



Towards reproducible neuroimaging across different analysis pipelines

Camille Maumet

► To cite this version:

Camille Maumet. Towards reproducible neuroimaging across different analysis pipelines. CogBases 2023, Oct 2023, Paris, France. 2023. hal-04236791

HAL Id: hal-04236791

<https://inria.hal.science/hal-04236791>

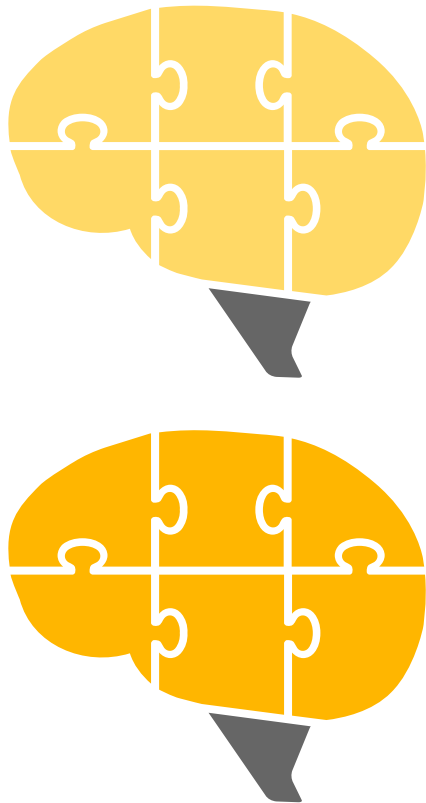
Submitted on 11 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Towards reproducible neuroimaging

