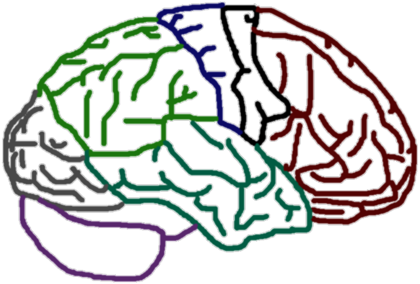
International Workshop on Pattern Recognition in Neuroimaging
June 22, 2016

MAX-PLANCK-GESELLSCHAFT

Recovery of non-linear cause-effect relationships from linearly mixed neuroimaging data

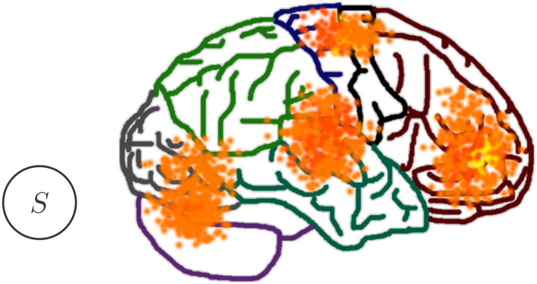
Sebastian Weichwald, Arthur Gretton*, Bernhard Schölkopf, Moritz Grosse-Wentrup
MPI for Intelligent Systems, *Gatsby Computational Neuroscience Unit

Motivation



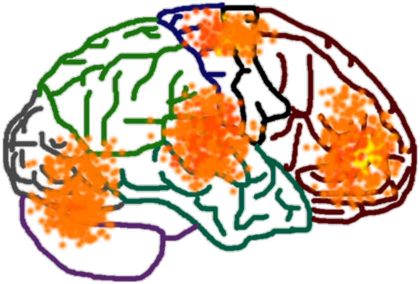

S

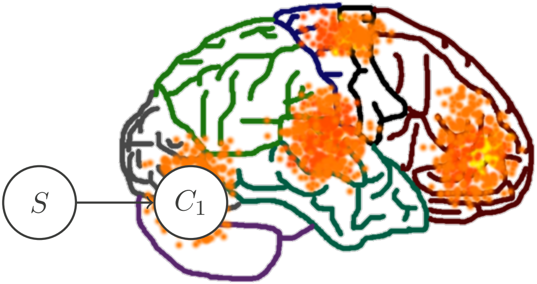
encoding analysis identifies effects of S

