LIINC, Columbia University
October 2, 2015

Causal interpretation rules for encoding and decoding models in neuroimaging

Sebastian Weichwald, Timm Meyer, Ozan Özdenizci[§],
Bernhard Schölkopf, Tonio Ball[‡], Moritz Grosse-Wentrup
MPI for Intelligent Systems, [§]Sabanci University, [‡]University of Freiburg

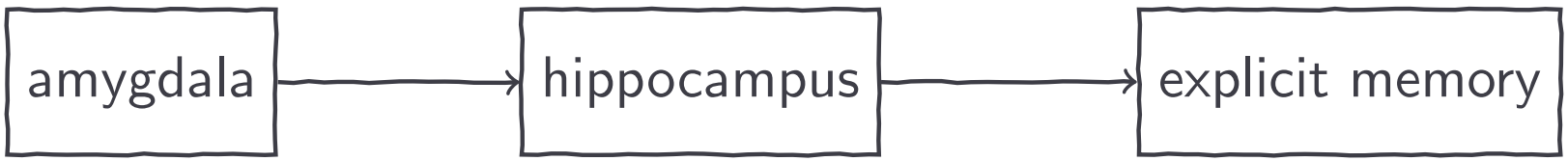
---

sweichwald.de/neuroimage2015  brain-computer-interfaces.net

Motivation

We tested [...] whether pre-stimulus alpha oscillations measured with electroencephalography (EEG) **influence** *the encoding of items into working memory.*

(Myers et al., *Journal of Neuroscience*, 2014)

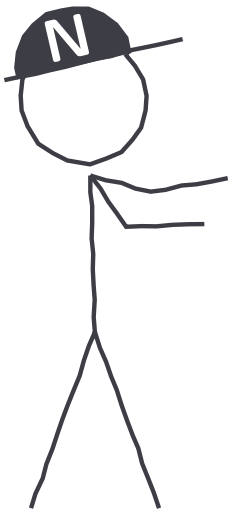


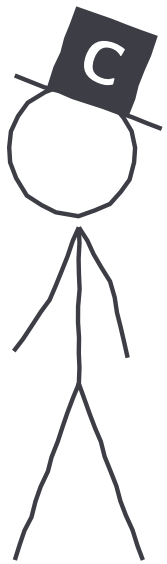
Hippocampal activity in this study was correlated with amygdala activity, supporting the view that the amygdala **enhances** *explicit memory by* **modulating** *activity in the hippocampus.*

(S. Hamann, *Trends in Cognitive Sciences*, 2001)

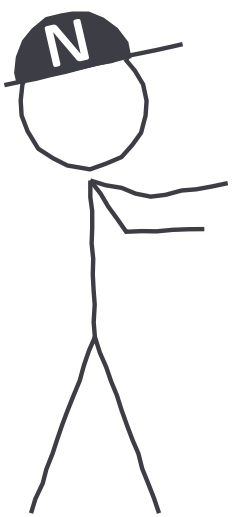
1. Motivation
2. Approach
3. Encoding and decoding models in neuroimaging
4. Causal Bayesian Networks
5. Causal interpretation of encoding and decoding models
6. Empirical example
7. Wrap-up

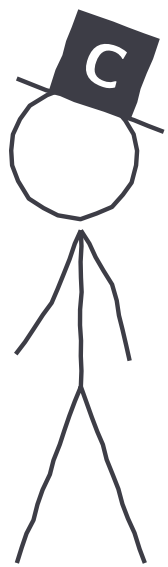
Approach


N


C

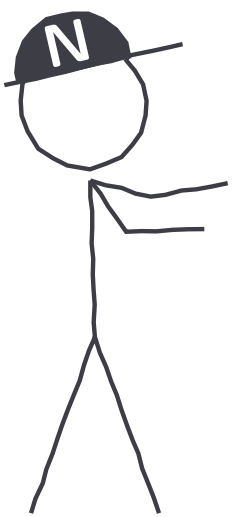
I'm interested in how neural activity gives rise to cognition.

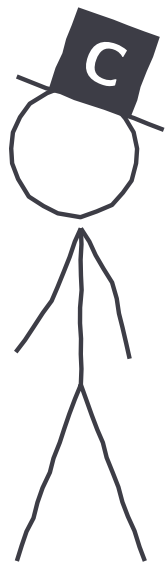




I'm interested in how neural activity gives rise to cognition.

That sounds intriguing! So, what do you do?

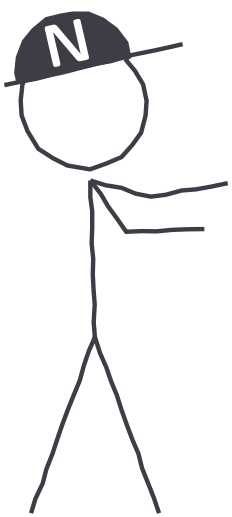


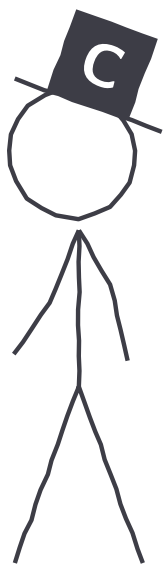


I'm interested in how neural activity gives rise to cognition.

That sounds intriguing! So, what do you do?

I present stimuli to subjects or observe their behaviour while recording their brain activity. ...[explains common analysis methods]...