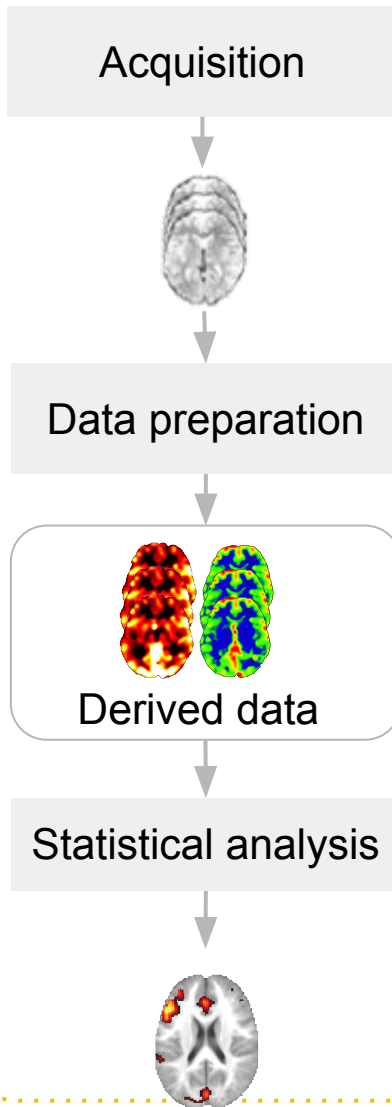
across different analysis pipelines

CogBases workshop 2023

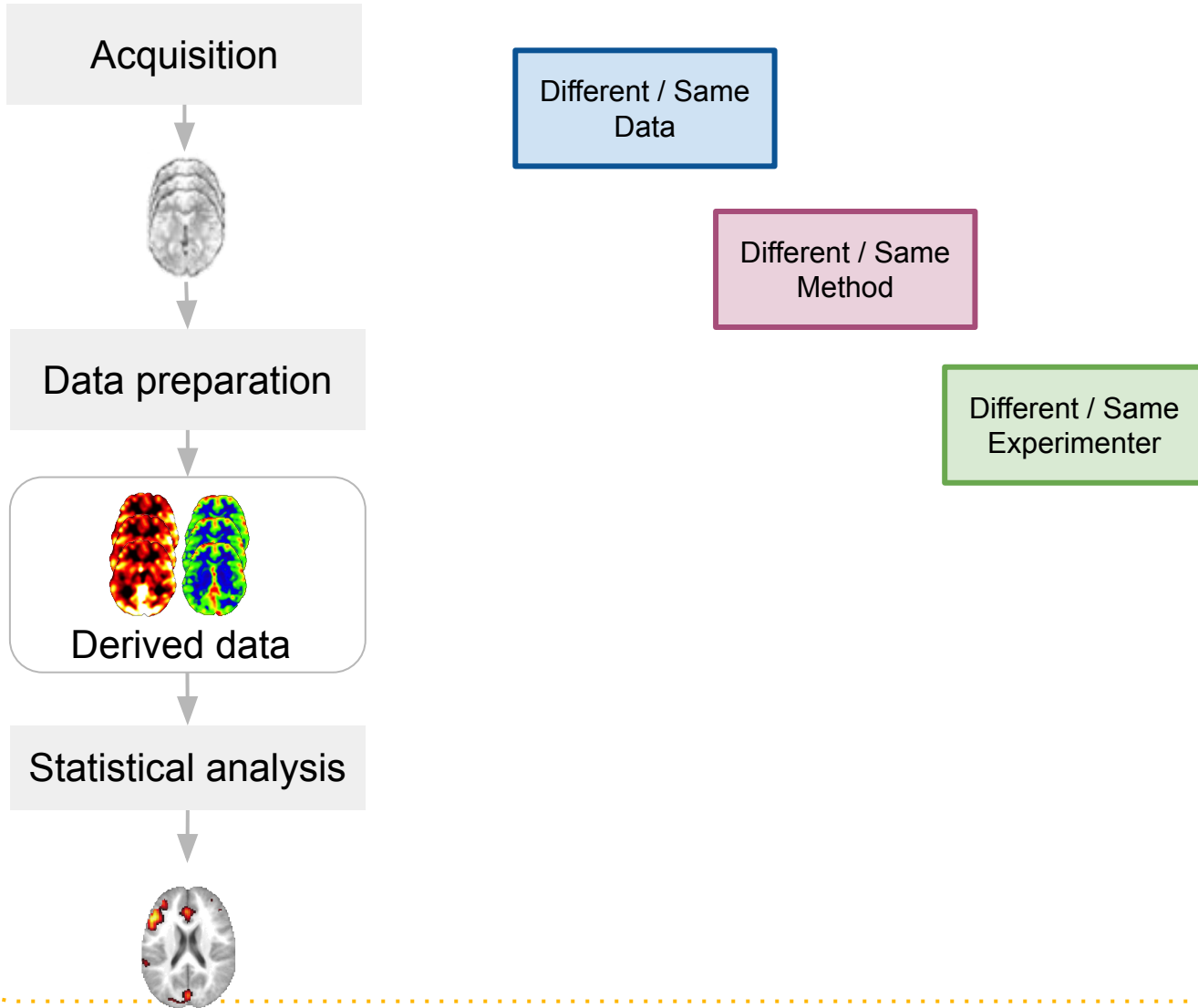
Camille Maumet

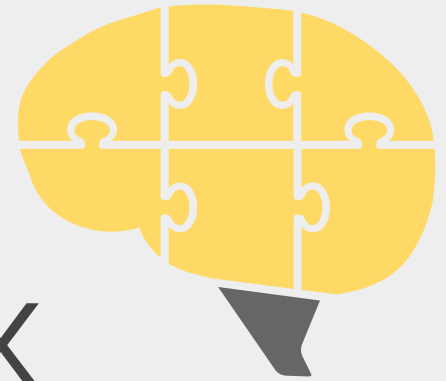
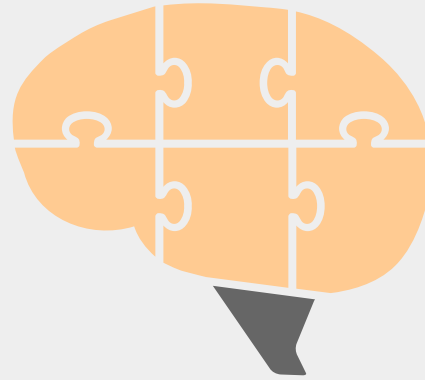
Inria, Univ Rennes, Inserm, CNRS, IRISA

Reproducible brain imaging



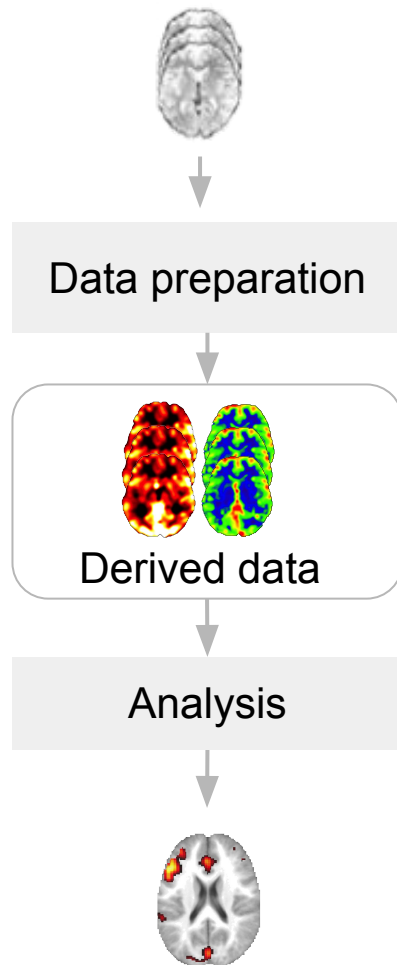
Reproducible brain imaging





On our way to fix
irreproducibility...