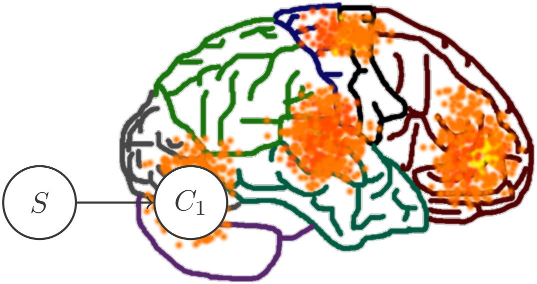


encoding analysis identifies effects of S

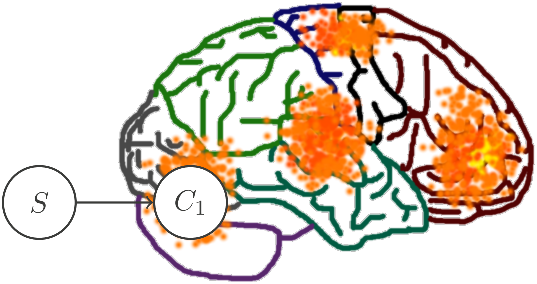
S

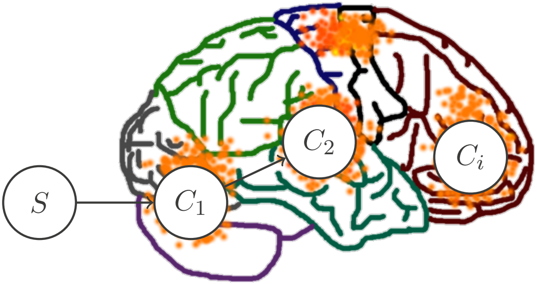



S
C_1

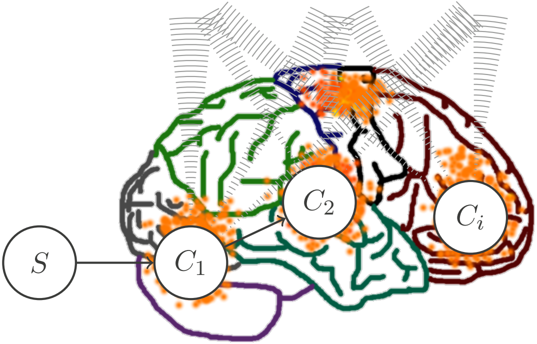


SCI algorithm: test whether a given variable is an effect of C_1


S
C_1

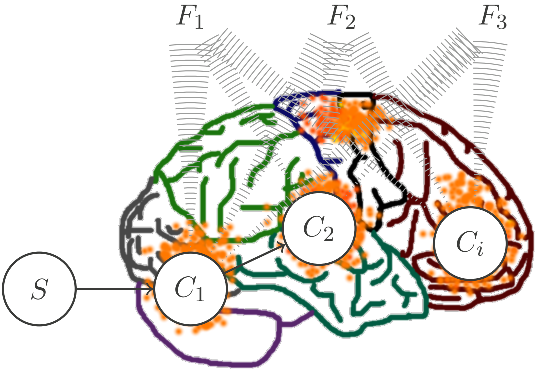


causal neural variables



linear mixing

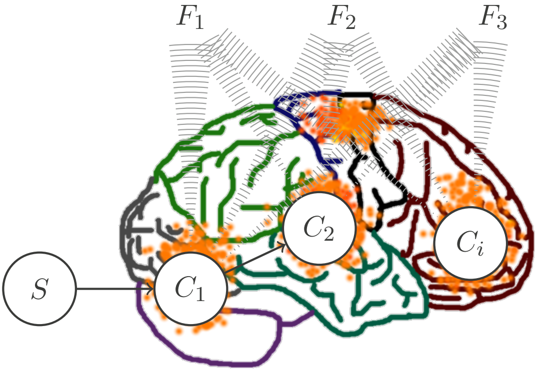
causal neural variables



observed linear mixture

linear mixing

causal neural variables



observed linear mixture

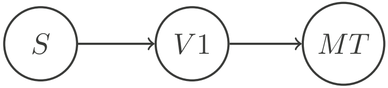
linear mixing

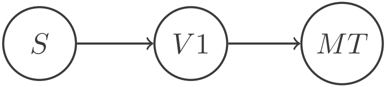
causal neural variables

find linear combination $\boldsymbol{w}$ such that $S \to C_1 \to \overbrace{[F_1, F_2, F_3]\boldsymbol{w}}^{C_2}$

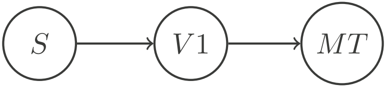
1. Motivation
2. Causal Bayesian Networks
3. Problem description
4. Non-linear MERLiN algorithm
5. Empirical validation
6. Wrap-up & Outlook

Causal Bayesian Networks


S
$V1$
MT

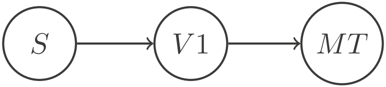


- causation predicts the impact of interventions



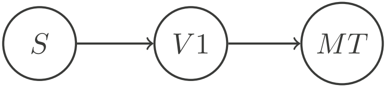
- causation predicts the impact of interventions

$$V1 | \mathrm{do}(S = \mathsf{cat})$$



- causation predicts the impact of interventions

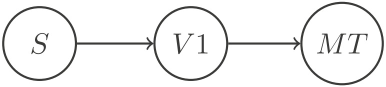
$$V1|\mathrm{do}(S = \mathsf{cat}) \not\sim V1|\mathrm{do}(S = \mathsf{dog})$$



- causation predicts the impact of interventions

$$V1|\,\mathrm{do}(S = \mathsf{cat}) \not\sim V1|\,\mathrm{do}(S = \mathsf{dog})$$

