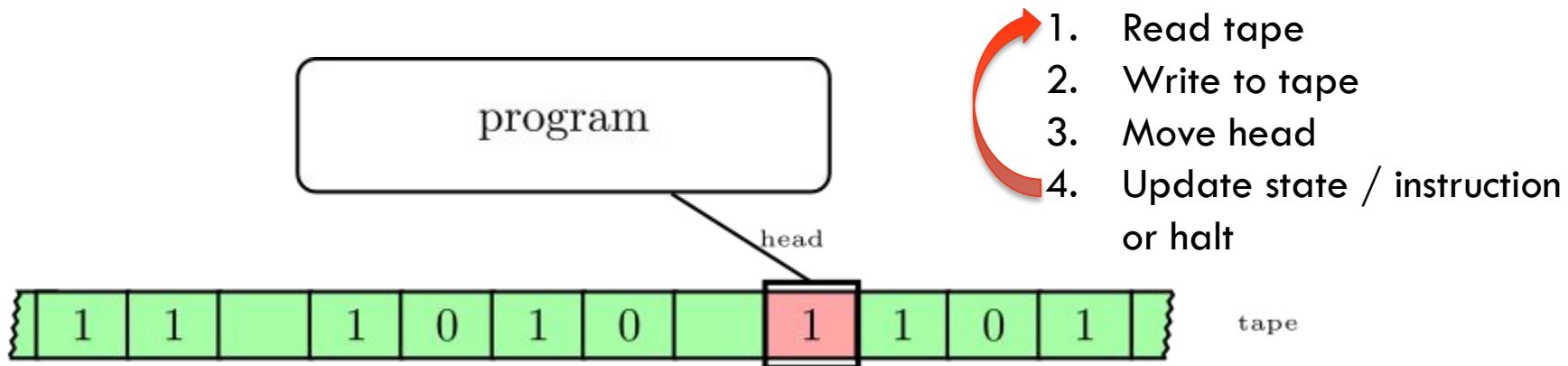


My research focus: how to model information processing in complex systems?

Computation



- How is computation approached in Computer Science?
 - Primary theoretical (abstract) model is a **Turing Machine**
 - A deterministic state machine operating on an infinite tape
 - Well-defined **inputs, outputs, algorithm/task, terminating condition**



Mitchell: For complex systems, the “*Language of dynamical systems may be more useful than language of computation.*”

M. Sipser “Introduction to the Theory of Computation”, PWS Publishing Company, Boston, 1997

Image by Wdvorak (Own work) [CC BY-SA 4.0], via Wikimedia Commons; Turing image (public domain) via Wikimedia Commons

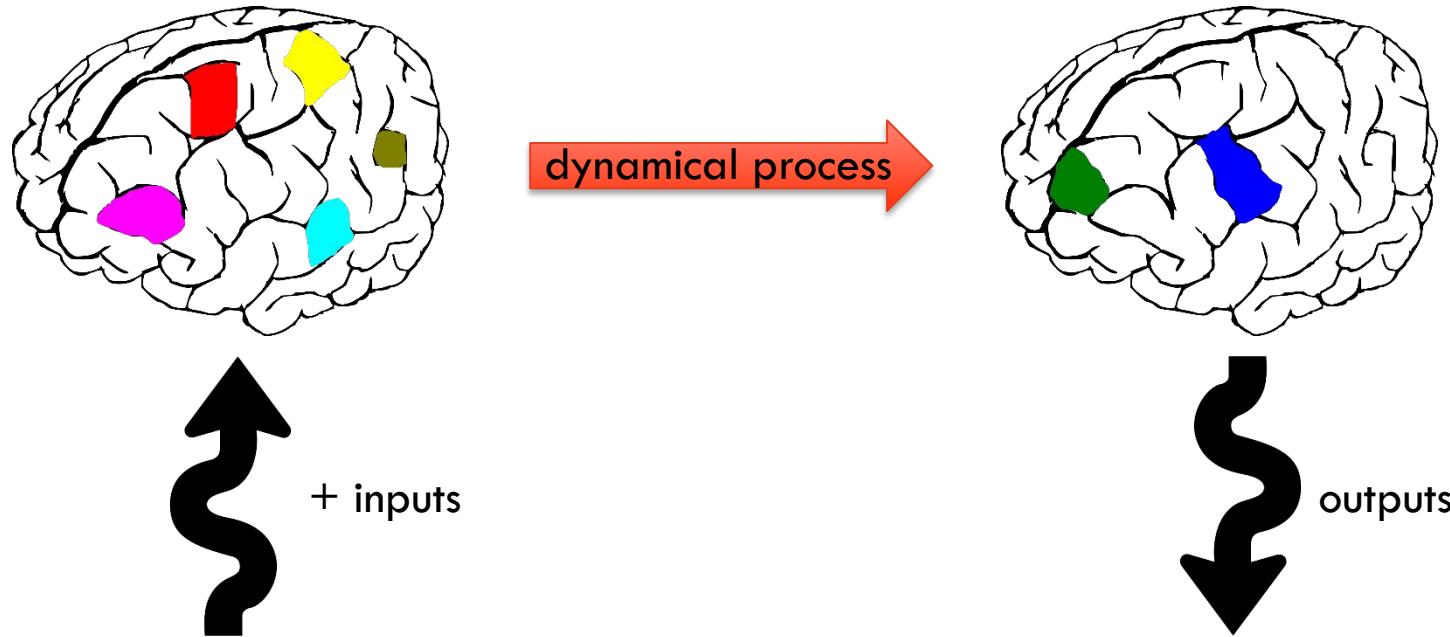
M. Mitchell, “Introduction to Complexity”, Lecture 7

The University of Sydney

Page 4

Biological computation: we need a new perspective

- Mitchell: For complex systems, the “*language of dynamical systems may be more useful than language of computation.*”

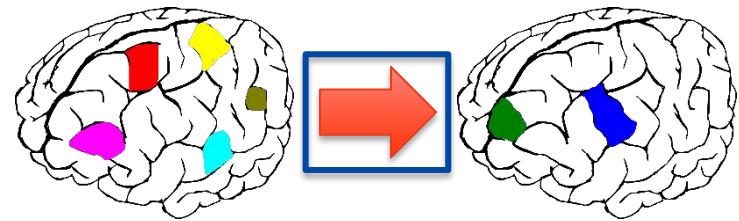


- Intrinsic information processing (Feldman et al., 2008) occurs whenever a system undergoes a dynamical process changing its initial state (+inputs) into some later state (+outputs)

D. P. Feldman, C. S. McTague, and J. P. Crutchfield. “The organization of intrinsic computation: Complexity-entropy diagrams and the diversity of natural information processing”, Chaos, 18(4):043106, 2008.

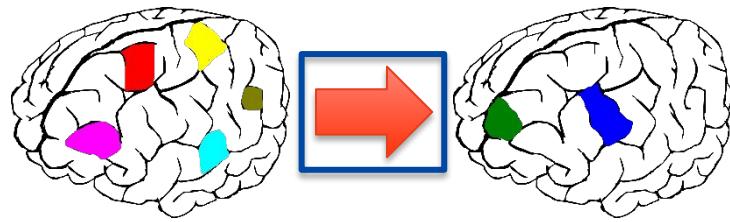
Intrinsic information processing

- We talk about computation as:
 - Memory
 - Signalling
 - Processing
- Intrinsic information processing is any process involving these features:



- How can we make these notions more precise?

Intrinsic information processing



- We talk about computation as:
 - Memory
 - Signalling
 - Processing
- Intrinsic information processing is any process involving these features:



- Idea: Model computation via:
 - Information storage
 - Information transfer
 - Information modification



- Goal: by modelling intrinsic computation in the language it is normally described in, we can better understand e.g. neural information processing and its deficiencies.

Contents

- What is information?
- How do we model information processing?
- What can this reveal about the brain?

A game about information: Guess Who? (Hasbro)

You can play [online](#)

What are the best/worst questions to ask?



Image credit: Matěj Baťha [CC BY-SA 3.0], from [Wikimedia Commons](#)

The University of Sydney

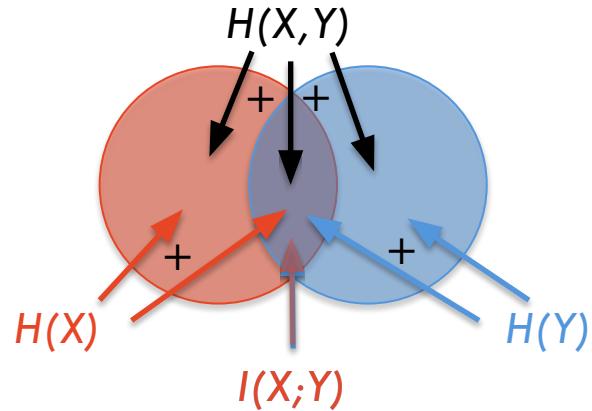
Page 9

Defining information – first pass

- JL: “*Information is all about questions and answers*”

- **Information** is the amount by which
 - one variable (an answer/signal/measurement)
 - reduces our **uncertainty** or **surprises** us
 - about another variable.

- We need to quantify both:
 - Uncertainty (**entropy**)
 - Uncertainty reduction (**information**)



C. E. Shannon. A mathematical theory of communication. Bell System Technical Journal, 27(3–4):379–423, 623–656, 1948.

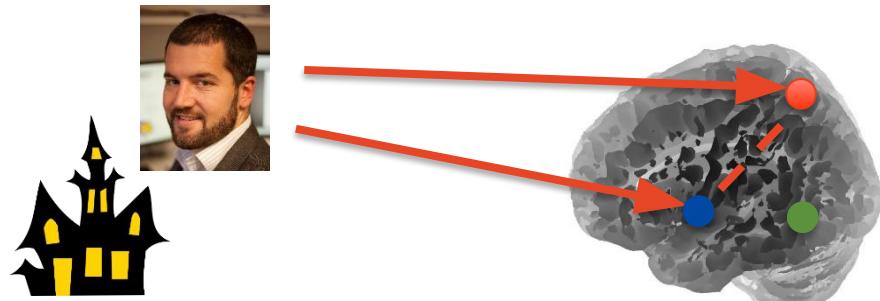
T. M. Cover and J. A. Thomas. Elements of Information Theory. Wiley-Interscience, New York, 1991.

D. J. C. MacKay. Information Theory, Inference, and Learning Algorithms. Cambridge University Press, Cambridge, 2003

The University of Sydney



What can you measure with information theory?

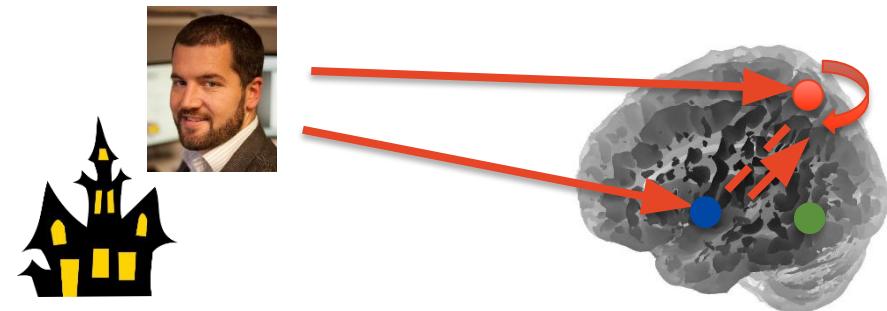


- How surprising is sex = F?
- How much uncertainty is there about the identity?
- Uncertainty reduction about identity from hair colour?
- Information about identity from hair = red?
- Information about sex = F from hair = red?
- Information about sex = F from hair = red given hat = yes?
- How surprising is object = face?
- How much uncertainty is there about the object?
- Uncertainty reduction about object from area-1?
- Information about object from area-1 = 0.8?
- Information about area-1 = 0.8 from area-2 = 0.2?
- Information about area-1 = 0.8 from area-2 = 0.2 given area-3 = -0.2?