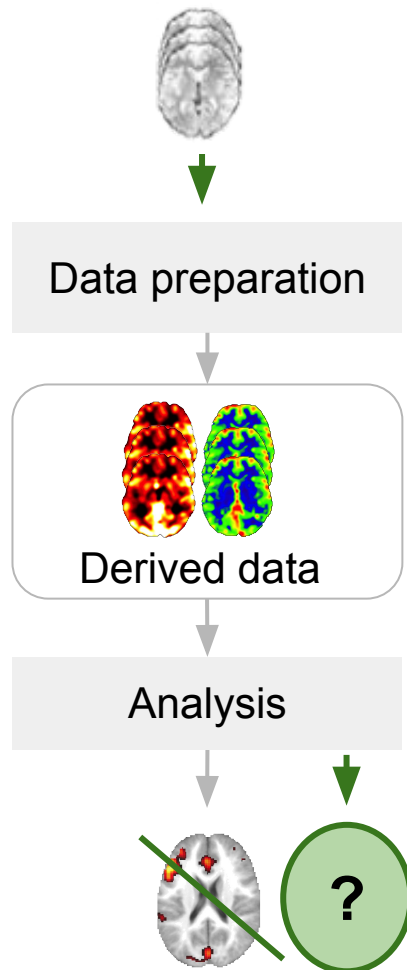
Towards **reproducible** neuroimaging



Towards **reproducible** neuroimaging

Irreproducible with...

Same Data, Same Methods

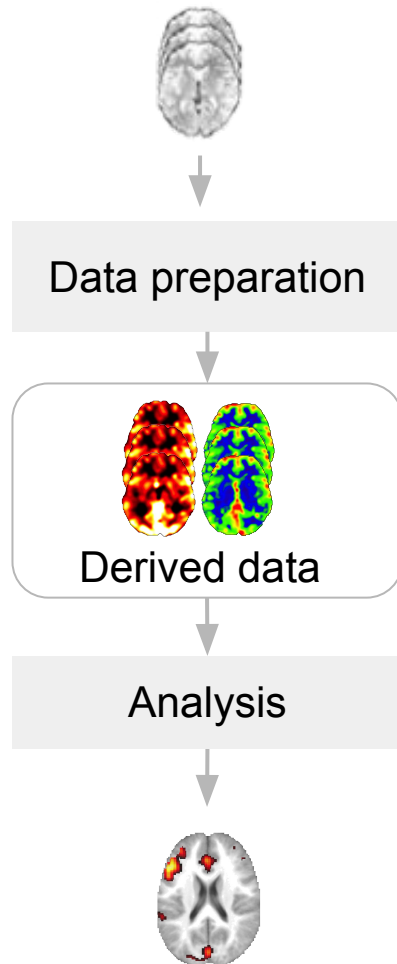


- Sharing code
- Sharing environments: Docker, guix

Towards **reproducible** neuroimaging

Irreproducible with...

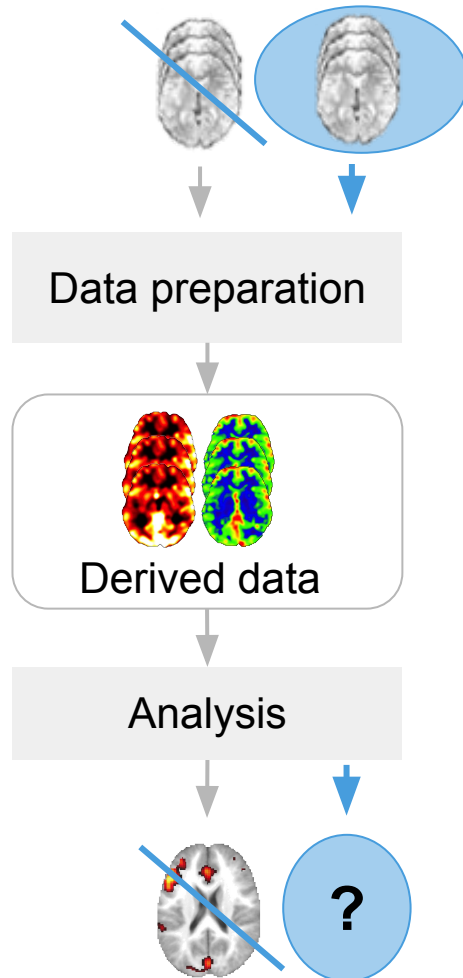
Different Data



Towards **reproducible** neuroimaging

Irreproducible with...