- infer the causal graph



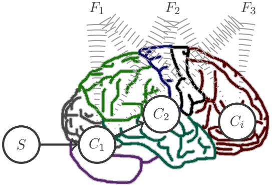
- causation predicts the impact of interventions

$$V1|\,\mathrm{do}(S = \text{cat}) \not\sim V1|\,\mathrm{do}(S = \text{dog})$$

- infer the causal graph

causal structure $\leftrightsquigarrow$ (conditional) (in)dependence

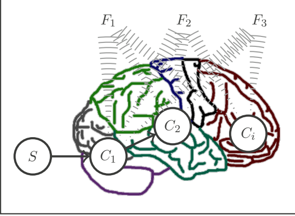
Problem description



Given

samples of S, C_1 and F

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



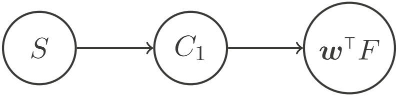
Given

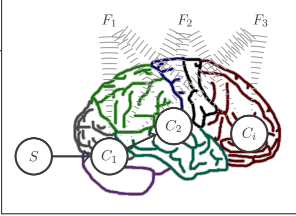
samples of S, C_1 and F

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

Goal

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





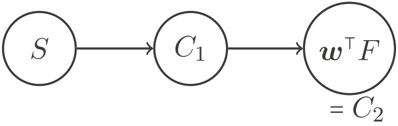
Given

samples of S, C_1 and F

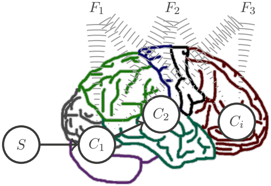
$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_d \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} C_1 \\ \vdots \\ C_d \end{bmatrix} = \boldsymbol{A}C$$

Goal

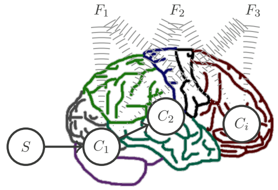
find linear combination $\boldsymbol{w}$ such that



Non-linear MERLiN algorithm

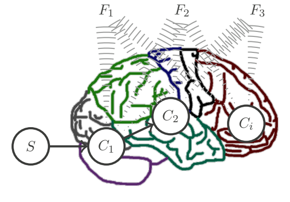


Sufficient conditions



Sufficient conditions

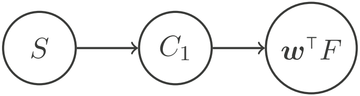
Given S (randomised), C_1, $\boldsymbol{w}^\top F$, and $S \to C_1$

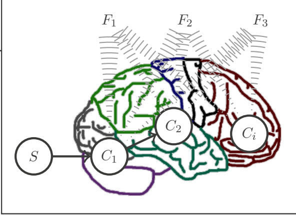


Sufficient conditions

Given S (randomised), C_1, $\boldsymbol{w}^\top F$, and $S \to C_1$, then

$C_1 \not\perp\!\!\!\perp \boldsymbol{w}^\top F$ and $S \perp\!\!\!\perp \boldsymbol{w}^\top F \,|\, C_1 \implies$

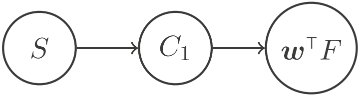




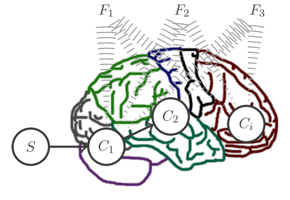
Sufficient conditions

Given S (randomised), C_1, $\boldsymbol{w}^\top F$, and $S \to C_1$, then

$C_1 \not\perp\!\!\!\perp \boldsymbol{w}^\top F$ and $S \perp\!\!\!\perp \boldsymbol{w}^\top F \,|\, C_1 \implies$



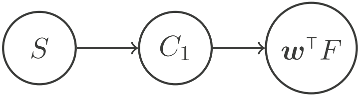
Idea



Sufficient conditions

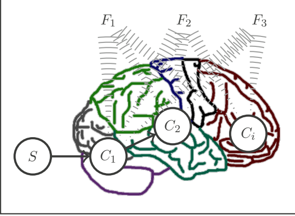
Given S (randomised), C_1, $\boldsymbol{w}^\top F$, and $S \rightarrow C_1$, then