Why use information theory for neural data analysis?

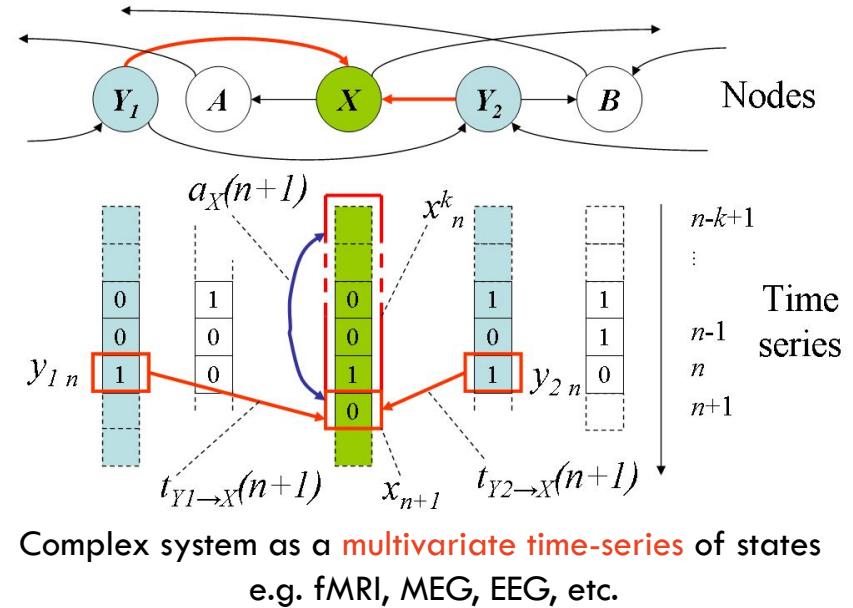
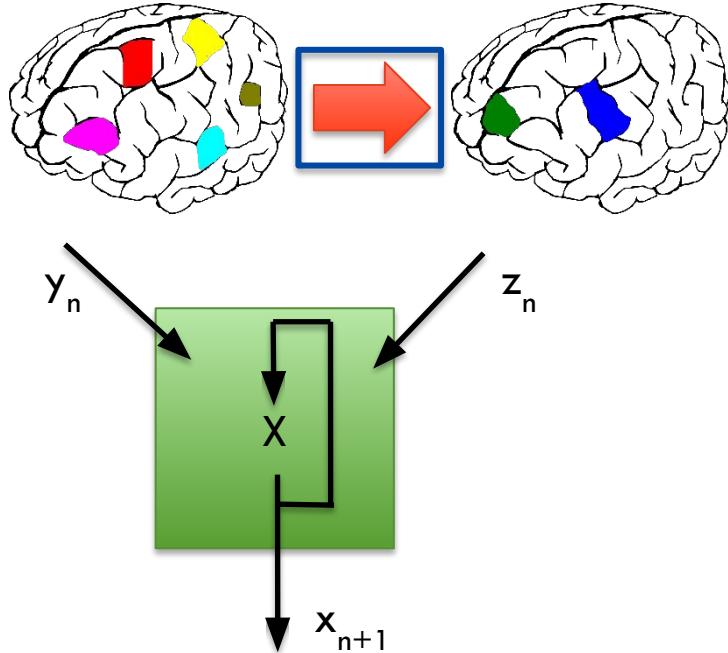
- Allows us to answer questions that are naturally phrased:
 - In a model-free way
 - Captures non-linearities
 - Estimators for different data types, and multivariates

1. The nature of neural codes
 - a. Bridging **Marr's task and implementation levels**
 - b. Functional relationships
2. Modelling information processing dynamics

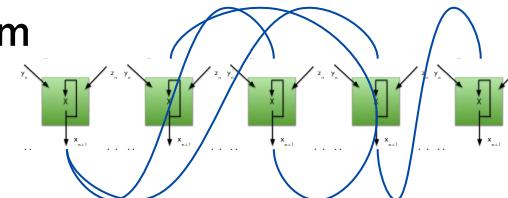


- How surprising is object = face?
- How much uncertainty is there about the object?
- Uncertainty reduction about object from **area-1**?
- Information about object from **area-1** = 0.8?
- Information about **area-1** = 0.8 from **area-2** = 0.2?
- Information about **area-1** = 0.8 from **area-2** = 0.2 given **area-3** = -0.2?

Modelling information processing dynamics



- How can we *model* neural **information processing dynamics**?
 - It is the **output of a local computation** within the system
 - Inputs, outputs, rules/algorithm, halt
 - Model in terms of **information storage and transfer**
 - Establishing an **information-theoretic footprint**

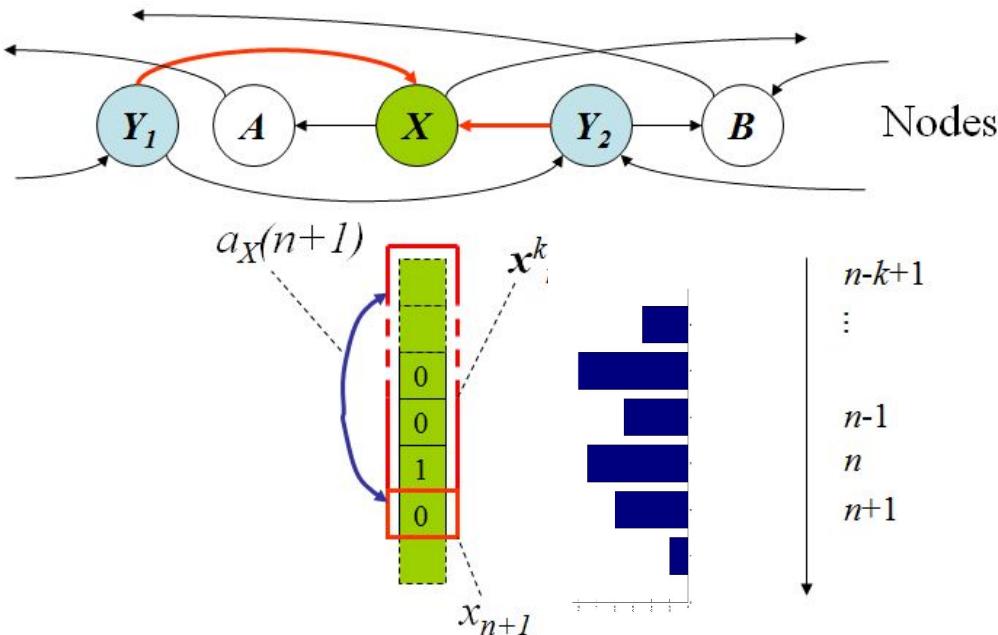


J.T. Lizier, "The local information dynamics of distributed computation in complex systems", Springer: Berlin/Heidelberg, 2013

M. Wibral, J.T. Lizier and V. Priesemann, "Bits from Brains for Biologically-inspired Computing", Frontiers in Robotics and AI, vol. 2, 5, 2015

Information storage

- How much information about the next observation X_n of process X can be found in its past state $\mathbf{X}_n^{(k)} = \{X_{n-k+1}, \dots, X_{n-1}, X_n\}$?



Active information storage (AIS)

$$A_X(k) = I(\mathbf{X}_n^{(k)}; X_{n+1})$$

- Average information from past state that is in use in predicting the next value
- Storage underpins periodic behaviour, stability and concept of memory.
- Model captures: Internally stored, distributed storage, input driven storage

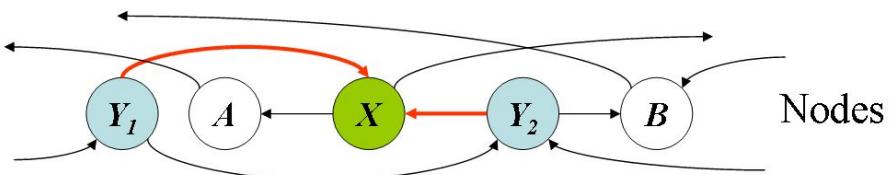
J.T. Lizier, M. Prokopenko, A.Y. Zomaya, "Detecting Non-trivial Computation in Complex Dynamics", Proc. ECAL, pp. 895-904 (2007).

J.T. Lizier, M. Prokopenko, & A.Y. Zomaya, "Local measures of information storage in complex distributed computation", Information Sciences 208, 39 (2012)

Obst, O., Boedecker, J., Schmidt, B., and Asada, M. (2013), "On active information storage in input-driven systems", arXiv:1303.5526

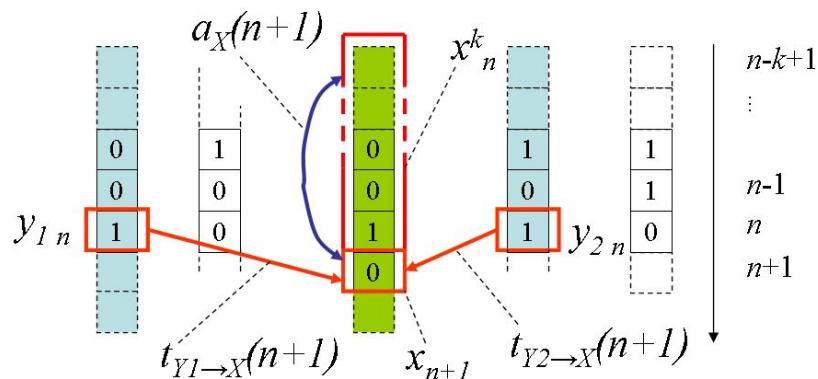
Information transfer

- How much information about the next observation X_n of process X can be found in observation Y_n of process Y , in the context of the past state $X_n^{(k)} = \{X_{n-k+1}, \dots, X_{n-1}, X_n\}$?



Transfer entropy (TE)

$$T_{Y \rightarrow X}(k) = I(Y_n; X_{n+1} | X_n^{(k)})$$



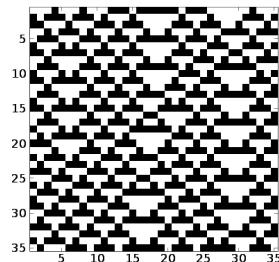
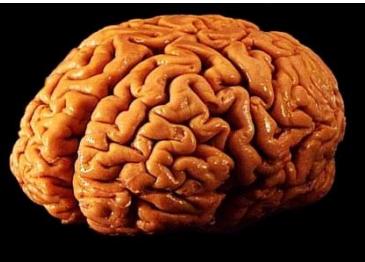
- Average info from source that helps predict next value in context of past.
- Granger-like
- Models strength of predictive effect or statistical dependence.

