PRNI 2017
21 June 2017

An introduction to the different causal frameworks in neuroimaging

Sebastian Weichwald
Max Planck Institute for Intelligent Systems,
Max Planck ETH Center for Learning Systems

sweichwald.de/prni2017 neural.engineering

Why causality?

To paraphrase a old joke, there are two types of statisticians: those who do causal inference and those who lie about it.

(L Wasserman, *Journal of the American Statistical Association*, 1999)

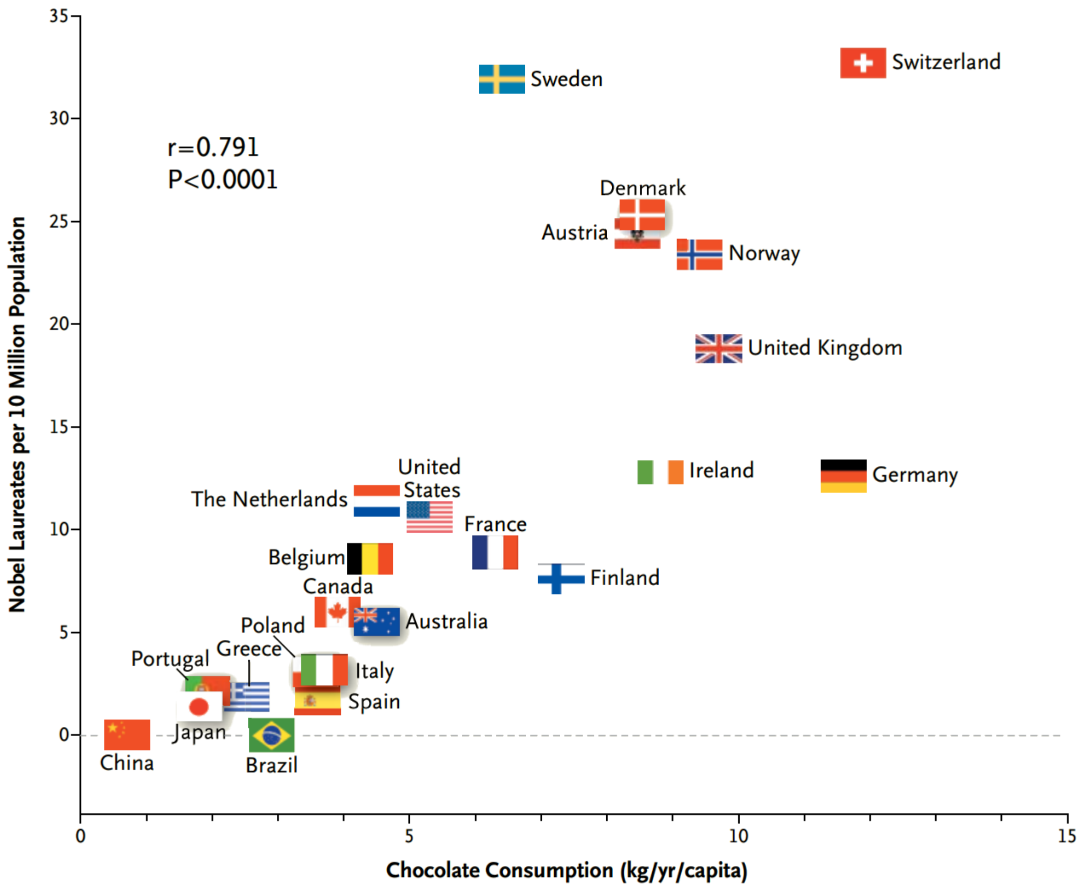
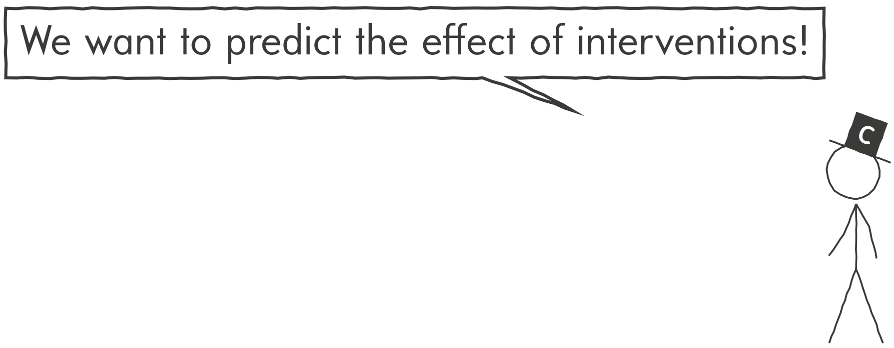


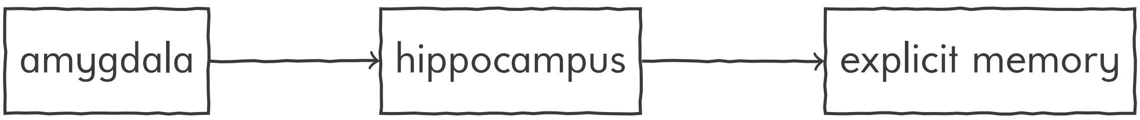
Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

(FH Messerli, Chocolate Consumption, Cognitive Function, and Nobel Laureates. *New England Journal of Medicine*, 2012)

A scientific theory should

- Explain already observed data
- Predict future observations
  - of a *passively* observed system
  - of a system that is *actively* intervened upon





Hippocampal activity in this study was correlated with amygdala activity, supporting the view that the amygdala ***enhances*** *explicit memory by* ***modulating*** *activity in the hippocampus.*

(Anonymous Authors, *Trends in Cognitive Sciences,* 2001)

Common causal frameworks

Common causal frameworks

- Potential Outcomes Framework
- Granger Causality
- Dynamic Causal Modelling
- Causal Bayesian Networks and Structural Equation Models

Potential Outcomes Framework

Ingredients:

- Population $\mathcal{U}$ of units $u \in \mathcal{U}$,

  e. g. a patient group

- Treatment variable $S : \mathcal{U} \to \{\mathrm{t}, \mathrm{c}\}$,

  e. g. assignment to treatment/control

- Potential outcomes $Y : \mathcal{U} \times \{\mathrm{t}, \mathrm{c}\} \to \mathbb{R}$,

  e. g. survival times $Y_{\mathrm{t}}(u)$ and $Y_{\mathrm{c}}(u)$ of patient u

(PW Holland, Statistics and Causal Inference. *Journal of the American Statistical Association,* 1986)

Fundamental problem of causal inference:
For each unit u we get to observe *either* $Y_{\mathrm{t}}(u)$ *or* $Y_{\mathrm{c}}(u)$ and hence the treatment effect $Y_{\mathrm{t}}(u) - Y_{\mathrm{c}}(u)$ cannot be computed.