$C_1 \not\perp\!\!\!\perp \boldsymbol{w}^\top F$ and $S \perp\!\!\!\perp \boldsymbol{w}^\top F \,|\, C_1 \implies$



Idea

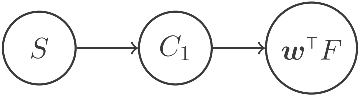
Optimise $\boldsymbol{w}$ such that



Sufficient conditions

Given S (randomised), C_1, $\boldsymbol{w}^\top F$, and $S \to C_1$, then

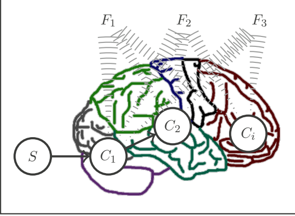
$C_1 \not\perp\!\!\!\perp \boldsymbol{w}^\top F$ and $S \perp\!\!\!\perp \boldsymbol{w}^\top F \,|\, C_1 \implies$



Idea

Optimise $\boldsymbol{w}$ such that

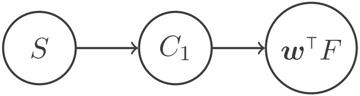
(a) $\operatorname{dep}(C_1, \boldsymbol{w}^\top F)$ is high



Sufficient conditions

Given S (randomised), C_1, $\boldsymbol{w}^\top F$, and $S \rightarrow C_1$, then

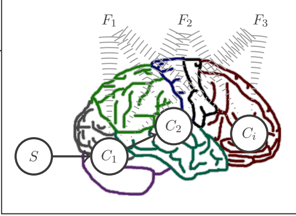
$C_1 \not\perp\!\!\!\perp \boldsymbol{w}^\top F$ and $S \perp\!\!\!\perp \boldsymbol{w}^\top F \,|\, C_1 \implies$



Idea

Optimise $\boldsymbol{w}$ such that

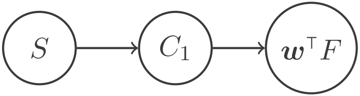
(a) $\mathrm{dep}\,(C_1, \boldsymbol{w}^\top F)$ is high
(b) $\mathrm{dep}\,(S,\ \boldsymbol{w}^\top F \,|\, C_1)$ is low



Sufficient conditions

Given S (randomised), C_1, $\boldsymbol{w}^\top F$, and $S \to C_1$, then

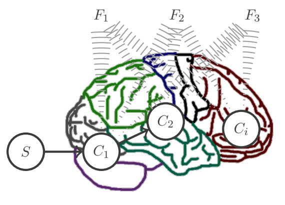
$C_1 \not\perp\!\!\!\perp \boldsymbol{w}^\top F$ and $S \perp\!\!\!\perp \boldsymbol{w}^\top F \,|\, C_1 \implies$



Idea

Optimise $\boldsymbol{w}$ such that

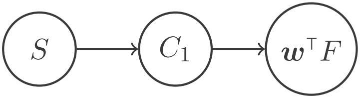
(a) $\mathrm{dep}\,(C_1, \boldsymbol{w}^\top F)$ is high $\rightsquigarrow$ HSIC
(b) $\mathrm{dep}\,(S,\ \boldsymbol{w}^\top F \,|\, C_1)$ is low



Sufficient conditions

Given S (randomised), C_1, $\boldsymbol{w}^\top F$, and $S \rightarrow C_1$, then

$C_1 \not\perp\!\!\!\perp \boldsymbol{w}^\top F$ and $S \perp\!\!\!\perp \boldsymbol{w}^\top F \,|\, C_1 \implies$