Total information: $H_X = A_X + T_{Y1 \rightarrow X} + T_{Y2 \rightarrow X|Y1} + H_{X|X^k, Y1, Y2}$

T. Schreiber, "Measuring Information Transfer", Physical Review Letters, 85(2), pp. 461-4, 2000.

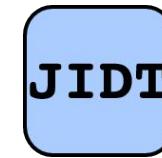
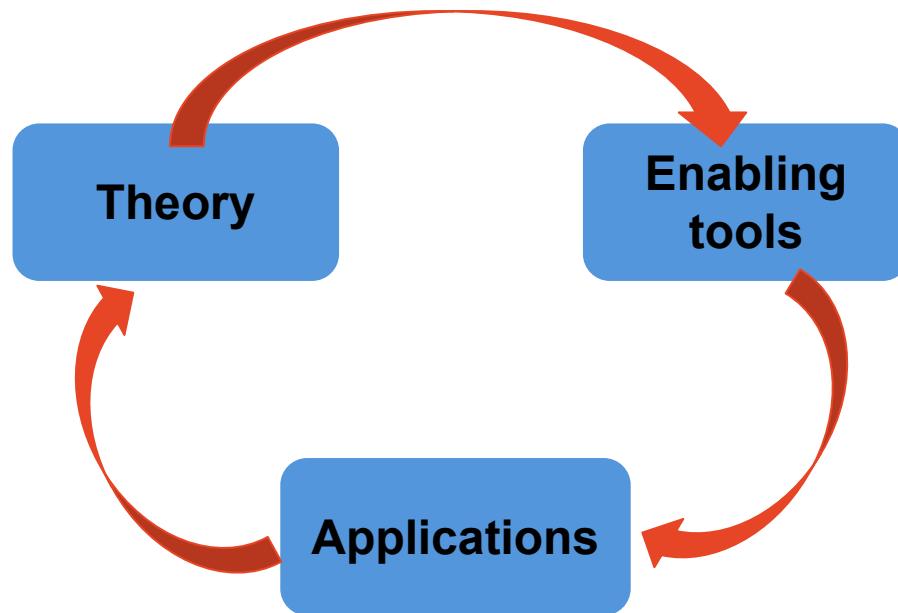
Boszomaier T, Barnett L, Harré M, Lizier JT. 2016 An introduction to transfer entropy. Cham, Switzerland: Springer

Studying information processing in complex systems



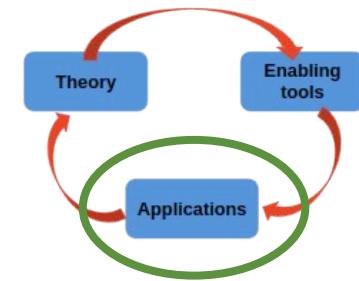
Also see:

Wibral / Priesemann
/ Wollstadt
Mediano / Rosas
Marinazzo
Battaglia
Faes
Stramaglia
Beggs
Barnett / Seth
Shimono
...



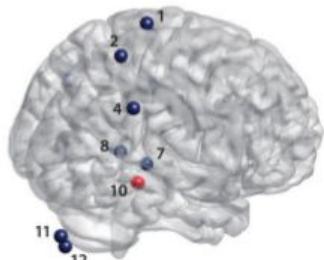
Open source
tools:
JIDT and IDTxl

Analysing neural information processing

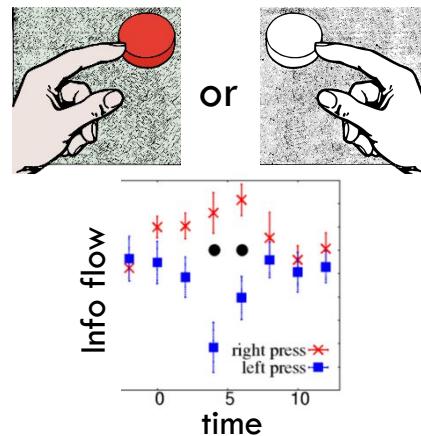
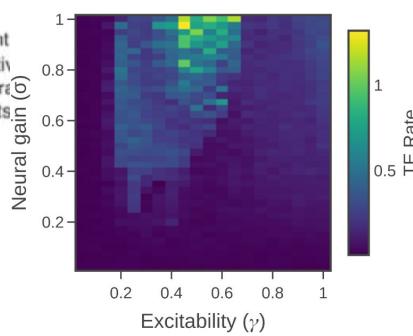


Key question: What does such modelling tell us about neural information processing?

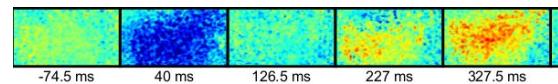
1. Characterising different regimes of behavior



● significant
● no significant decrease of active information storage for ASD patients

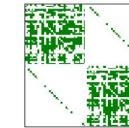


2. Space-time dynamics

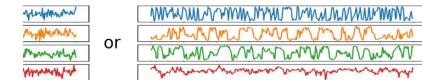


3. Functional and effective network modelling

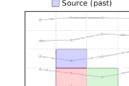
Adjacency matrix (ground truth)



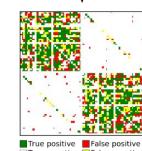
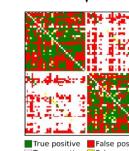
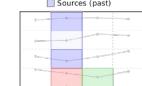
Neural mass model

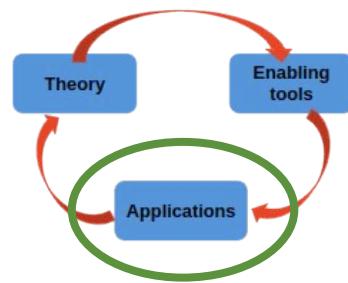


Bivariate TE



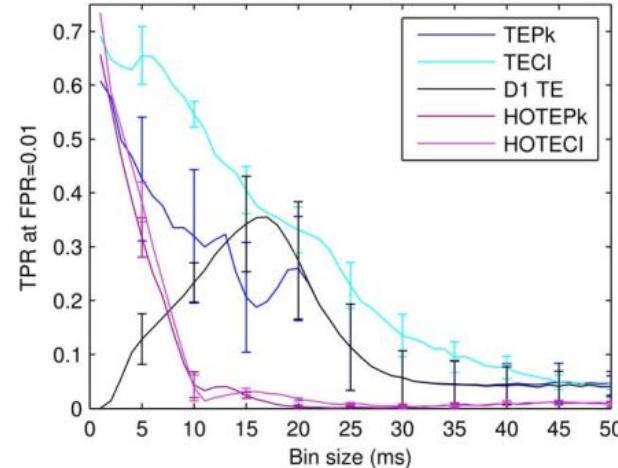
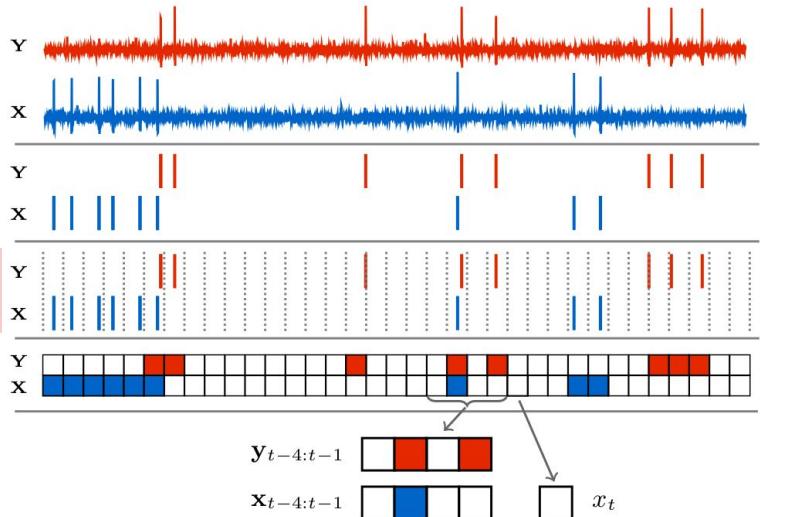
Multivariate TE



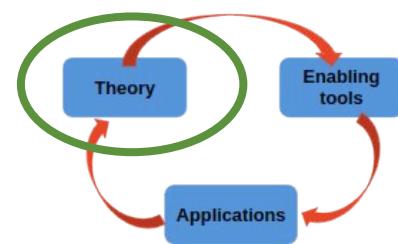


1. Information flow between spike trains

- Promise of high spatial / temporal precision recordings to facilitate study of information flows at the neural level
- Prior use of discrete-time TE estimator to infer directed functional networks and study their properties:
 - Long-tailed degree distribution
 - Rich club
 - Relation of hubs to specific time-scales



But TE is highly sensitive to discrete bin size Δt (Ito et al 2011)



1. Information flow between spike trains

Established theoretical formulation for TE between spike trains:

- Properly formulated in terms of **rates** and **path integrals**
- If a limiting rate exists, a consistent discrete-time estimate of should scale with Δt :

$$\dot{T}_{y \rightarrow x}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} T_{y \rightarrow x}^{\Delta t}(t)$$

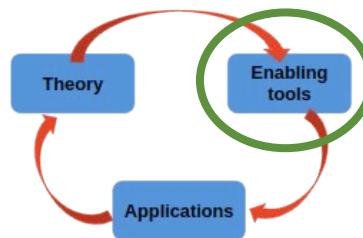
- Only interactions at target spikes contribute to the average rate.

$$\dot{T}_{Y \rightarrow X} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \sum_{i=1}^{N_X} \ln \frac{\lambda_{x|\mathbf{x}_{<t}, \mathbf{y}_{<t}}[\mathbf{x}_{<x_i}, \mathbf{y}_{<x_i}]}{\lambda_{x|\mathbf{x}_{<t}}[\mathbf{x}_{<x_i}]}$$

Sampling time ↑

Expected target spike rate,
given source and target past

Expected target spike rate,
given target past

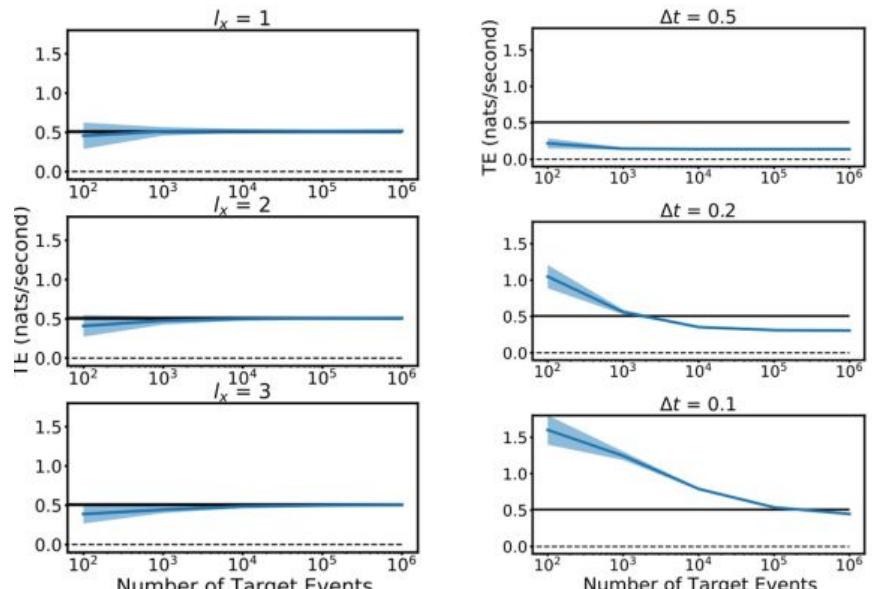
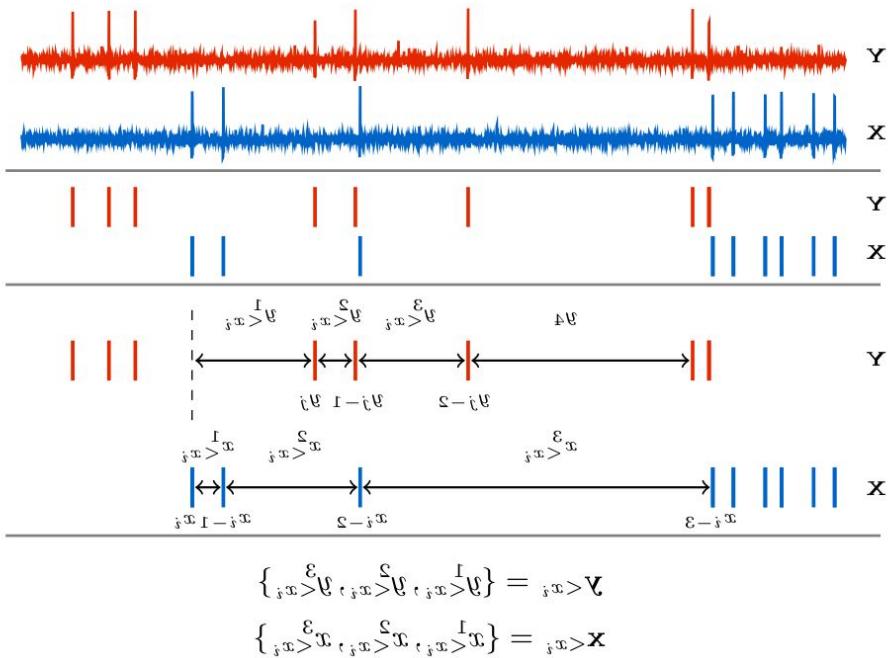