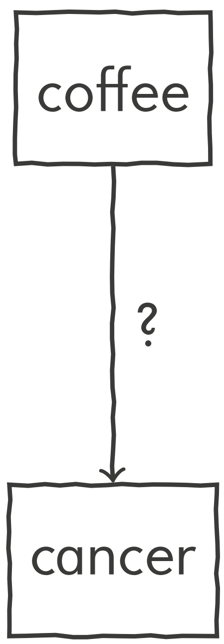
Possible remedy assumptions:

- Unit homogeneity: $Y_{\mathrm{t}}(u_1) = Y_{\mathrm{t}}(u_2)$ and $Y_{\mathrm{c}}(u_1) = Y_{\mathrm{c}}(u_2)$
- Causal transience: can measure $Y_{\mathrm{t}}(u)$ and $Y_{\mathrm{c}}(u)$ sequentially

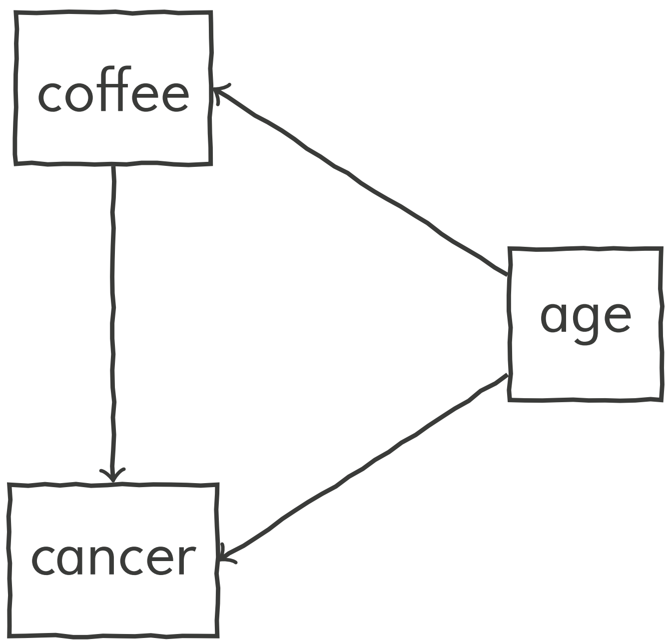
"Statistical solution": Average Treatment Effect $\mathbb{E}[Y_{\mathrm{t}}] - \mathbb{E}[Y_{\mathrm{c}}]$

- Can observe $\mathbb{E}[Y_{\mathrm{t}}|S = \mathrm{t}]$ and $\mathbb{E}[Y_{\mathrm{c}}|S = \mathrm{c}]$
- which, when randomly assigning treatments, i. e. $(Y_{\mathrm{t}}, Y_{\mathrm{c}}) \perp\!\!\!\perp S$,
- is equal to $\mathbb{E}[Y_{\mathrm{t}}]$ and $\mathbb{E}[Y_{\mathrm{c}}]$.

(PW Holland, Statistics and Causal Inference. *Journal of the American Statistical Association,* 1986)



- Split population $\mathcal{U}$ into
  - 'consumed little': $S(u) = \square$
  - 'consumed lots': $S(u) = \blacksquare$
- Observe whether they suffer from cancer or not, $Y \in \{0, 1\}$
- Assume older units have higher cumulative coffee consumption as well as an increased risk of cancer



- Split population $\mathcal{U}$ into
  - 'consumed little': $S(u) = \square$
  - 'consumed lots': $S(u) = \blacksquare$
- Observe whether they suffer from cancer or not, $Y \in \{0,1\}$
- Assume older units have higher cumulative coffee consumption as well as an increased risk of cancer
  - $(Y_\square, Y_\blacksquare) \not\perp S$
  - $\mathbb{E}[Y_\square | S = \square] < \mathbb{E}[Y_\square]$

$\Longrightarrow$ $\mathbb{E}[Y_\blacksquare] - \mathbb{E}[Y_\square]$ systematically overestimates the effect of cumulative coffee consumption on cancer

- Potential Outcomes Framework

  may work under certain (untestable) assumptions
- Granger Causality
- Dynamic Causal Modelling
- Causal Bayesian Networks and Structural Equation Models