Different Data



Towards **reproducible** neuroimaging

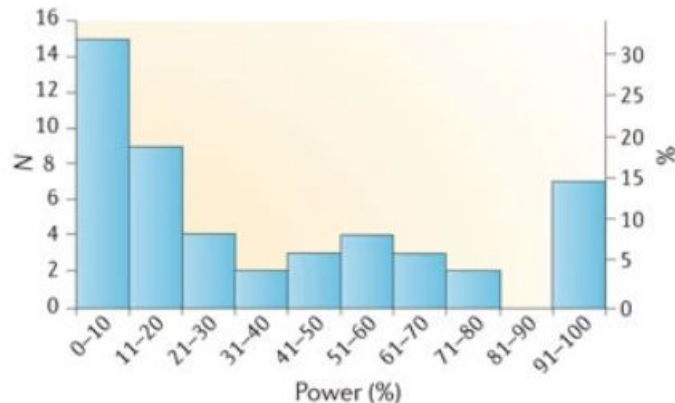
Irreproducible with...

Different Data

Lack of power

Power of neuroscience studies

Power = Prob. to correctly find a significant effect when the alternative hypothesis is true.



[Button et. al, Nat Rev Neurosci 2013]

We need bigger datasets

Towards **reproducible** neuroimaging

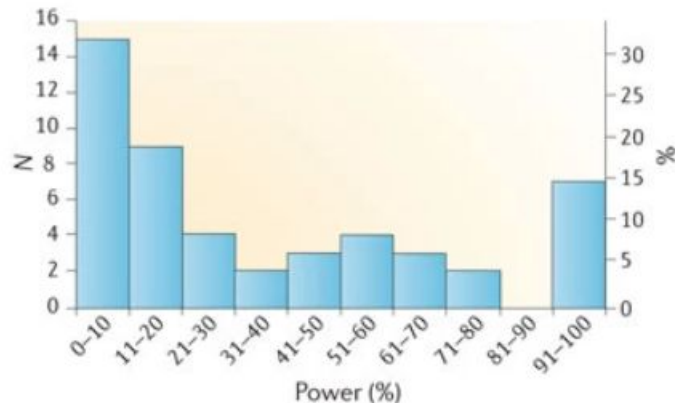
Irreproducible with...

Different Data

Lack of power

Power of neuroscience studies

Power = Prob. to correctly find a significant effect when the alternative hypothesis is true.



[Button et. al, Nat Rev Neurosci 2013]

Generalizability and fairness

X-ray: Lung opacity detection

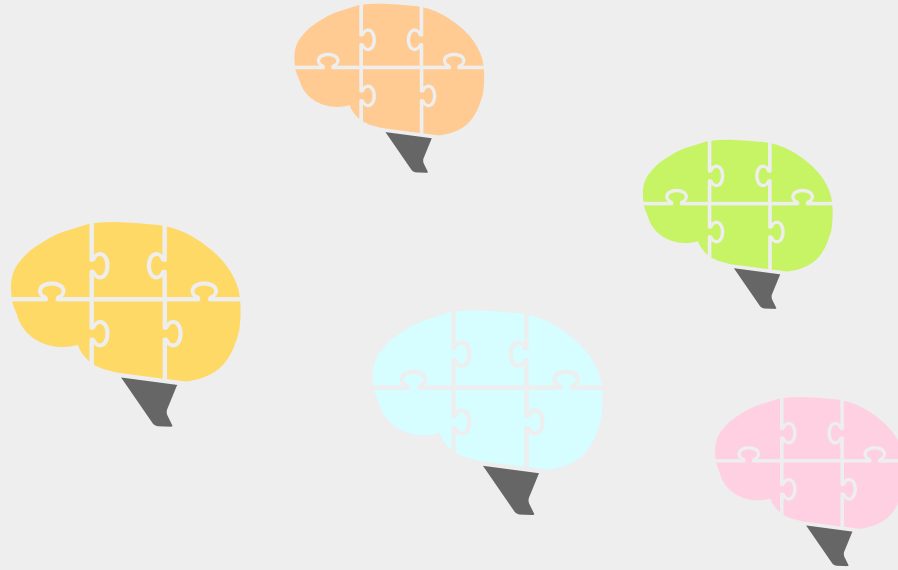
Model trained on male images, tested on female images



[Larrazabal et. al, PNAS 2020]