1. Information flow between spike trains

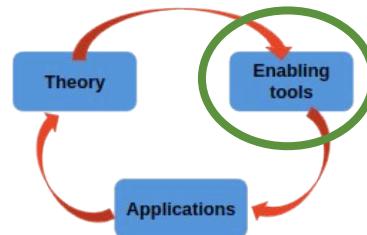
Translated theoretical formulation into estimation algorithm:

- Utilised nearest-neighbour matching on source-target ISI vectors to provide an efficient estimator, which:

is consistent, not sensitive to parameters

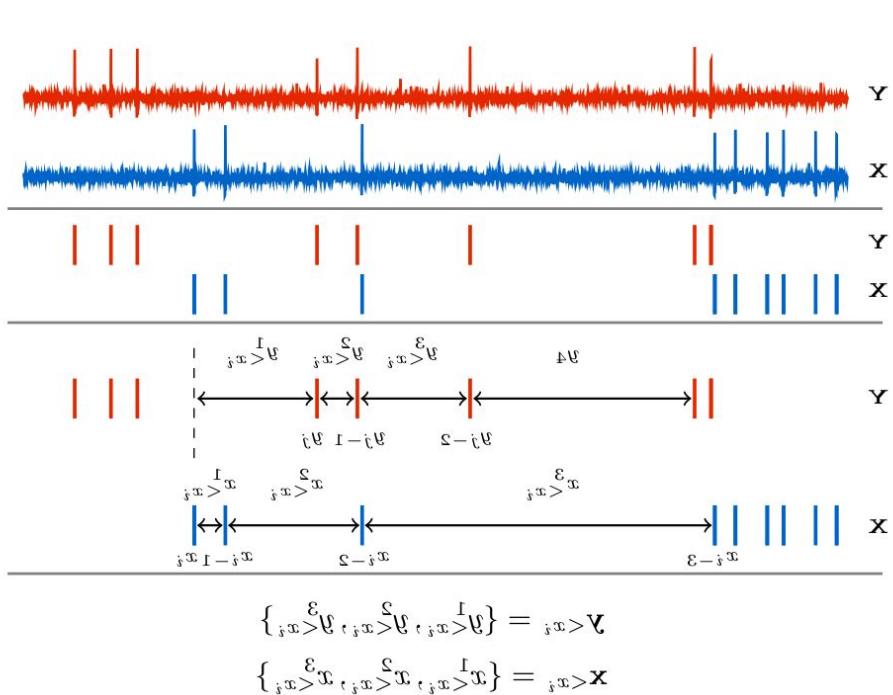


1. Information flow between spike trains

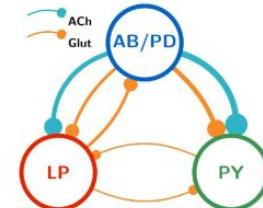


Translated theoretical formulation into estimation algorithm:

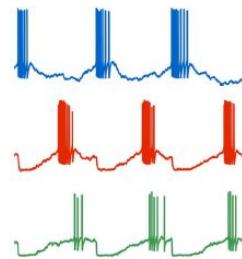
- Utilised nearest-neighbour matching on source-target ISI vectors to provide an efficient estimator, which:



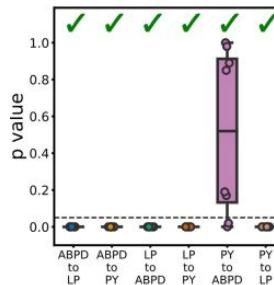
reliably distinguishes conditional in/dependence



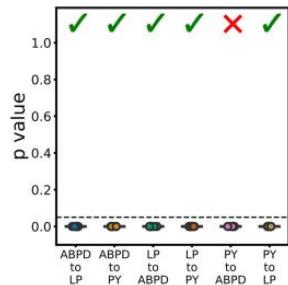
(A) Circuit connectivity diagram



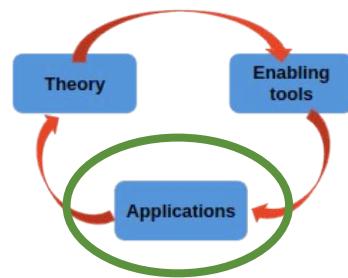
(B) Example membrane potential traces produced by the circuit.



(C) Distribution of p values from the continuous-time estimator and the local permutation surrogate generation method.



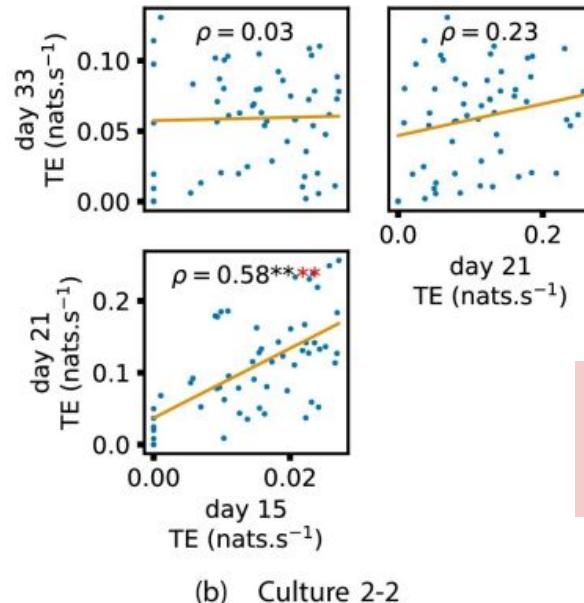
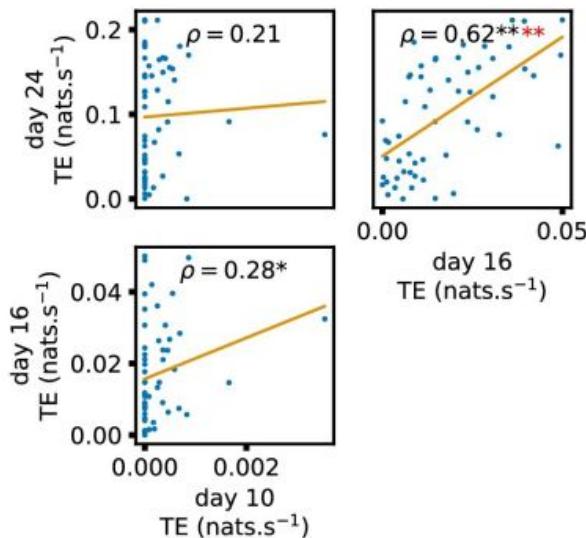
(D) Distribution of p values from the discrete-time estimator and the source time-shift surrogate generation method



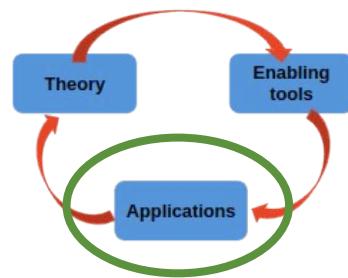
1. Information flow between spike trains

Studied changes in TE during development of neural cell cultures:

- Examined all pairs directed functional network, as function of culture age (DIV)



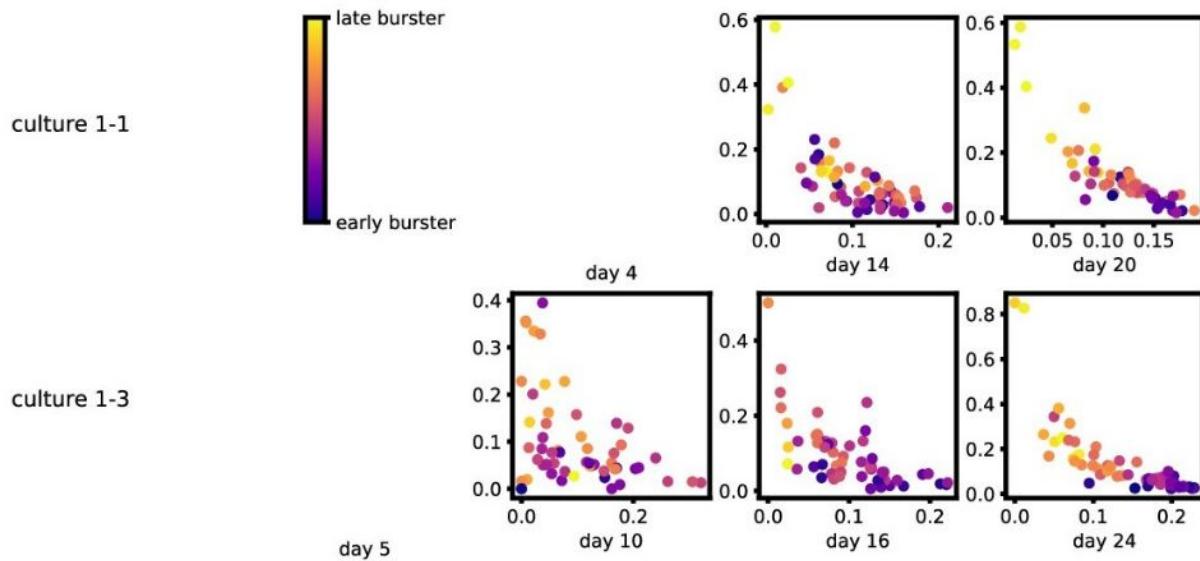
Information flows lock
in early in development



1. Information flow between spike trains

Studied changes in TE during development of neural cell cultures:

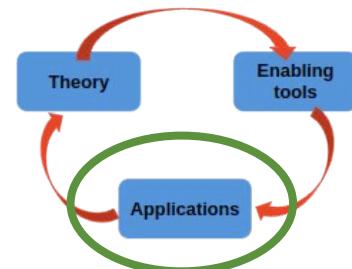
- Examined all pairs directed functional network, as function of culture age (DIV)