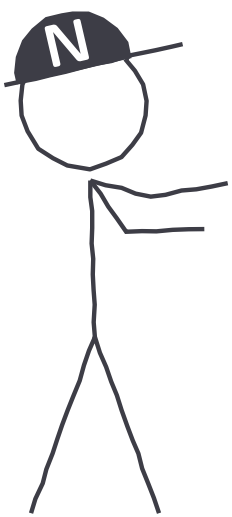


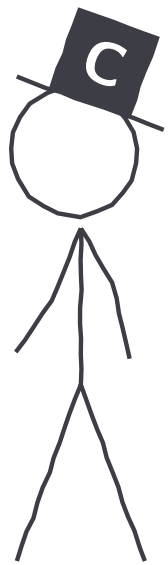
I'm interested in how neural activity gives rise to cognition.

That sounds intriguing! So, what do you do?

I present stimuli to subjects or observe their behaviour while recording their brain activity. ...[explains common analysis methods]...

Ha, interesting! ...[rephrases what has been said in causal inference slang]...





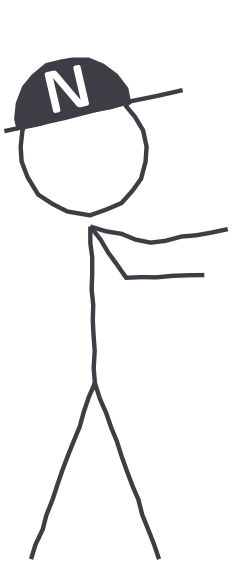
I'm interested in how neural activity gives rise to cognition.

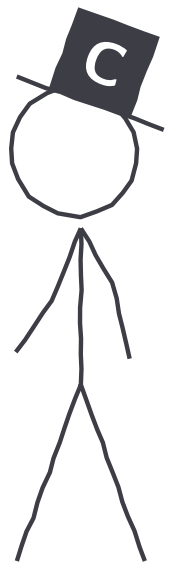
That sounds intriguing! So, what do you do?

I present stimuli to subjects or observe their behaviour while recording their brain activity. ...[explains common analysis methods]...

Ha, interesting! ...[rephrases what has been said in causal inference slang]...

Yeah, a solid basis for our interpretations! It also clarifies problems that we recently discussed in the community.





Encoding and decoding models in neuroimaging

Trial 3

L

Trial 4

L

Trial 5

R

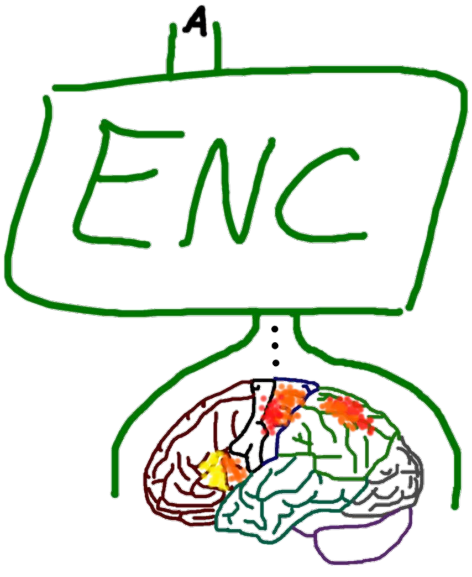
Trial 6

R

Trial

R

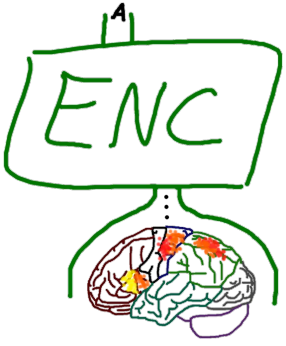
encoding



e.g. mean difference between conditions

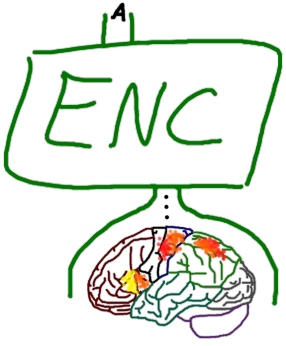
decoding

e.g. classifier for experimental conditions



"Significant variation across experimental conditions?"

"Does removal impair decoding performance?"



"Significant variation across experimental conditions?"

"Does removal impair decoding performance?"

?

relevant feature ⇝ cognitive process

Causal Bayesian Networks

- $X \to Y \iff p(Y) \neq p(Y|do\{X = x\})$
- Causal Markov Condition: d-separation $\rightsquigarrow$ independence
- Faithfulness: d-separation $\leftsquigarrow$ independence

chain	fork	collider
$X \to Y \to Z$	$X \leftarrow Y \to Z$	$X \to Y \leftarrow Z$

- $X \to Y \iff p(Y) \neq p(Y|do\{X = x\})$
- Causal Markov Condition: d-separation $\rightsquigarrow$ independence
- Faithfulness: d-separation $\leftsquigarrow$ independence

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

- $X \to Y \iff p(Y) \neq p(Y|do\{X = x\})$
- Causal Markov Condition: d-separation $\rightsquigarrow$ independence
- Faithfulness: d-separation $\leftsquigarrow$ independence

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

- $X \to Y \iff p(Y) \neq p(Y|do\{X = x\})$
- Causal Markov Condition: d-separation $\rightsquigarrow$ independence
- Faithfulness: d-separation $\leftsquigarrow$ independence

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

- $X \to Y \iff p(Y) \neq p(Y|do\{X = x\})$
- Causal Markov Condition: d-separation $\rightsquigarrow$ independence
- Faithfulness: d-separation $\leftsquigarrow$ independence

