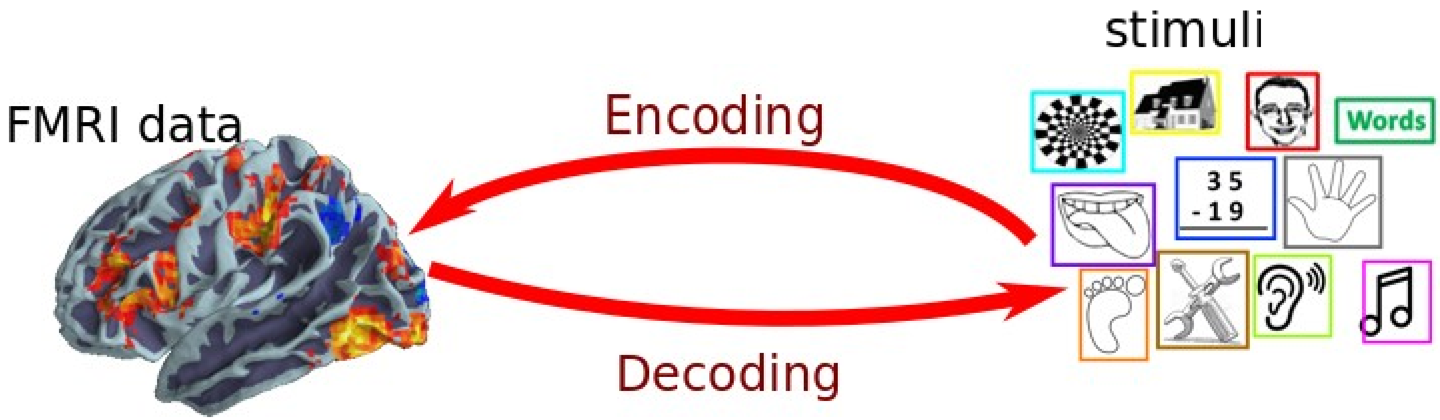
Analyzing brain function: decoding and reverse inference

Deep learning summer school

Bertrand Thirion, bertrand.thirion@inria.fr

Introduction



- Encoding and decoding brain activity: why are both interesting ?
- Decoding with pre-defined categories of decoding at large scale ? Functional specificity
- Decoding as an inverse problem: ill-posedness and (some) solution
-

Outline

- Why is decoding important ? A causal perspective
- Intermezzo: Learning representations for brain images with dictionary learning
- Large-scale brain activity decoding
  - coordinated decoding
  - Joint decoding
- Decoding an in inverse problem: what can we learn from it ?

What analysis is necessary to gain insights into brain-function relationships ?

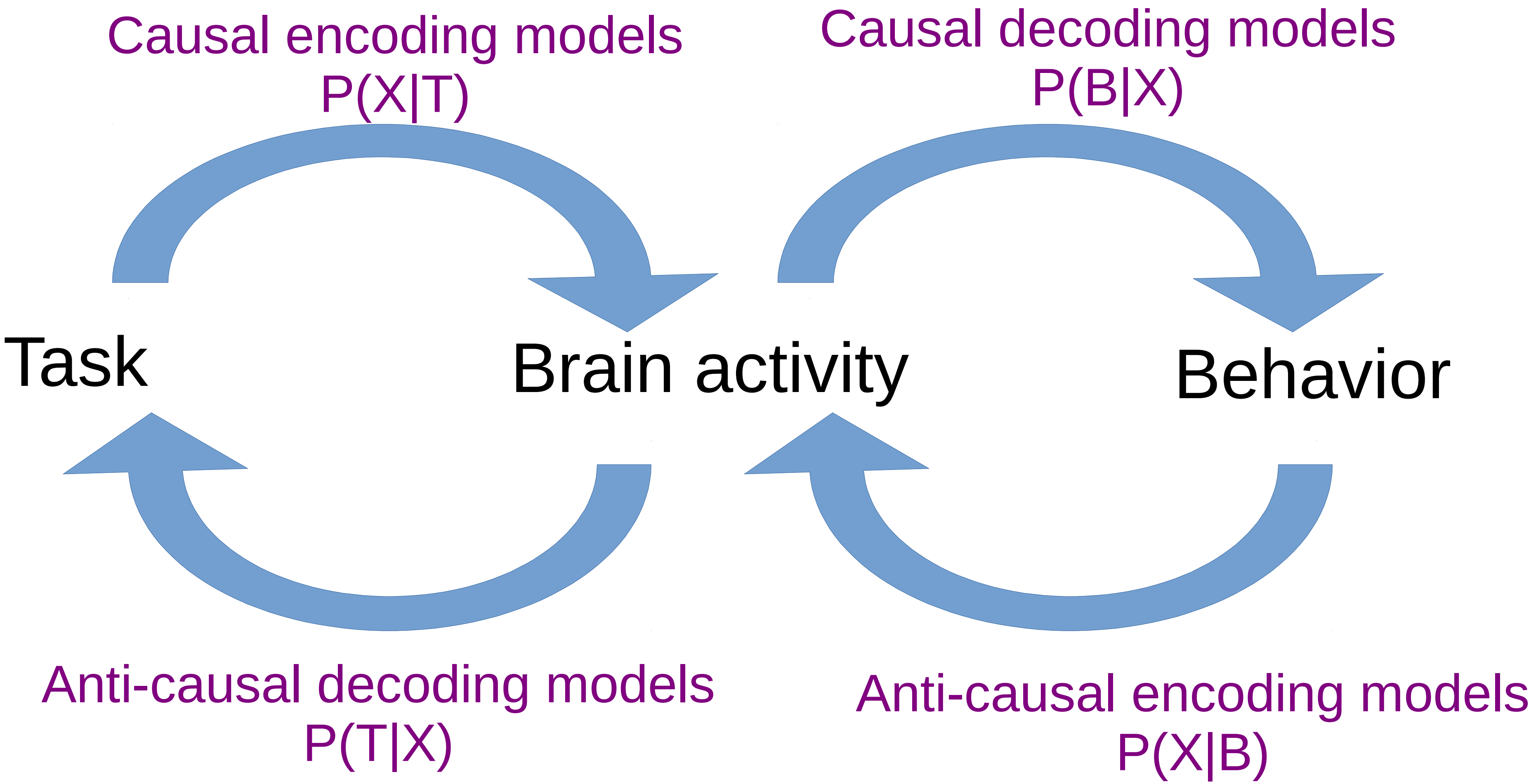
A) Encoding: study the association of brain activity in any given region with the stimuli/tasks

B) Decoding: assess whether a regions is important or not in the prediction of stimulus occurrence

C) Both

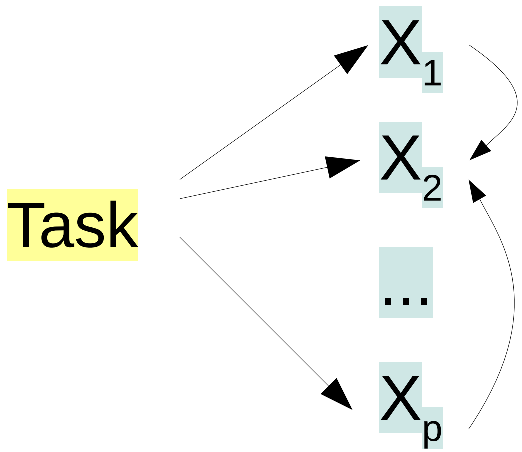
D) Other suggestions

Causal reasoning on encoding/decoding



[Weichwald et al Nimg 2015]

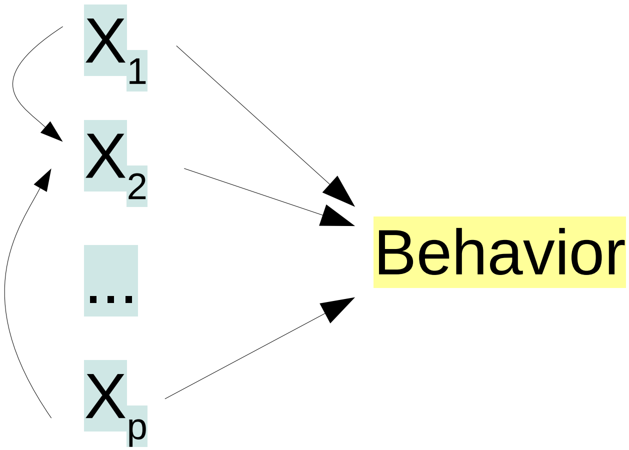
Causal interpretation



$$\mathbf{X_i} \perp\!\!\!\perp \mathbf{T}$$

Encoding: causal
Decoding: anti-causal

$$\mathbf{X_i} \perp\!\!\!\perp \mathbf{T} | (\mathbf{X_j}, \mathbf{j} \neq \mathbf{i})$$



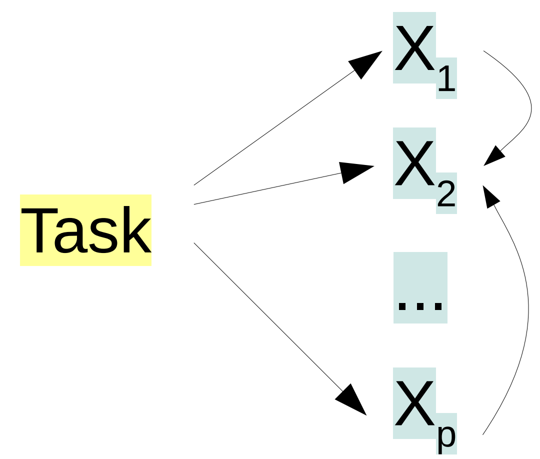
$$\mathbf{X_i} \perp\!\!\!\perp \mathbf{B}$$

Encoding: anti-causal
Decoding: causal

$$\mathbf{X_i} \perp\!\!\!\perp \mathbf{B} | (\mathbf{X_j}, \mathbf{j} \neq \mathbf{i})$$

Causal reasoning on encoding/decoding

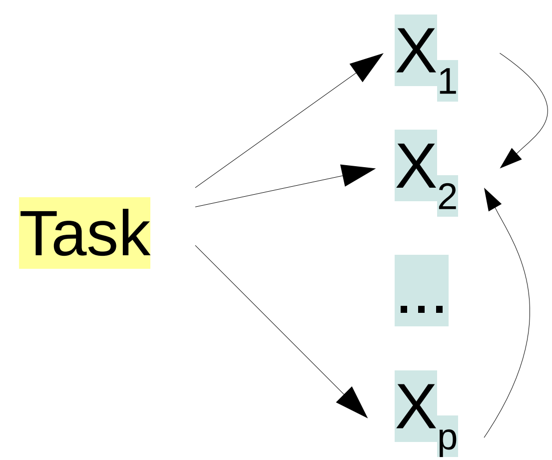
		Feature X_i relevant?		
		Encoding	Decoding	Causal interpretation
Experimental setting	Task	$\times$		$T \perp\!\!\!\perp X_i \Rightarrow X_i$ is no effect of T
		$\surd$		$T \not\perp\!\!\!\perp X_i \Rightarrow X_i$ is an effect of T
	Behaviour			



[Weichwald et al. NIMG 2015]

Causal reasoning on encoding/decoding

Experimental setting		Feature X_i relevant? Encoding	Decoding	Causal interpretation
	Task	$\times$		$T \perp\!\!\!\perp X_i \Rightarrow X_i$ is no effect of T
	Task	$\surd$		$T \not\!\perp\!\!\!\perp X_i \Rightarrow X_i$ is an effect of T
	Task		$\times$	$T \perp\!\!\!\perp X_i \mid \boldsymbol{X} \setminus X_i \Rightarrow$ inconclusive
	Task		$\surd$	$T \not\!\perp\!\!\!\perp X_i \mid \boldsymbol{X} \setminus X_i \Rightarrow$ inconclusive
	Behaviour			
	Behaviour			



[Weichwald et al. NIMG 2015]

Causal reasoning on encoding/decoding

Experimental setting		Feature X_i relevant? Encoding	Decoding	Causal interpretation
Experimental setting	Task	$\times$		$T \perp\!\!\!\perp X_i \Rightarrow X_i$ is no effect of T
	Task	$\surd$		$T \not\!\perp\!\!\!\perp X_i \Rightarrow X_i$ is an effect of T
	Task		$\times$	$T \perp\!\!\!\perp X_i \mid \boldsymbol{X} \setminus X_i \Rightarrow$ inconclusive
	Task		$\surd$	$T \not\!\perp\!\!\!\perp X_i \mid \boldsymbol{X} \setminus X_i \Rightarrow$ inconclusive
	Behaviour	$\times$		$B \perp\!\!\!\perp X_i \Rightarrow X_i$ is no cause of B
	Behaviour	$\surd$		$B \not\!\perp\!\!\!\perp X_i \Rightarrow$ inconclusive
	Behaviour		$\times$	$B \perp\!\!\!\perp X_i \mid \boldsymbol{X} \setminus X_i \Rightarrow$ inconclusive
	Behaviour		$\surd$	$B \not\!\perp\!\!\!\perp X_i \mid \boldsymbol{X} \setminus X_i \Rightarrow$ inconclusive

[Weichwald et al. NIMG 2015]

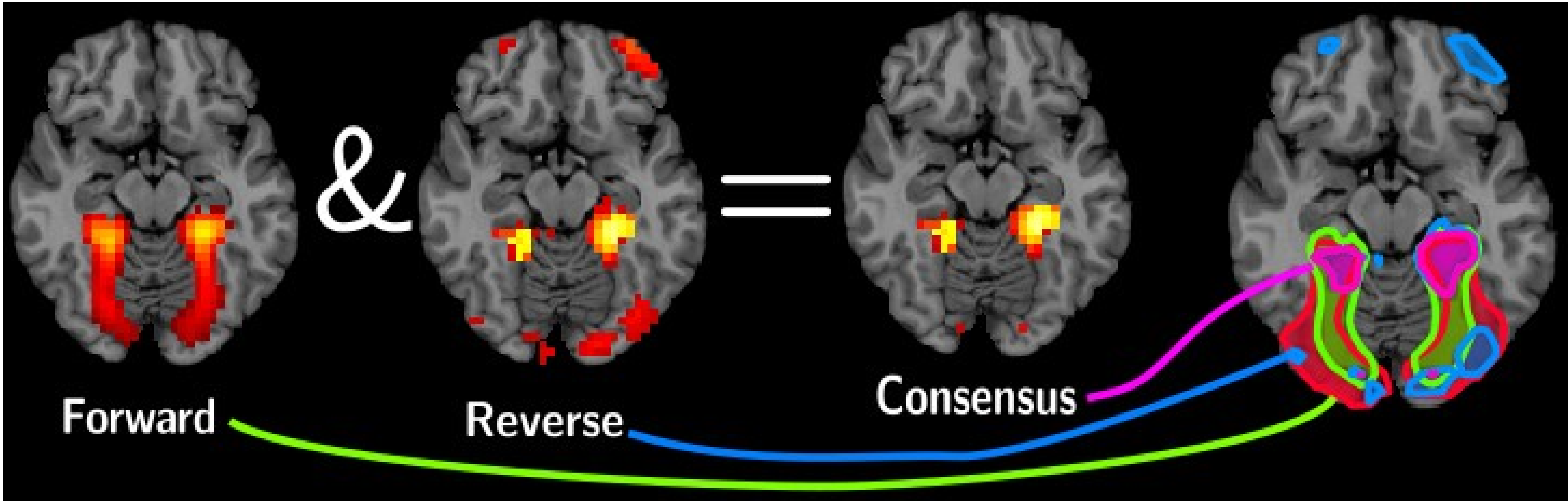
Causal reasoning on encoding/decoding

		Feature X_i relevant? Encoding	Decoding	Causal interpretation
Experimental paradigm	Task	$\times$	$\times$	X_i is no effect of T
		$\surd$	$\times$	X_i is an indirect effect of T
		$\times$	$\surd$	X_i provides context
		$\surd$	$\surd$	X_i is an effect of T
	Behaviour	$\times$	$\times$	X_i is no cause of B
		$\surd$	$\times$	X_i is no direct cause of B
		$\times$	$\surd$	X_i provides context
		$\surd$	$\surd$	inconclusive

[Weichwald et al. NIMG 2015]

Joint encoding and decoding

Brain response to the task: “view a place image”



“Encoding” “Decoding”

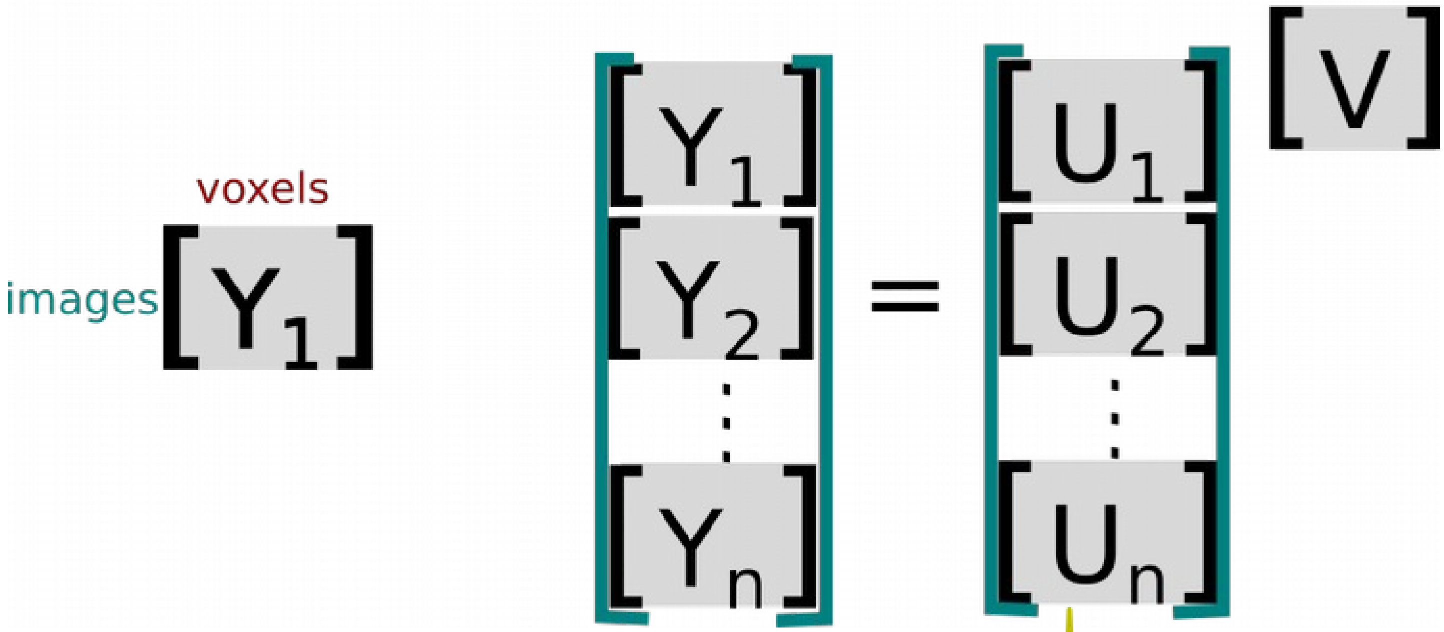
Decoding solves a difficult inverse problem: noisy estimates [Haufe et al. Nimg 2014, Chevalier et al. MICCAI 2018]

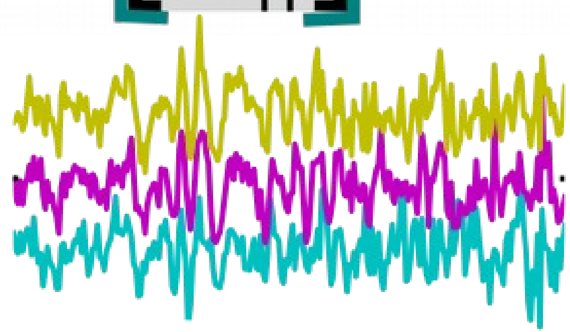
[Schwartz et al. NIPS 2013, Varoquaux et al. PCB 2018]

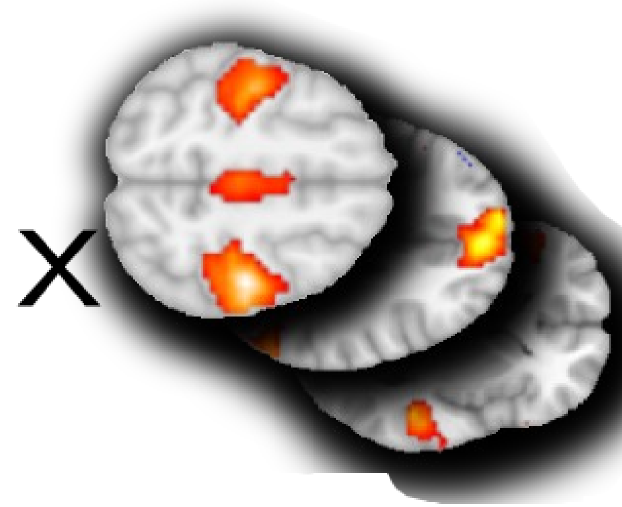
Outline

- Why is decoding important ? A causal perspective
- Intermezzo: Learning representations for brain images with dictionary learning
- Large-scale brain activity decoding
  - coordinated decoding
  - Joint decoding
- Decoding an in inverse problem: what can we learn from it ?