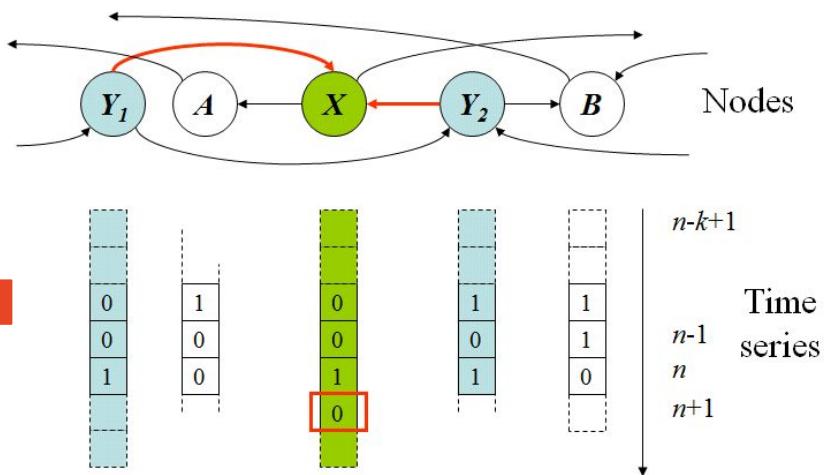
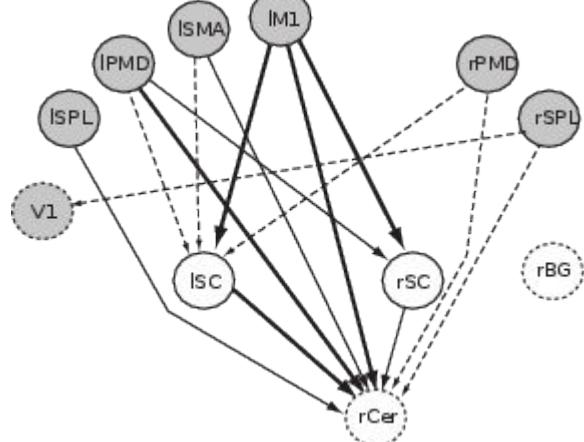
Nodes take on different computational roles during burts

Similar results observed in synthetic model with STDP

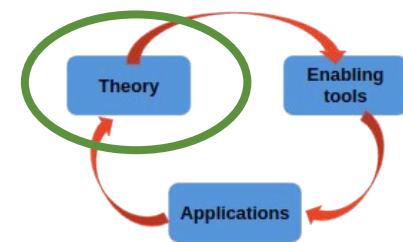
2. Effective network modelling



- Information transfer is ideally placed for the “*inverse problem*”
 - effective connectivity analysis – inferring a “*minimal circuit model*” that can explain observed dynamics
 - Aim: **Model the computation** taking place in the system, revealing emergent computational structure.

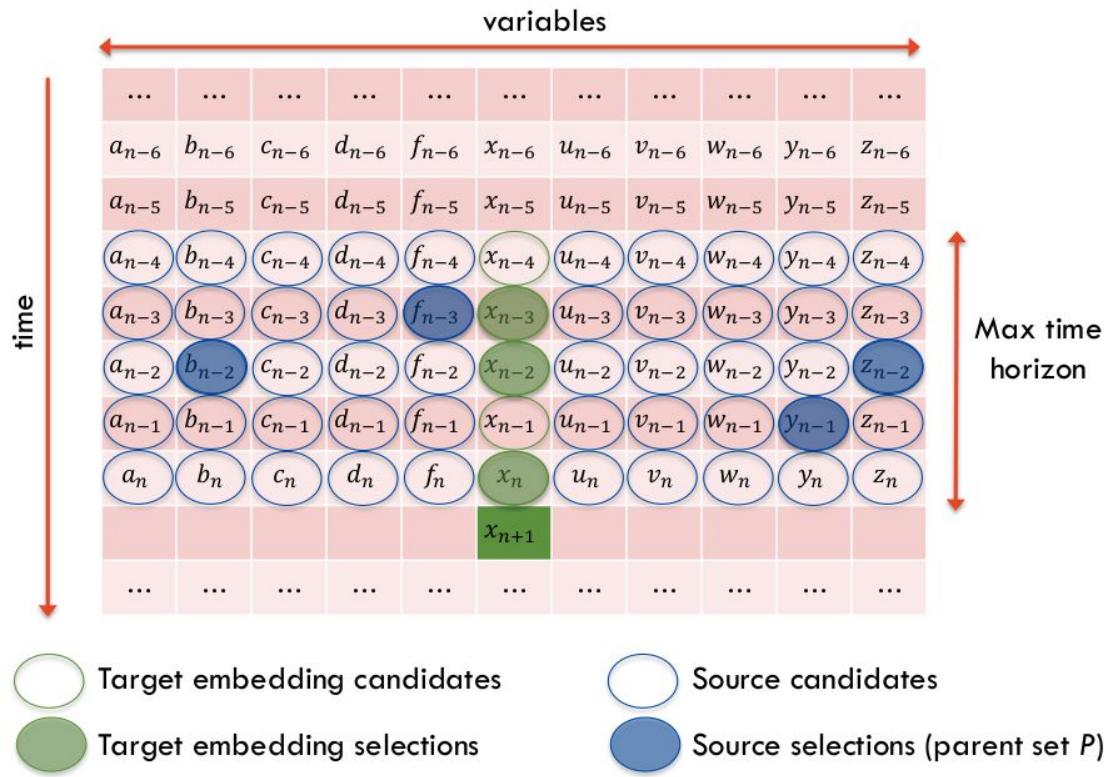


2. Effective network modelling



Real goal here: to infer this **parent set** $\{Y, Z\}$ for X

$$H(X_{n+1}) = I\left(\mathbf{X}_n^{(k)}; X_{n+1}\right) + I\left(Y_n, Z_n; X_{n+1} \mid \mathbf{X}_n^{(k)}\right) + H\left(X_{n+1} \mid \mathbf{X}_n^{(k)}, Y_n, Z_n\right)$$



Vlachos, I., & Kugiumtzis, D., Physical Review E, 82(1), 016207, 2010

Faes, L., Nollo, G., & Porta, A., Physical Review E, 83(5), 051112, 2011

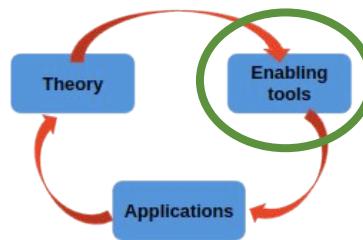
J. T. Lizier and M. Rubinov. Technical Report 25/2012, Max Planck Institute for Mathematics in the Sciences, 2012.

S. Stramaglia, G.-R. Wu, M. Pellicoro, and D. Marinazzo, Physical Review E, 86(6):066211+, 2012

Montalto, A., Faes, L., & Marinazzo, D., PLoS ONE, 9(10), e109462, 2014

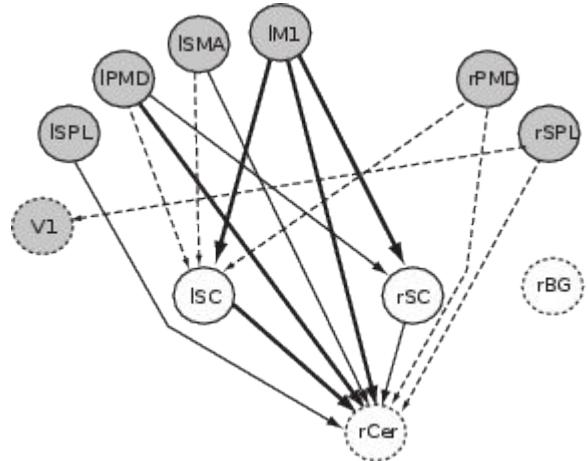
Sun, J., Taylor, D., & Boltt, E. M., SIAM Journal on Applied Dynamical Systems, 14(1), 73–106, 2015

L. Novelli, P. Wollstadt, P. Mediano, M. Wibral, J.T. Lizier, Network Neuroscience, 3(3), 827-847, 2019



2. Effective network modelling

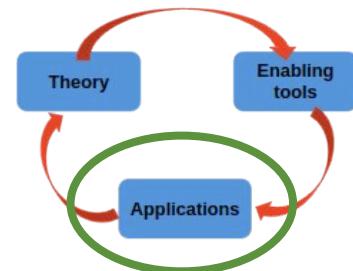
- Information transfer is ideally placed for the “*inverse problem*”
- effective connectivity analysis – inferring a “*minimal circuit model*” that can explain observed dynamics
 - Aim: **Model the computation** taking place in the system, revealing emergent computational structure.



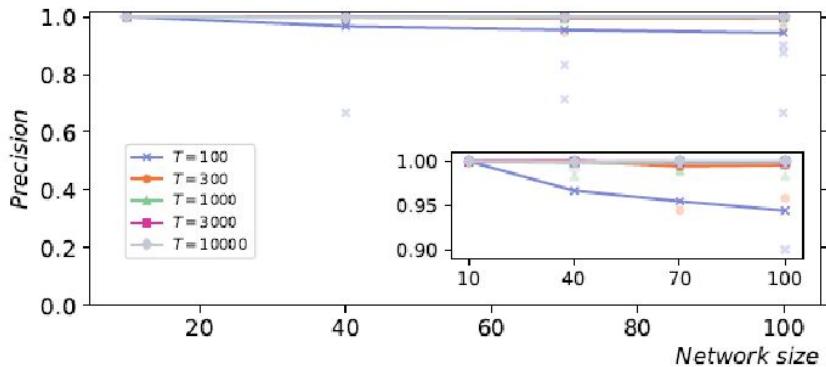
+ Multivariate extensions, GPU & efficiencies
 = <https://github.com/pwollstadt/IDTxl>

P. Wollstadt, J.T. Lizier, R. Vicente, C. Finn, M. Martinez-Zarzuela, P. Mediano, L. Novelli and M. Wibral, "IDTxl: The Information Dynamics Toolkit xl: a Python package for the efficient analysis of multivariate information dynamics in networks", Journal of Open Source Software, 4(3), 1081, 2019
 L. Novelli, P. Wollstadt, P. Mediano, M. Wibral, J.T. Lizier, Network Neuroscience, 3(3), 827-847, 2019