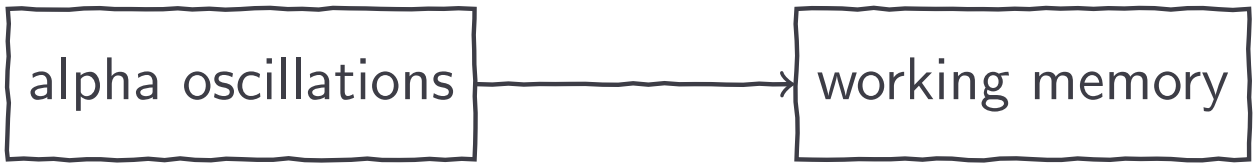
chain	fork	collider
$X \to Y \to Z$	$X \leftarrow Y \to Z$	$X \to Y \leftarrow Z$
$X \not\perp\!\!\!\perp Z$	$X \not\perp\!\!\!\perp Z$	
$X \perp\!\!\!\perp Z \mid Y$	$X \perp\!\!\!\perp Z \mid Y$	

- $X \to Y \iff p(Y) \neq p(Y|do\{X = x\})$
- Causal Markov Condition: d-separation $\rightsquigarrow$ independence
- Faithfulness: d-separation $\leftsquigarrow$ independence

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

- $X \to Y \iff p(Y) \neq p(Y|do\{X = x\})$
- Causal Markov Condition: d-separation $\rightsquigarrow$ independence
- Faithfulness: d-separation $\leftsquigarrow$ independence

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



We tested [...] whether pre-stimulus alpha oscillations measured with electroencephalography (EEG) **influence** *the encoding of items into working memory.*

(Myers et al., *Journal of Neuroscience*, 2014)

We tested [...] whether pre-stimulus alpha oscillations measured with electroencephalography (EEG) **influence** *the encoding of items into working memory.*

(Myers et al., *Journal of Neuroscience*, 2014)

$\alpha \not\perp \text{WM}$

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.*

(S. Hamann, *Trends in Cognitive Sciences*, 2001)

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.*

(S. Hamann, *Trends in Cognitive Sciences*, 2001)

$$\text{AM} \not\perp\!\!\!\perp \text{EM}$$
$$\text{AM} \perp\!\!\!\perp \text{EM} \mid \text{HC}$$

Causal interpretation of encoding and decoding models

Let's set out the causal component of already performed analyses..

Let's set out the causal component of already performed analyses..

stimulus- vs response-based

feature relevance $\nsim$ marginal/conditional dependence

$\leadsto$ 16 causal interpretation rules

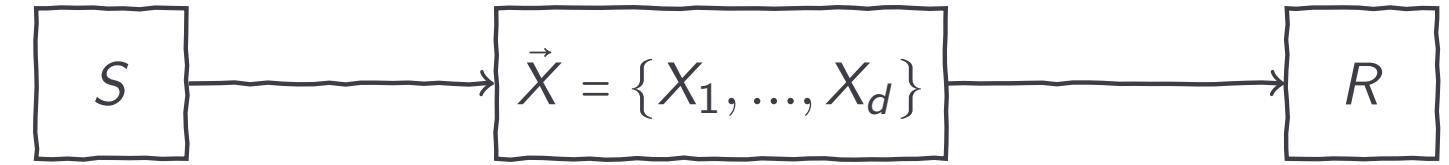
Let's set out the causal component of already performed analyses..

stimulus- vs response-based

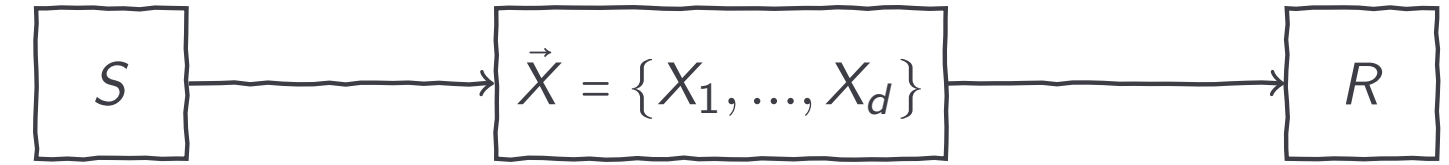
feature relevance $\nsim$ marginal/conditional dependence

$\rightsquigarrow$ 16 causal interpretation rules

simple

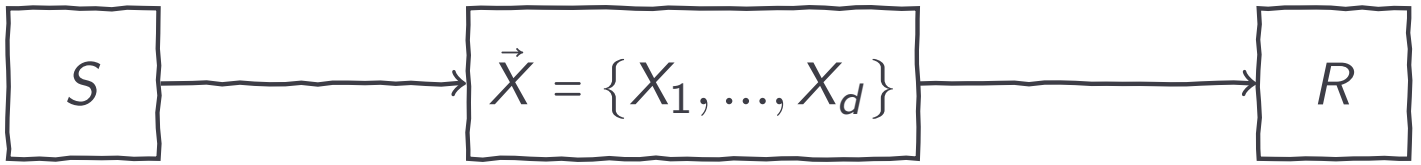


stimulus brain state features response



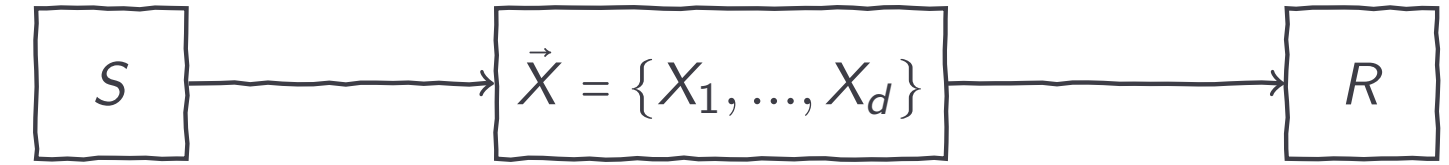
stimulus | brain state features | response

stimulus-based		response-based
$p(\vec{X}\|S)$	encoding	$p(\vec{X}\|R)$
$p(S\|\vec{X})$	decoding	$p(R\|\vec{X})$