Learning representation of fMRI data







$$\min_{\mathbf{V},\mathbf{U}:\|\mathbf{U}_k\|\leq 1,\forall k} \|\mathbf{Y} - \mathbf{U}\mathbf{V}^T\|^2 + \lambda\|\mathbf{V}\|_1$$

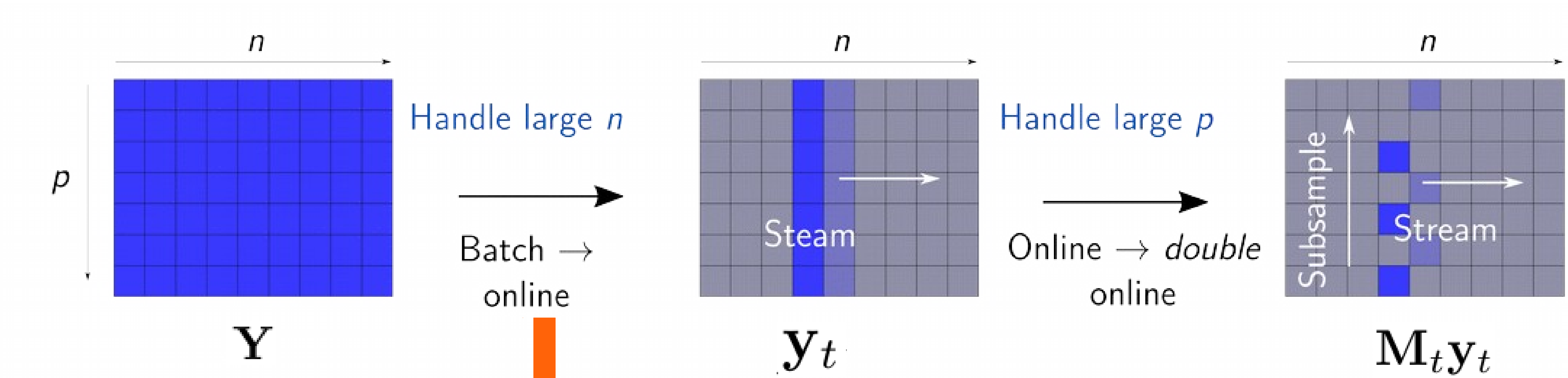
Concatenation mode for multi-subject → long “time courses”

Huge ?

- Human Connectome project $n=4.10^6$, $p=2.10^5$, 4**TB** of data
- Spatially-concatenated DL [varoquaux IPMI 2013] does not scale well:
  - Requires loading the data [Mensch et al. ISBI 2016]
- Solution
  - Work on batches of images **and** voxels
    - Online method in both samples and feature dimensions

Stochastic gradient approaches

http://amensch.fr/research/2016/06/10/modl.html



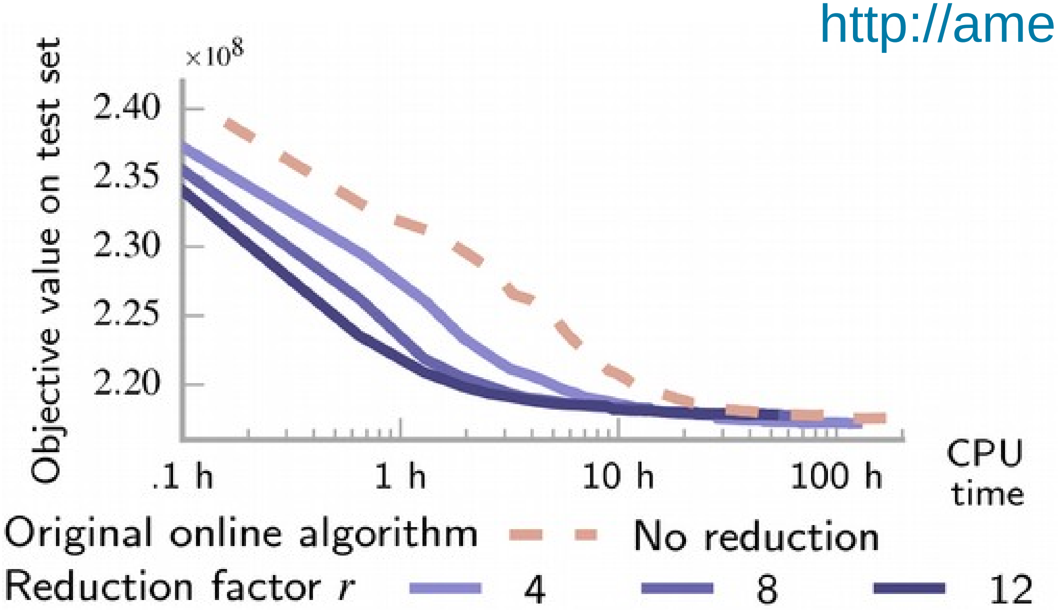
$$\mathbf{u}_t(\mathbf{V}) = \operatorname{argmin}_{\mathbf{u}\in\mathbb{R}^{n\times k}} \|\mathbf{y}_t - \mathbf{u}\mathbf{V}\|_F^2 + \lambda\Omega(\mathbf{u})$$

$$\mathbf{V}_t = \operatorname{argmin}_{\mathbf{V}\in\mathcal{C}} \sum_{i=1}^{t} \|\mathbf{y}_i - \mathbf{u}_i\mathbf{V}\|_F^2$$

$$\mathbf{u}_t(\mathbf{V}) = \operatorname{argmin}_{\mathbf{u}\in\mathbb{R}^{n\times k}} \|\mathbf{M}_t\left(\mathbf{y}_t - \mathbf{u}\mathbf{V}\right)\|_F^2$$
$$\lambda\frac{s}{p}\Omega(\mathbf{u})$$

Stochastic gradient approaches

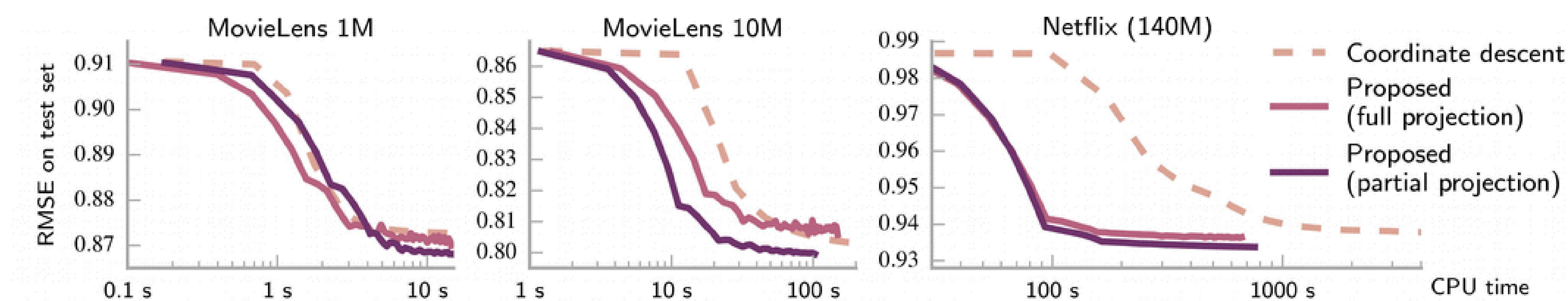
http://amensch.fr/research/2016/06/10/modl.html



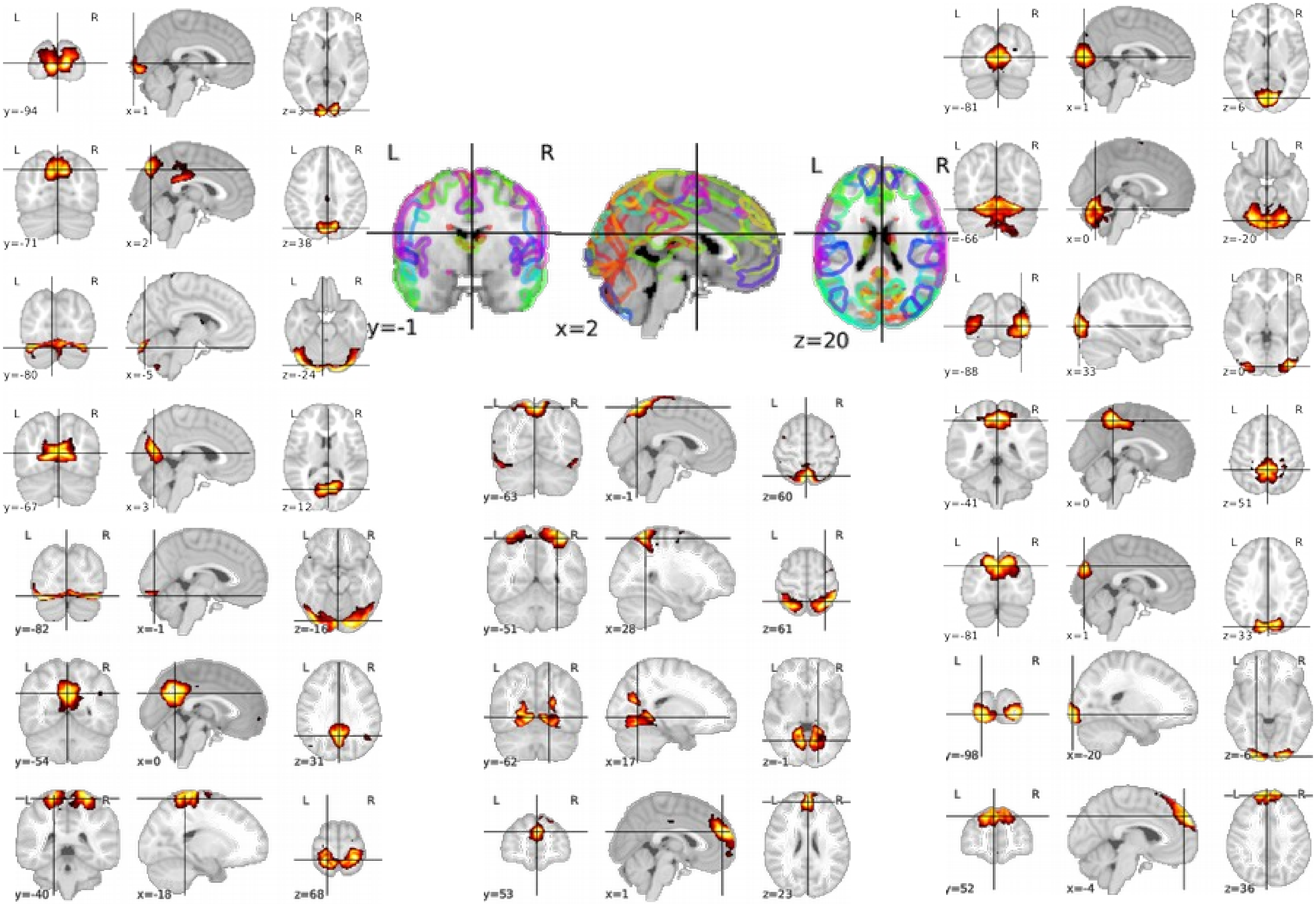
10-fold gain in CPU time
without loss in accuracy

[Mensch et al. IEEE TSP 2018]

Can be used for recommender systems



Brain atlases



[Mensch et al. ICML 2016]

Brain atlases from fMRI data

- Relax clustering into spatio-temporal decompositions
- Sparse-/smooth- PCA algorithms perform well
  - Fine-tuning to individual configuration is expensive
  - Sophisticated penalties (TV + l1…) are expensive
- Development of online methods to scale to large data [Mairal et al. JMLR 2010, Mensch et al. ICML 2016, IEEE TSP 2018]

Outline

- Why is decoding important ? A causal persepective
- Intermezzo: Learning representations for brain images with dictionary learning
- Large-scale brain activity decoding
  - coordinated decoding
  - Joint decoding
- Decoding an in inverse problem: what can we learn from it ?

The big data revolution is ongoing – in neuroimaging also !

Nature Reviews Neuroscience | AOP, published online 10 April 2013; doi:10.1038/nrn3475

nature REVIEWS

Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button[1,2], John P. A. Ioannidis[3], Claire Mokrysz[1], Brian A. Nosek[4], Jonathan Flint[5], Emma S. J. Robinson[6] and Marcus R. Munafò[1]

https://en.wikipedia.org/wiki/Replication_crisis

The big data revolution is ongoing – in neuroimaging also !

Nature Reviews Neuroscience | AOP, published online 10 April 2013; doi:10.1038/nrn3475

nature REVIEWS

Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button[1,2], John P. A. Ioannidis[3], Claire Mokrysz[1], Brian A. Nosek[4], Jonathan Flint[5], Emma S. J. Robinson[6] and Marcus R. Munafò[1]

nature REVIEWS NEUROSCIENCE

Analysis | Published: 05 January 2017

Scanning the horizon: towards transparent and reproducible neuroimaging research

Russell A. Poldrack, Chris I. Baker, Joke Durnez, Krzysztof J. Gorgolewski, Paul M. Matthews, Marcus R. Munafò, Thomas E. Nichols, Jean-Baptiste Poline, Edward Vul & Tal Yarkoni

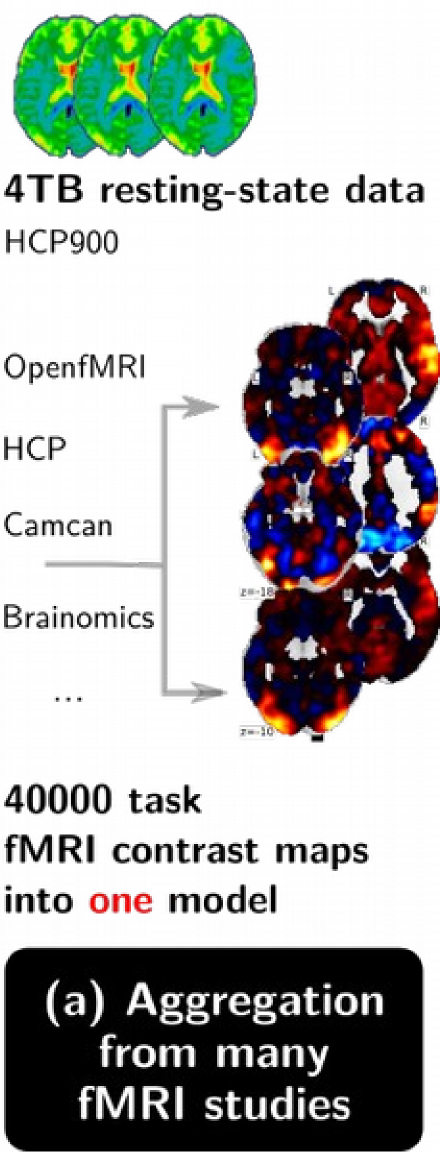
Problem: generalization across studies

"You cannot play 20 questions with nature and win"

- Coordinated analysis: Use large studies to inform small studies ("transfer learning")
  - Principle: co-analyse studies, leverage joint representations
- joint/mega-analysis: find semantic commonalities across studies
  - Difficulty: what common vocabulary across studies?

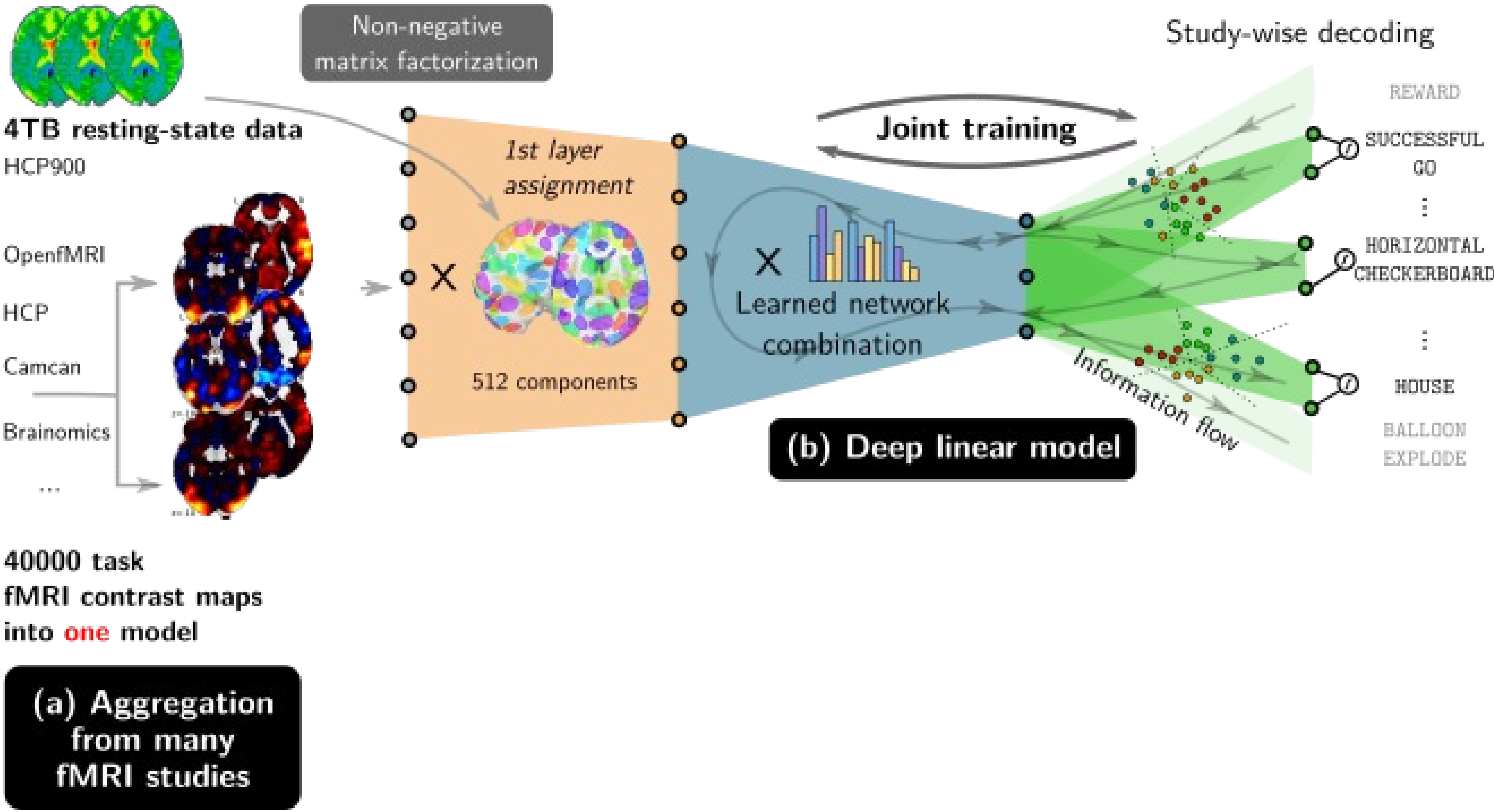
Large studies to inform analysis of small studies: coordinated decoding

Predictive modeling across datasets



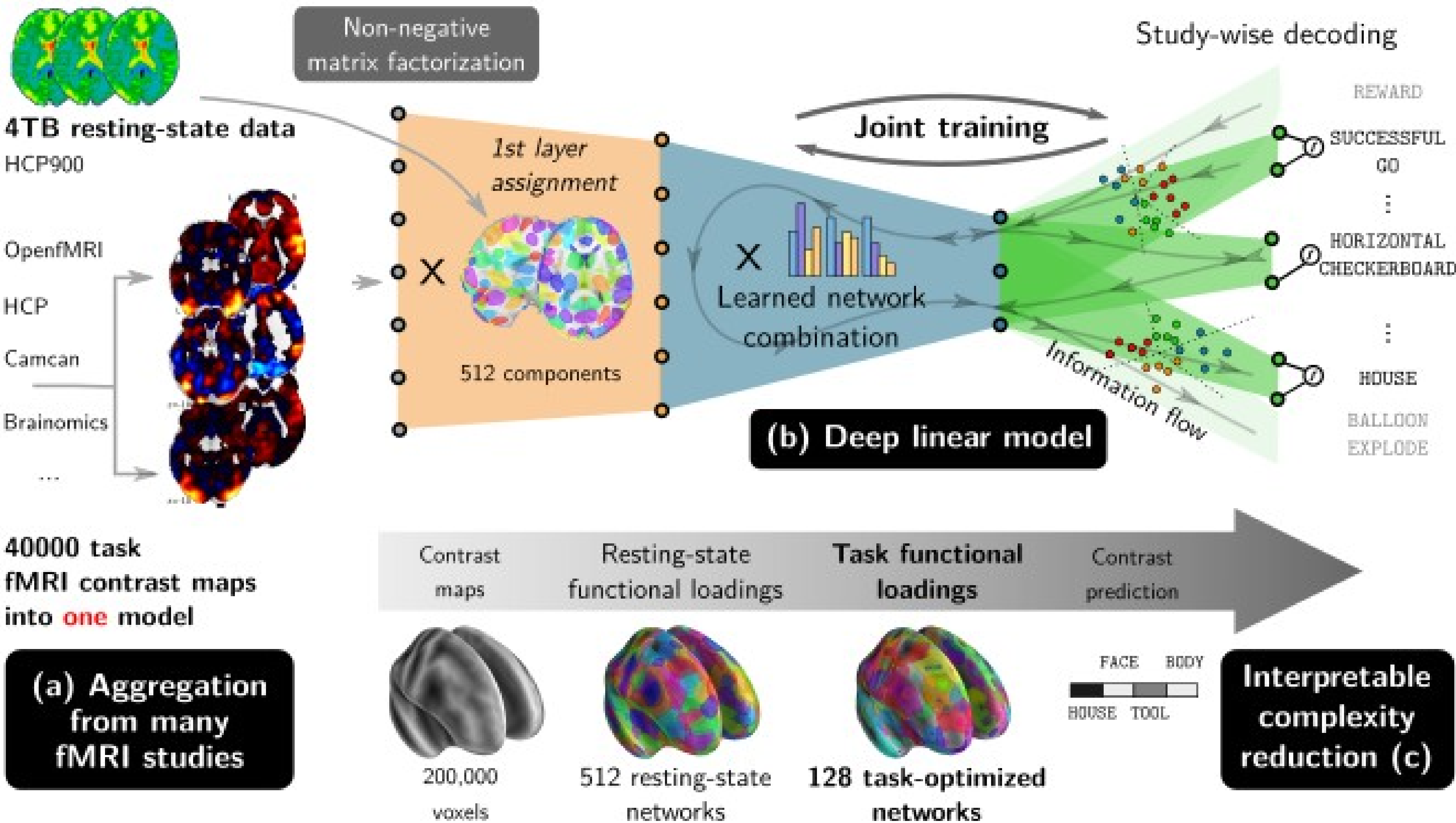
[Bzdok et al. Plos Comp Biol 2016, Mensch et al NIPS 2017]