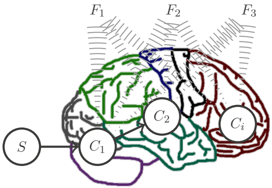


Idea

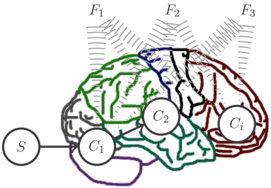
Optimise $\boldsymbol{w}$ such that

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

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

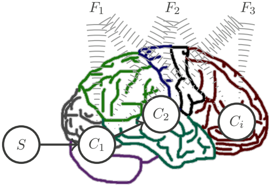


Regression-based conditional independence criterion



Regression-based conditional independence criterion

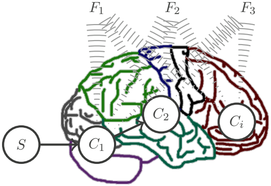
If there exists a regression function r with $\boldsymbol{w}^\top F - r(C_1) \perp\!\!\!\perp (S, C_1)$,



Regression-based conditional independence criterion

If there exists a regression function r with $\boldsymbol{w}^\top F - r(C_1) \perp\!\!\!\perp (S, C_1)$,

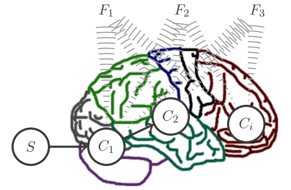
then $S \perp\!\!\!\perp \boldsymbol{w}^\top F \,|\, C_1$.



Regression-based conditional independence criterion

If there exists a regression function r with $\boldsymbol{w}^\top F - r(C_1) \perp\!\!\!\perp (S, C_1)$, then $S \perp\!\!\!\perp \boldsymbol{w}^\top F \,|\, C_1$.

Implementation

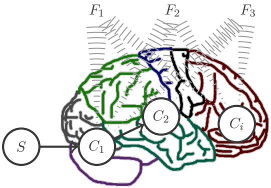


Regression-based conditional independence criterion

If there exists a regression function r with $\boldsymbol{w}^\top F - r(C_1) \perp\!\!\!\perp (S, C_1)$, then $S \perp\!\!\!\perp \boldsymbol{w}^\top F \,|\, C_1$.

Implementation

If we can find kernel ridge regression parameters (σ, θ) such that $\mathrm{dep}\big(\ \boldsymbol{w}^\top F - \mathrm{krr}_{\sigma,\theta}(C_1)\ ,\ (S, C_1)\ \big)$ is low,

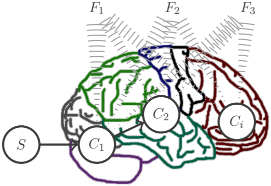


Regression-based conditional independence criterion

If there exists a regression function r with $\boldsymbol{w}^\top F - r(C_1) \perp\!\!\!\perp (S, C_1)$, then $S \perp\!\!\!\perp \boldsymbol{w}^\top F \,|\, C_1$.

Implementation

If we can find kernel ridge regression parameters (σ, θ) such that $\mathrm{dep}\big(\ \boldsymbol{w}^\top F - \mathrm{krr}_{\sigma,\theta}(C_1)\ ,\ (S, C_1)\ \big)$ is low, then $S \perp\!\!\!\perp \boldsymbol{w}^\top F \,|\, C_1$.

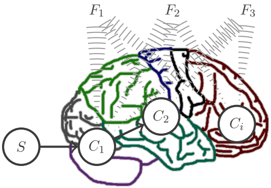


Idea

Optimise $\boldsymbol{w}$ such that

Implementation

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



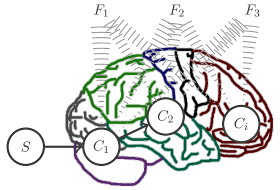
Idea

Optimise $\boldsymbol{w}$ such that

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

Implementation

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



Idea

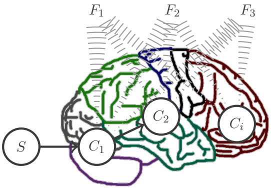
Optimise $\boldsymbol{w}$ such that

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

Implementation

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

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



Idea

Optimise $\boldsymbol{w}$ such that

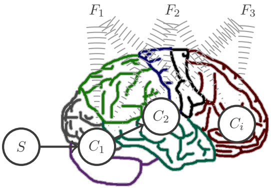
(a) $\mathrm{dep}\,(C_1, \boldsymbol{w}^\top F)$ is high