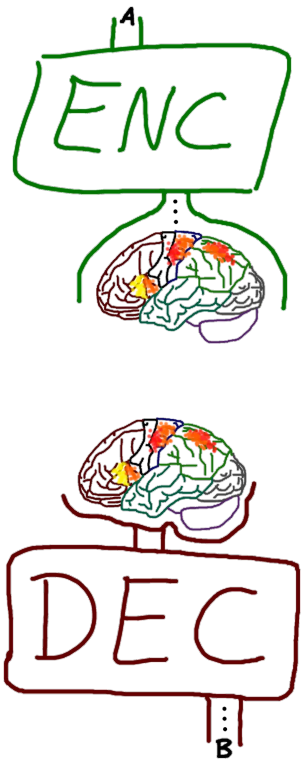
stimulus brain state features response

stimulus-based			response-based	
$p(\vec{X}\|S)$	causal	encoding		$p(\vec{X}\|R)$
$p(S\|\vec{X})$		decoding	causal	$p(R\|\vec{X})$

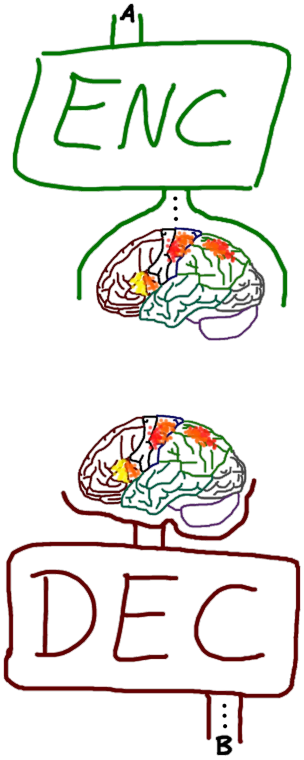


stimulus brain state features response

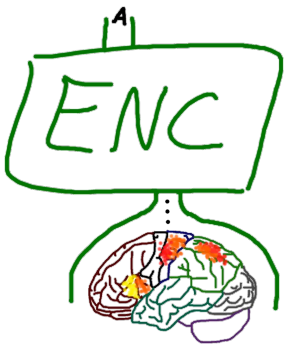
	stimulus-based		response-based	
$p(\vec{X}\|S)$	causal	encoding	*anti*-causal	$p(\vec{X}\|R)$
$p(S\|\vec{X})$	*anti*-causal	decoding	causal	$p(R\|\vec{X})$


A
ENC
DEC
B

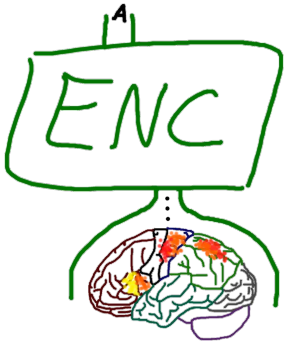

C



$$p(X_i|C = c_1) \stackrel{?}{\neq} p(X_i|C = c_2)$$

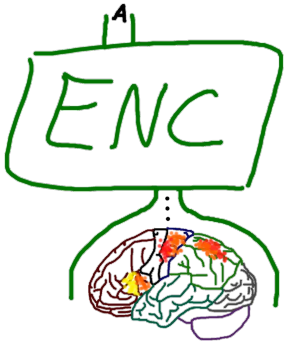


$$p(X_i|C = c_1) \stackrel{?}{\neq} p(X_i|C = c_2) \qquad X_i \not\perp C$$



$$p(X_i|C = c_1) \stackrel{?}{\neq} p(X_i|C = c_2) \qquad X_i \not\perp C$$

$$p(C|\vec{X}) \stackrel{?}{\neq} p(C|\vec{X} \setminus X_i)$$



$$p(X_i|C = c_1) \stackrel{?}{\neq} p(X_i|C = c_2) \qquad X_i \not\perp C$$

$$p(C|\vec{X}) \stackrel{?}{\neq} p(C|\vec{X} \setminus X_i) \qquad X_i \not\perp C|\vec{X} \setminus X_i$$

Causal interpretation rules (1)

	Feature X_i relevant?		
	Encoding	Decoding	Causal interpretation
Stimulus-based	×		
	√		
		×	
		√	
Response-based	×		
	√		
		×	
		√	

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

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

	Feature X_i relevant?		
	Encoding	Decoding	Causal interpretation
Stimulus-based	×		no effect of S
	√		effect of S
		×	$X_i \perp\!\!\!\perp S \vert \vec{X} \setminus X_i$
		√	$S \rightarrow X_j \rightarrow X_i$
Response-based	×		$S \quad X_i$
	√		
		×	
		√	

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

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

	Feature X_i relevant?		
	Encoding	Decoding	Causal interpretation
Stimulus-based	×		no effect of S
	√		effect of S
		×	inconclusive
		√	$S \rightarrow X_i$
Response-based	×		$S \rightarrow X_j \leftarrow X_i$
	√		
		×	
		√	

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

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

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

	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
	√		
		×	
		√	

	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
	√		
	×		
	√		

$$X_i \not\perp R$$

$$X_i \leftarrow h \rightarrow R$$

$$X_i \rightarrow R$$

example (A)
$\alpha \rightarrow$ WM

	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
	√		
		×	
		√	

	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
		×	
		✓	

	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
		×	
		✓	

$$X_i \perp\!\!\!\perp R | \vec{X} \setminus X_i$$

$$X_i \rightarrow X_j \rightarrow R$$

$$X_i \qquad R$$

	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
		×	
		✓	

	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
		√	

	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
		√	

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

$$X_i \rightarrow R$$

$$X_i \leftarrow h \rightarrow R$$

$$\downarrow$$

$$X_j$$

	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
		√	

	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

	Feature X_j 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

Are decoding models useful at all?