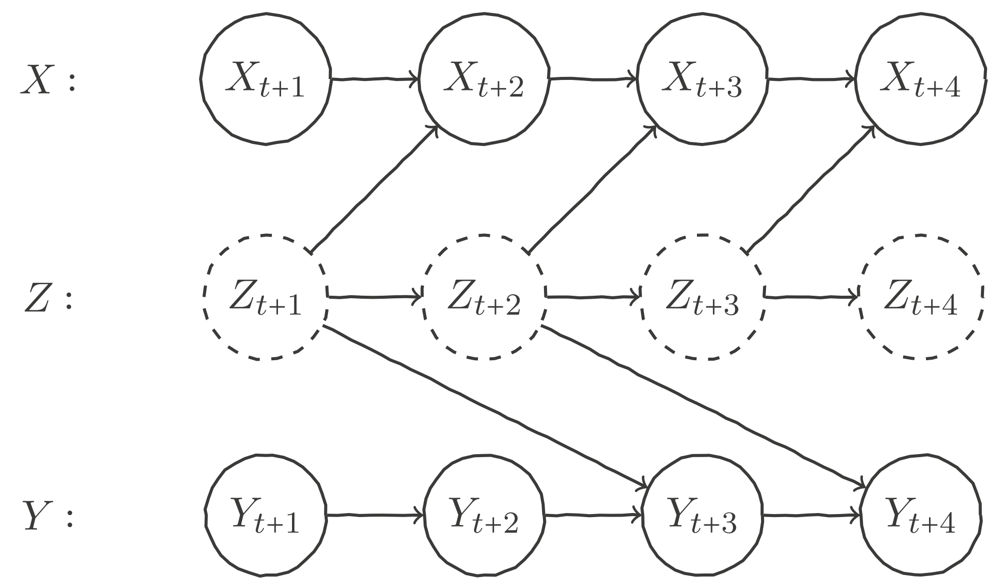
Granger Causality

Simplified Definition: One stochastic process X is causal to a second Y if the autoregressive predictability of the second process at a given time point is improved by *including* measurements from the past of the first, i. e. if

$$\mathrm{PredAcc}[Y_t | Y_{<t}] < \mathrm{PredAcc}[Y_t | Y_{<t}, X_{<t}]$$

(*not* by C Granger)

(CWJ Granger, Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica,* 1969)



$$\mathrm{PredAcc}[Y_t|Y_{<t}] < \mathrm{PredAcc}[Y_t|Y_{<t}, X_{<t}]$$

Granger causality erroneously infers causal influence from X to Y!

(J Peters et al. Causal discovery on time series using restricted structural equation models. *NIPS,* 2013)

Simplified Definition: One stochastic process X is causal to a second Y if the autoregressive predictability of the second process at a given time point is improved by *including* measurements from the past of the first, i. e. if

$$\mathrm{PredAcc}[Y_t|Y_{<t}] < \mathrm{PredAcc}[Y_t|Y_{<t}, X_{<t}]$$

(*not* by C Granger)

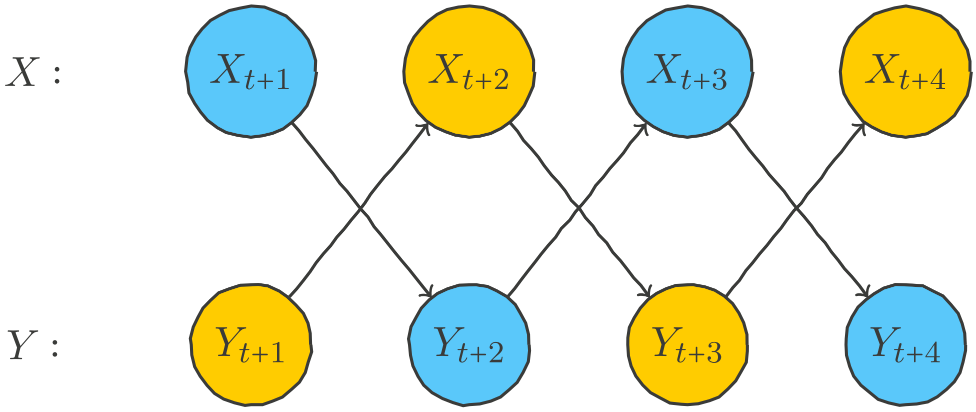
Granger's Definition: One stochastic process X is causal to a second Y if the predictability of the second process at a given time point is worsened by *removing* past measurements of the first from the universe's past, i. e. if

$$\mathrm{PredAcc}[Y_t|\text{🌍}_{<t}] > \mathrm{PredAcc}[Y_t|\text{🌍}_{<t} \setminus X_{<t}]$$

(by C Granger)

(CWJ Granger, Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica,* 1969)

Granger Causality



$$\mathrm{PredAcc}[Y_t|\text{🌍}_{<t}] = \mathrm{PredAcc}[Y_t|\text{🌍}_{<t} \setminus X_{<t}]$$

Granger causality fails to predict the effects of interventions!

(N Ay and D Polani, Information flows in causal networks. *Advances in Complex Systems*, 2008)

Common causal frameworks

- Potential Outcomes Framework

  may work under certain (untestable) assumptions
- Granger Causality

  problems with confounding

  may fail to predict effects of interventions
- Dynamic Causal Modelling
- Causal Bayesian Networks and Structural Equation Models

Dynamic Causal Modelling

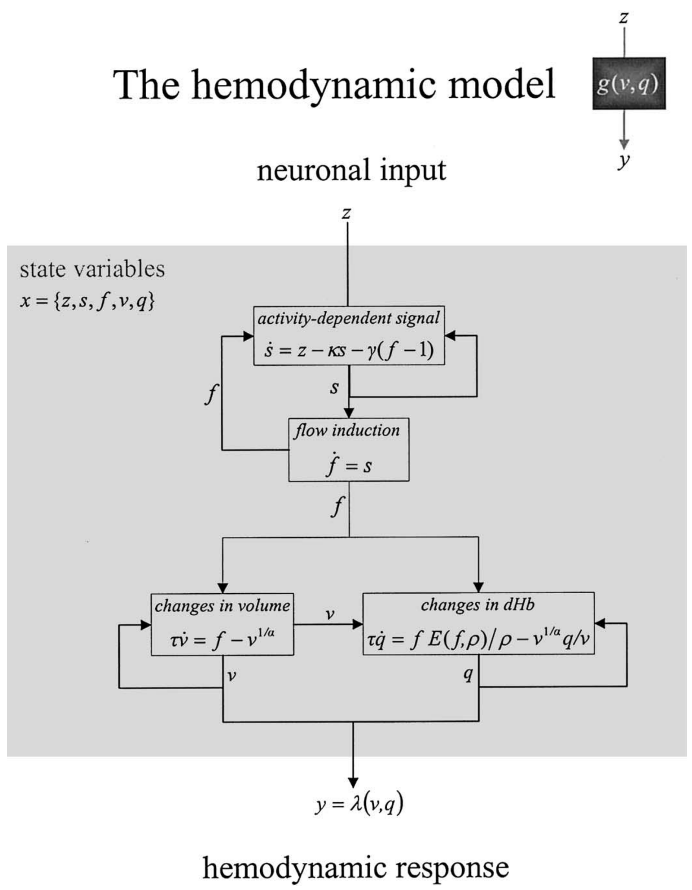
Causality in DCM is used in a control theory sense and means that, under the model, activity in one brain area causes dynamics in another, and that these dynamics cause the observations.