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

Implementation

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

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



Idea

Optimise $\boldsymbol{w}$ such that

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

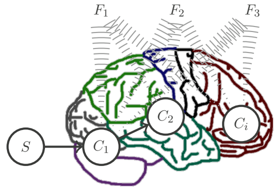
(b) $\operatorname{dep}(S, \boldsymbol{w}^\top F \,|\, C_1)$ is low

Implementation

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

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

(b) $\operatorname{HSIC}\left(\boldsymbol{w}^\top F - \operatorname{krr}_{\sigma,\theta}(C_1) \ , \ (S, C_1) \right)$ is low



Idea

Optimise $\boldsymbol{w}$ such that

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

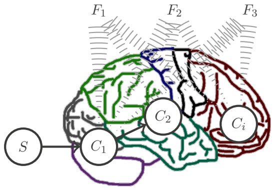
(b) $\mathrm{dep}\,(S,\ \boldsymbol{w}^\top F \,|\, C_1)$ is low

Implementation

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

$$\mathrm{HSIC}\,(C_1, \boldsymbol{w}^\top F)$$
$$-\ \mathrm{HSIC}\left(\ \boldsymbol{w}^\top F - \mathrm{krr}_{\sigma,\theta}(C_1)\ ,\ (S, C_1)\ \right)$$

is being maximised.



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 (Non-linear MERLiN algorithm)

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

$$\mathrm{HSIC}\,(C_1, \boldsymbol{w}^\top F)$$
$$-\ \mathrm{HSIC}\left(\ \boldsymbol{w}^\top F - \mathrm{krr}_{\sigma,\theta}(C_1)\ ,\ (S, C_1)\ \right)$$

is being maximised.

Empirical validation

S :
C_1:
F :

S : instruction to up-/downregulate C_1 $\{\pm 1\}$
C_1:
F :

S : instruction to up-/downregulate C_1 $\{\pm 1\}$
C_1: γ-bandpower in superior parietal cortex $\mathbb{R}$
F :

S : instruction to up-/downregulate C_1 $\{\pm 1\}$
C_1: γ-bandpower in superior parietal cortex $\mathbb{R}$
F : EEG electrode signals $\mathbb{R}^{\text{channels} \times \text{time}}$

S : instruction to up-/downregulate C_1 $\{\pm 1\}$
C_1: γ-bandpower in superior parietal cortex $\mathbb{R}$
F : EEG electrode signals $\mathbb{R}^{\text{channels} \times \text{time}}$

SCI algorithm

S : instruction to up-/downregulate C_1 $\{\pm 1\}$
C_1: γ-bandpower in superior parietal cortex $\mathbb{R}$
F : EEG electrode signals $\mathbb{R}^{\text{channels} \times \text{time}}$

SCI algorithm

test

$S \to C_1 \to \gamma\text{-bp}\left(\text{dipole}_j\right)$

for $j = 1, ..., 15028$

S : instruction to up-/downregulate C_1 $\{\pm 1\}$
C_1: γ-bandpower in superior parietal cortex $\mathbb{R}$
F : EEG electrode signals $\mathbb{R}^{\text{channels} \times \text{time}}$

SCI algorithm *Non-linear MERLiN algorithm*

test

$S \rightarrow C_1 \rightarrow \gamma\text{-bp}\left(\text{dipole}_j\right)$

for $j = 1, ..., 15028$

S : instruction to up-/downregulate C_1 $\{\pm 1\}$
C_1: γ-bandpower in superior parietal cortex $\mathbb{R}$
F : EEG electrode signals $\mathbb{R}^{\text{channels} \times \text{time}}$

SCI algorithm

test

$S \to C_1 \to \gamma\text{-bp}\left(\text{dipole}_j\right)$

for $j = 1, \ldots, 15028$

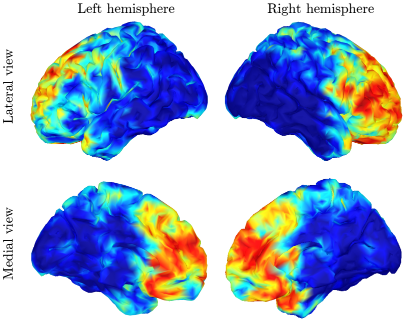
Non-linear MERLiN algorithm

optimise $\boldsymbol{w}$ such that

$S \to C_1 \to \gamma\text{-bp}\left(\boldsymbol{w}^\top F\right)$

S : instruction to up-/downregulate C_1 $\{\pm 1\}$
C_1: γ-bandpower in superior parietal cortex $\mathbb{R}$
F : EEG electrode signals $\mathbb{R}^{\text{channels} \times \text{time}}$