	Feature X_i relevant?		
	Encoding	Decoding	Causal interpretation
Stimulus-based	×	×	
	√	×	
	×	√	
	√	√	
Response-based	×	×	
	√	×	
	×	√	
	√	√	

	Feature X_i relevant?		
	Encoding	Decoding	Causal interpretation
Stimulus-based	×	×	
	√	×	
	×	√	
	√	√	
Response-based	×	×	
	√	×	
	×	√	
	√	√	

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

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

$$X_i \not\perp\!\!\!\perp S \quad \text{and} \quad X_i \perp\!\!\!\perp S | \vec{X} \setminus X_i$$

$$S \dashrightarrow X_i \quad \text{indirectly}$$

e.g. $S \rightarrow X_j \rightarrow X_i$

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

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

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

$$X_i \not\perp\!\!\!\perp R \quad \text{and} \quad X_i \perp\!\!\!\perp R | \vec{X} \setminus X_i$$

$$X_i \rightarrow X_j \rightarrow R$$

$$X_i \leftarrow X_j \rightarrow R$$

example (B)

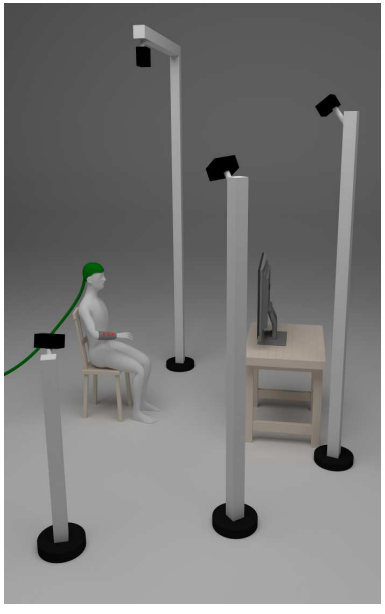
AM $\rightarrow$ HC $\rightarrow$ EM

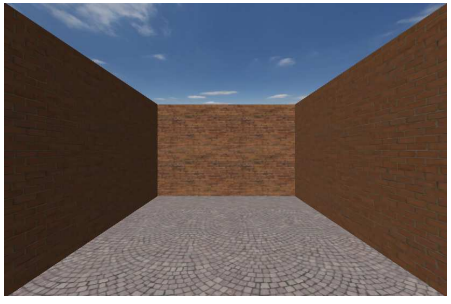
	Feature X_i relevant?		
	Encoding	Decoding	Causal interpretation
Stimulus-based	×	×	no effect of S
	√	×	indirect effect of S
	×	√	provides context
	√	√	effect of S
Response-based	×	×	no cause of R
	√	×	
	×	√	provides context
	√	√	inconclusive

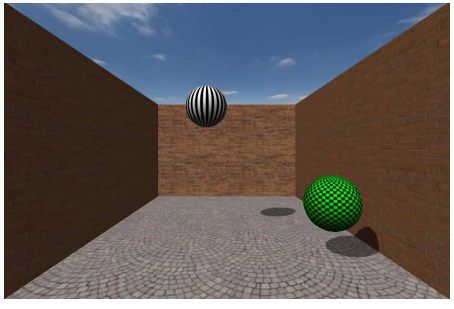
	Feature X_i relevant?		
	Encoding	Decoding	Causal interpretation
Stimulus-based	×	×	no effect of S
	√	×	indirect effect of S
	×	√	provides context
	√	√	effect of S
Response-based	×	×	no cause of R
	√	×	no direct cause of R
	×	√	provides context
	√	√	inconclusive