(Friston, *PLOS Biology,* 2009)

Inference procedure:

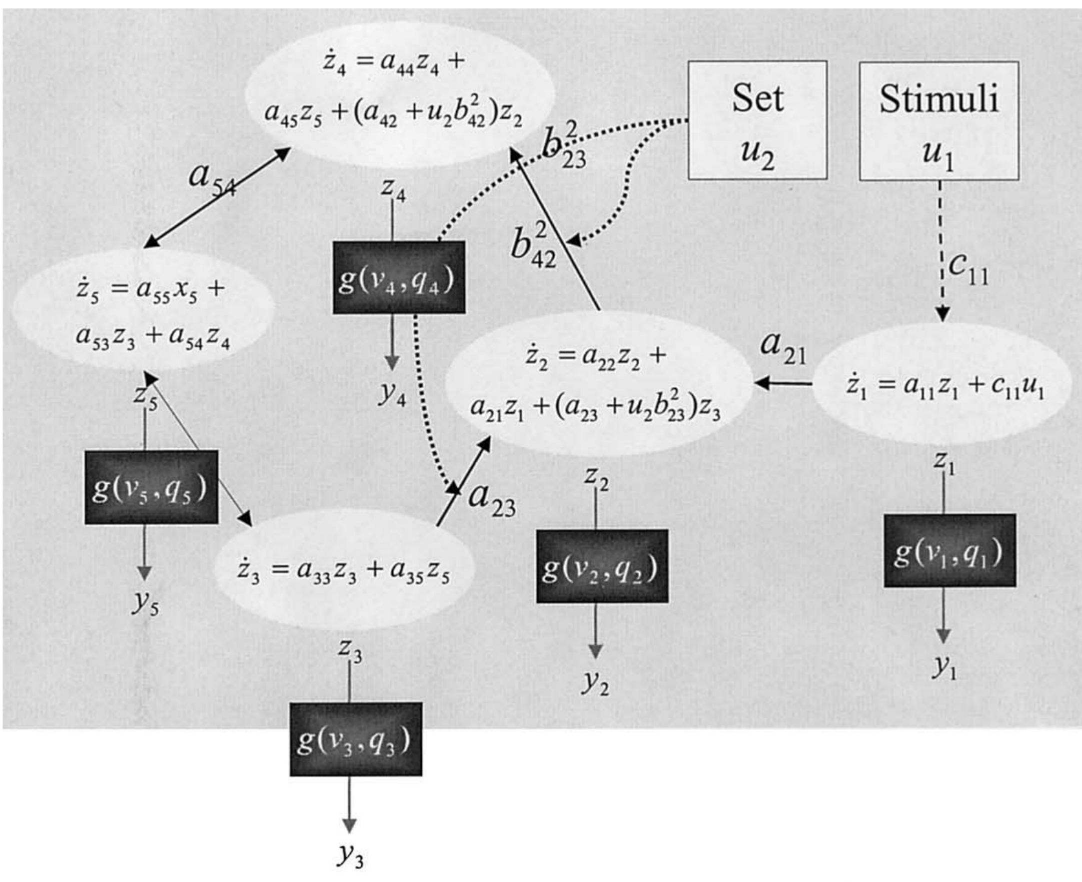
- Observe
- Define models $\mathcal{M} = \{M_1, \ldots, M_N\}$
- Fit models to observed data
- Best fitting model $\widehat{M}$ wins

(KJ Friston et al., Dynamic Causal Modelling. *NeuroImage,* 2003)

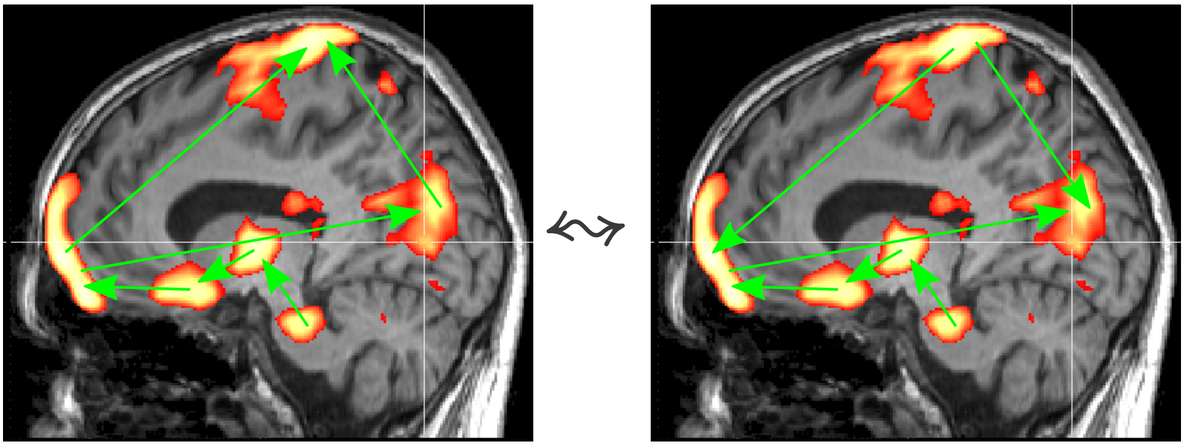


(KJ Friston et al., Dynamic Causal Modelling. *NeuroImage,* 2003)

Dynamic Causal Modelling



(KJ Friston et al., Dynamic Causal Modelling. *NeuroImage,* 2003)