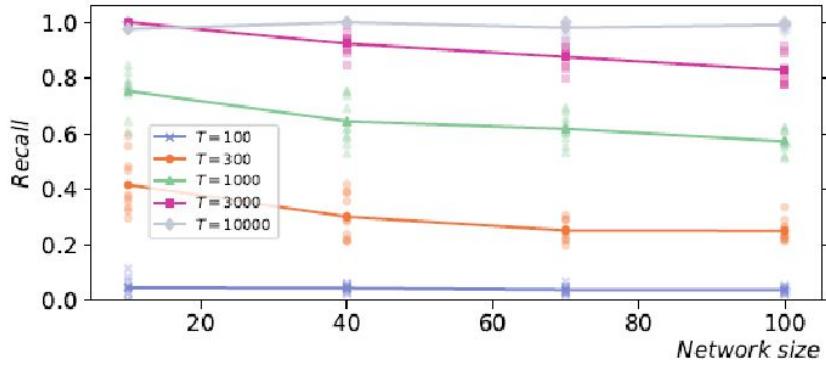
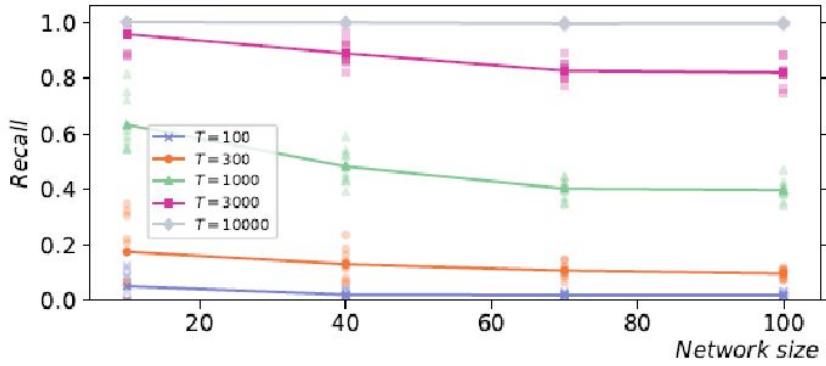
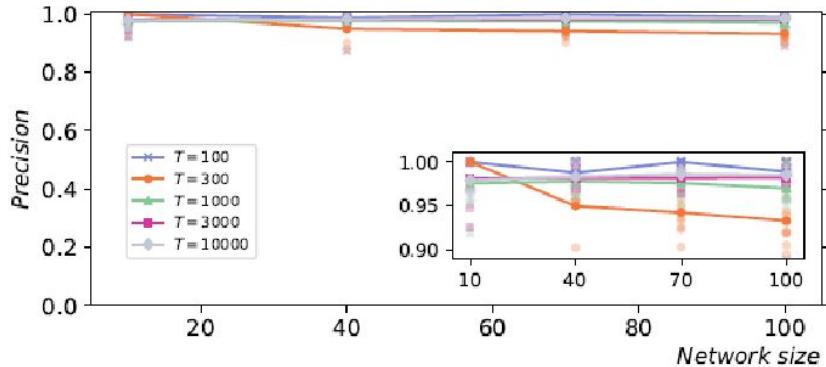
2. Effective network modelling



Vector auto-regressive

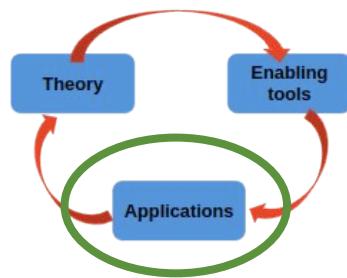


Coupled logistic maps

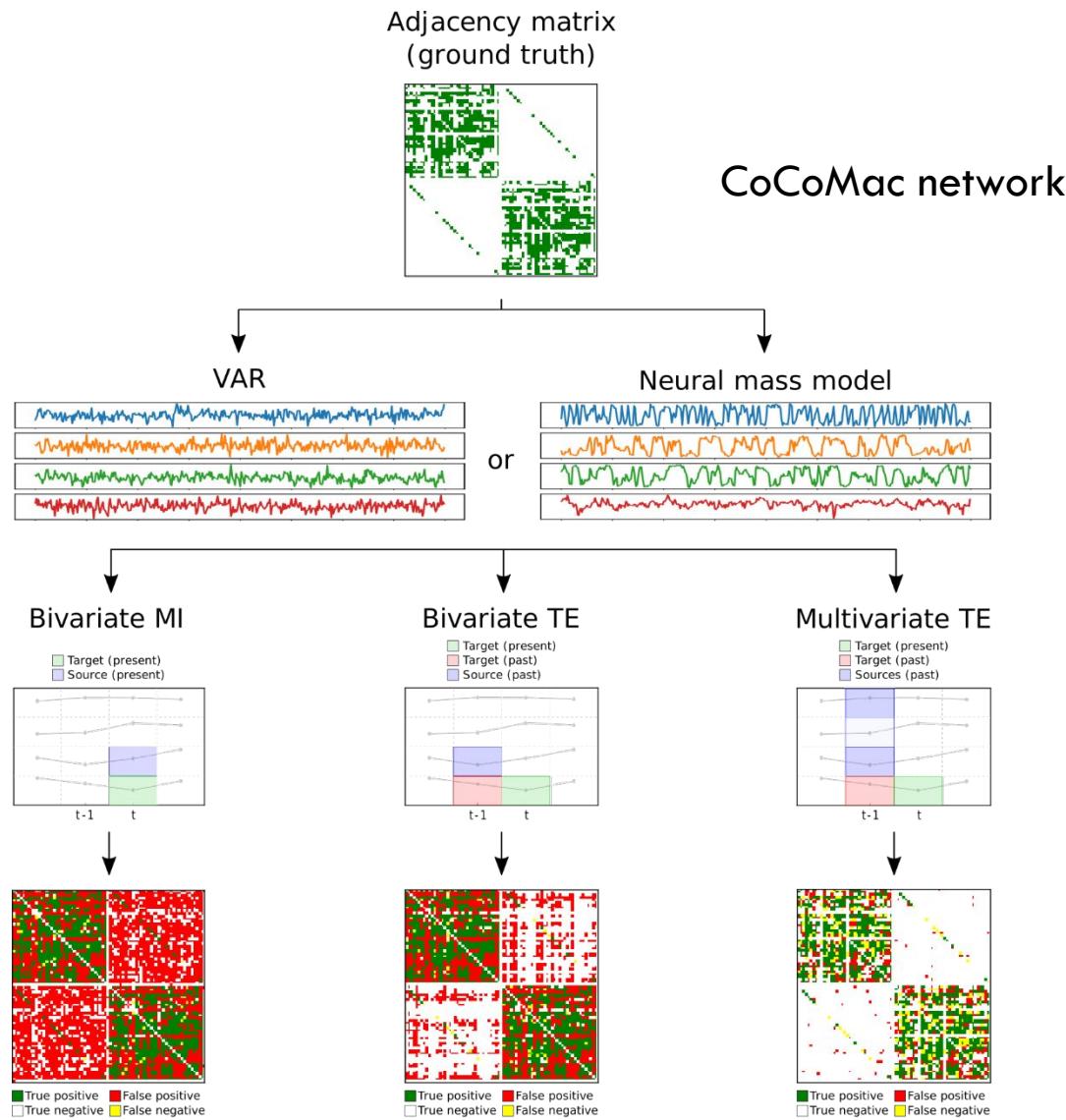


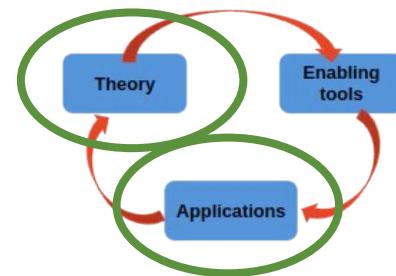
First time test of realistic size data sets with non-linear estimator

$$\alpha_{\max} = 0.001, S = 1000 \text{ surrogates}, \max \bar{\delta}_x = 5$$



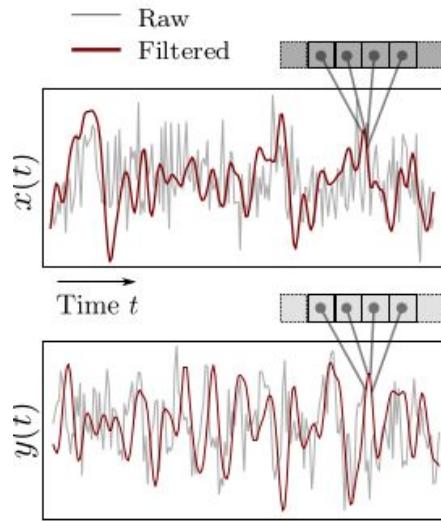
2. Effective network modelling



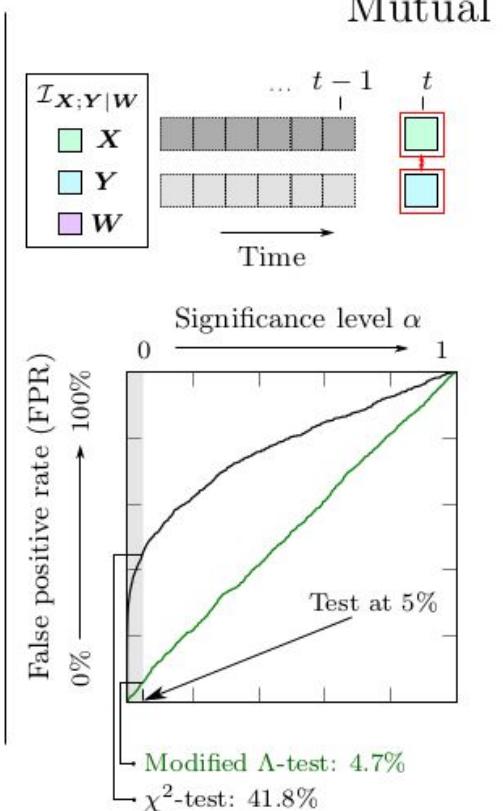


2. Effective network modelling

Time series data



Mutual information



E. Florin, J. Gross, J. Pfeifer, G. R. Fink, and L. Timmermann, *NeuroImage* 50, 577 (2010).

L. Barnett and A. K. Seth, *J. Neurosci. Methods* 201, 404 (2011).

C. E. Davey, D. B. Grayden, G. F. Egan, and L. A. Johnston, *NeuroImage* 64, 728 (2013).

S. Afyouni, S. M. Smith, and T. E. Nichols, *NeuroImage* 199, 609 (2019).

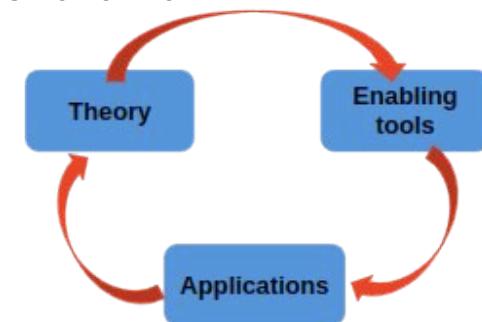
M. S. Bartlett, *J. R. Stat. Soc.* 98, 536 (1935).

The University of Sydney

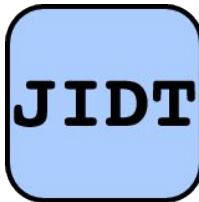
Cliff, O. M., Novelli, L., Fulcher, B. D., Shine, J. M., & Lizier, J. T., *Phys. Rev. Research* 3, 013145 (2021)

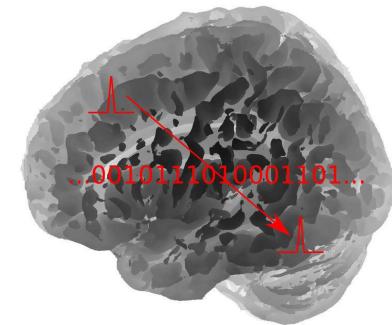
Summary

- How to model information processing in complex systems
 - How this is naturally applied to answer meaningful questions in a neuroscience setting
 - The integration between theory, empirical enablers and applications is helping forge the path
-
- What can it do for us in a neuroscience setting?
 - Characterising different **regimes** of behaviour;
 - **Space-time** characterisation of information processing **dynamics**;
 - Effective **network** inference;
 - etc. ...