We need bigger datasets

We need representative
and diverse datasets

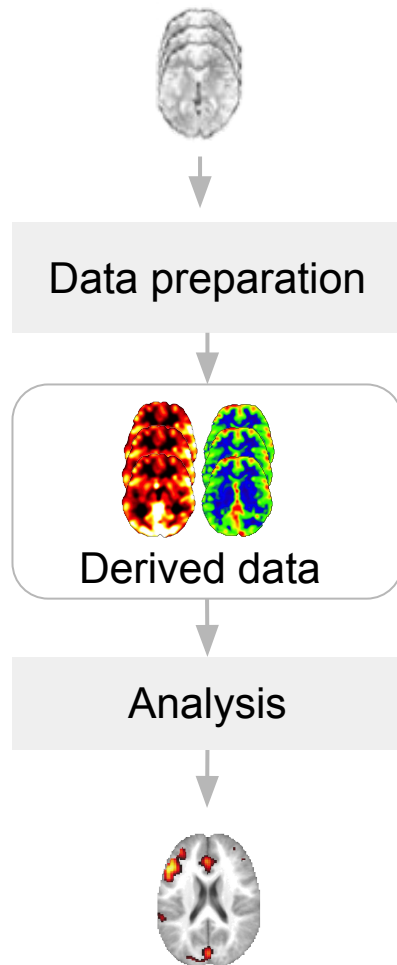


Different pipelines...
different results?

Towards **reproducible** neuroimaging

Irreproducible with...

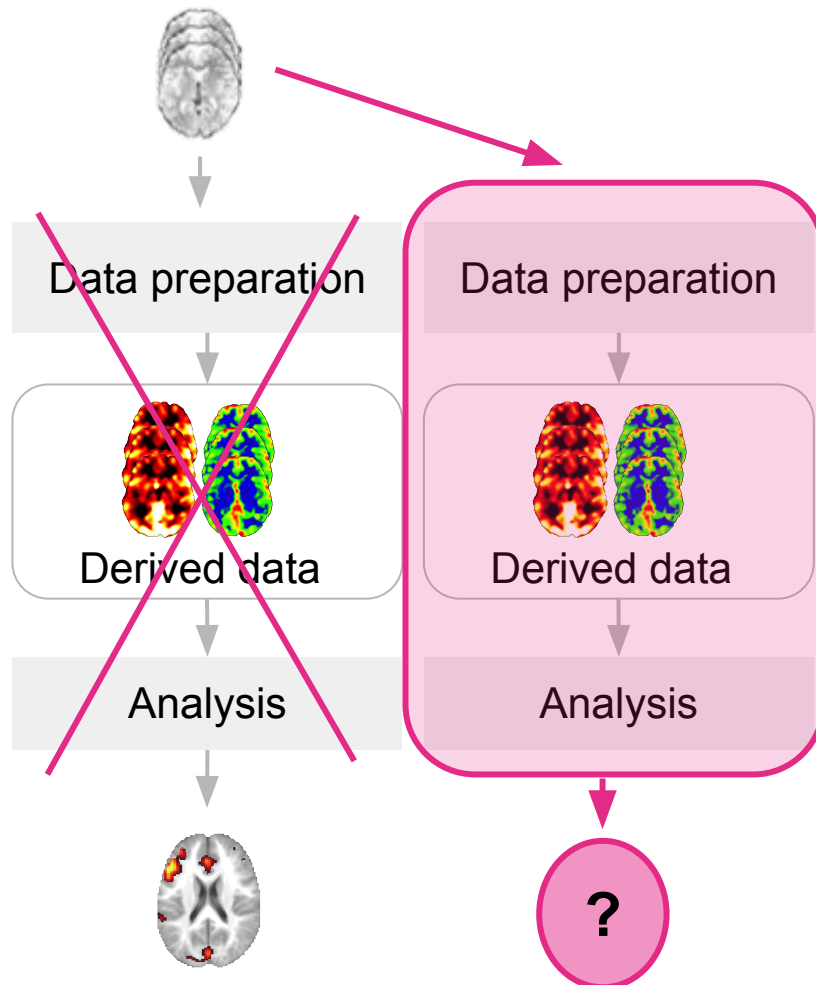
Different Analysis



Towards **reproducible** neuroimaging

Irreproducible with...