Predictive modeling across datasets



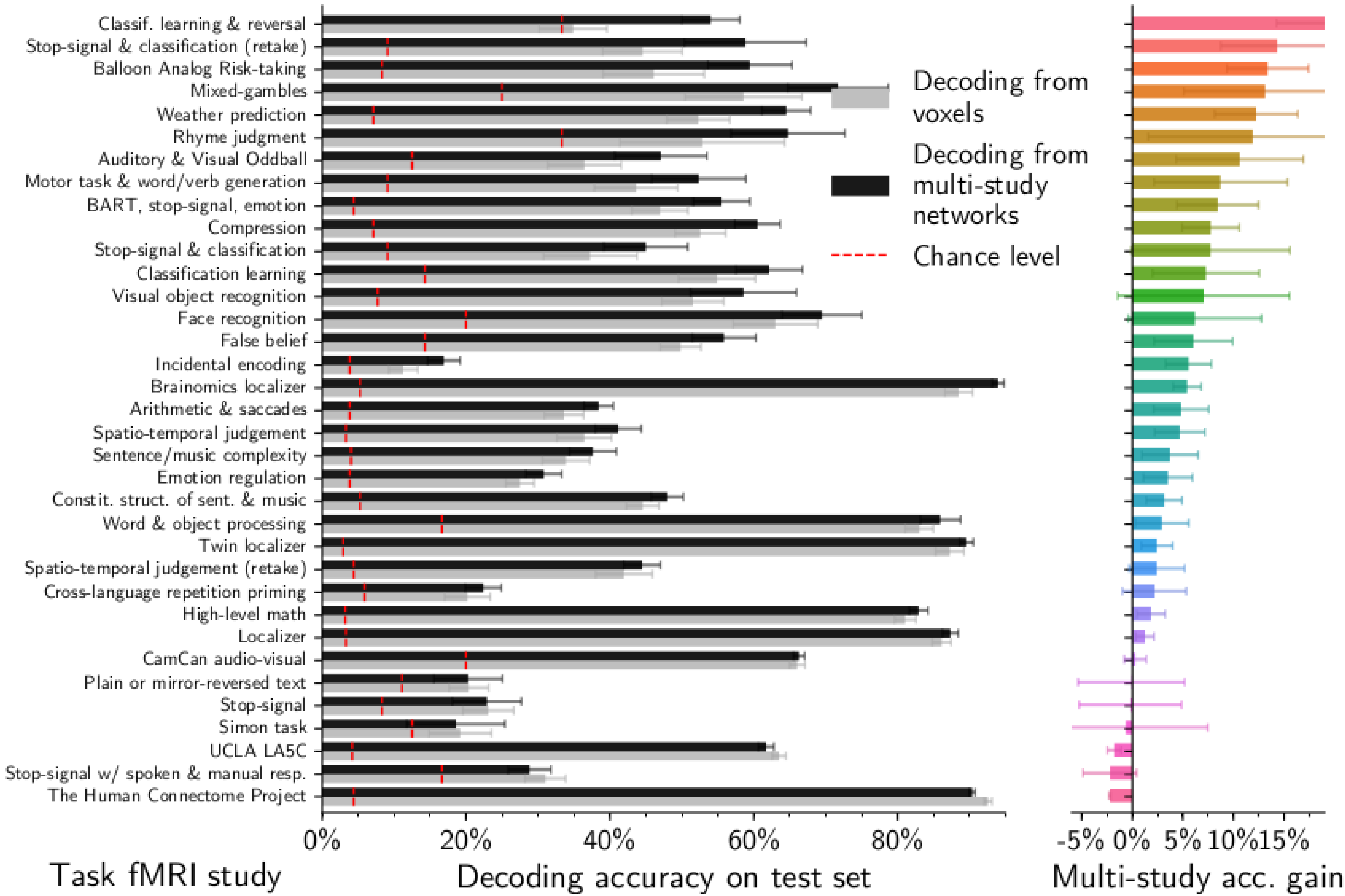
[Bzdok et al. Plos Comp Biol 2016, Mensch et al NIPS 2017]

Predictive modeling across datasets

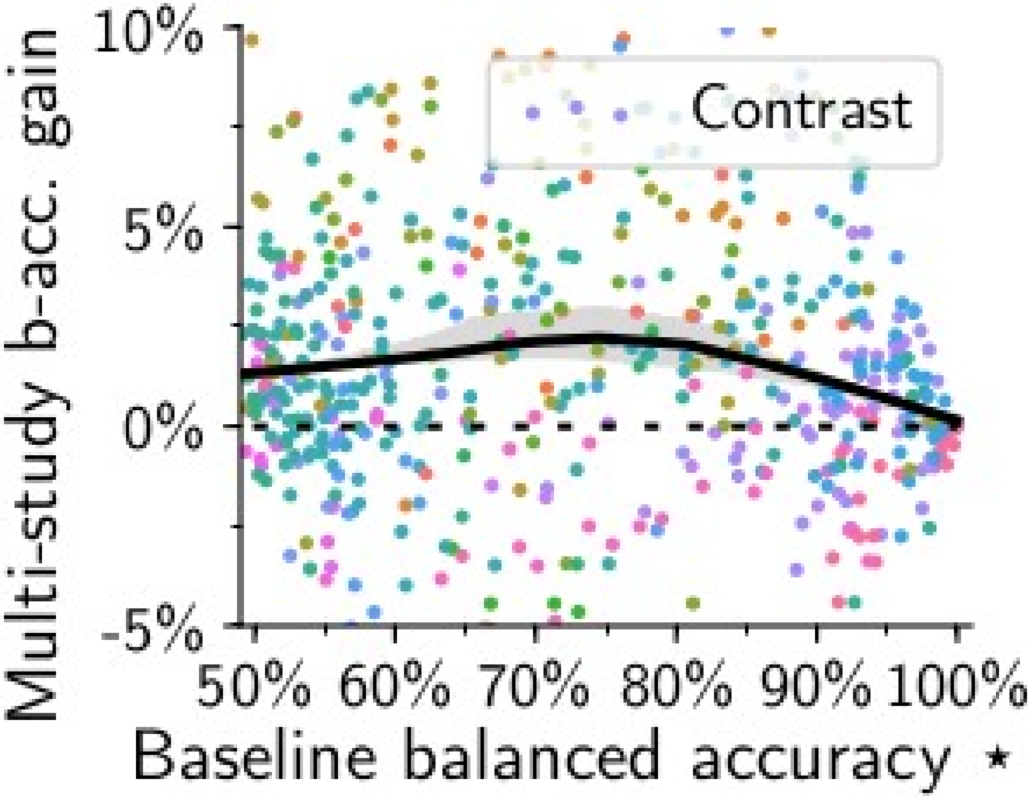


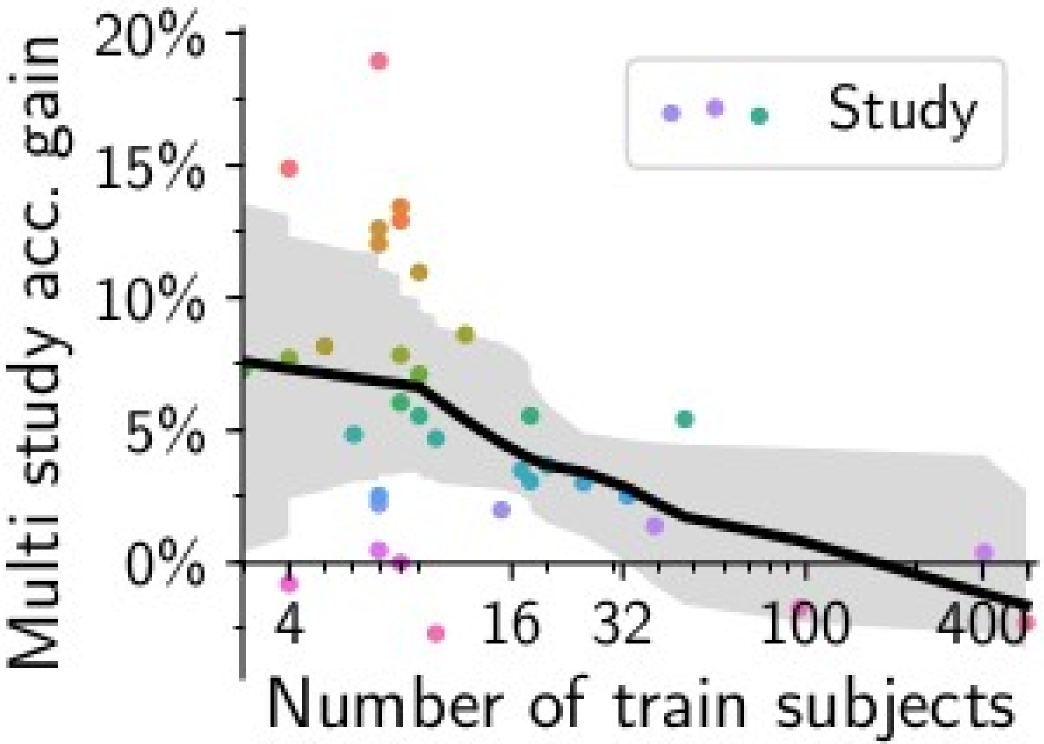
[Bzdok et al. Plos Comp Biol 2016, Mensch et al NIPS 2017]

Transfer learning

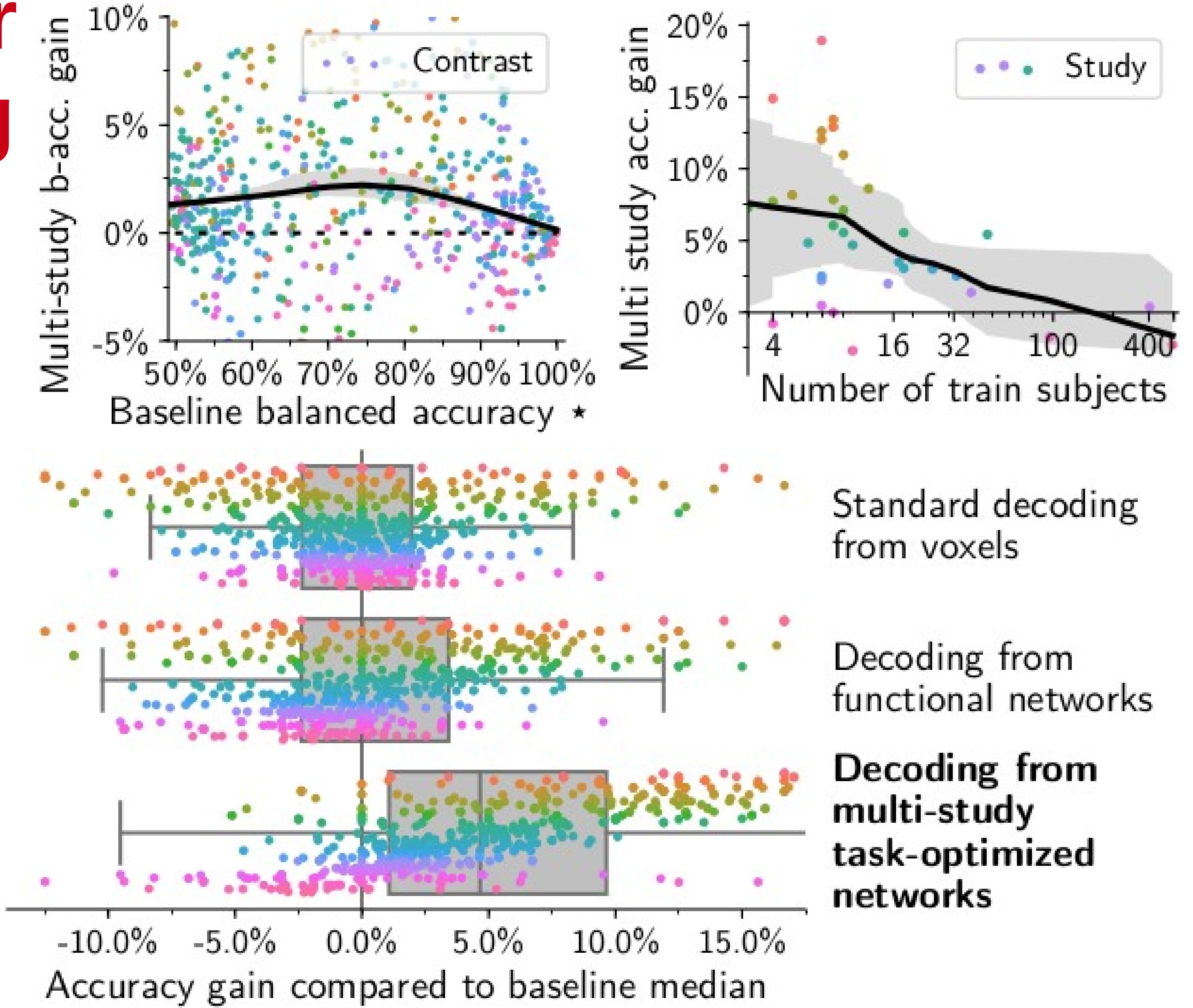


Transfer learning

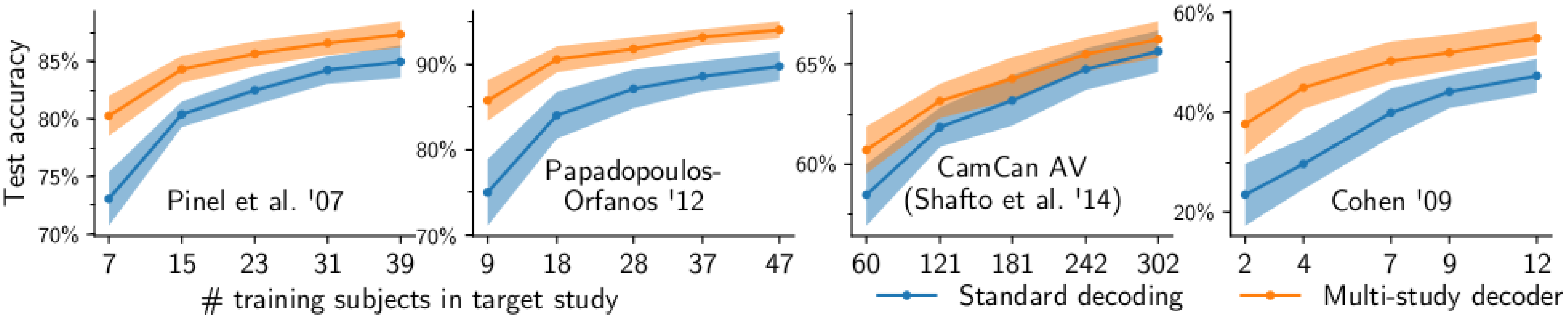




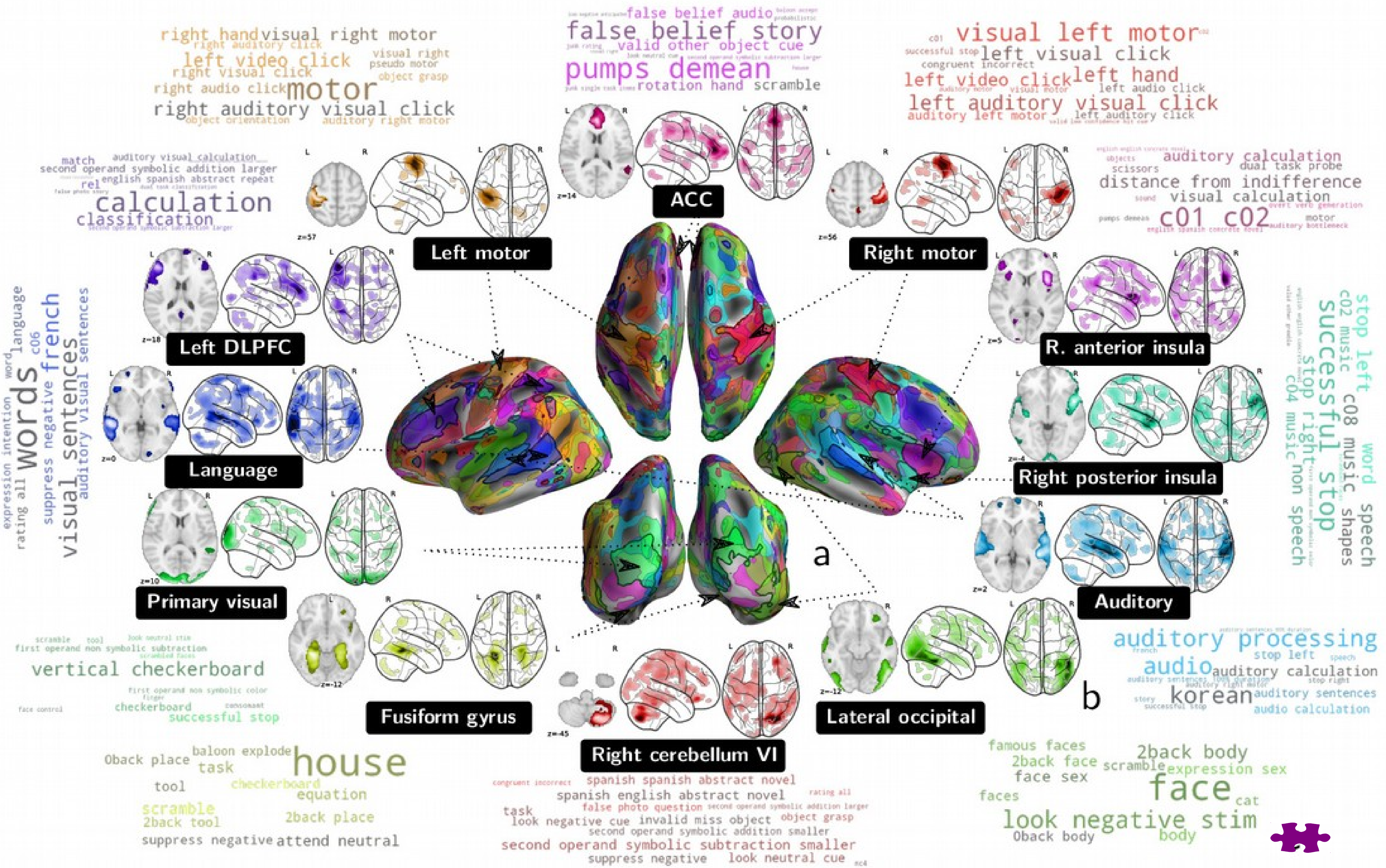
Transfer learning



Small studies benefit more than large studies



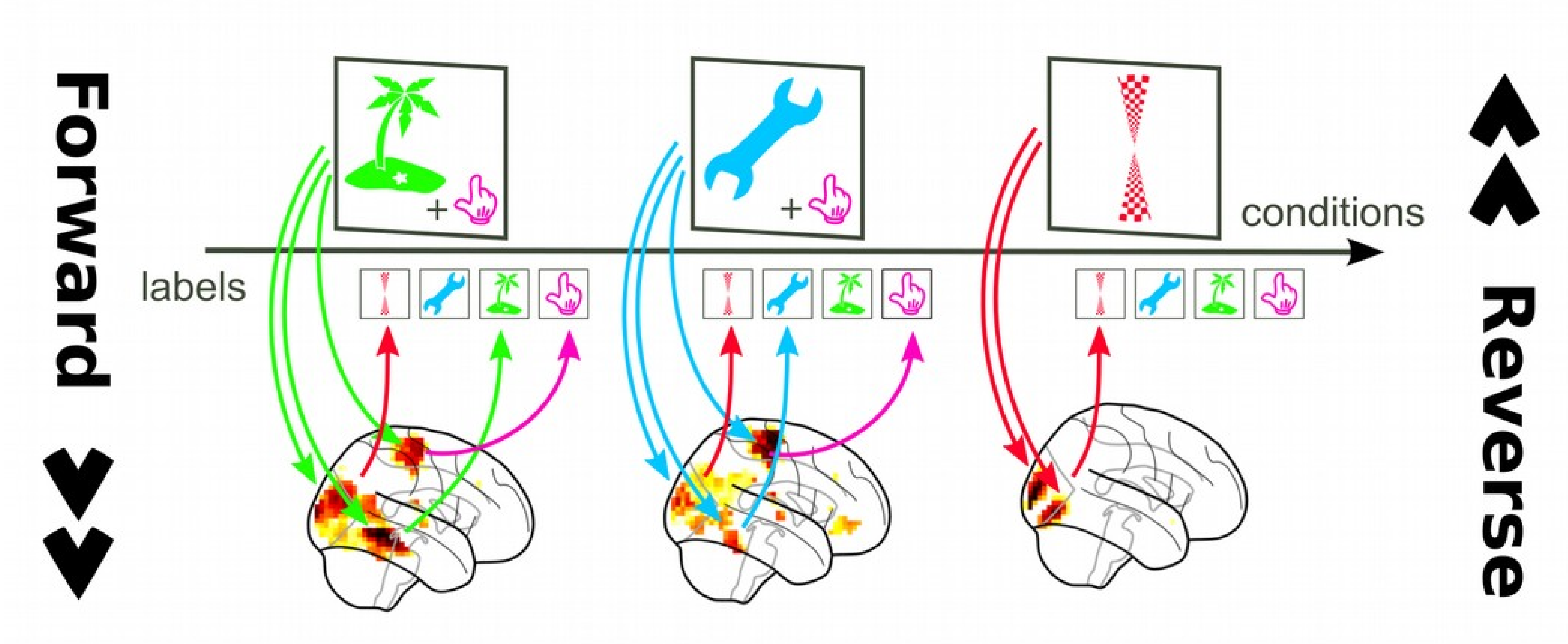
Resulting atlas



Joint decoding aka Mega-analyses

[Costafreda et al. Front. Neuroinf. 2009]

Indentifying cognitive tasks in brain activity



An image database

Task fMRI repository
[Gorgolewski et al. 2015]

Currently 48k independent usable fMRIs

[Poldrack 2011], knowledge-base

- *concepts*: cognitive activity/state (e.g. working memory)
- *tasks*: standard experiment to probe it (e.g. n-back task)

Forward inference

What is the brain response common to these stimuli?