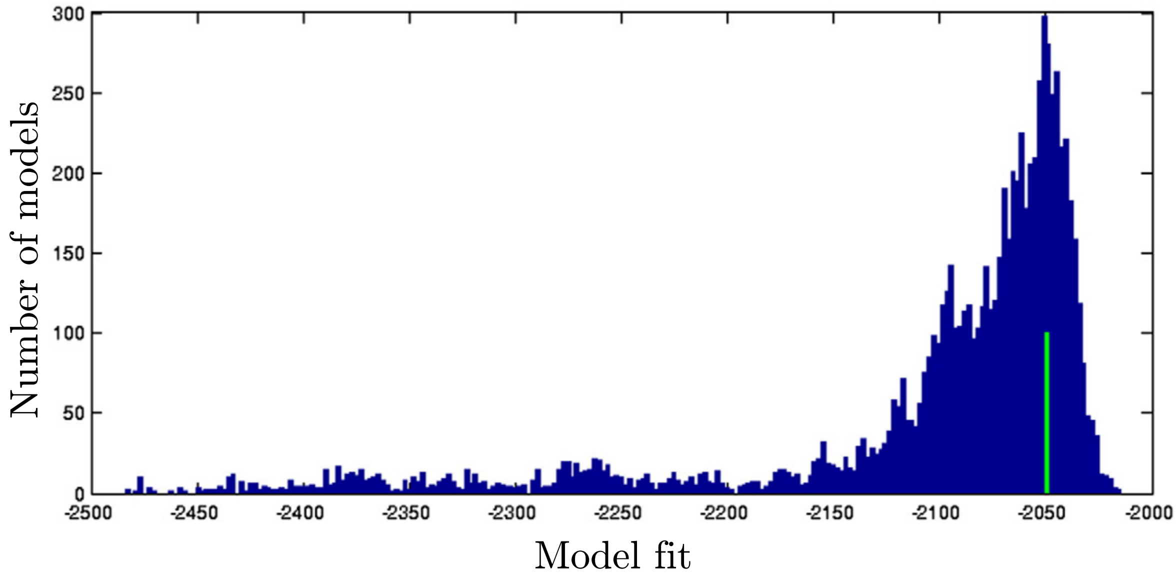
Causality in DCM is used in a control theory sense and means that, under the model, activity in one brain area causes dynamics in another, and that these dynamics cause the observations.

(Friston, *PLOS Biology,* 2009)

Inference procedure:

- Observe
- Define models $\mathcal{M} = \{M_1, \dots, M_N\}$
- Fit models to observed data
- Best fitting model $\widehat{M}$ wins

Is $\widehat{M}$ guaranteed to reflect the true connectivities?



$\Longrightarrow$ Similar model fit does not translate into similar connectivities!

(Lohmann et al., Critical comments on dynamic causal modelling. *NeuroImage,* 2012)

- Potential Outcomes Framework

  may work under certain (untestable) assumptions

- Granger Causality

  problems with confounding

  may fail to predict effects of interventions

- Dynamic Causal Modelling

  unclear how it predicts interventional setting

  inference procedure provably correct?

- Causal Bayesian Networks and Structural Equation Models

Causal Bayesian Networks and Structural Equation Models

A Structural Equation Model (SEM) $\mathcal{M}_X = (\mathcal{S}_X, \mathcal{I}_X, \mathbb{P}_{E_X})$ with

- structural equations $\mathcal{S}_X$;
- a set of interventions $\mathcal{I}_X$;
- exogenous variables distributed according to $\mathbb{P}_{E_X}$

induces distributions $\mathbb{P}_X$ over the X variables for each $i \in \mathcal{I}_X$.

(J Pearl, *Causality: Models, reasoning, and inference,* 2000; P Spirtes et al., *Causation, Prediction, and Search,* 2001)

$\mathcal{M}_X = (\mathcal{S}_X, \mathcal{I}_X, \mathbb{P}_{E_X})$

- $\mathcal{S}_X = \begin{cases} X_1 = E_1 \\ X_2 = X_1 + E_2 \end{cases}$
- $\mathcal{I}_X = \{\varnothing,\ \mathrm{do}(X_1 = 5),\ \mathrm{do}(X_2 = 3)\}$
- $E \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$

observational	intervention on X_1	intervention on X_2
$\mathbb{P}^{\varnothing}_{X_1} \sim \mathbb{N}(0,1)$	$\mathbb{P}^{\mathrm{do}(X_1=5)}_{X_1} \equiv 5$	$\mathbb{P}^{\mathrm{do}(X_2=3)}_{X_1} \sim \mathbb{N}(0,1)$
$\mathbb{P}^{\varnothing}_{X_2} \sim \mathbb{N}(0,2)$	$\mathbb{P}^{\mathrm{do}(X_1=5)}_{X_2} \sim \mathbb{N}(5,1)$	$\mathbb{P}^{\mathrm{do}(X_2=3)}_{X_2} \equiv 3$

Definition of Cause and Effect

$$X \to Y \iff \mathbb{P}^{\mathrm{do}(X=x)}_Y \neq \mathbb{P}^{\varnothing}_Y \text{ for some } x$$

Causal Markov Condition

d-separation $\rightsquigarrow$ independence

Faithfulness

d-separation $\leftsquigarrow$ independence

chain	fork	collider
$X \to Y \to Z$	$X \leftarrow Y \to Z$	$X \to Y \leftarrow Z$
$X \not\perp\!\!\!\perp Z$	$X \not\perp\!\!\!\perp Z$	$X \perp\!\!\!\perp Z$
$X \perp\!\!\!\perp Z \mid Y$	$X \perp\!\!\!\perp Z \mid Y$	$X \not\perp\!\!\!\perp Z \mid Y$