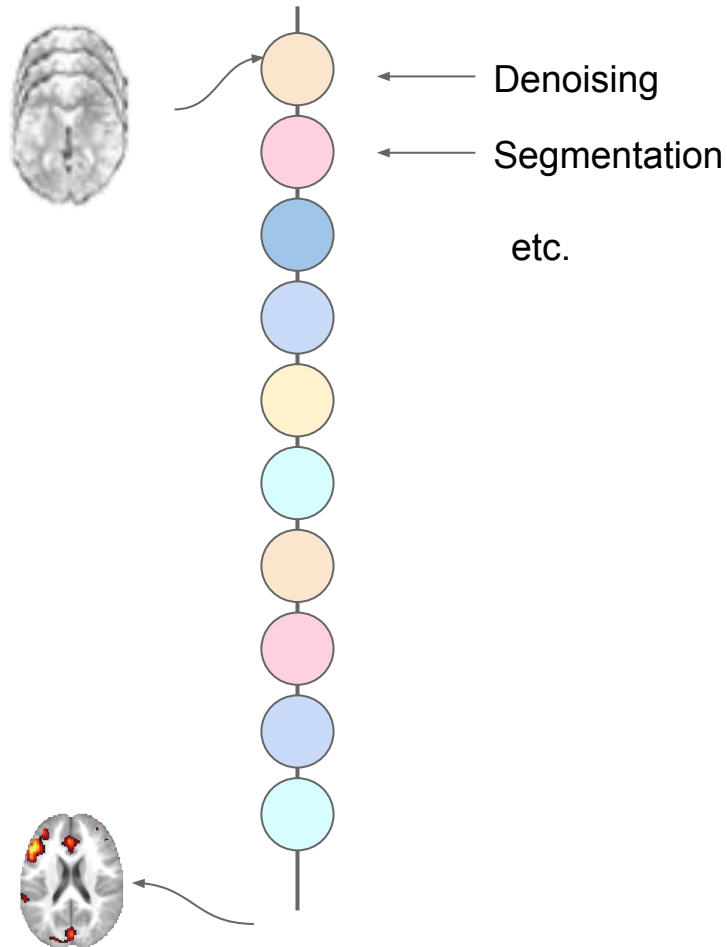
Different Analysis



Towards **reproducible** neuroimaging

Irreproducible with...

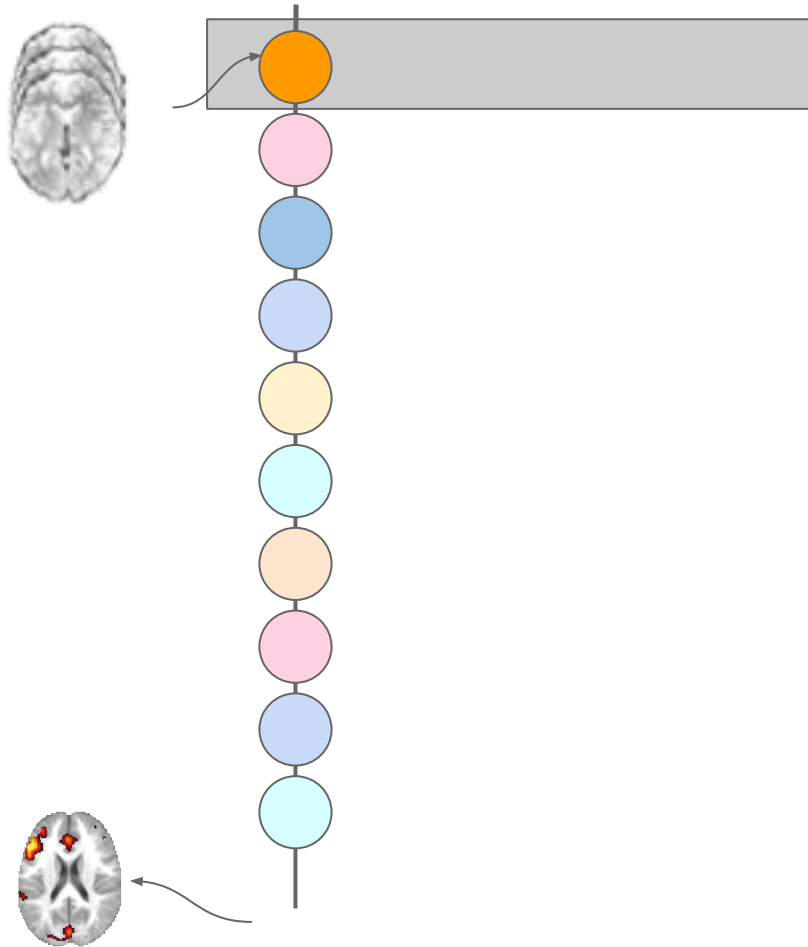
Different Analysis



Towards **reproducible** neuroimaging

Irreproducible with...