Which regions are recruited by tasks containing a given term?

- Linear Model for terms effects

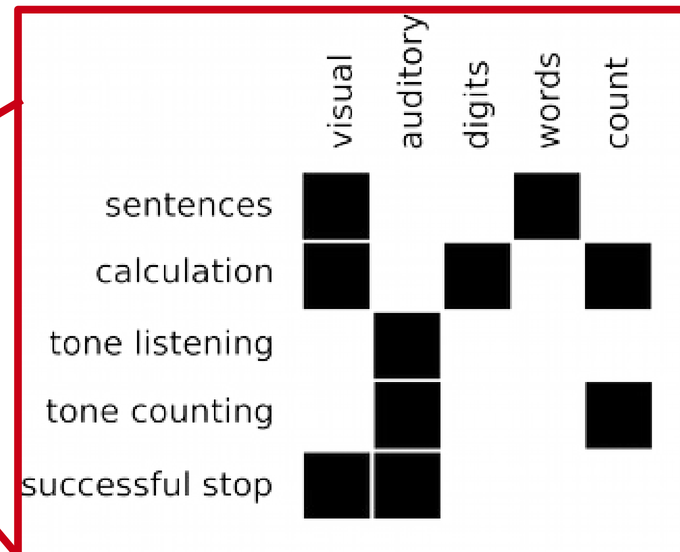
$$X = Y\beta + \varepsilon$$

X Conditions images

Y Design matrix

β Terms effect

ε Error

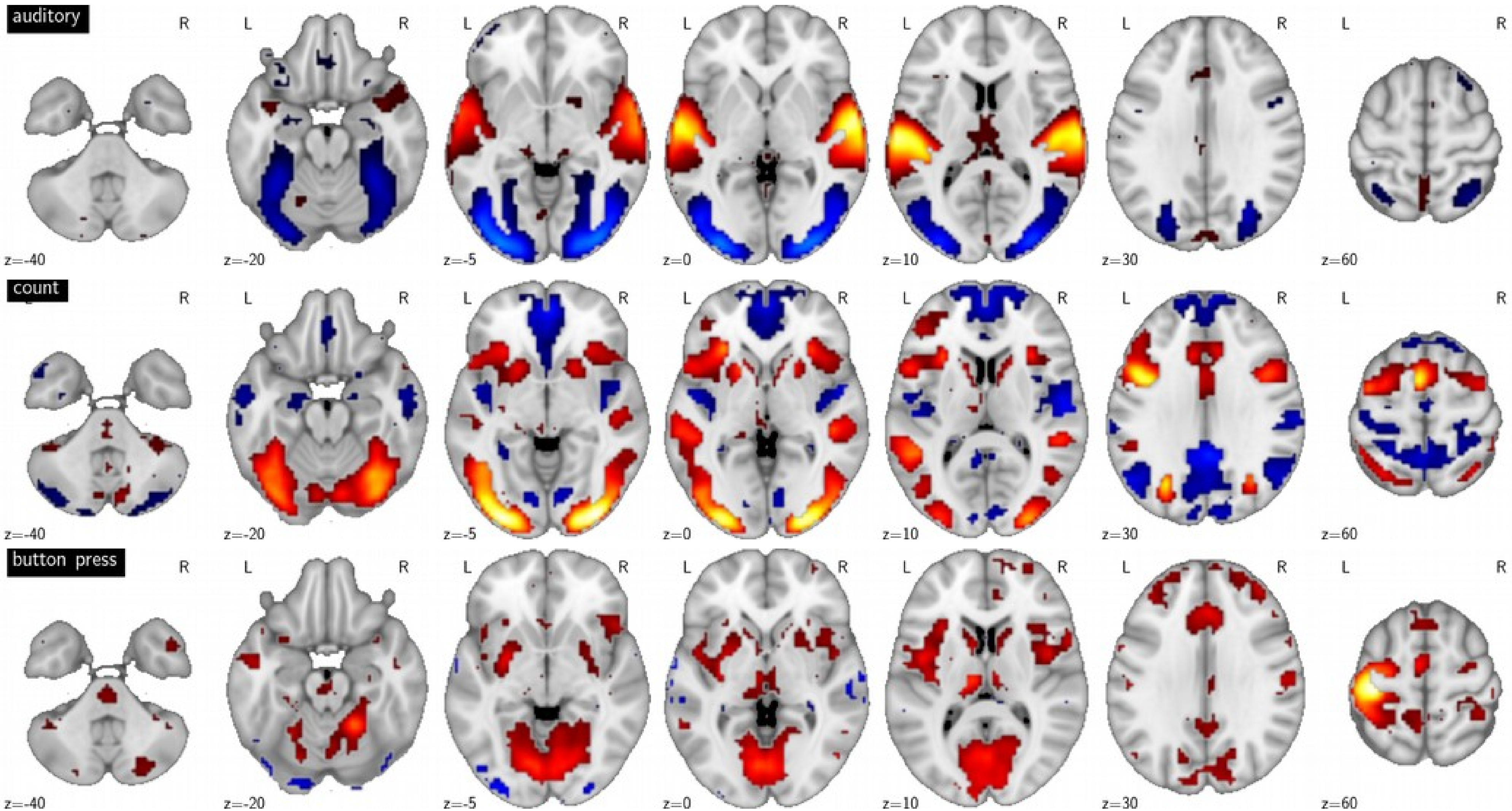


Forward inference

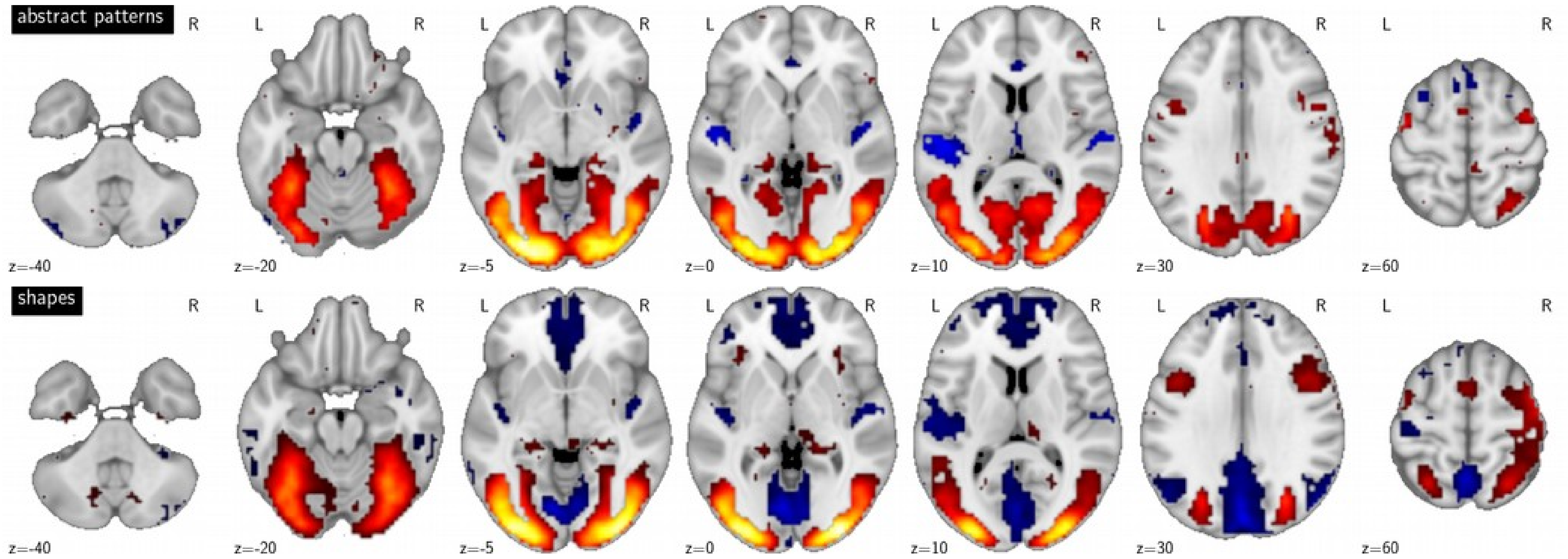
Correlation of the design matrix: difficulties from the heavily correlated terms (database bias)

	visual	auditory	digits	face	patterns	scramble	saccades	none	button press	count	inhibit	discriminate	read	move	track	sounds	shapes	attend	words
visual	+1.0	-0.9	+0.0	+0.1	+0.1	+0.1	+0.1	-0.1	+0.1	-0.1	-0.0	+0.2	+0.4	-0.2	+0.1	-0.3	+0.2	-0.3	-0.1
auditory	-0.9	+1.0	-0.0	-0.1	-0.2	-0.1	-0.1	+0.1	-0.2	+0.0	+0.2	-0.3	-0.2	+0.2	-0.1	+0.4	-0.2	+0.3	+0.2
digits	+0.0	-0.0	+1.0	-0.1	-0.1	-0.1	-0.1	+0.3	-0.3	+0.8	-0.1	-0.2	-0.1	-0.2	-0.1	-0.1	-0.1	-0.2	-0.4
face	+0.1	-0.1	-0.1	+1.0	-0.1	-0.0	-0.0	+0.1	-0.1	-0.0	-0.0	-0.0	-0.1	-0.1	-0.0	-0.0	-0.0	+0.2	-0.1
patterns	+0.1	-0.2	-0.1	-0.1	+1.0	-0.1	-0.1	-0.0	+0.1	-0.1	+0.1	+0.3	-0.2	-0.2	-0.1	+0.0	-0.1	+0.1	-0.4
scramble	+0.1	-0.1	-0.1	-0.0	-0.1	+1.0	-0.0	-0.0	+0.0	-0.1	-0.0	+0.1	-0.1	-0.1	-0.0	-0.1	-0.1	+0.1	-0.2
saccades	+0.1	-0.1	-0.1	-0.0	-0.1	-0.0	+1.0	-0.2	-0.1	-0.1	-0.0	-0.1	-0.1	-0.1	+0.9	-0.1	+0.5	-0.1	-0.2
none	-0.1	+0.1	+0.3	+0.1	-0.0	-0.0	-0.2	+1.0	-0.9	+0.2	-0.0	-0.5	+0.1	-0.5	-0.2	+0.1	-0.1	+0.6	-0.2
button press	+0.1	-0.2	-0.3	-0.1	+0.1	+0.0	-0.1	-0.9	+1.0	-0.2	-0.1	+0.6	-0.1	+0.5	-0.1	-0.2	+0.0	-0.5	+0.2
count	-0.1	+0.0	+0.8	-0.0	-0.1	-0.1	-0.1	+0.2	-0.2	+1.0	-0.1	-0.1	-0.2	-0.1	-0.1	+0.0	-0.1	-0.2	-0.4
inhibit	-0.0	+0.2	-0.1	-0.0	+0.1	-0.0	-0.0	-0.0	-0.1	-0.1	+1.0	+0.1	+0.1	-0.1	-0.0	+0.5	-0.1	-0.1	+0.0
discriminate	+0.2	-0.3	-0.2	-0.0	+0.3	+0.1	-0.1	-0.5	+0.6	-0.1	+0.1	+1.0	+0.0	-0.2	-0.1	-0.0	+0.1	-0.4	-0.1
read	+0.4	-0.2	-0.1	-0.1	-0.2	-0.1	-0.1	+0.1	-0.1	-0.2	+0.1	+0.0	+1.0	-0.2	-0.1	-0.0	-0.2	-0.4	+0.5
move	-0.2	+0.2	-0.2	-0.1	-0.2	-0.1	-0.1	-0.5	+0.5	-0.1	-0.1	-0.2	-0.2	+1.0	-0.1	-0.1	-0.1	-0.3	+0.4
track	+0.1	-0.1	-0.1	-0.0	-0.1	-0.0	+0.9	-0.2	-0.1	-0.1	-0.0	-0.1	-0.1	-0.1	+1.0	-0.1	+0.5	-0.1	-0.2
sounds	-0.3	+0.4	-0.1	-0.0	+0.0	-0.1	-0.1	+0.1	-0.2	+0.0	+0.5	-0.0	-0.0	-0.1	-0.1	+1.0	-0.1	+0.2	-0.2
shapes	+0.2	-0.2	-0.1	-0.0	-0.1	-0.1	+0.5	-0.1	+0.0	-0.1	-0.1	+0.1	-0.2	-0.1	+0.5	-0.1	+1.0	+0.0	-0.3
attend	-0.3	+0.3	-0.2	+0.2	+0.1	+0.1	-0.1	+0.6	-0.5	-0.2	-0.1	-0.4	-0.4	-0.3	-0.1	+0.2	+0.0	+1.0	-0.2
words	-0.1	+0.2	-0.4	-0.1	-0.4	-0.2	-0.2	-0.2	+0.2	-0.4	+0.0	-0.1	+0.5	+0.4	-0.2	-0.2	-0.3	-0.2	+1.0

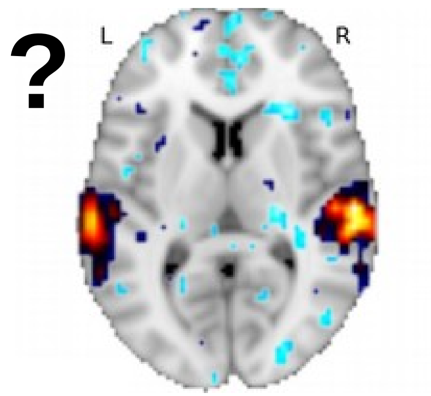
Results of forward inference



Limittaions of forward inference



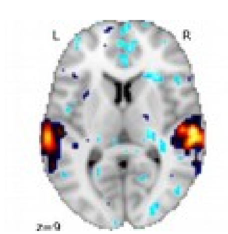
Open-ended brain decoding



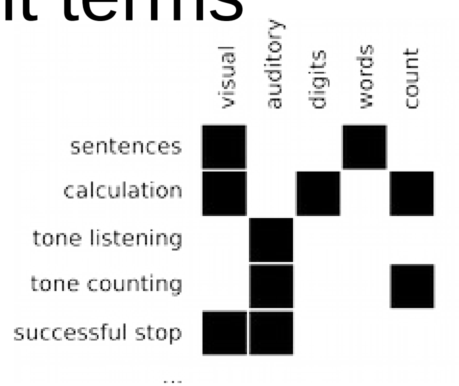
What is
this brain doing?

Which regions are predictive of tasks containing a given term?

- **Multilabel** classification problem
  - more than one class may be associated with each sample
- Predict occurrence of frequent terms