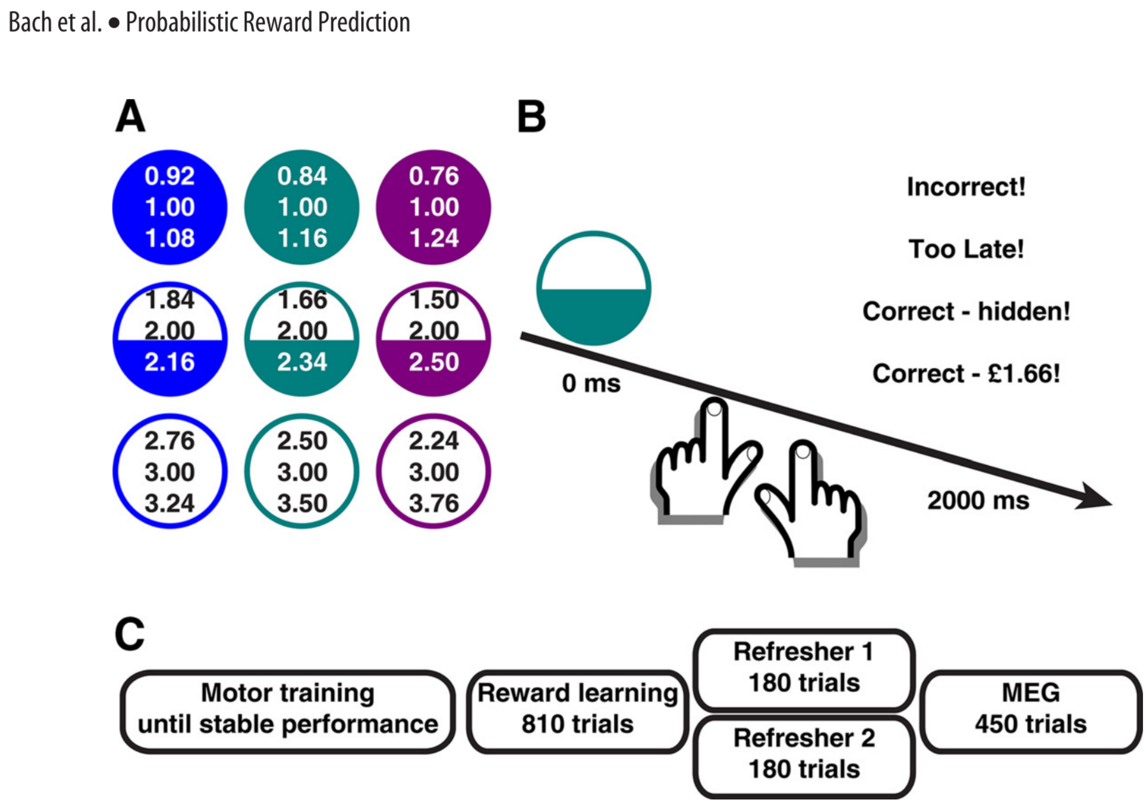
- Randomised stimulus S
- Observe neural activity X and Y

$\rightsquigarrow$ Estimate $\mathbb{P}^{\varnothing}_{S,X,Y}$

- Assume we find
  - $S \not\perp\!\!\!\perp X \implies$ existence of path between S and X w/o collider
  - $S \not\perp\!\!\!\perp Y \implies$ existence of path between S and Y w/o collider
  - $S \perp\!\!\!\perp Y | X \implies$ all paths between S and Y blocked by X
- Can rule out cases such as $S \rightarrow X \leftarrow h \rightarrow Y$
- Can formally prove that X indeed is a cause of Y

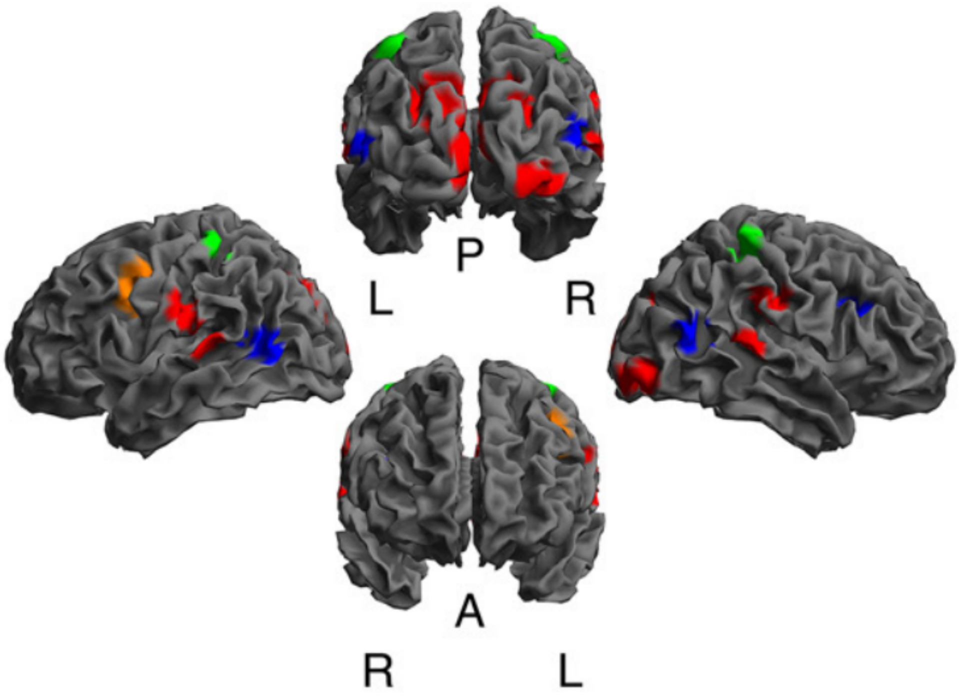
$\implies$ Robust against hidden confounding

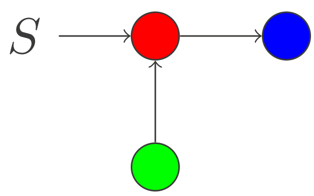
(M Grosse-Wentrup et al., *NeuroImage*, 2015; S Weichwald et al., *IEEE Journal of Selected Topics in Signal Processing*, 2016)