Plugs

- JIDT / IDTxI 
- Info Theory workshop on Tuesday / Wednesday
→ <https://bit.ly/cns2022itw>
- Come study with me!
- Slides:



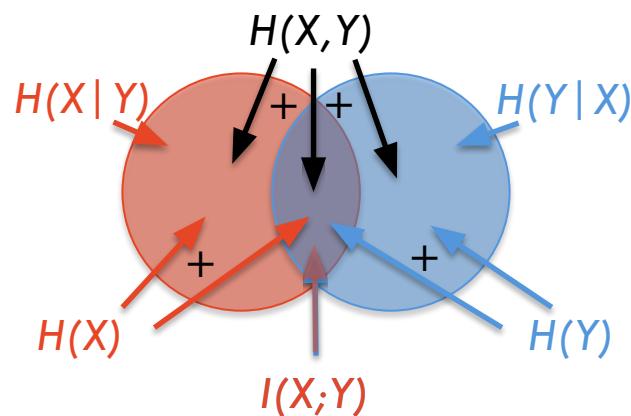
Questions



THE UNIVERSITY OF
SYDNEY

Fundamental measures

	Entropy (uncertainty)	Mutual information
Average		
Pointwise/local		



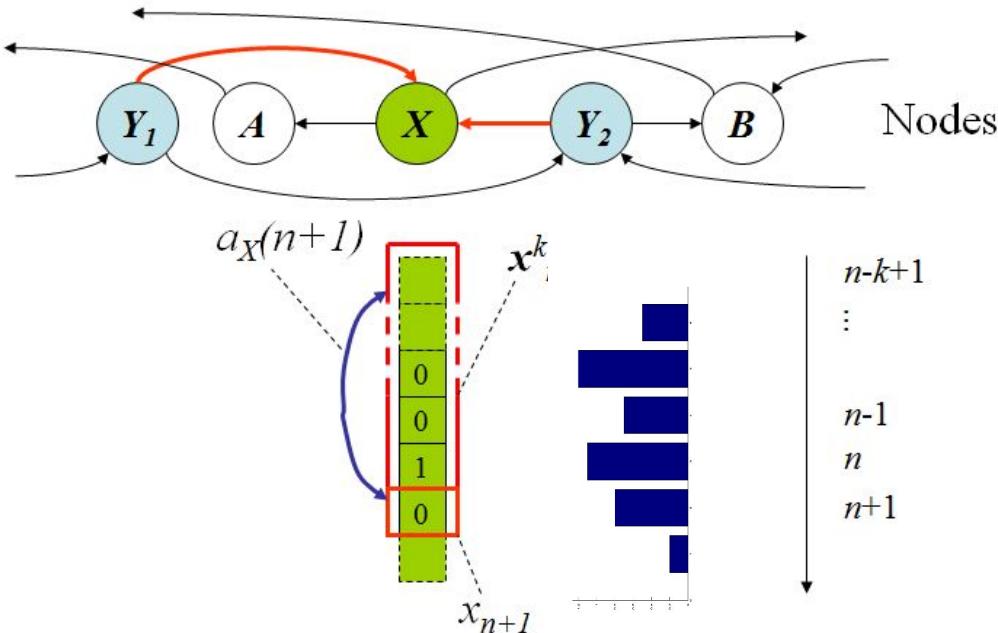
Can define:

- Joint versions,
 - e.g. $H(X, Y)$ or $H(X)$; and
- Conditional versions,
 - e.g. $H(X|Y)$, $I(X;Y|Z)$

Uncertainties and information in Guess Who?

Active information storage

- How much information about the next observation X_n of process X can be found in its past state $\mathbf{X}_n^{(k)} = \{X_{n-k+1}, \dots, X_{n-1}, X_n\}$?



Active information storage (AIS)

$$A_X(k) = I(\mathbf{X}_n^{(k)}; X_{n+1})$$

- Average information from past state that is in use in predicting the next value

Local AIS

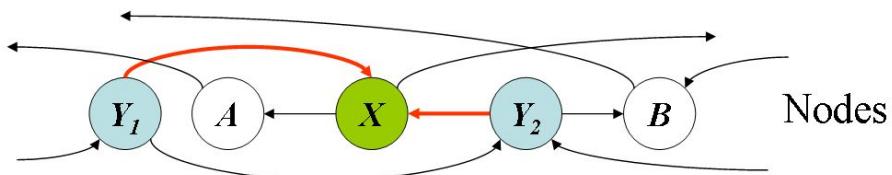
- Information from a specific past state in use in predicting specific next value

J.T. Lizier, M. Prokopenko, A.Y. Zomaya, “Detecting Non-trivial Computation in Complex Dynamics”, Proc. ECAL, pp. 895-904 (2007).

J.T. Lizier, M. Prokopenko, & A.Y. Zomaya, “Local measures of information storage in complex distributed computation”, Information Sciences 208, 39 (2012)

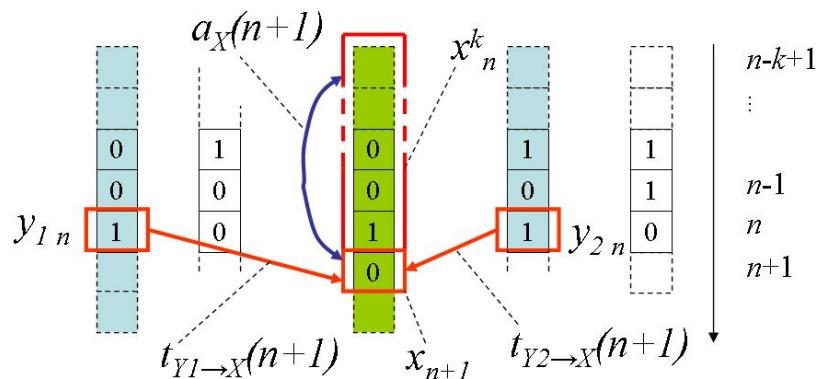
Transfer entropy

- How much information about the next observation X_n of process X can be found in observation Y_n of process Y , in the context of the past state $X_n^{(k)} = \{X_{n-k+1}, \dots, X_{n-1}, X_n\}$?



Transfer entropy (TE)

$$T_{Y \rightarrow X}(k) = I(Y_n; X_{n+1} | X_n^{(k)})$$



- Average info from source that helps predict next value in context of past.
- Granger-like

Local TE

- Local TE: info from a specific source value to predict specific next value in context of past.

T. Schreiber, "Measuring Information Transfer", Physical Review Letters, 85(2), pp. 461-4, 2000.

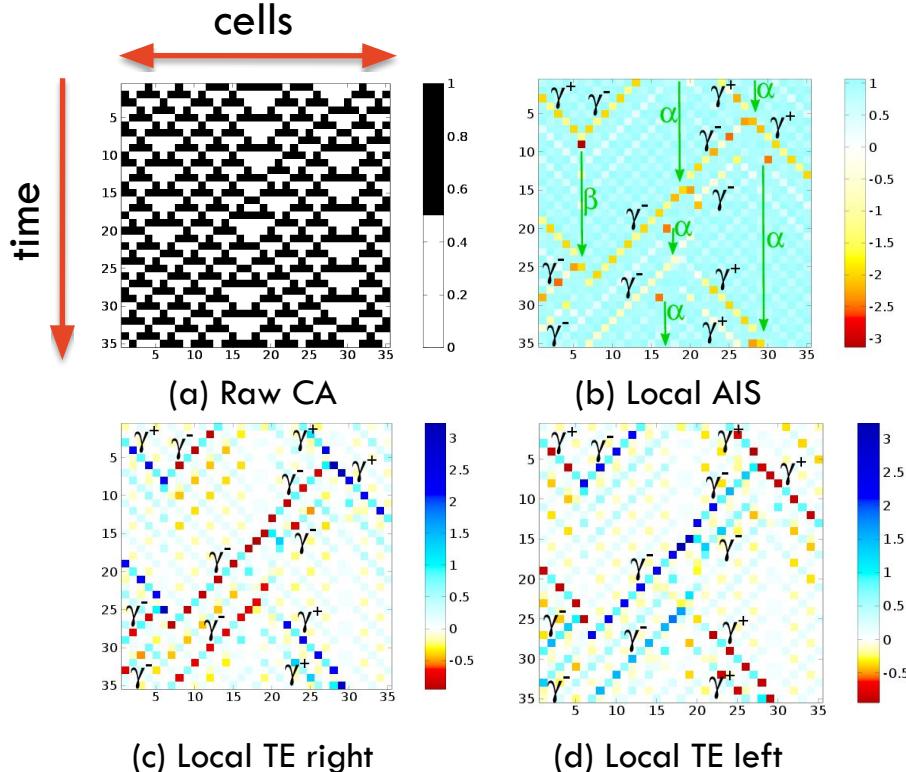
J. T. Lizier, M. Prokopenko, & A. Y. Zomaya. "Local information transfer as a spatiotemporal filter for complex systems". Physical Review E, 77(2):026110, 2008.
The University of Sydney

A critique of a critique of transfer entropies

- James et al. (2016) criticise 3 aspects of transfer entropies:
 1. It is not causal;
 2. Includes synergies as well as unique information; and changes on conditioning.
 3. “Shoehorns” dyadic relationships into network representation
- Resolution: It is a **model** of information flow, in juxtaposition to information storage. Responses:
 1. Known (Ay and Polani, 2008; Chicharro and Ledberg, 2012; Lizier and Prokopenko, 2010): but only this modelling perspective aligns with how we describe information processing.
 2. Known (Lizier et al., 2010; Williams and Beer, 2011): Synergies are an important part of how we describe information flow (e.g. gliders). Different models (pairwise and conditional) of flow are important components of a wholistic view of information processing; they are complementary.
 3. Disagree on interpretation (Bossomaier et al., 2016; Lizier and Rubinov, 2012): More later ...

Example: modelling information processing in CAs

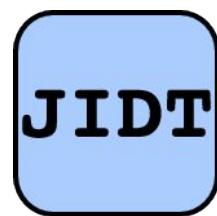
- Transfer entropy perspective aligns with how we describe complex systems
 - Emergent glider structures have strongest information transfer values.



Blinkers and background domains are dominant storage entities!

Links algorithmic and implementation levels

Gliders are dominant transfer entities!



J. T. Lizier, M. Prokopenko, & A. Y. Zomaya.
“Local information transfer as a spatiotemporal filter for complex systems”. Physical Review E, 77(2):026110, 2008.

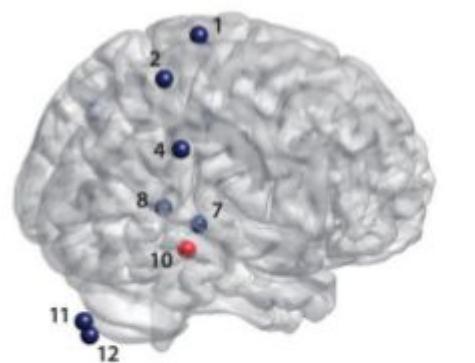
J.T. Lizier, “JIDT: An Information-Theoretic toolkit for studying the dynamics of complex systems”. Frontiers in Robotics and AI, 1:11, 2014.

1. Characterising different regimes of behaviour

Aim: to characterise behaviour and responses in terms of information processing

- For example:

W.r.t neural conditions:
ASD vs controls



● significant
● no significant
decrease of active
information storage
for ASD patients

W.r.t different stimuli

Regarding critical dynamics

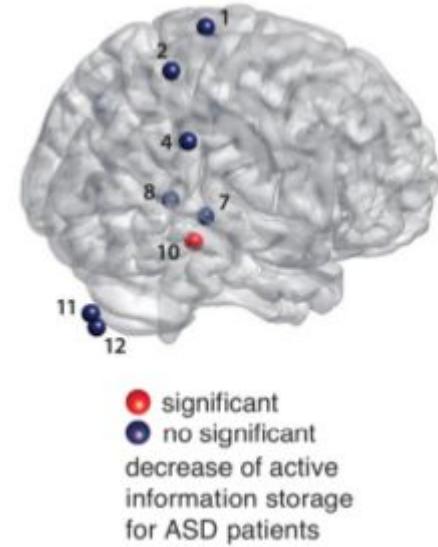
Does information storage use change in ASD?