Data: experimental condition images



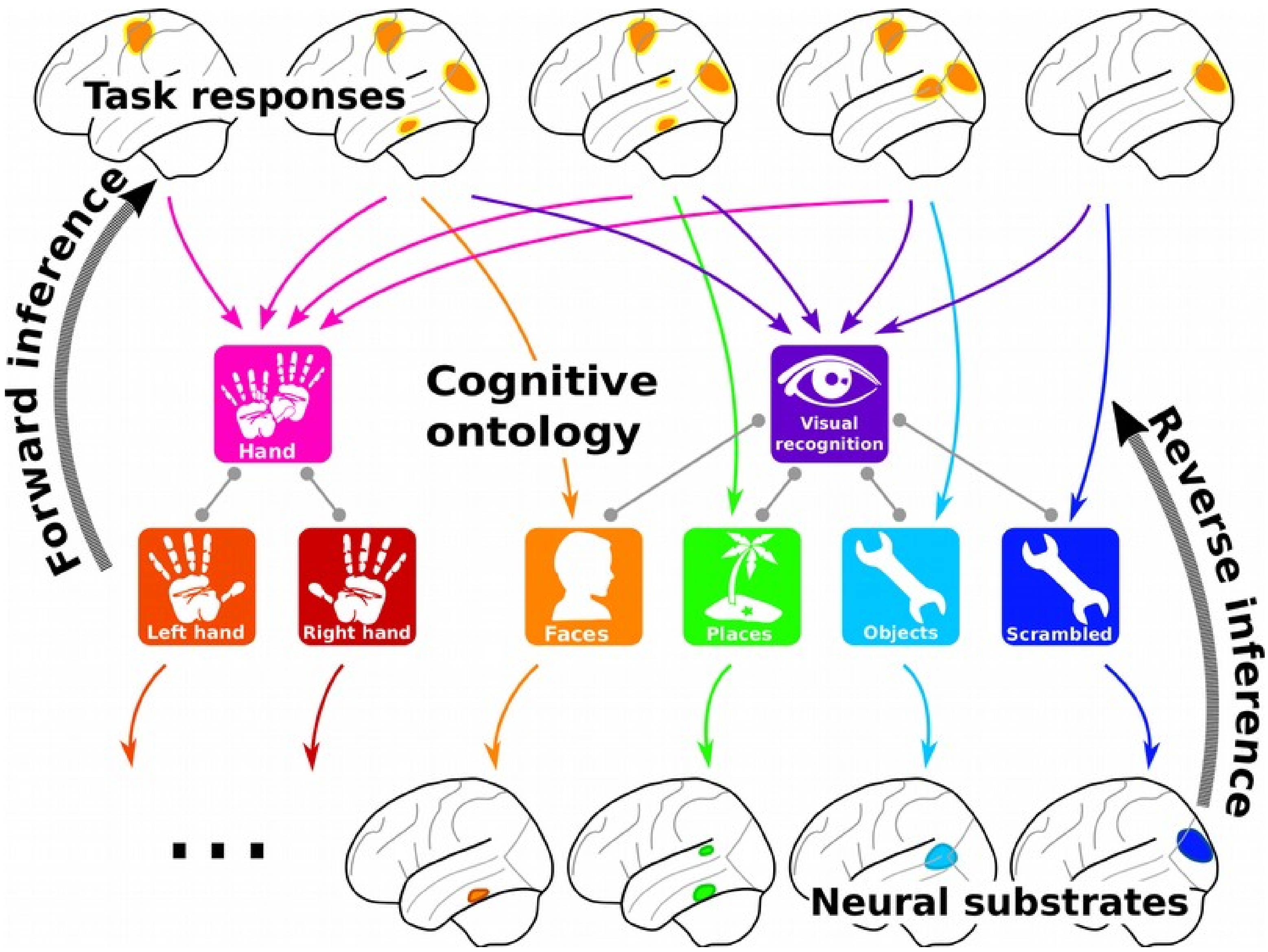
Target

Beyond cognitive tasks: decoding concepts

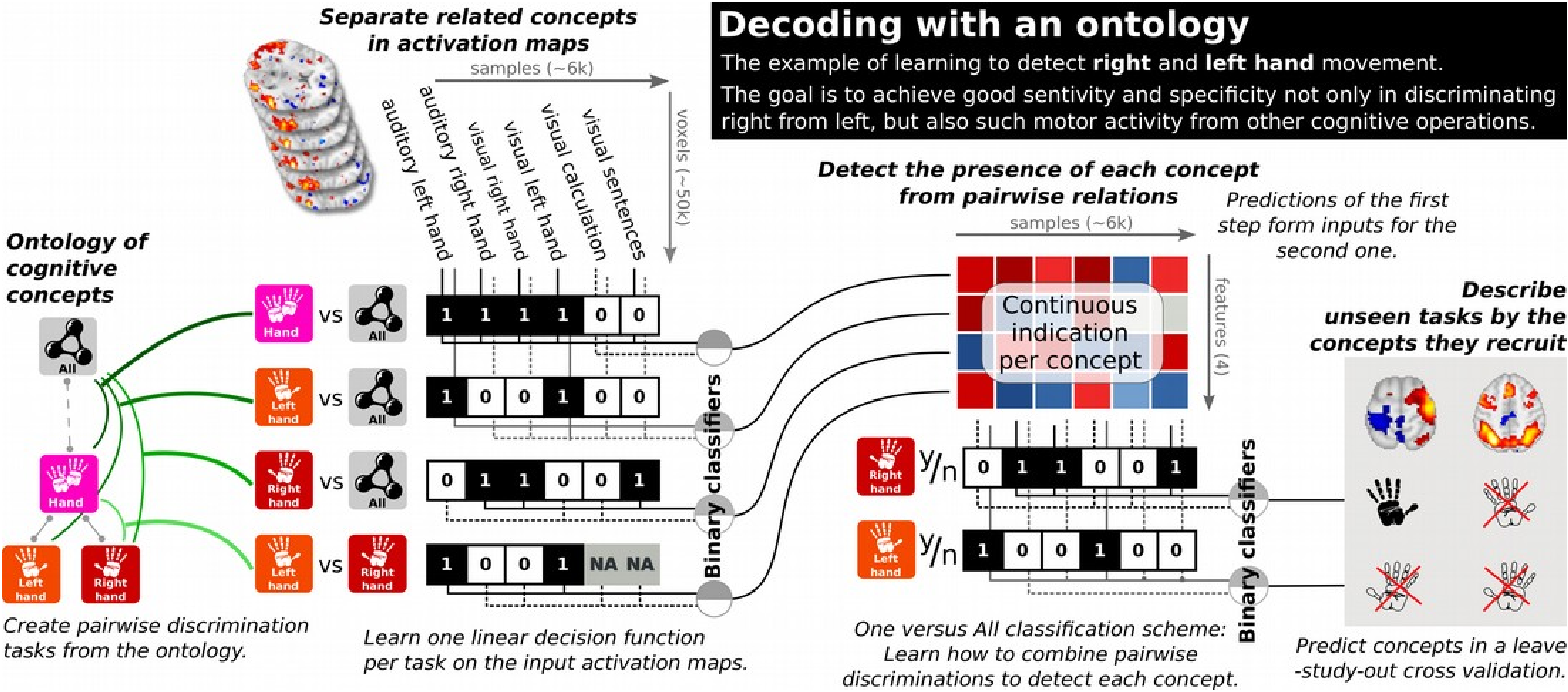
Task Level

Concept Level

Cognitive modules

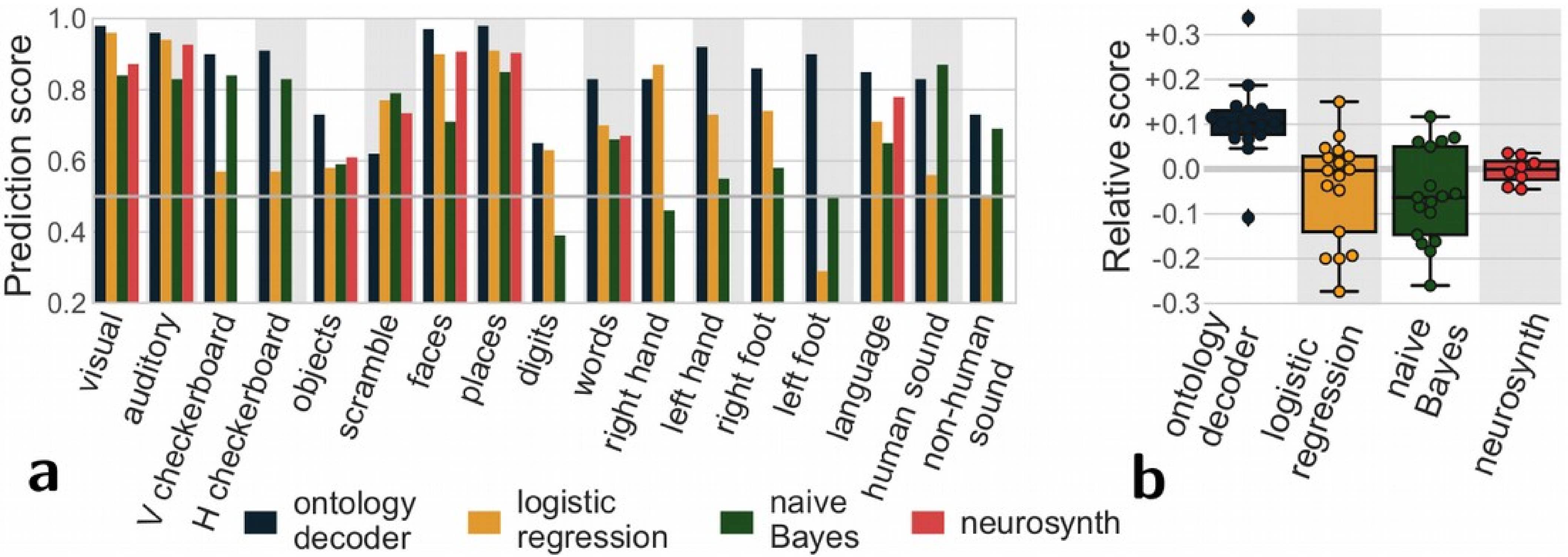


A classifier for mega-analyses



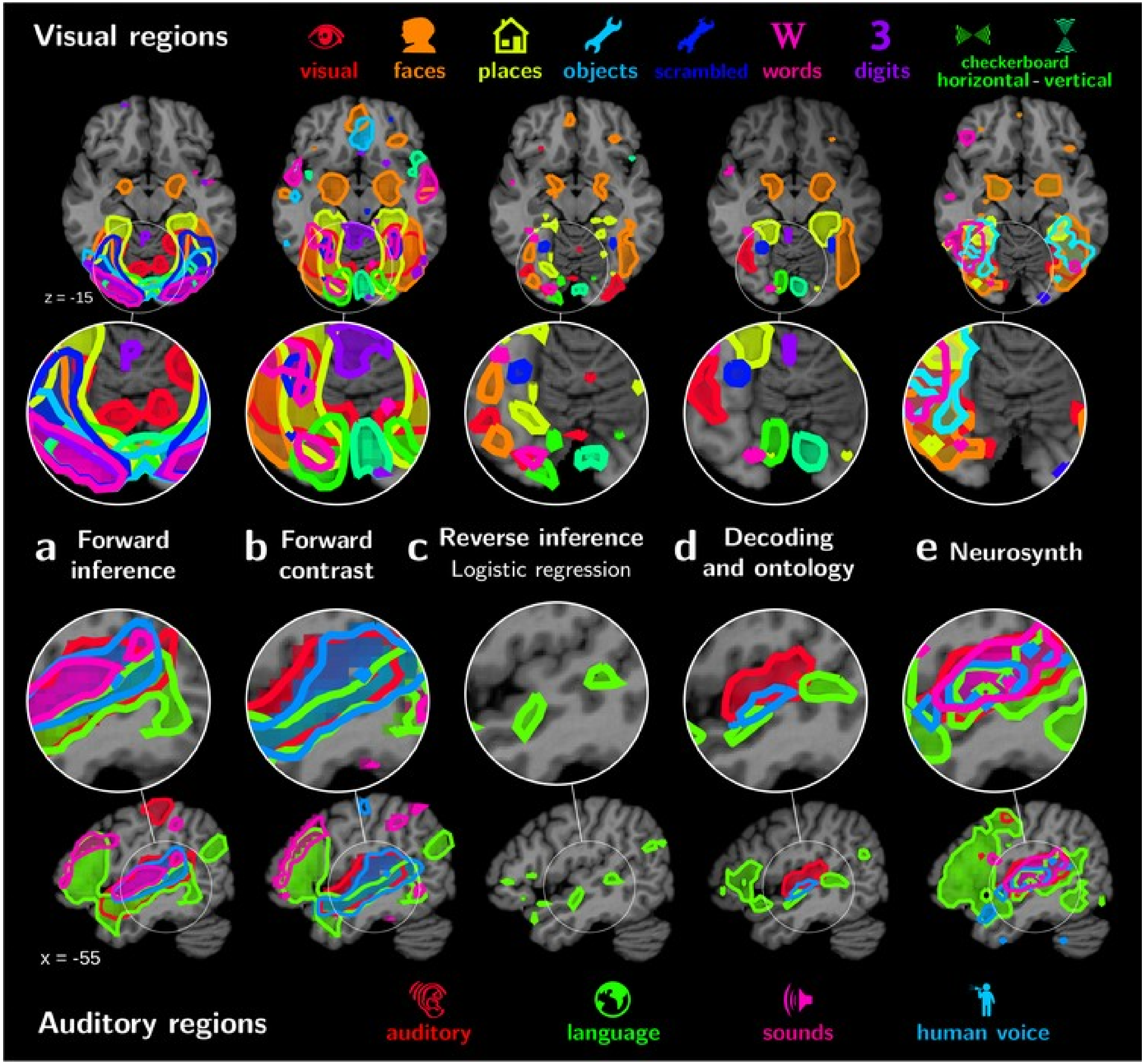
[Schwartz et al. NIPS 2013, Varoquaux et al. PCB 2018]

Classification results

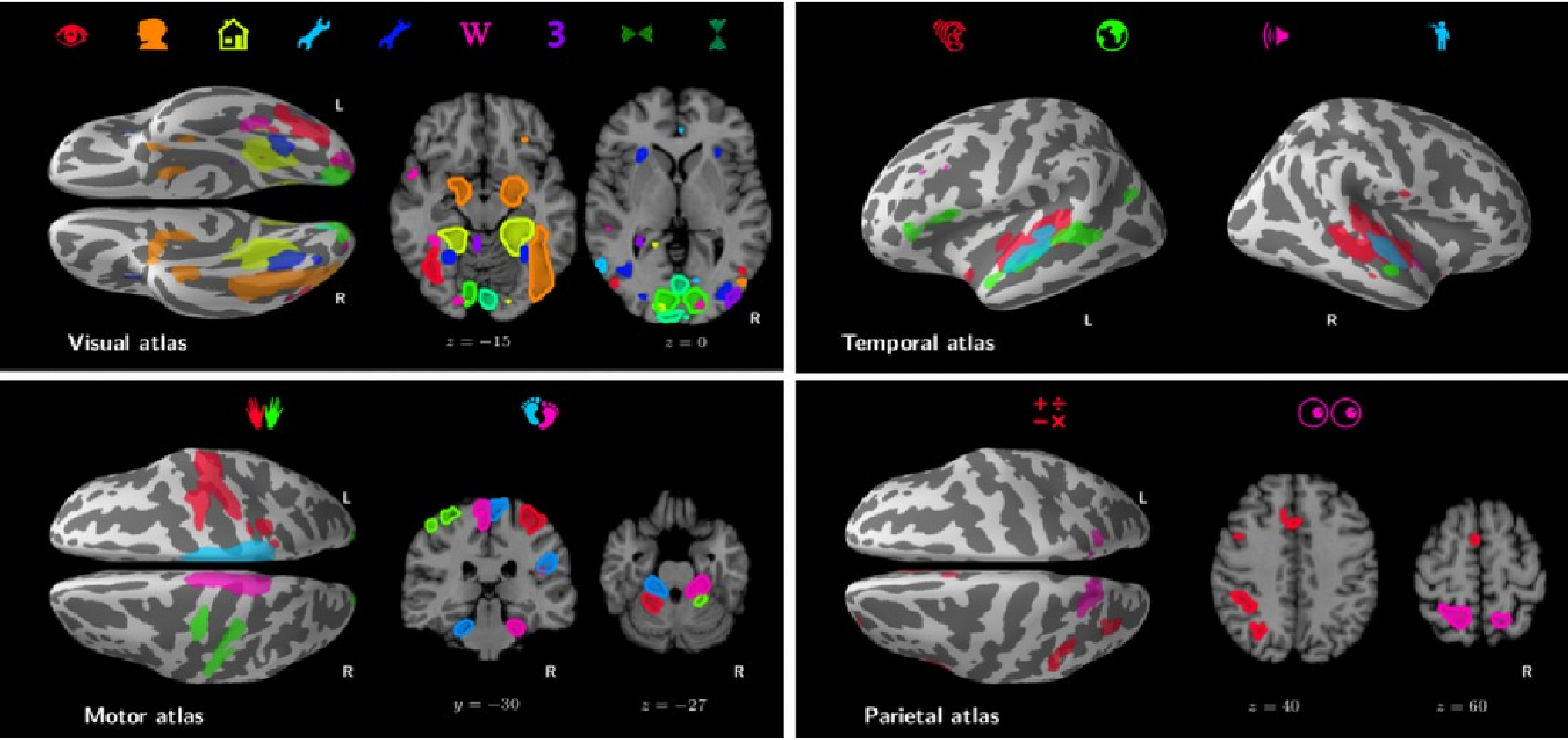


[Schwartz et al. NIPS 2013, Varoquaux et al. PCB 2018]

Decoding maps



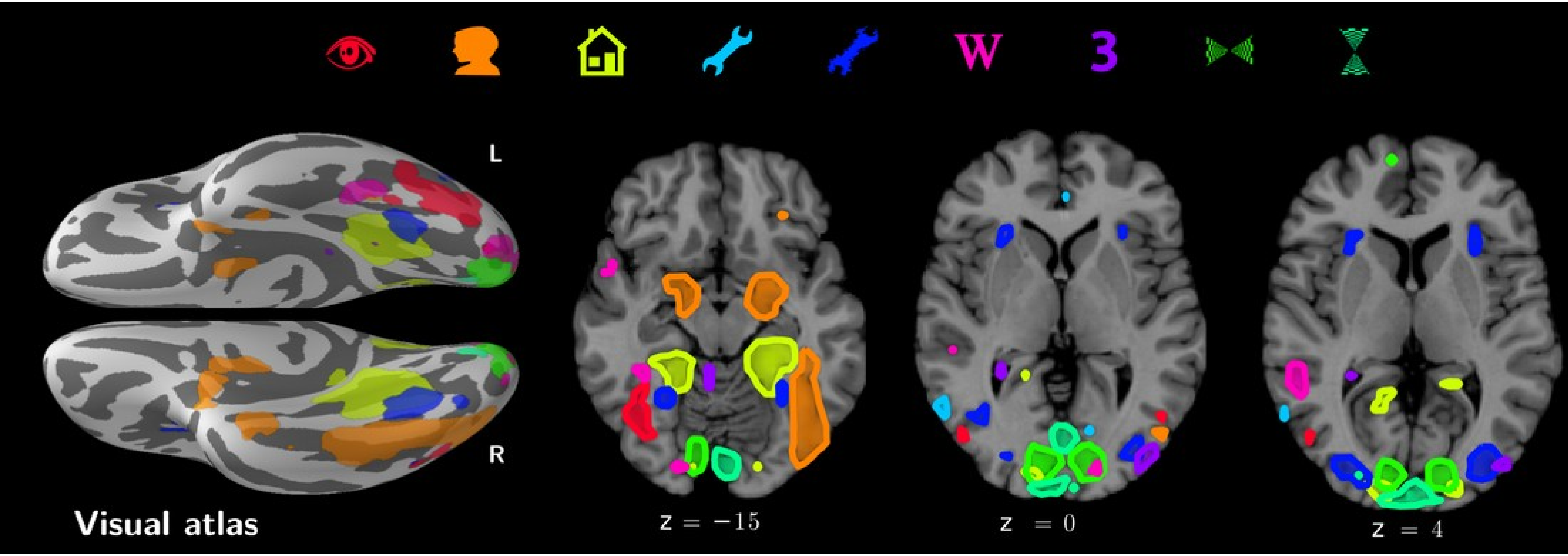
Discriminative patterns



[Schwartz et al. NIPS 2013, Varoquaux et al. PCB 2018.]

Predicting versus identifying

- Unlike BCI, classification performance is not the main goal
- The task you're interested in is unsupervised:
  - Understand the functional specificity of brain regions



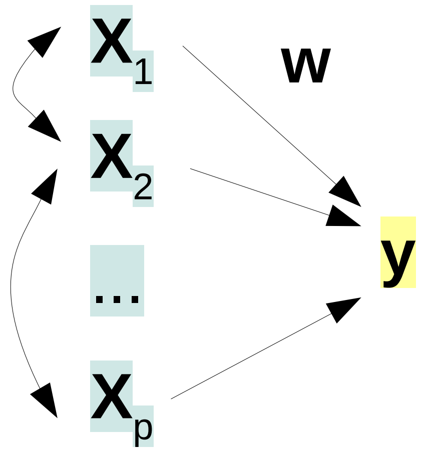
Outline

- Why is decoding important ? A causal persepective
- Intermezzo: Learning representations for brain images with dictionary learning
- Large-scale brain activity decoding
  - coordinated decoding
  - Joint decoding
- Decoding an in inverse problem: what can we learn from it ?
  - Identifiability issues
  - Statistical control

(Non-)identifiability of the model ?

- When solving the inverse problem, you don't recover the true pattern but a poor approximation thereof [Haufe et al. Neuroimage 2013]
- True model: $\mathbf{y} = \mathbf{X}\,\mathbf{w}_0 + e$
- Estimated model: $\hat{\mathbf{w}} = \mathrm{argmin}_{\mathbf{w}}(l(\mathbf{w}) + R(\mathbf{w}))$
- $\hat{\mathbf{w}}$ cannot be equal to $\mathbf{w_0}$, as $\mathbf{X}$ is non-invertible

Brain activity decoding



- behavior = f (brain activity)

$$\mathbf{y} = \mathbf{X}\boldsymbol{w}^* + \sigma_* \varepsilon$$

- error vector: $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$
- noise magnitude: $\sigma_* > 0$

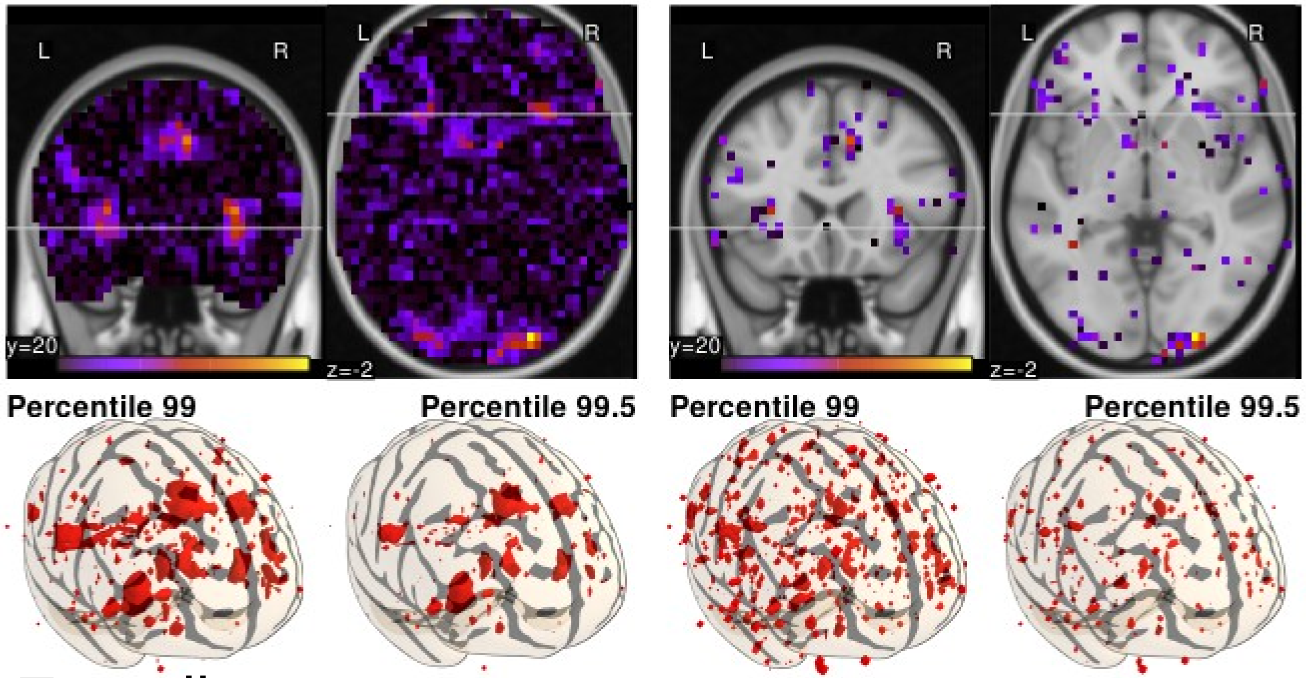
- prediction: find $\hat{\boldsymbol{w}}$ that minimizes $\|\mathbf{X}\hat{\boldsymbol{w}} - \mathbf{X}\boldsymbol{w}^*\|_2$
- estimation: find $\hat{\boldsymbol{w}}$ with control on $|\hat{w}_j - w_j^*|$ for all $j \in [p]$

(Non-)identifiability of the model ?