SCI algorithm



Non-linear MERLiN algorithm

optimise $\boldsymbol{w}$ such that

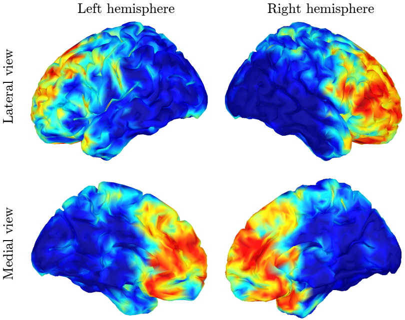
$$S \to C_1 \to \gamma\text{-bp}\left(\boldsymbol{w}^\top F\right)$$

S : instruction to up-/downregulate C_1 — $\{\pm 1\}$

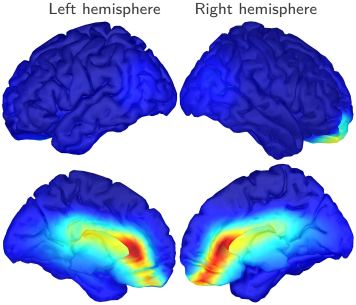
C_1: γ-bandpower in superior parietal cortex — $\mathbb{R}$

F : EEG electrode signals — $\mathbb{R}^{\text{channels} \times \text{time}}$

SCI algorithm

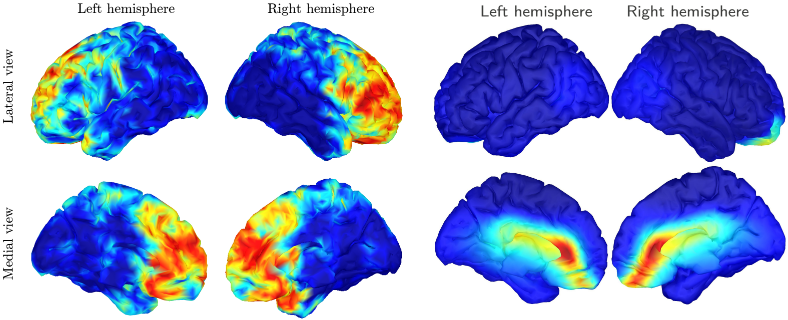


Non-linear MERLiN algorithm



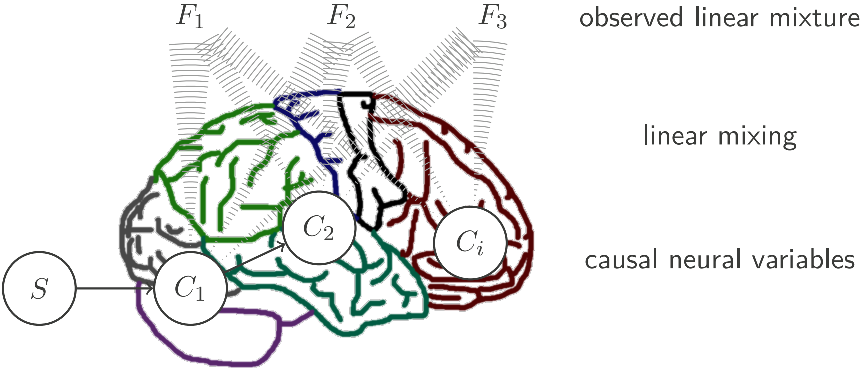
S : instruction to up-/downregulate C_1 $\{\pm 1\}$
C_1: γ-bandpower in superior parietal cortex $\mathbb{R}$
F : EEG electrode signals $\mathbb{R}^{\text{channels} \times \text{time}}$