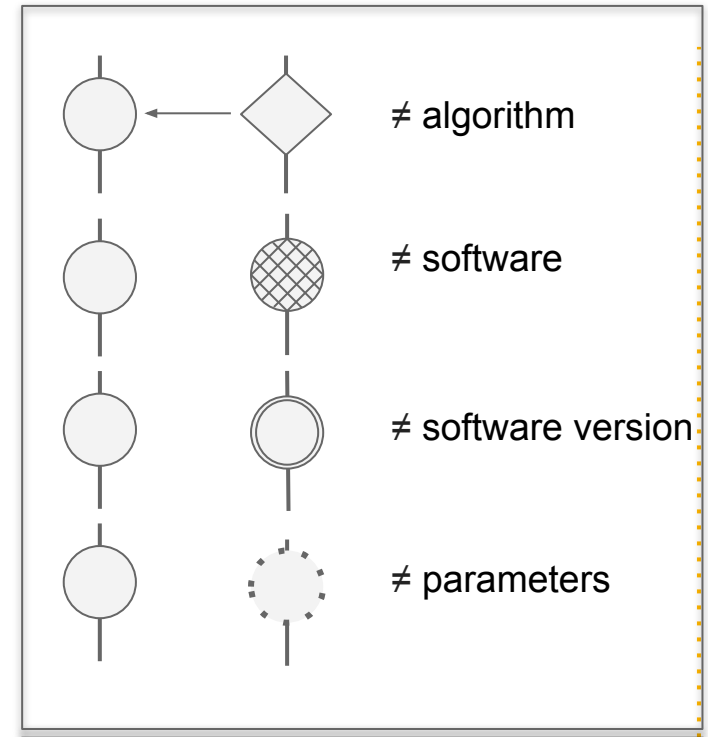
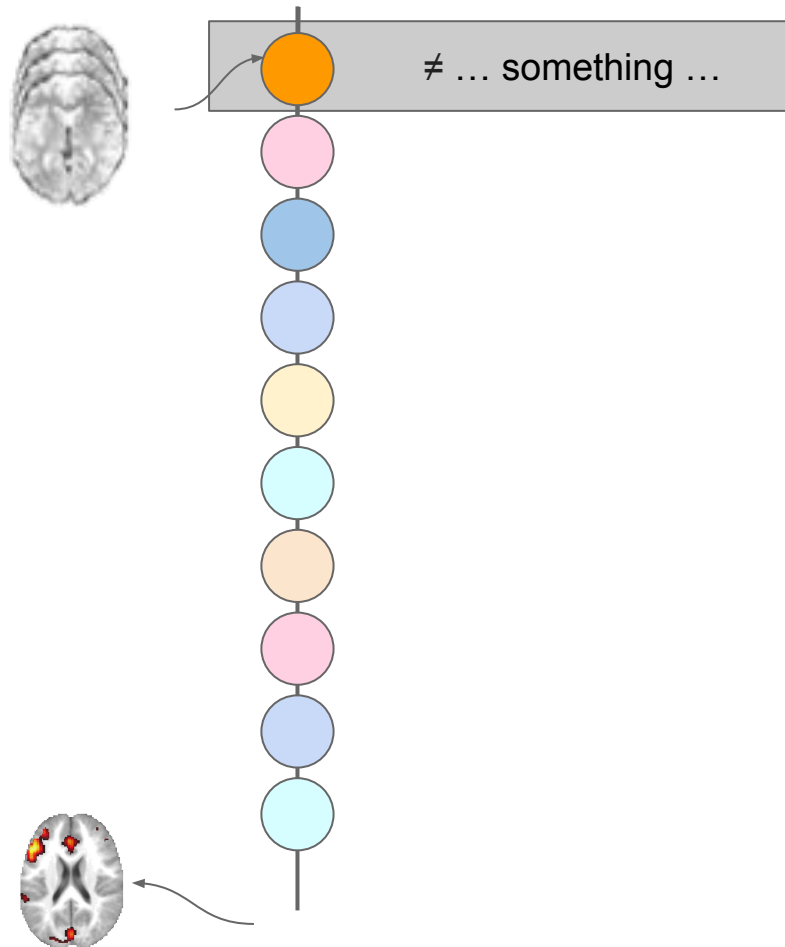
Different Analysis



Towards **reproducible** neuroimaging

Irreproducible with...