(Bach, Symmonds, Barnes, and Dolan. *Journal of Neuroscience*, 2017)

Application: Neural Dynamics of Probabilistic Reward Prediction





(Bach, Symmonds, Barnes, and Dolan, Whole-brain neural dynamics of probabilistic reward prediction. *Journal of Neuroscience,* 2017)

Common causal frameworks

- Potential Outcomes Framework

  may work under certain (untestable) assumptions

- Granger Causality

  problems with confounding

  may fail to predict effects of interventions

- Dynamic Causal Modelling

  unclear how it predicts interventional setting

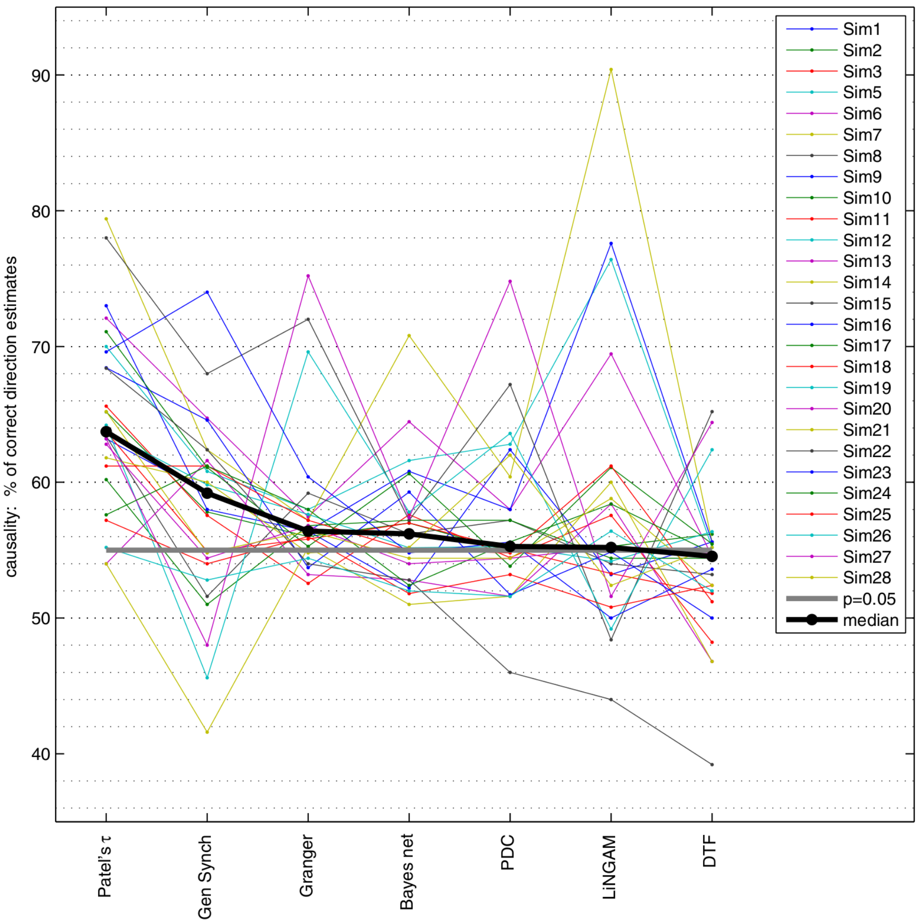
  inference procedure provably correct?

- Causal Bayesian Networks and Structural Equation Models

  may work under certain (untestable) assumptions

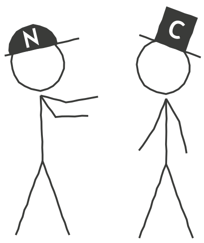
  not finding dependence is not evidence for independence

Wrap-Up



(Smith et al., Network modelling methods for fMRI. *NeuroImage,* 2011)

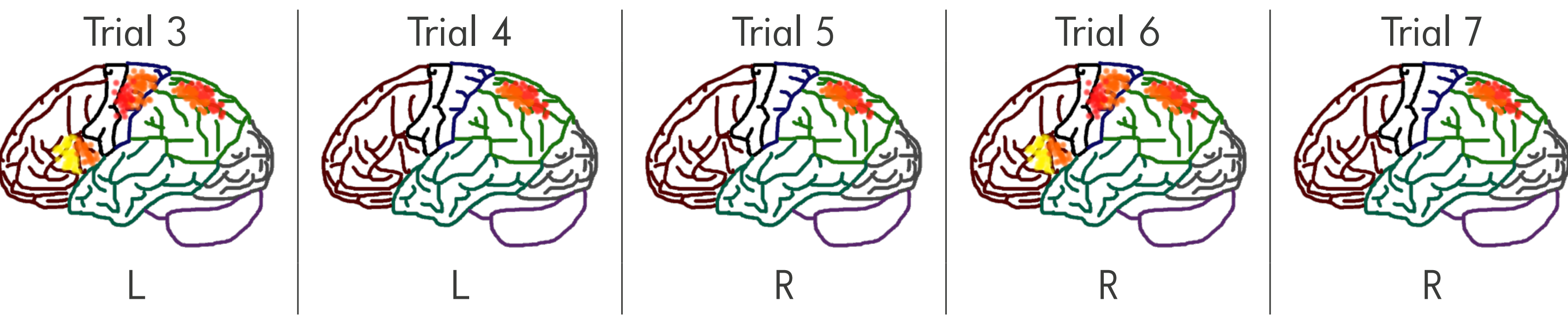
- (Causal) Inference rests on *untestable* assumptions.
- Causal inference algorithms appear to perform above chance-level.
- Causal inference may be useful to guide the design of interventional studies.



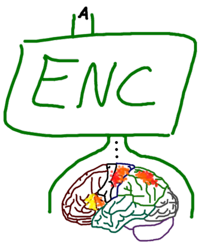
ADDENDA

Causal interpretation of encoding and decoding models

Relevance in encoding and decoding models



(S Weichwald et al., Causal interpretation rules for encoding and decoding models in neuroimaging. *NeuroImage,* 2015)



"Significant variation explained by experimental condition?"