- Can $\hat{\mathbf{w}}$ have at least the correct support ?
- No: the encoding model violates the conditions for accurate reconstruction [Varoquaux et al. 2012]
- Better support recovery by introducing relevant priors on the decoder [Varoquaux et al. 2012]
  - Sparsity
  - Small variations

(Non-)identifiability of the model

Can nevertheless be improved with adapted techniques



Encoding

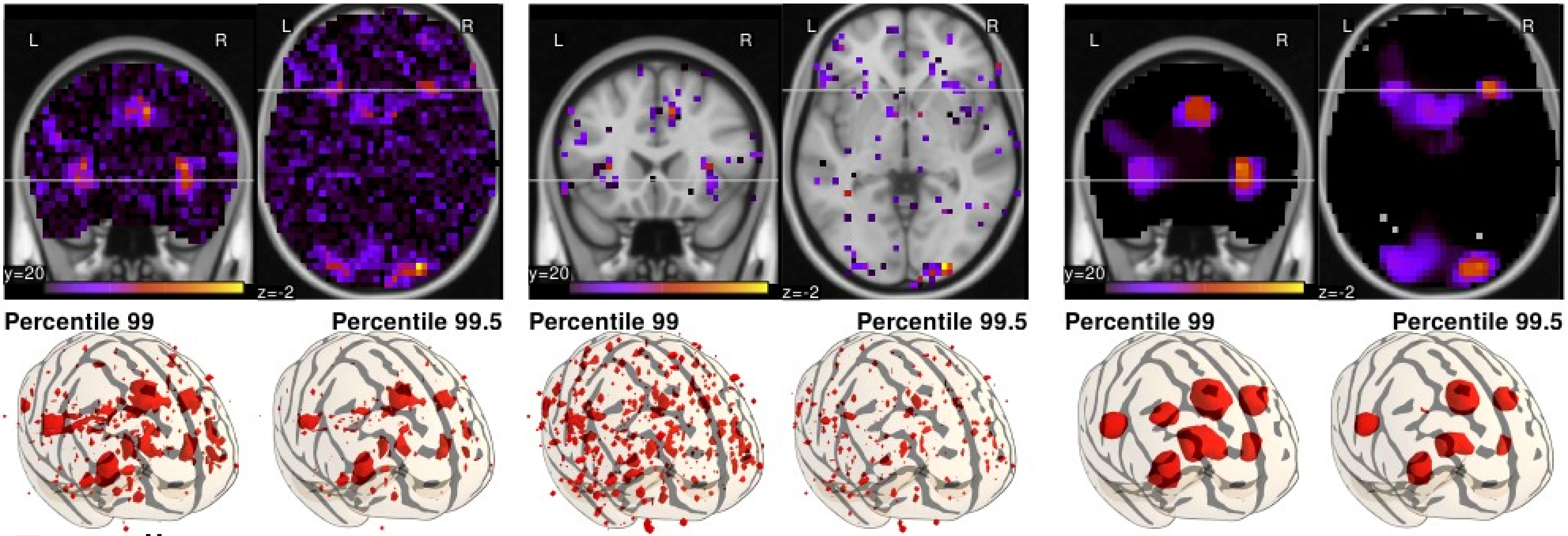
Elastic net decoding

$$(\hat{\mathbf{w}}) = \operatorname{argmin}_{\mathbf{w}} \ell(\mathbf{X}, \mathbf{Y}\mathbf{w}) + \lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \|\mathbf{w}\|_2^2$$

[Gramfort et al PRNI 2013]

(Non-)identifiability of the model

Can nevertheless be improved with adapted techniques



Encoding

Elastic net decoding

Piecewise cst decoding

$$(\hat{\mathbf{w}}) = \text{argmin}_{\mathbf{w}} \ell(\mathbf{X}, \mathbf{Y}\mathbf{w}) + \lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \|\mathbf{w}\|_2^2$$

$$(\hat{\mathbf{w}}) = \text{argmin}_{\mathbf{w}} \ell(\mathbf{X}, \mathbf{Y}\mathbf{w}) + \lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \|\nabla \mathbf{w}\|_1$$

[Gramfort et al PRNI 2013]

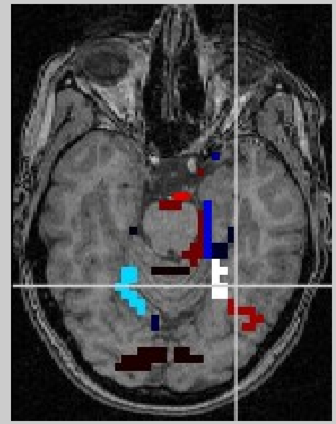
Bagging of sparse clustered models

X

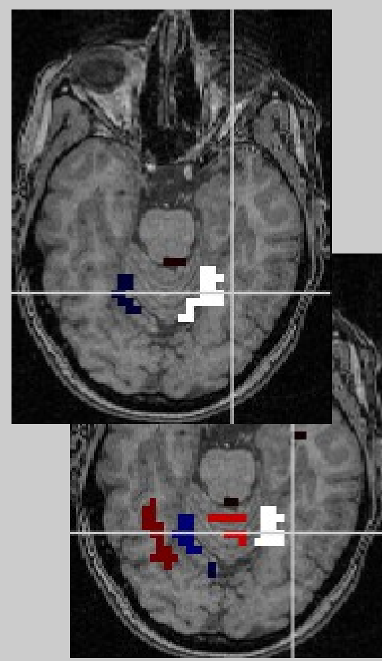
y

Clustering (create contiguous regions)

Solve **Lasso** on cluster-based representation

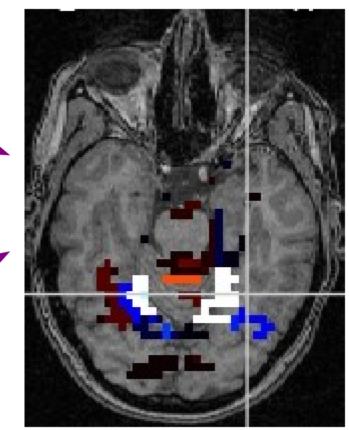


...



various clusterings

average

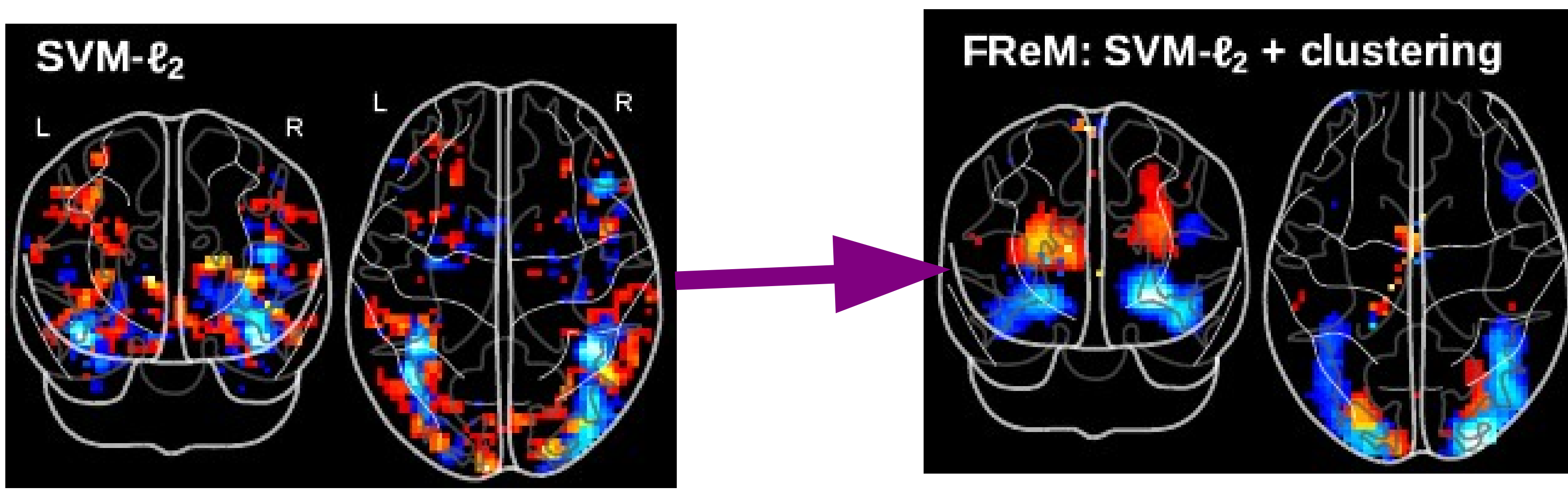


[Varoquaux et al ICML 2012
Hoyos idrobo et al. Nimg 2017]

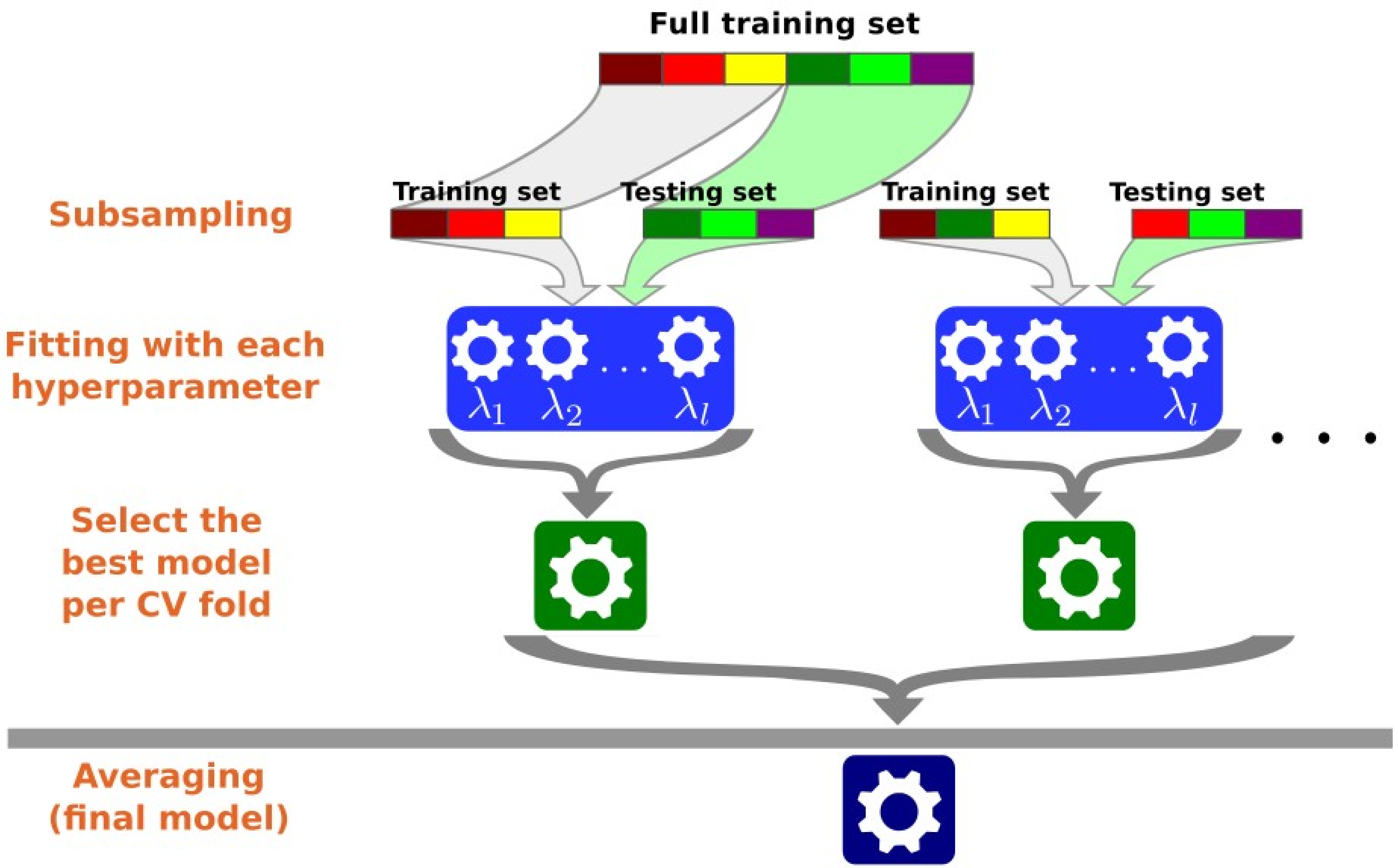
Computationally efficient structure

“fast regularized ensembles of models”

State of the art solution: not very stable, but cheap



Computationally efficient structure

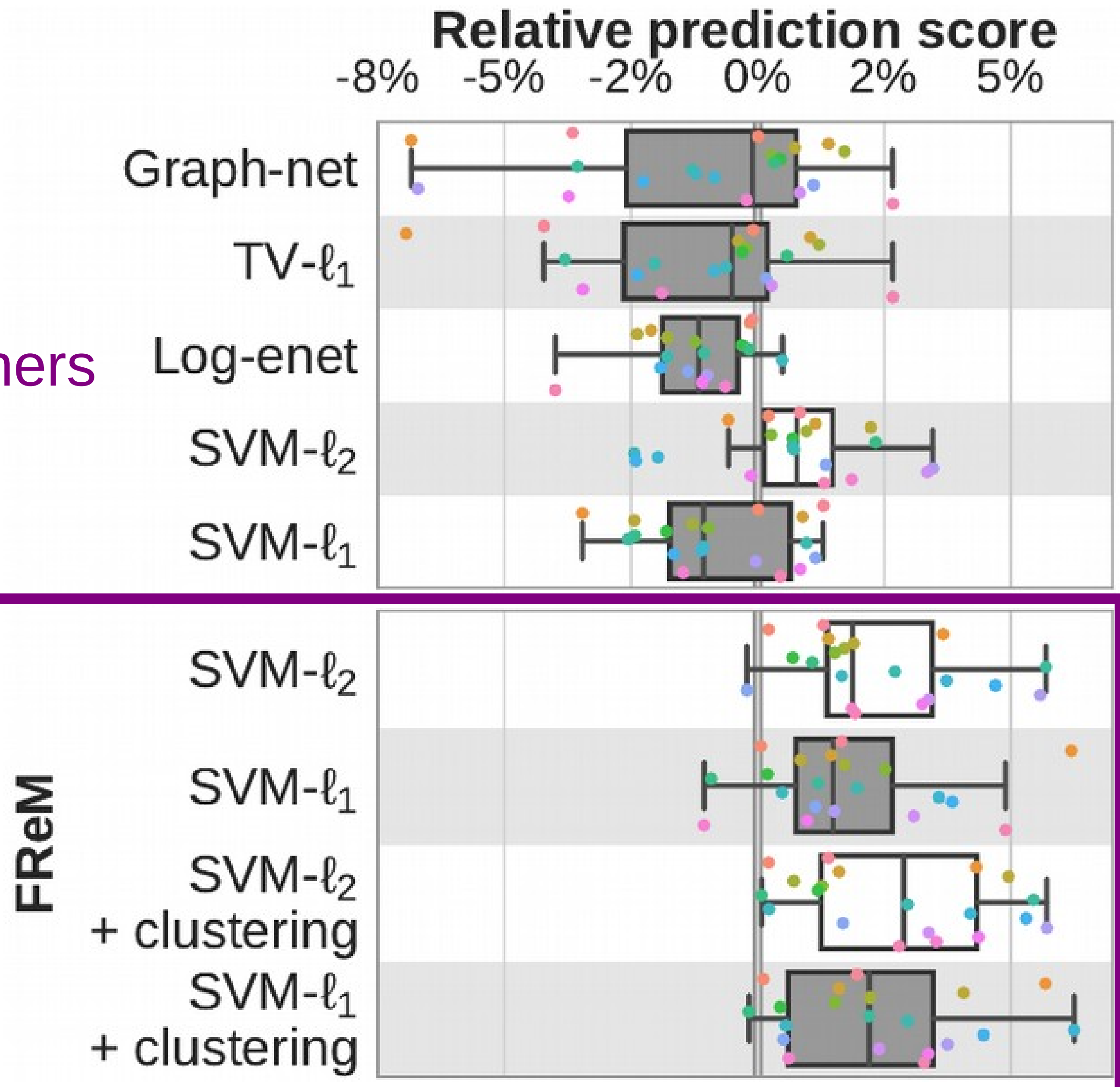


Effect on prediction accuracy

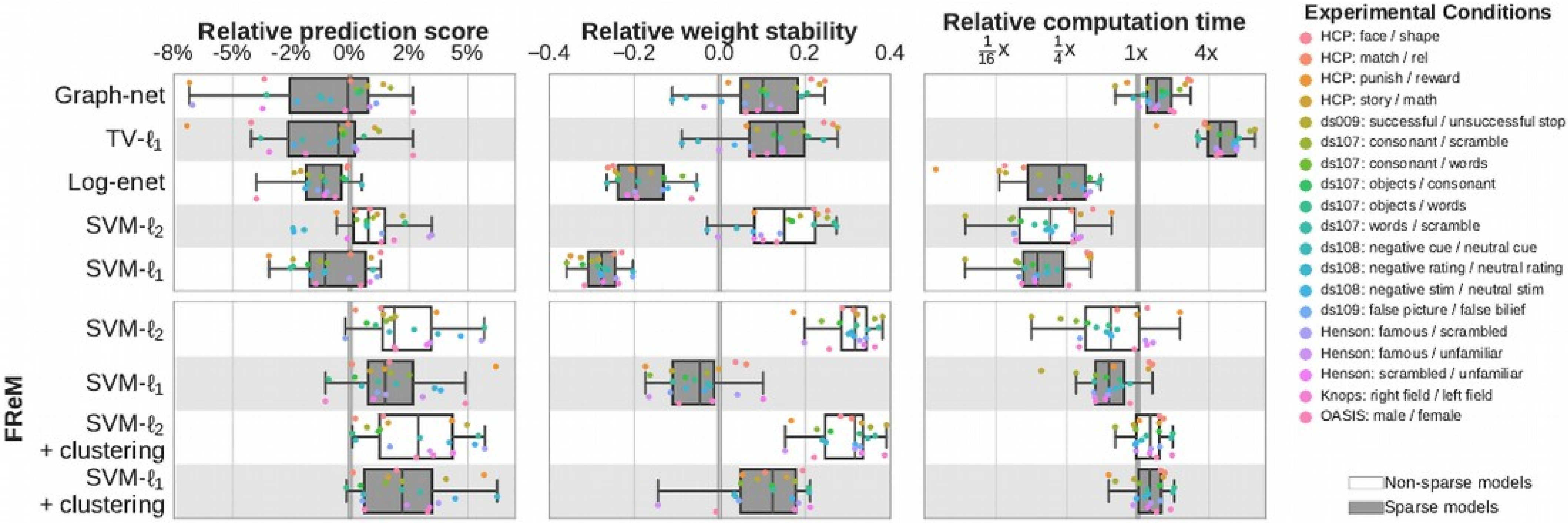
[Hoyos Idrobo et al PRNI 2015, Neuroimage 2017, PAMI 2018]

Convex learners

"fast regularized ensembles of models"

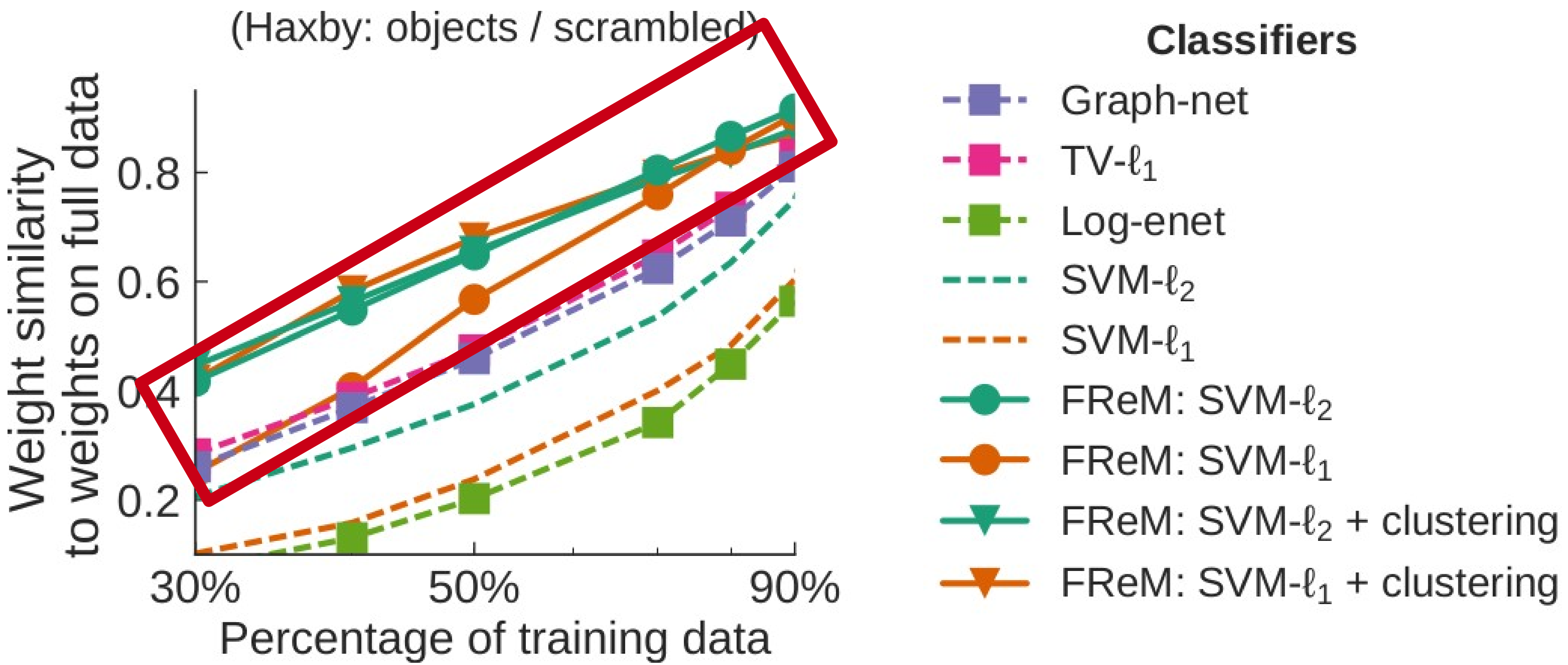


More results



[Hoyos Idrobo et al PRNI 2015, Neuroimage 2017, PAMI 2018]

Learning curve



[Hoyos Idrobo et al PRNI 2015, Neuroimage 2017]

Good recovery is great, but could we have probabilistic guarantees on the solution ?