Two MEG studies contrasting active information storage in resting state activity of ASD subjects vs controls:

- [1] 10x14, 12 sources;
- [2] 19x19, 478 sources.

Lower information storage overall and in:

- hippocampus [1],
 - precuneus, posterior cingulate cortex, supramarginal gyrus (all DMN) [2].
- of Autism Spectrum Disorder subjects.

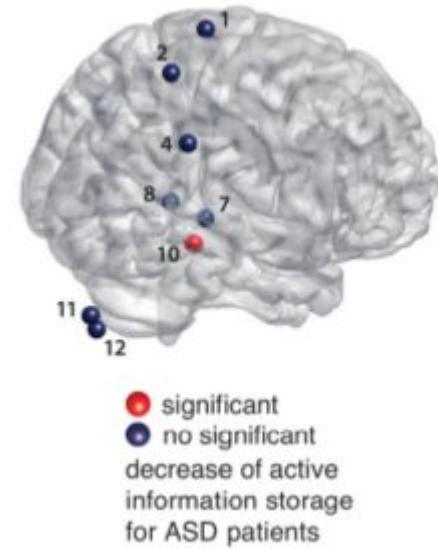


[1] C. Gómez, et al., "Reduced predictable information in brain signals in autism spectrum disorder", *Frontiers in Neuroinformatics*, 8:9+, 2014.

[2] A. Brodski-Guerniero, et al., "Predictable information in neural signals during resting state is reduced in autism spectrum disorder", *Human Brain Mapping*, 39(8):3227–3240, 2018

Does information storage use change in ASD?

- Lower information storage overall and in particular areas for ASD subjects.
- Use/precision of prior reduced at source level
- Decrease correlates with symptom severity (ADI-R rit) for cerebellum and precuneus.
- No significant changes for α power, ACT nor signal entropy.



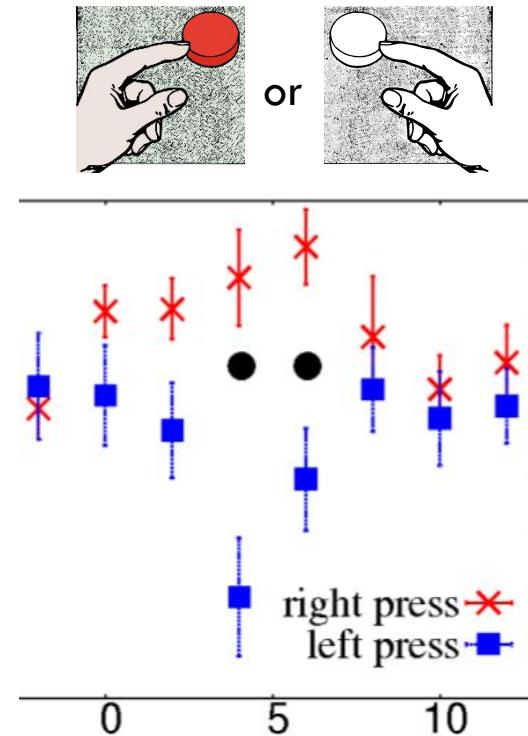
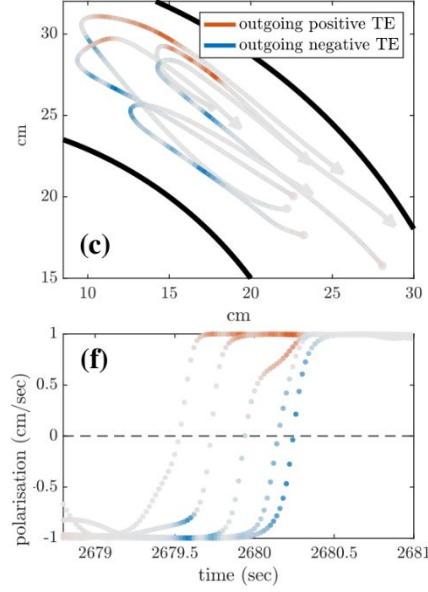
[1] C. Gómez, et al., "Reduced predictable information in brain signals in autism spectrum disorder", *Frontiers in Neuroinformatics*, 8:9+, 2014.

[2] A. Brodski-Guerniero, et al., "Predictable information in neural signals during resting state is reduced in autism spectrum disorder", *Human Brain Mapping*, 39(8):3227–3240, 2018

2. Space-time characterization of info processing

Aim:

- Highlight info processing hot-spots;
- Use info processing to explain dynamics;
- Validate conjectures on neural information processing.



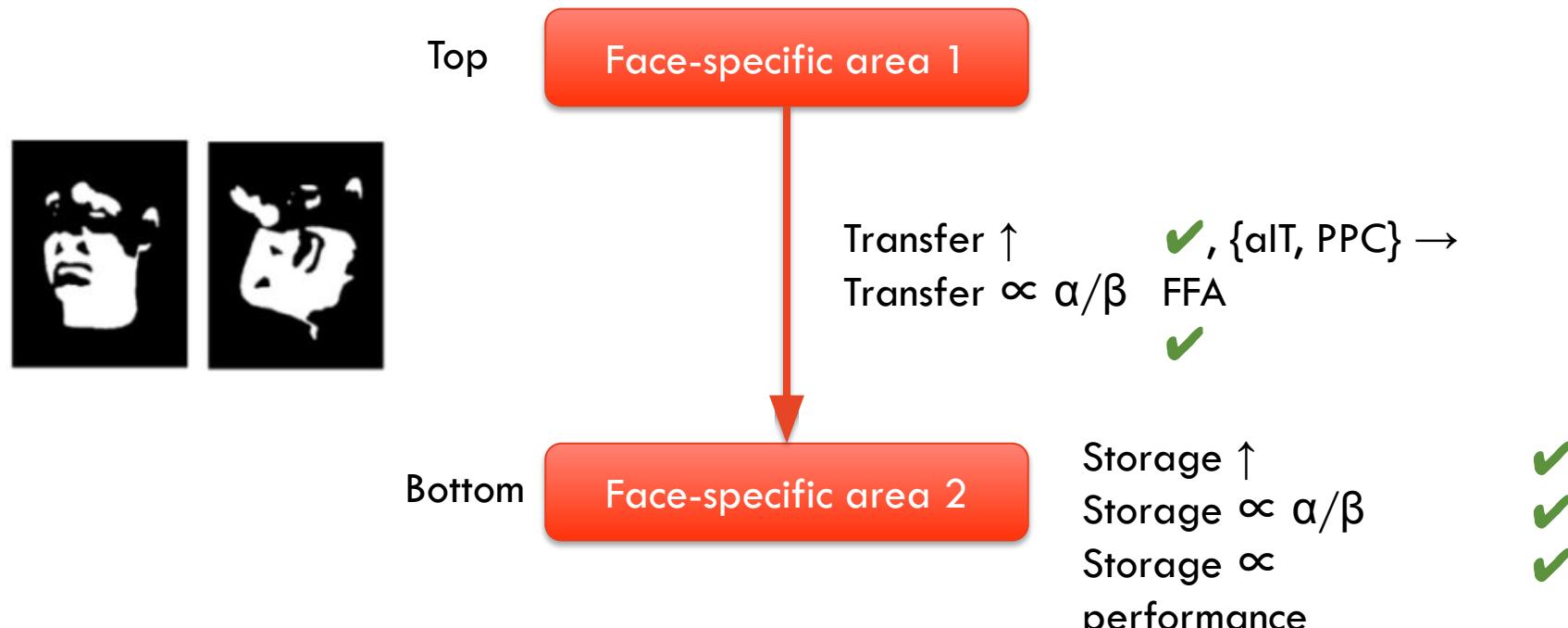
E. Crosato et al., "Informative and misinformative interactions in a school of fish", Swarm Intelligence, 12(4), pp.283-305, 2018

J.T. Lizier et al., "Spatiotemporal information transfer pattern differences in motor selection", BMC Neuroscience 12 (Suppl. 1): P261, 2011

2. Space-time characterization of info processing

Predictive coding makes conjectures regarding when and where information should be stored and transferred in layered structure.

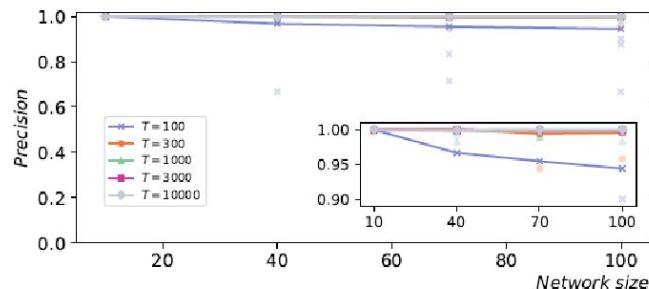
In our MEG recordings from a Mooney face/house detection experiment (52 subjects, 960 trials), suggests that when priming for a face:



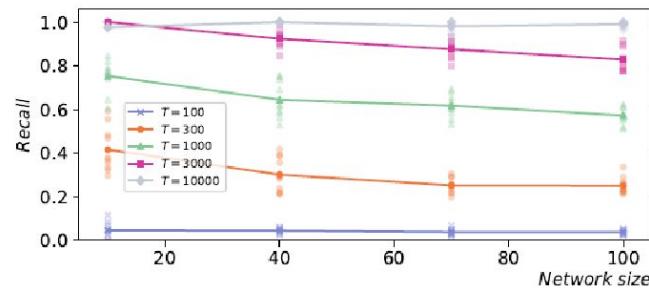
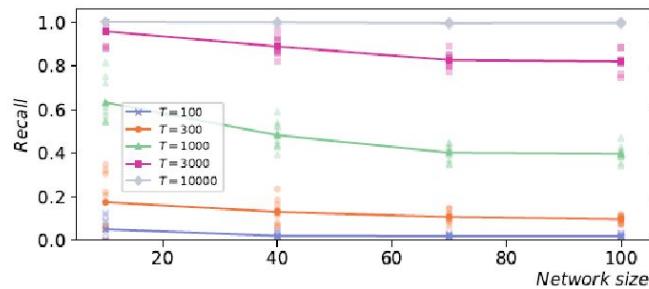
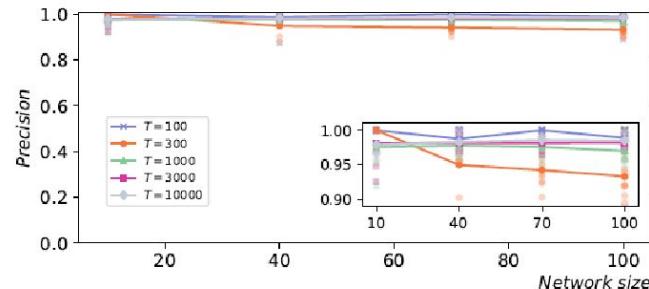
A. Brodski-Guerniero, et al., "Information theoretic evidence for predictive coding in the face processing system". J. Neuroscience, 37(34):8273–8283, 2017

2. Effective network modelling

Vector auto-regressive



Coupled logistic maps



Group differences