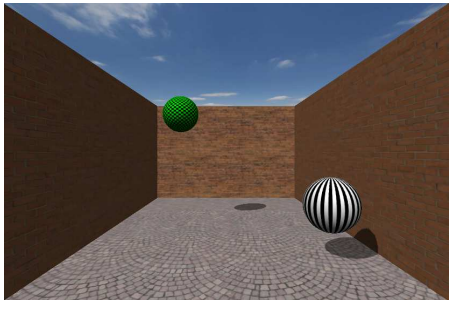
Empirical example

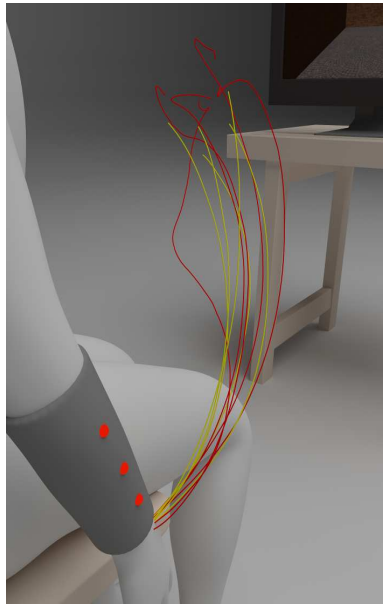
Visuo-motor learning task



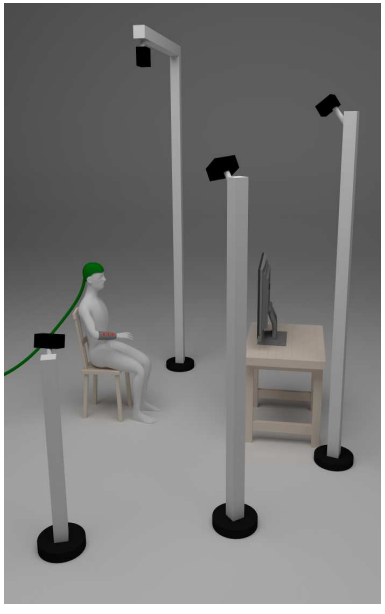


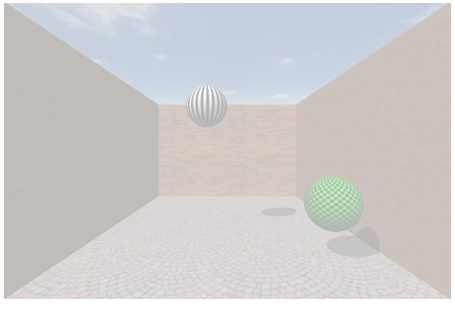


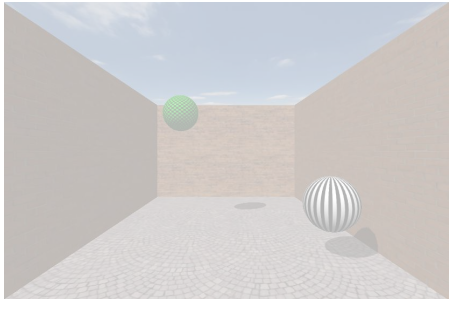


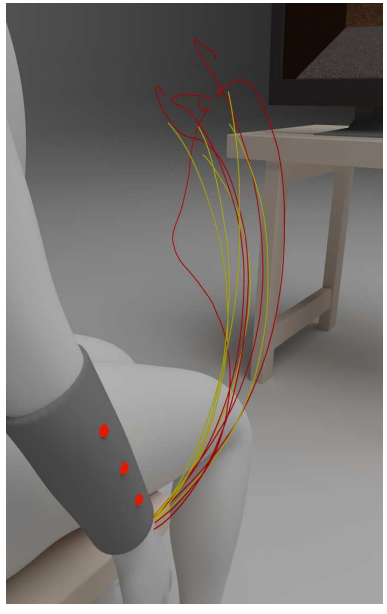


Visuo-motor learning task

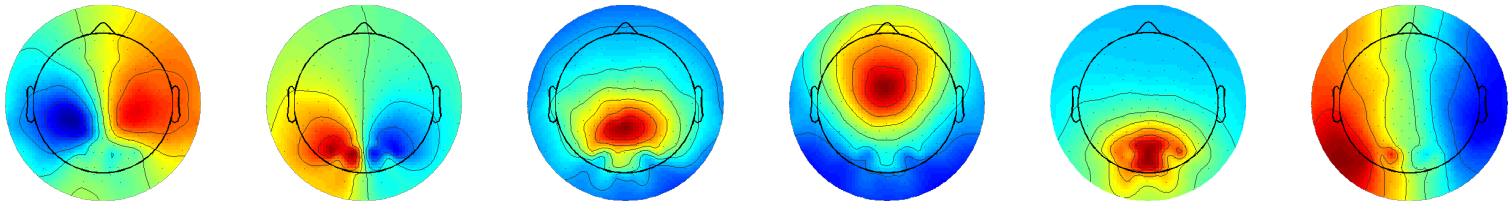




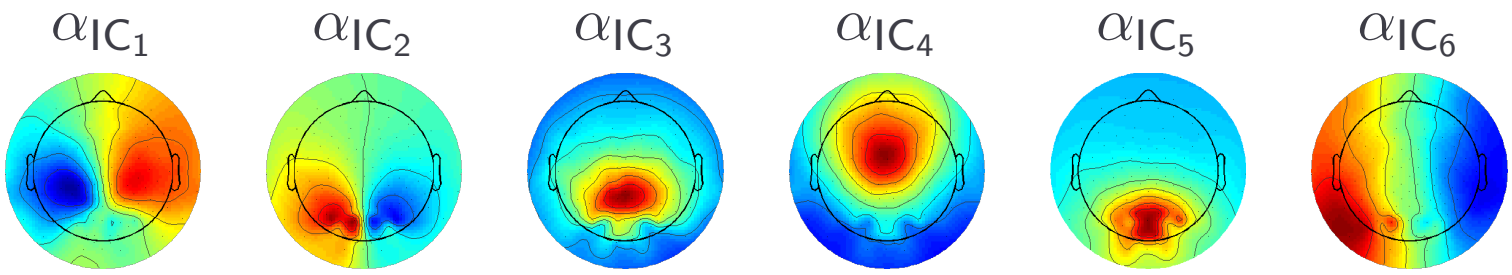


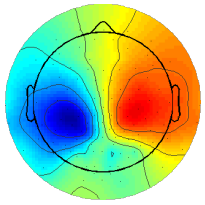
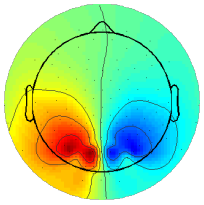
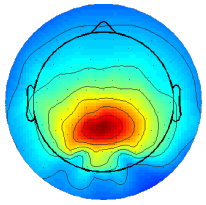
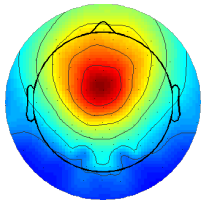
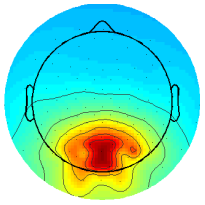
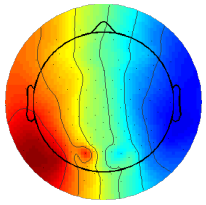