Statistical inference on w

- Inference: find {j: $\mathbf{w}_j > 0$} with some statistical guarantees

- Standard solutions for high-dimensional linear models (p > n)
  - Corrected ridge [Bühlmann 2013]
  - Desparsified Lasso [Zhang & Zhang 2014, Montanari 2014]
  - Multi-split [Meinshausen 2009], knockoffs [Candès 2015+]
- **Fail for $p \gg n$**

Desparsified Lasso

- **Objective:** construct confidence bounds on the coefficients of $\boldsymbol{w}^*$
- **Principle:** [Zhang & Zhang 2014 Series B Stat Meth]
  - construct an unbiased estimator of $\boldsymbol{w}^*$ (generalization of $\hat{\boldsymbol{w}}^{\text{OLS}}$)
  - compute its covariance matrix
- **Heuristic argument:** in low dimension we can prove that:

$$\hat{w}_j^{\text{OLS}} = \frac{\mathbf{z}_j^\top \mathbf{y}}{\mathbf{z}_j^\top \mathbf{x}_j} \quad ,$$

where $\mathbf{z}_j$ is the residual of the OLS regression of $\mathbf{x}_j$ versus $\mathbf{X}^{(-j)}$:

$$\mathbf{z}_j = \mathbf{x}_j - \mathbf{P}_{\mathbf{X}^{(-j)}} \mathbf{x}_j \quad ,$$

where $\mathbf{P}_{\mathbf{X}^{(-j)}}$ is the projection onto $\text{Span}(\mathbf{X}^{(-j)}) \subset \mathbb{R}^{p-1}$

Desparsified Lasso

- **Desparsified Lasso estimator:** when $n < p$, $\mathbf{z}_j$ is the residual of a Lasso-CV regression of $\mathbf{x}_j$ vs $\mathbf{X}^{(-j)}$ and the debiased estimator is:

$$\hat{w}_j = \frac{\mathbf{z}_j^\top \mathbf{y}}{\mathbf{z}_j^\top \mathbf{x}_j} - \sum_{k \neq j} \frac{\mathbf{z}_j^\top \mathbf{x}_k \hat{w}_k^{(init)}}{\mathbf{z}_j^\top \mathbf{x}_j} \quad ,$$

  where $\hat{\boldsymbol{w}}^{(init)}$ is an initial non linear estimator of $\boldsymbol{w}^*$ (*e.g.*, Lasso)

- **Covariance:** the covariance matrix of this estimator is:

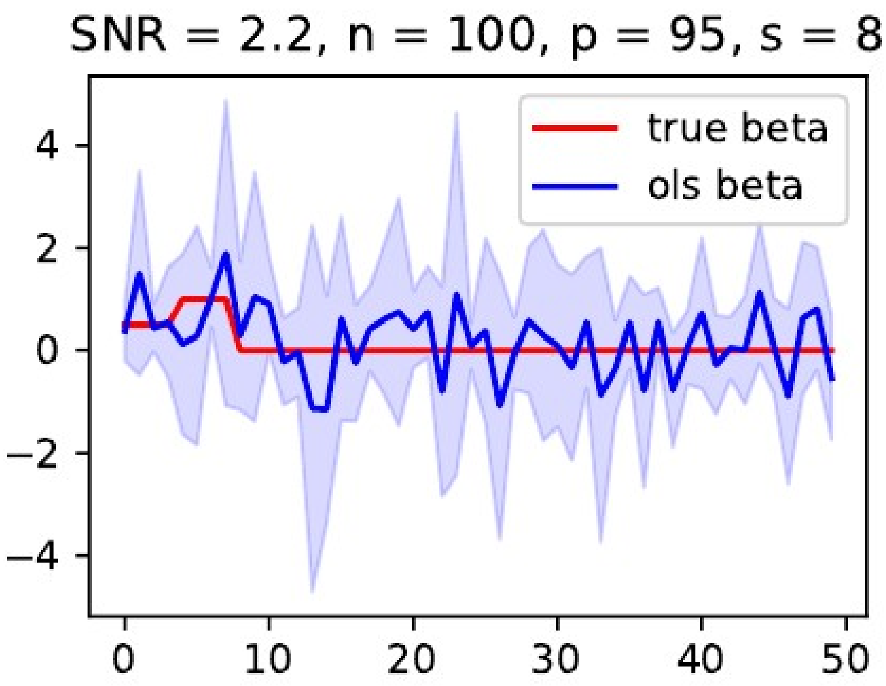
$$\Omega_{jk} = \frac{n \mathbf{z}_j^\top \mathbf{z}_k}{(\mathbf{z}_j^\top \mathbf{x}_j)(\mathbf{z}_k^\top \mathbf{x}_k)}$$

- **Confidence bounds:** under few assumptions (Dezeure et al. [2015]):

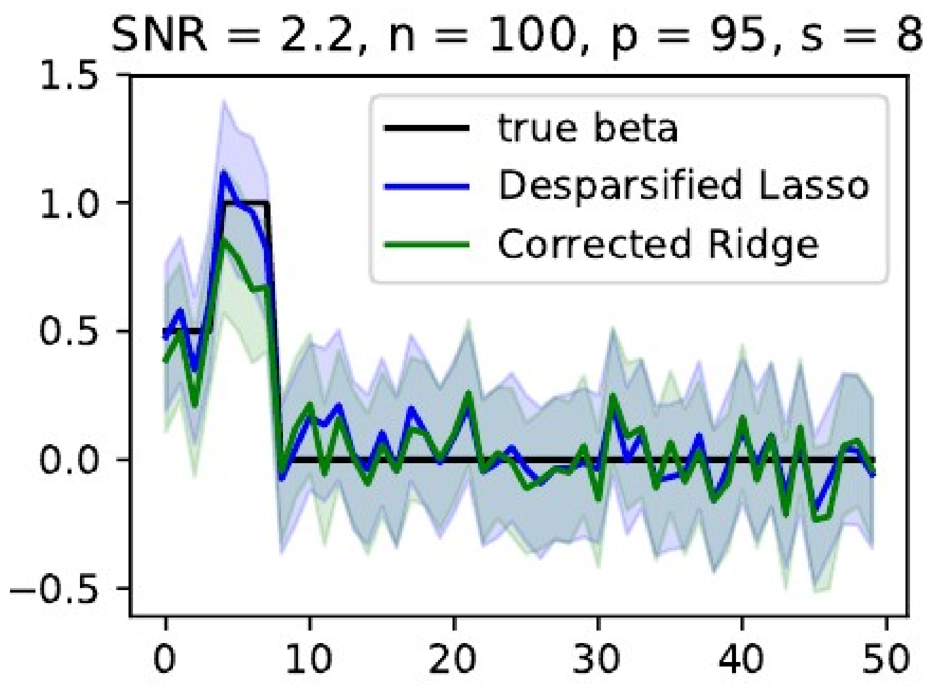
$$\boxed{\sigma_*^{-1}(\Omega_{jj})^{-1/2}(\hat{w}_j - w_j^*) \sim \mathcal{N}(0, 1)}$$

Preliminary assessment

- **Low dimension:** $n = 100$ and $p = 95$
- **OLS versus corrected Ridge and desparsified Lasso:**



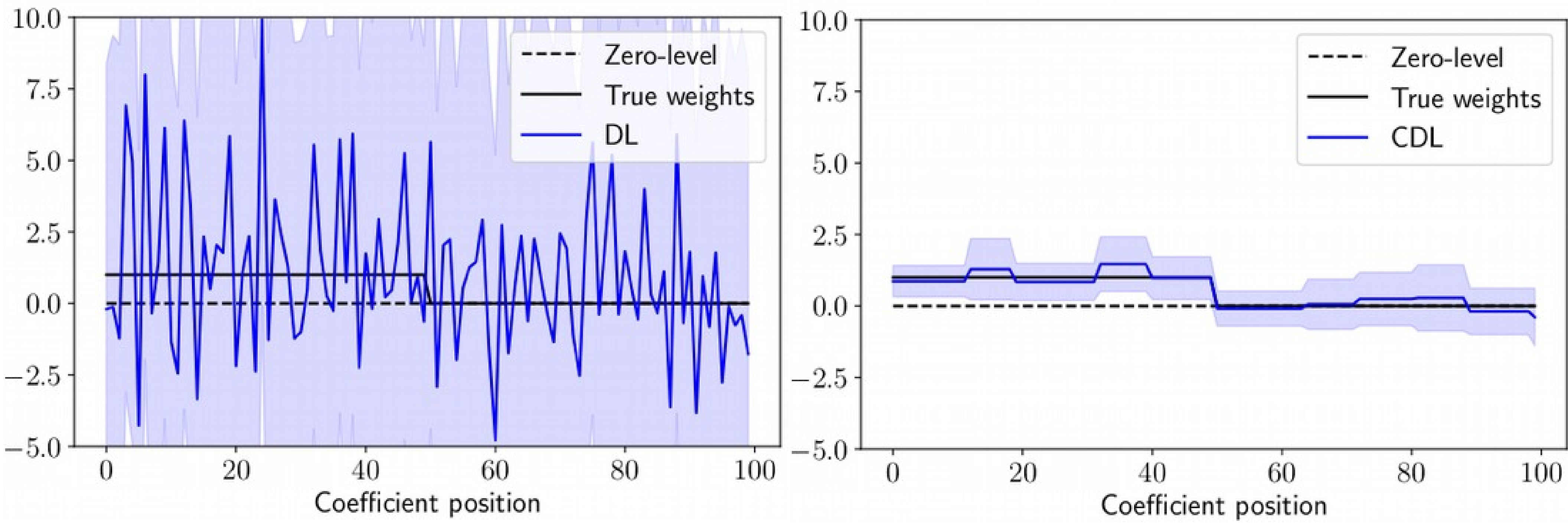
OLS regression when $p \approx n$



Corrected Ridge and **Desparsified Lasso** when $p \approx n$

Large p → need dimension reduction

p=2000, n=100



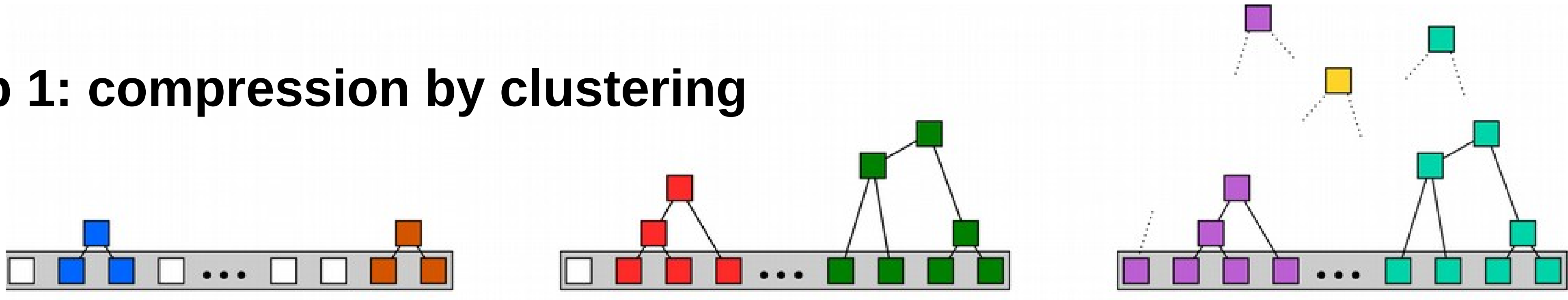
Large p kills statistical power

Clustering features tames the variance

[Chevalier et al. MICCAI 2018]

Adaptation to brain imaging

Step 1: compression by clustering



Step 2: inference on compressed representations

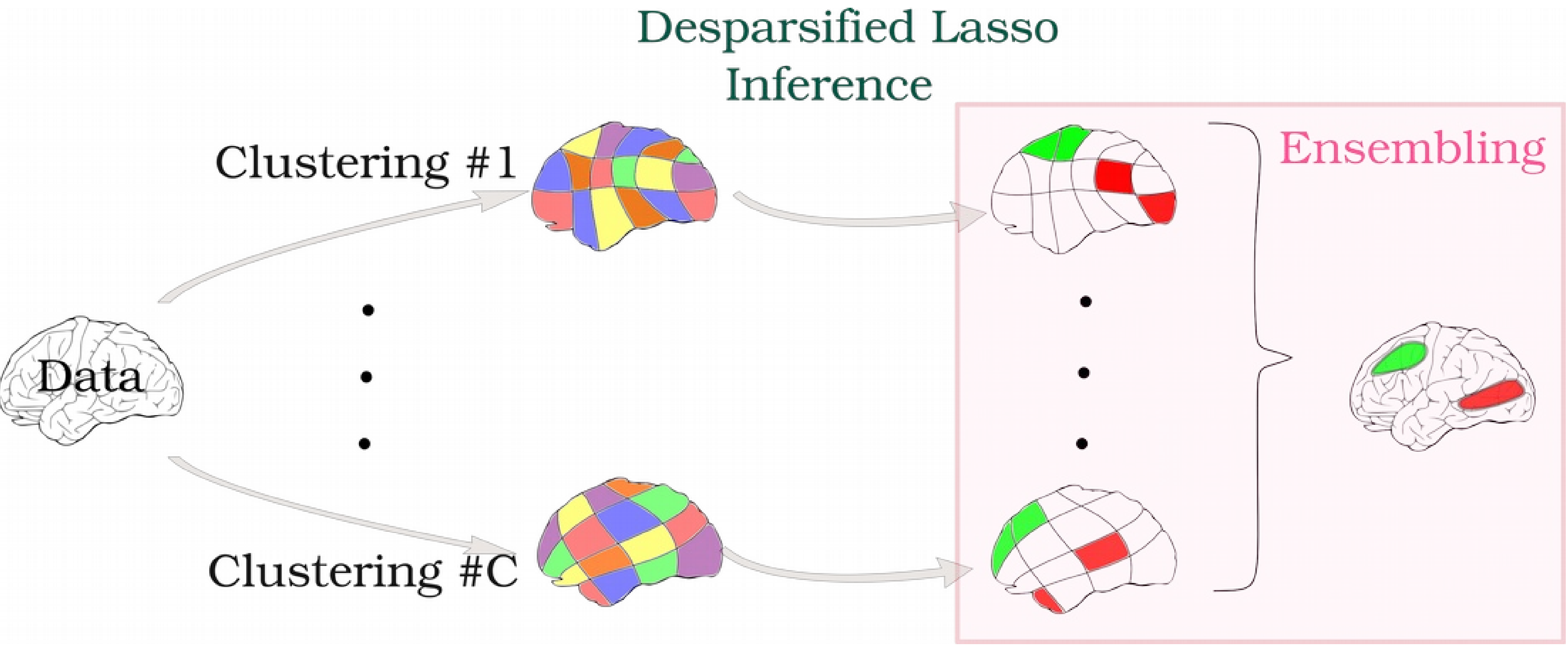
$$\sigma_*^{-1}(\Omega_{jj})^{-1/2}(\hat{w}_j - w_j^*) \sim \mathcal{N}(0,1)$$

***C**lustered*
***D**esparsified*
***L**asso*

Step 3: ensembling iterate with different parcellations
→ **aggregate p-values** (FReM-like approach)

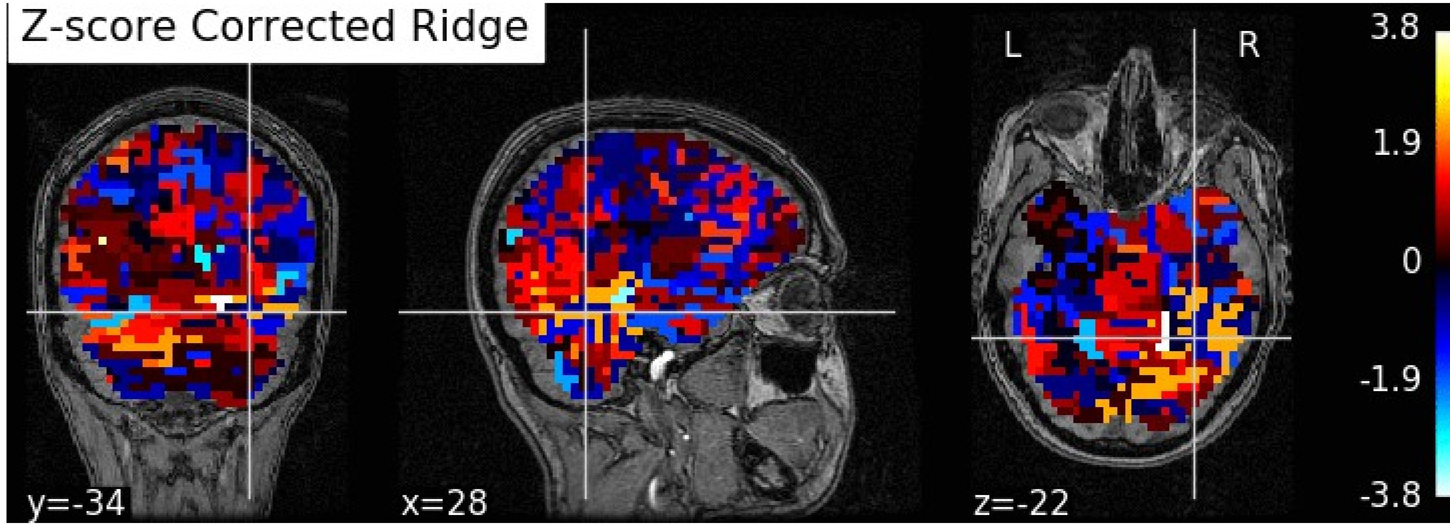
***E**nsemble of*
***C**lustered*
***D**esparsified*
***L**asso*