$X_i \not\perp C$

$X_i \not\perp C | \vec{X} \setminus X_i$

"Does removal impair decoding performance?"

relevant feature $\overset{?}{\leftrightsquigarrow}$ cognitive process

$$\boxed{S} \longrightarrow \boxed{\vec{X} = \{X_1, \ldots, X_d\}} \longrightarrow \boxed{R}$$

stimulus — brain state features — response

stimulus-based		response-based
causal	encoding	*anti*-causal
anti-causal	decoding	causal

Causal interpretation chart (1)

	Feature X_i relevant?		
	Encoding	Decoding	Causal interpretation
Stimulus-based	×		no effect of S
	√		effect of S
		×	inconclusive
		√	inconclusive
Response-based	×		no cause of R
	√		inconclusive
		×	inconclusive
		√	inconclusive

(S Weichwald et al., Causal interpretation rules for encoding and decoding models in neuroimaging. *NeuroImage,* 2015)

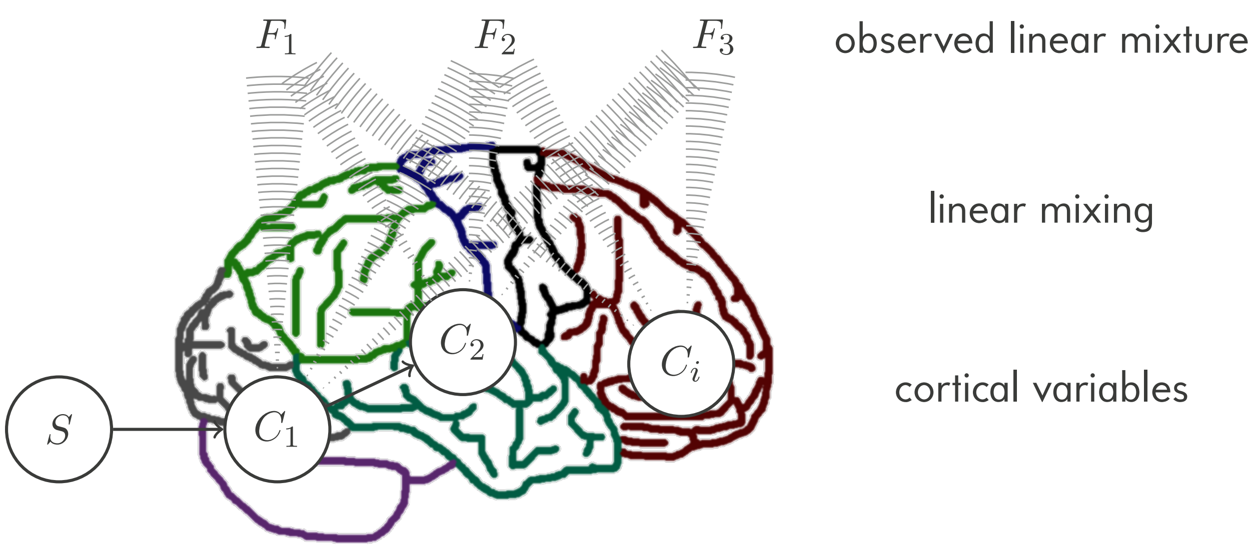
Causal interpretation chart (2)

	Feature X_i relevant?		
	Encoding	Decoding	Causal interpretation
Stimulus-based	√	√	effect of S
	√	×	indirect effect of S
	×	√	provides context
	×	×	no effect of S
Response-based	√	√	inconclusive
	√	×	no direct cause of R
	×	√	provides context
	×	×	no cause of R

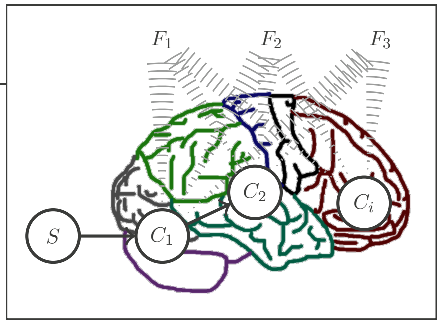
(S Weichwald et al., Causal interpretation rules for encoding and decoding models in neuroimaging. *NeuroImage,* 2015)

MERL$\star$N

Problem description



(S Weichwald et al., MERLiN: Mixture Effect Recovery in Linear Networks. *IEEE Journal of Selected Topics in Signal Processing,* 2016)



Given

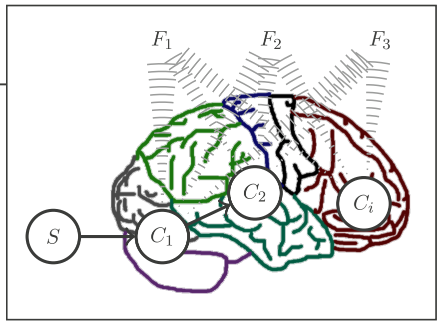
samples of S, C_1 and F

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_d \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} C_1 \\ \vdots \\ C_d \end{bmatrix} = \boldsymbol{A}C$$

Goal

find linear combination $\boldsymbol{w}$ such that

$$S \longrightarrow C_1 \longrightarrow \underbrace{\boldsymbol{w}^\top F}_{= C_2}$$



Idea

Optimise $\boldsymbol{w}$ such that

(a) $\mathrm{dep}\,(C_1, \boldsymbol{w}^\top F)$ is high

(b) $\mathrm{dep}\,(S,\ \boldsymbol{w}^\top F \,|\, C_1)$ is low

Implementation

Optimise $\boldsymbol{w}$ and σ, θ such that

$$\mathrm{HSIC}\,(C_1, \boldsymbol{w}^\top F) \quad \text{is high}$$
$$-\ \mathrm{HSIC}\Big(\ \boldsymbol{w}^\top F - \mathrm{krr}_{\sigma,\theta}(C_1)\ ,\ (S, C_1)\ \Big) \quad \text{is low}$$

is being maximised.