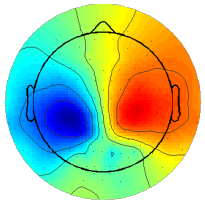
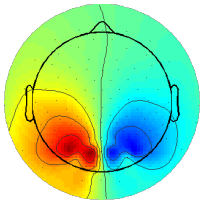
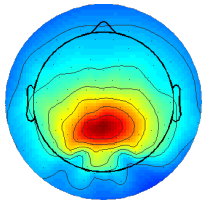
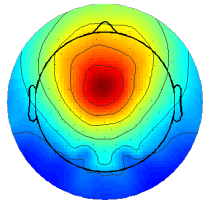
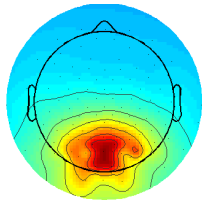
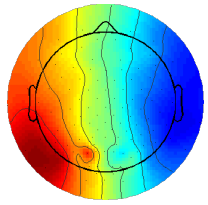
(Relevant) Features

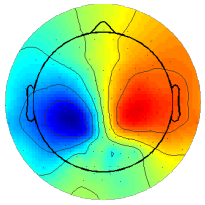
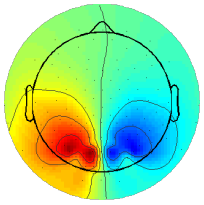
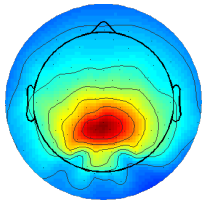
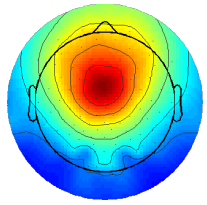
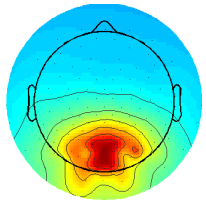
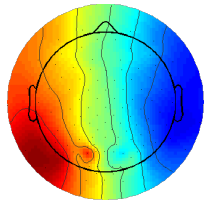


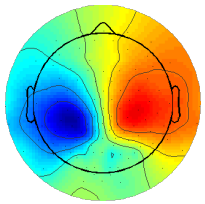
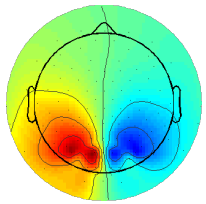
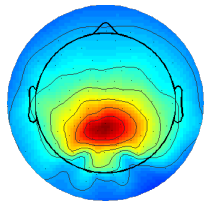
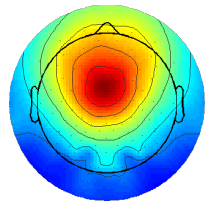
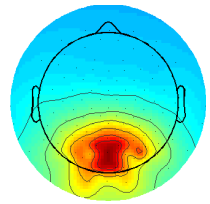
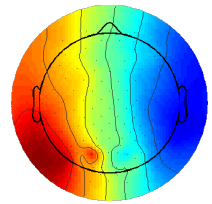
(Relevant) Features

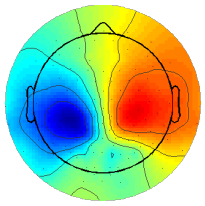
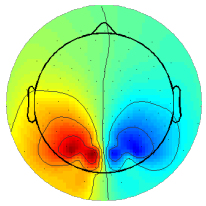
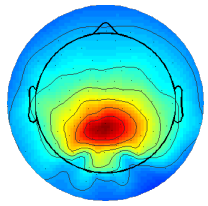
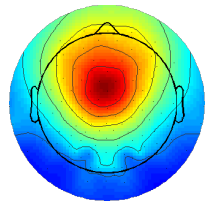
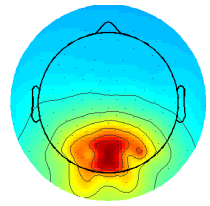
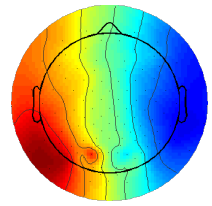


	α_{IC_1}	α_{IC_2}	α_{IC_3}	α_{IC_4}	α_{IC_5}	α_{IC_6}
p-values						
Encoding						
Decoding						

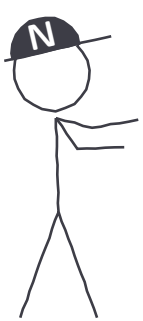
p-values	α_{IC_1} 	α_{IC_2} 	α_{IC_3} 	α_{IC_4} 	α_{IC_5} 	α_{IC_6} 
Encoding	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
Decoding	10^{-3}	10^{-3}	0.50	0.34	0.79	0.13