Adaptation to brain imaging

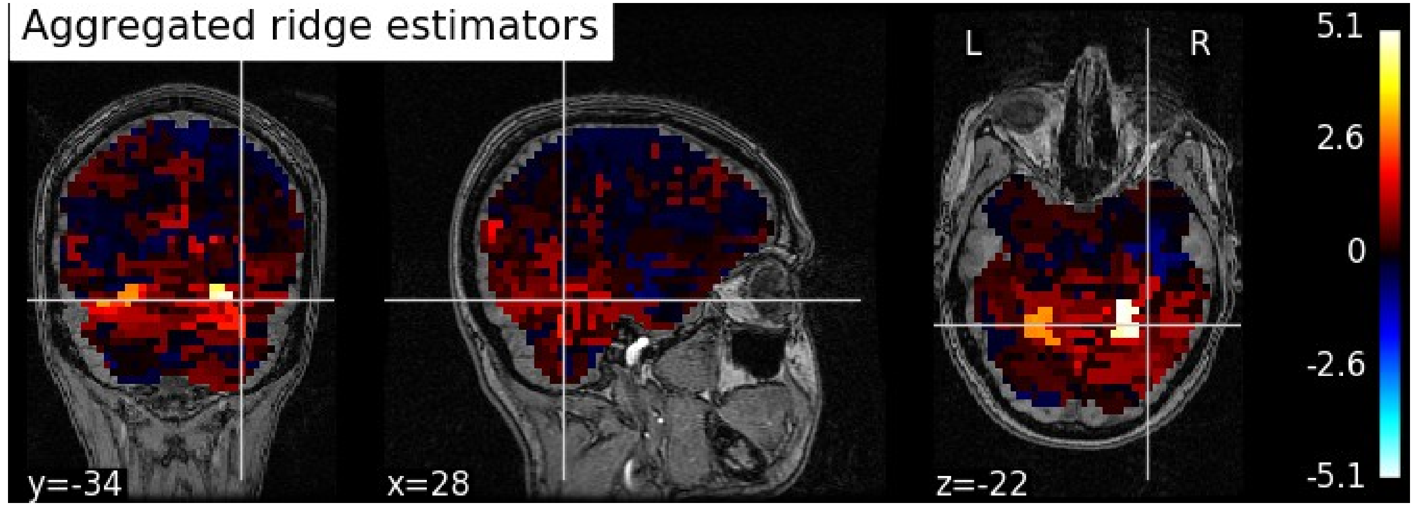


From CDL to ECDL

DL p-values from different clusterings



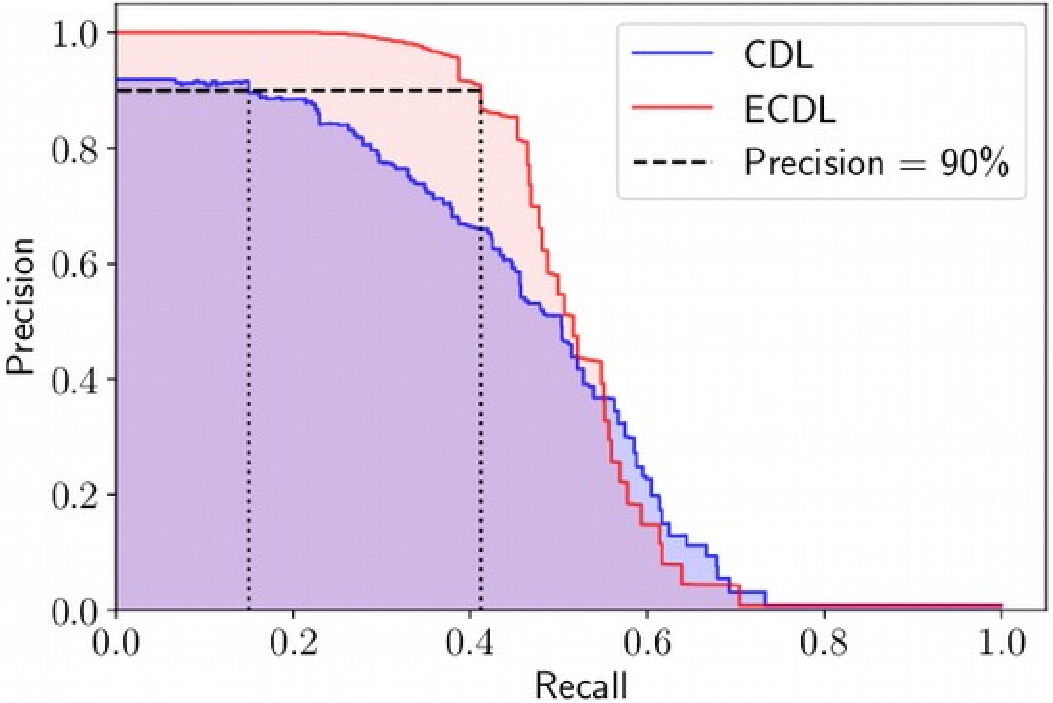
aggregation

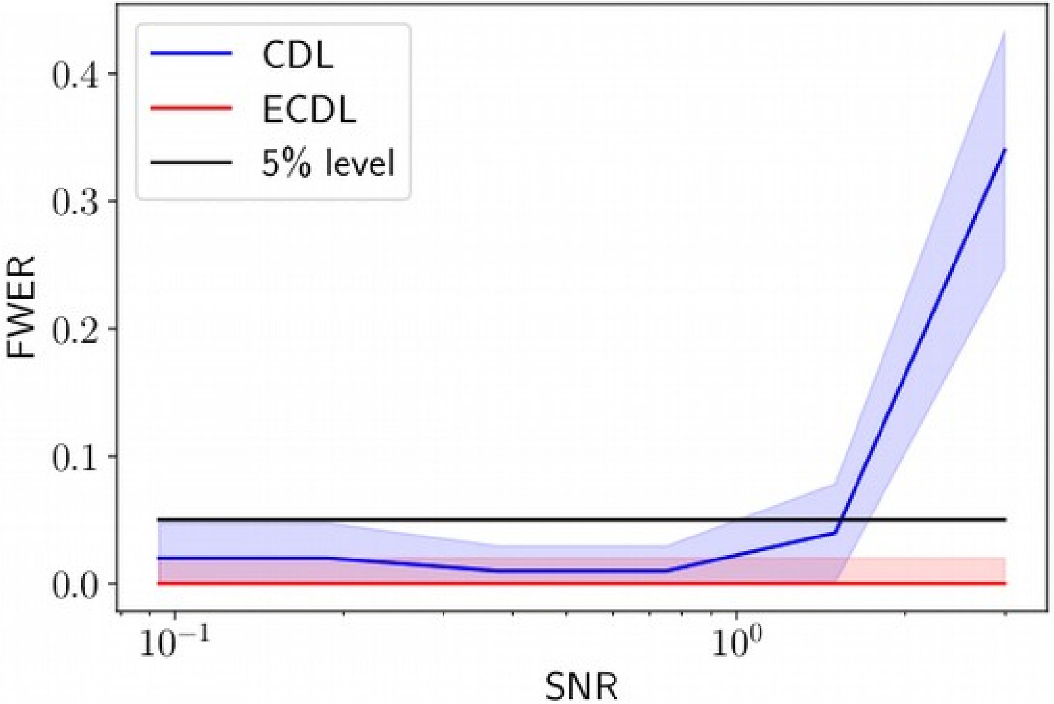


Experiments: PR and FWER control

$$\text{Recall} = \frac{\text{Number of true positive}}{\text{Size of the active set}} \quad \text{Precision} = \frac{\text{Number of true positive}}{\text{Number of discoveries}}$$

$$\text{FWER} = \text{Prob}(\text{Number of false positive} \geq 1)$$



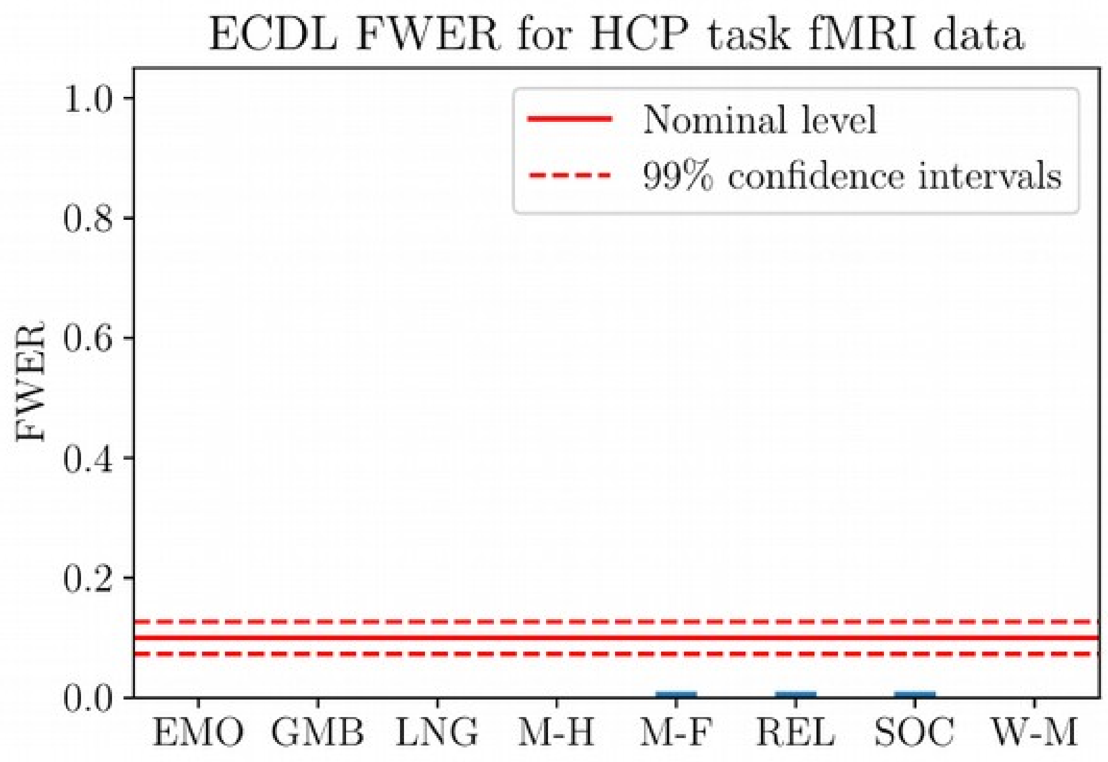


Better PR with ECDL

+ More accurate FWER control

[Chevalier et al. MICCAI 2018]

ECDL: Face validity on real data



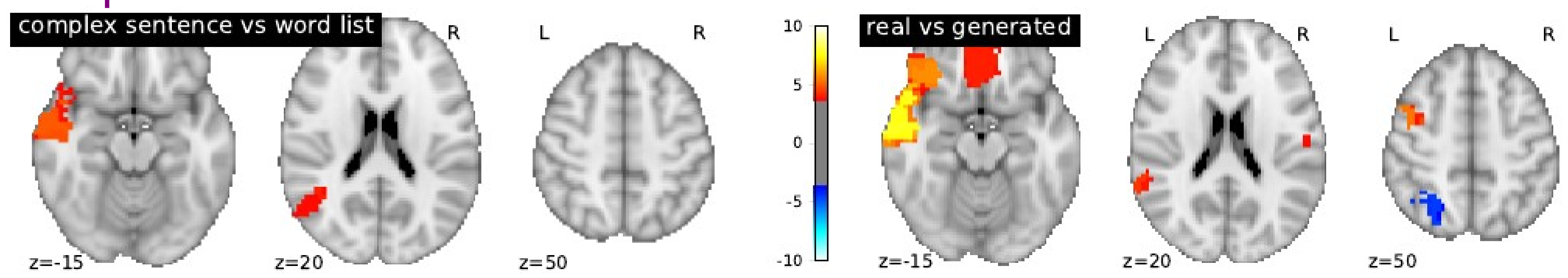
Task paradigm:

EMO = Emotion
GMB = Gambling
LNG = Language
M-H = Motor Hand
M-F = Motor Foot
REL = Relational
SOC = Social
W-M = Working Memory

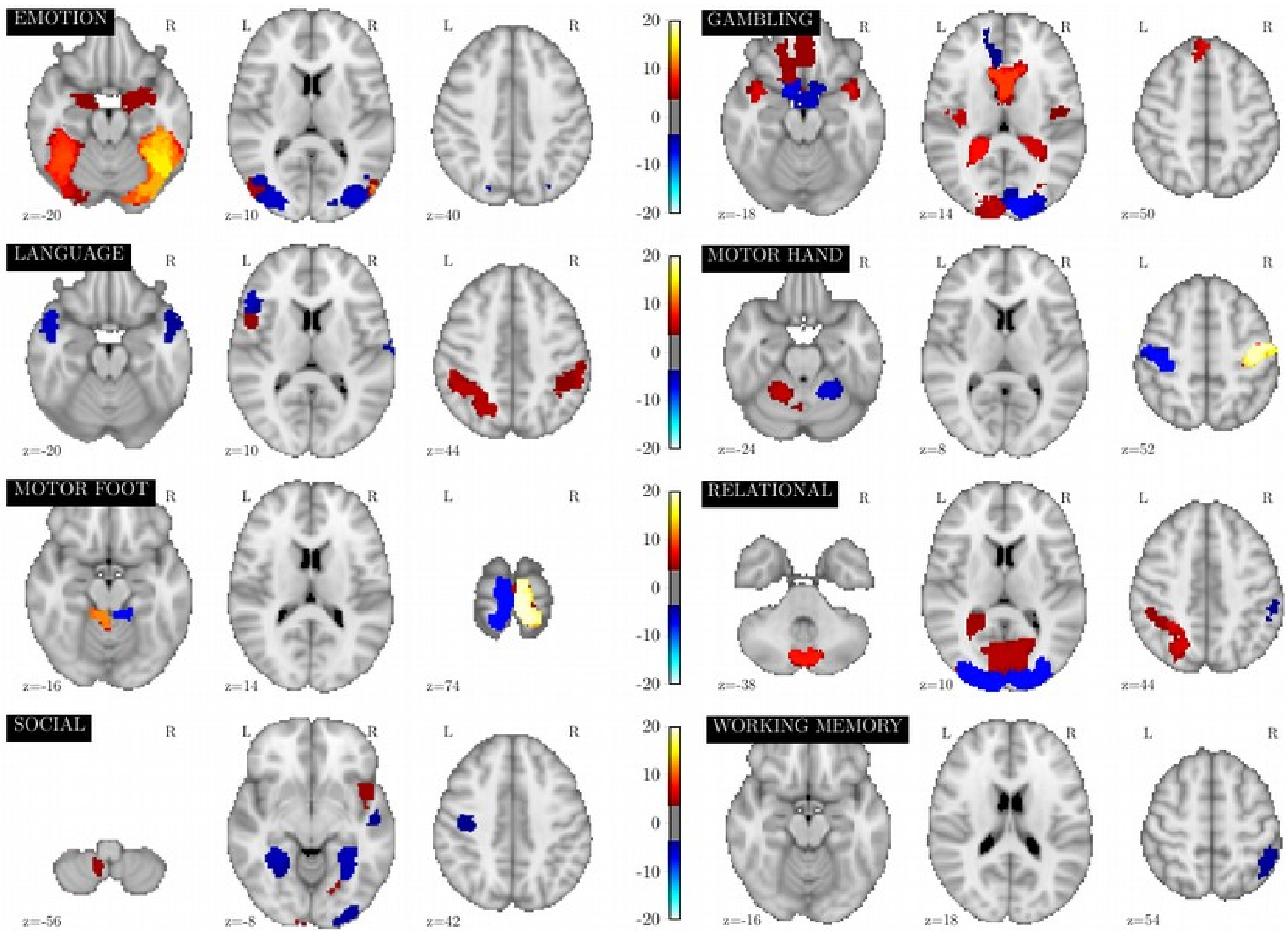
No spurious detection on **null** (permuted) data

[Chevalier et al. In prep]

No spurious detection on unmasked data



ECDL: Face validity on real data



On non-permuted data, the maps are (mostly) what one would expect

Conclusion

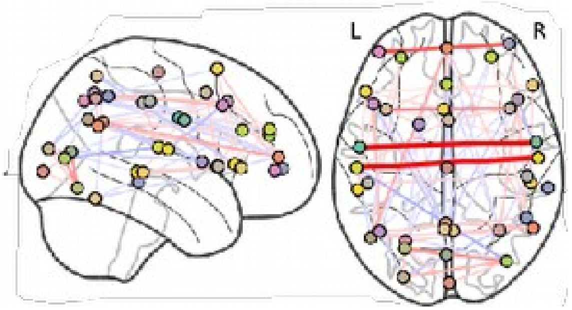
- Joint decoding/encoding for better functional specificity
- Finding commonalities across cognitive studies is hard
- Big data approach:
  - Extract weak signals from huge amounts of data
  - Common representation across datasets (*bottleneck*)

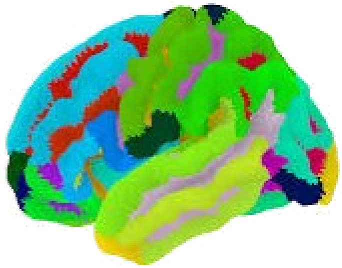
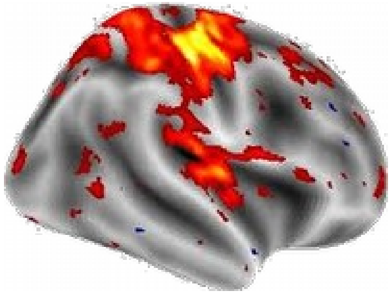
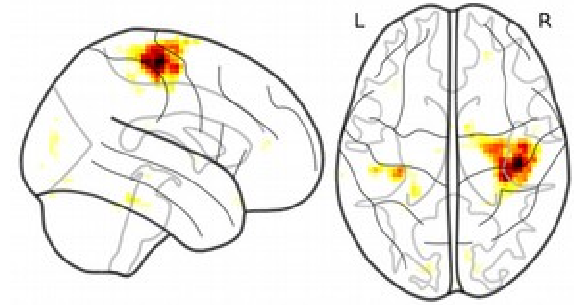
WIP

- Structure underlying cognitive concepts

From good ideas to good practices: software

- Machine learning in Python
- Machine learning for neuroimaging http://nilearn.github.io
- BSD, Python, OSS
  - Classification of (neuroimaging) data
  - Network analysis



Acknowledgements

Parietal

G. Varoquaux,
A. Gramfort,
P. Ciuciu,
D. Wassermann,
D. Engemann,
A. Manoel,
D. Chyzhyk
A.L. Grilo Pinho,
E. Dohmatob,
A. Mensch,
J.A. Chevalier,
A. Hoyos idrobo,
D. Bzdok,
J. Dockès,
P. Cerda,
C. Lazarus
D. La Rocca
G. Lemaitre
L. El Gueddari
O. Grisel
M. Massias
P. Ablin
H. Janati
J. Massich
K. Dadi
C. Petitot
JJ Torres
T. bazeille
B. Nguyen
H.Cherkaoui

Other collaborators
(thanks for the data)
S. Dehaene
R. Poldrack,
J. Haxby
C. F. Gorgolevski
J. Salmon

Human Brain Project
université PARIS-SACLAY
AGENCE NATIONALE DE LA RECHERCHE ANR

Bibliography: encoding vs decoding

- Weichwald S, Meyer T, Özdenizci O, Schölkopf B, Ball T, Grosse-Wentrup M. Causal interpretation rules for encoding and decoding models in neuroimaging. Neuroimage. 2015 Apr 15;110:48-59.
- Haufe S, Meinecke F, Görgen K, Dähne S, Haynes JD, Blankertz B, Bießmann F. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage. 2014 Feb 15;87:96-110.
- Varoquaux G, Schwartz Y, Poldrack RA, Gauthier B, Bzdok D, Poline JB, Thirion B. Atlases of cognition with large-scale human brain mapping. PLoS Comput Biol. 2018 Nov 29;14(11)
- Mapping cognitive ontologies to and from the brain. Yannick Schwartz, Bertrand Thirion, Gaël Varoquaux NIPS (Neural Information Processing Systems), Dec 2013,

Bilbiography: large-scale neuroimaging

- Gorgolewski KJ, Varoquaux G, Rivera G, Schwartz Y, Sochat VV, Ghosh SS, Maumet C, Nichols TE, Poline JB, Yarkoni T, Margulies DS, Poldrack RA. NeuroVault.org: A repository for sharing unthresholded statistical maps, parcellations, and atlases of the human brain. Neuroimage. 2016 Jan 1;124(Pt B):1242-1244.
- Bzdok D, Varoquaux G, Grisel O, Eickenberg M, Poupon C, Thirion B. Formal Models of the Network Co-occurrence Underlying Mental Operations. PLoS Comput Biol. 2016 Jun 16;12(6)
- Poldrack RA, Kittur A, Kalar D, Miller E, Seppa C, Gil Y, Parker DS, Sabb FW, Bilder RM. The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. Front Neuroinform. 2011 Sep 6;5:17.
- Costafreda SG. Pooling FMRI data: meta-analysis, mega-analysis and multi-center studies. Front Neuroinform. 2009 Sep 30;3:33.
- Learning Neural Representations of Human Cognition across Many fMRI Studies. Arthur Mensch, Julien Mairal, Danilo Bzdok, Bertrand Thirion, Gaël Varoquaux. Neural Information Processing Systems, Dec 2017,

Bibliography: recovery & statistics in brain activity decoding

- Hoyos-Idrobo A, Varoquaux G, Schwartz Y, Thirion B. FReM - Scalable and stable decoding with fast regularized ensemble of models. Neuroimage. 2018 Oct 15;180(Pt A):160-172.
- Gaël Varoquaux, Alexandre Gramfort, Bertrand Thirion. Small-sample brain mapping: sparse recovery on spatially correlated designs with randomization and clustering International Conference on Machine Learning, 2012
- Tuan-Binh Nguyen, Jérôme-Alexis Chevalier, Bertrand Thirion. ECKO: Ensemble of Clustered Knockoffs for robust multivariate inference on MRI data. IPMI 2019 – International Conference on Information Processing in Medical Imaging, Jun 2019,
- Jérôme-Alexis Chevalier, Joseph Salmon, Bertrand Thirion Statistical Inference with Ensemble of Clustered Desparsified Lasso. MICCAI, 2018, Grenade, Spain

Bibliography: high-dimensional statstical inference

- N. Meinshausen, L. Meier, and P. B⊖uhlmann. P-values for high-dimensional regression. J. Amer. Statist. Assoc., 104(488):1671{1681, 2009.
- P. Buhlmann. Statistical signicance in high-dimensional linear models. Bernoulli, 19(4):1212{1242, 09 2013.
- C.-H. Zhang and S. S. Zhang. Condence intervals for low dimensional parameters in high dimensional linear models. J. R. Stat. Soc. Ser. B Stat. Methodol., 76(1):217{242, 2014.
- Rina Foygel Barber and Emmanuel Candès. “Controlling the False Discovery Rate via Knockoffs”, Ann. Statist. 43 (2015).
- A. Javanmard and A. Montanari, Confidence Intervals and Hypothesis Testing for High-Dimensional Regression 2014.