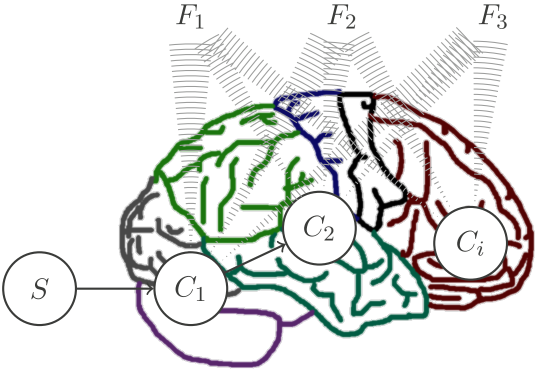
SCI algorithm *Non-linear MERLiN algorithm*



both find $S \to \gamma$-bp (SPC) $\to \gamma$-bp (MPFC)

Wrap-up & Outlook


F_1
F_2
F_3
observed linear mixture
linear mixing
C_2
C_i
causal neural variables
S
C_1

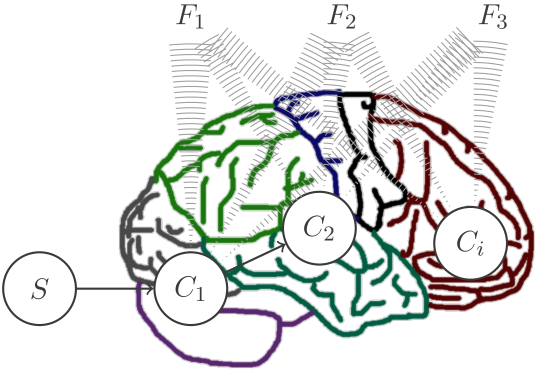


observed linear mixture

linear mixing

causal neural variables

feed samples of S, C_1 and F to the non-linear MERLiN algorithm



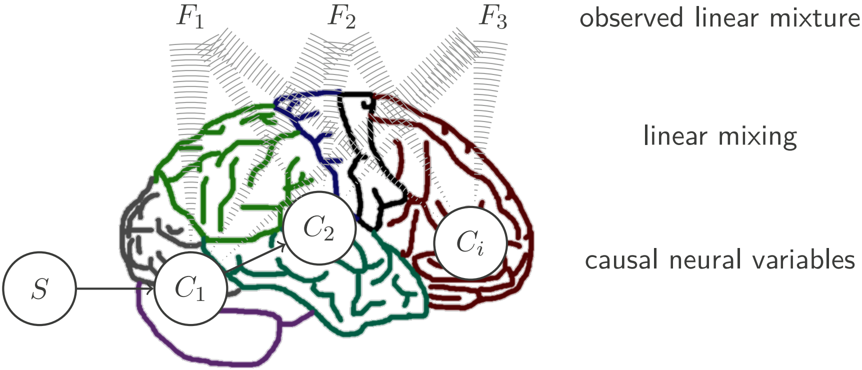
observed linear mixture

linear mixing

causal neural variables

feed samples of S, C_1 and F to the non-linear MERLiN algorithm

$\rightsquigarrow$ recovery of the (non-linear) causal effect $C_2 = \boldsymbol{w}^\top F$



feed samples of S, C_1 and F to the non-linear MERLiN algorithm

$\rightsquigarrow$ recovery of the (non-linear) causal effect $C_2 = \boldsymbol{w}^\top F$

"A general idea to learn causally meaningful features?"

- Recovery of non-linear cause-effect relationships from linearly mixed neuroimaging data. *PRNI*, 2016. e-print arxiv.org/pdf/1512.04808.
- MERLiN: Mixture Effect Recovery in Linear Networks. Under review. e-print arxiv.org/pdf/1512.01255.
- Identification of causal relations in neuroimaging data with latent confounders: An instrumental variable approach. *NeuroImage*, 2016. e-print mlin.kyb.tuebingen.mpg.de/Grosse-WentrupNI2015.pdf.
- Causal interpretation rules for encoding and decoding models in neuroimaging. *NeuroImage*, 2015. sweichwald.de/neuroimage2015.
- Causal and anti-causal learning in pattern recognition for neuroimaging. *PRNI*, 2014. e-print arxiv.org/pdf/1512.04808.