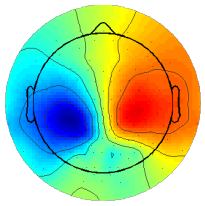
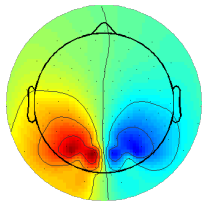
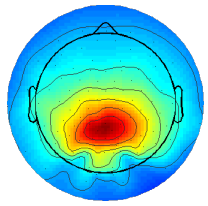
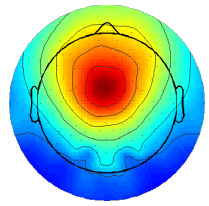
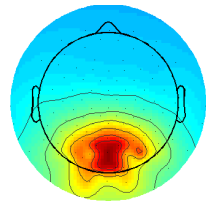
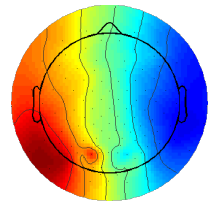
p-values	α_{IC_1} 	α_{IC_2} 	α_{IC_3} 	α_{IC_4} 	α_{IC_5} 	α_{IC_6} 
Encoding	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
Decoding	10^{-3}	10^{-3}	0.50	0.34	0.79	0.13

p-values	α_{IC_1} 	α_{IC_2} 	α_{IC_3} 	α_{IC_4} 	α_{IC_5} 	α_{IC_6} 
Encoding	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
Decoding	10^{-3}	10^{-3}	0.50	0.34	0.79	0.13

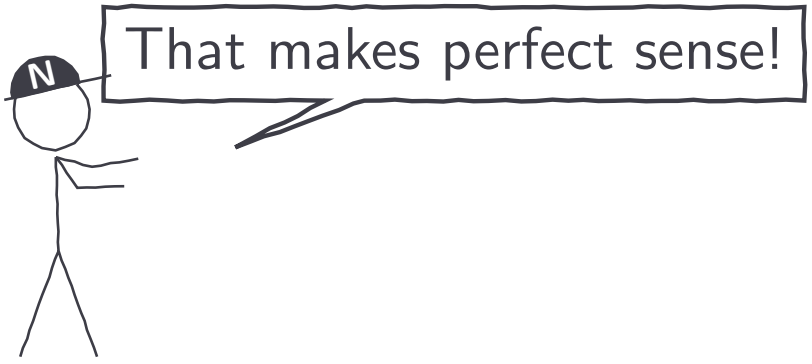
p-values	α_{IC_1} 	α_{IC_2} 	α_{IC_3} 	α_{IC_4} 	α_{IC_5} 	α_{IC_6} 
Encoding	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
Decoding	10^{-3}	10^{-3}	0.50	0.34	0.79	0.13

- instruction to plan a reaching movement is causal for all α_{IC_i}
- $\alpha_{IC_3}, \ldots, \alpha_{IC_6}$ are only indirect effects



p-values	α_{IC_1} 	α_{IC_2} 	α_{IC_3} 	α_{IC_4} 	α_{IC_5} 	α_{IC_6} 
Encoding	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
Decoding	10^{-3}	10^{-3}	0.50	0.34	0.79	0.13

- instruction to plan a reaching movement is causal for all α_{IC_i}
- $\alpha_{IC_3}, \ldots, \alpha_{IC_6}$ are only indirect effects



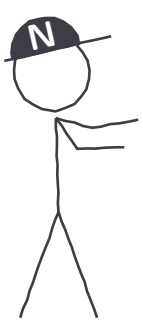
Wrap-up


N


C

feature relevance

feature relevance ⇄ (conditional) (in)dependence



feature relevance ⇝ (conditional) (in)dependence ⇝ causal structure

feature relevance ⇝ (conditional) (in)dependence ⇝ causal structure

- simple interpretation rules
- reinterpretation of previous results?
- resolve recently discussed issues

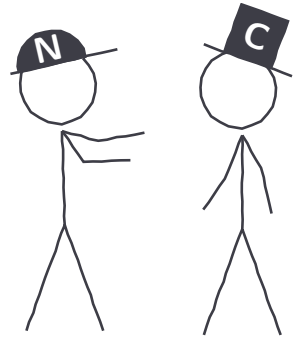
feature relevance ⇝ (conditional) (in)dependence ⇝ causal structure

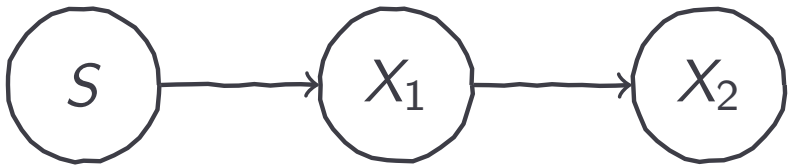
- simple interpretation rules
- reinterpretation of previous results?
- resolve recently discussed issues

It's an interesting application!

Sebastian Weichwald, Timm Meyer, Ozan Özdenizci, Bernhard Schölkopf, Tonio Ball, Moritz Grosse-Wentrup:

- Causal interpretation rules for encoding and decoding models in neuroimaging. *NeuroImage*, 2015.
- Causal and anti-causal learning in pattern recognition for neuroimaging. *PRNI*, 2014.





- Establish $X_1 \rightarrow X_2$ even in the presence of latent confounders

  (Grosse-Wentrup et al., under revision)

- Identify X_2 from a linear mixture of signals

- Problem of confounds in MVPA

  (Todd et al., *NeuroImage*, 2013; Woolgar et al., *NeuroImage*, 2014)

- Interpretation of weight vectors of linear models

  (Haufe et al., *NeuroImage*, 2014)

- Type II errors

- Conditional independence tests
  - permutation-based $\rightsquigarrow$ biased
    (Strobl et al., *BMC Bioinformatics*, 2008)
  - unbiased $\rightsquigarrow$ hard
    (Zhang et al., *UAI*, 2011)

- Untestable assumption: faithfulness