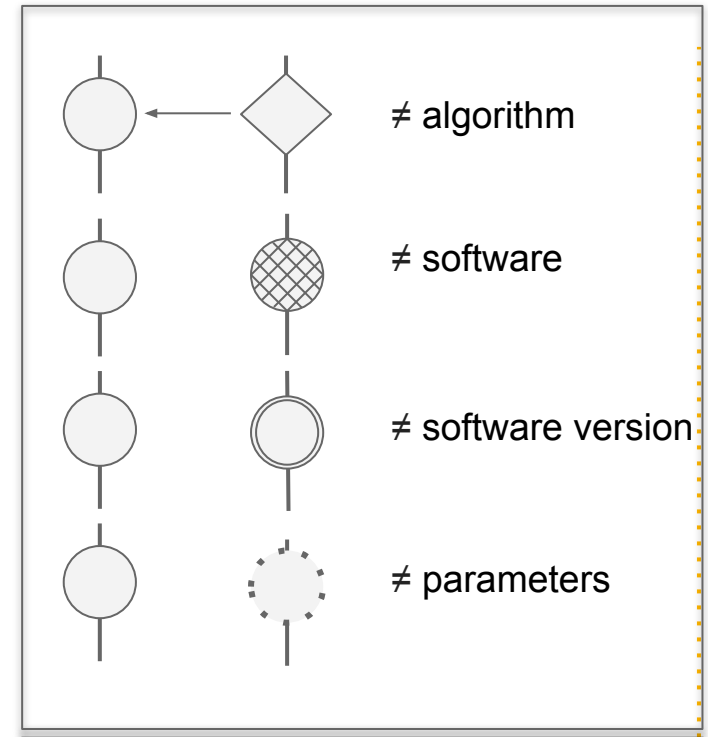
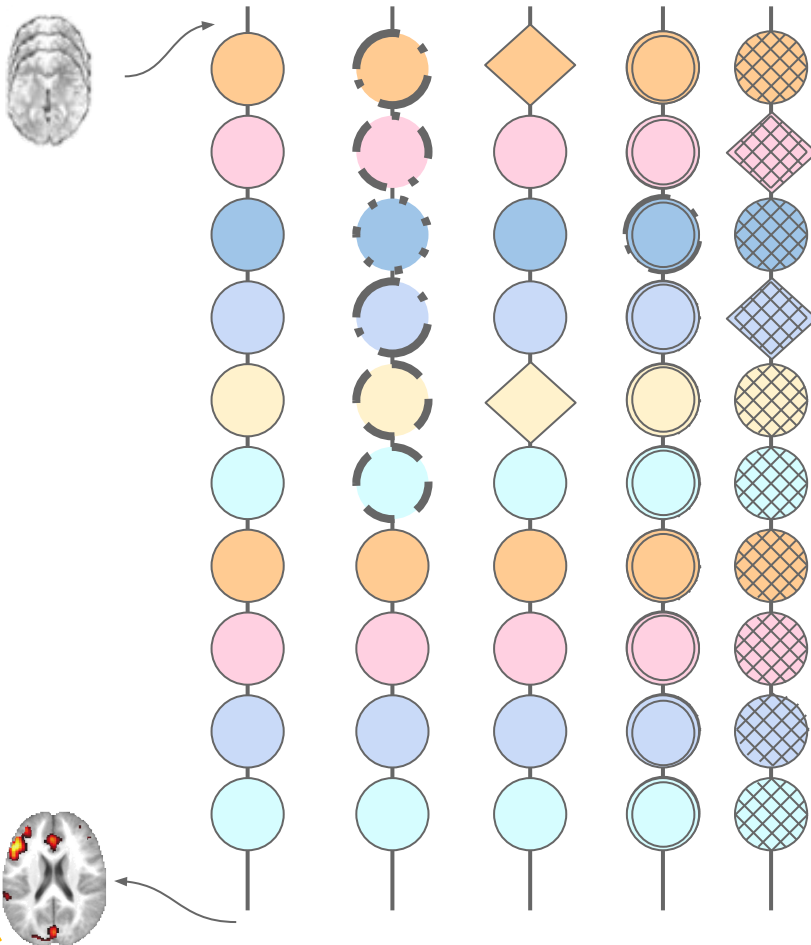
Different Analysis



Towards **reproducible** neuroimaging

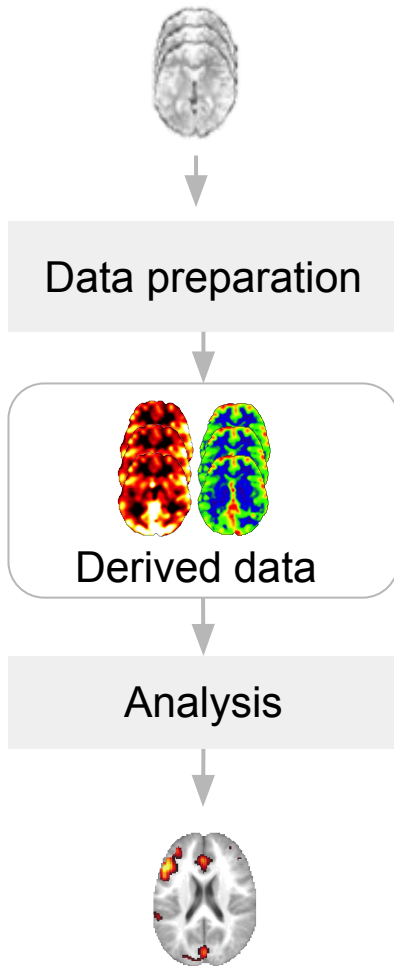
Irreproducible with...

Different Analysis

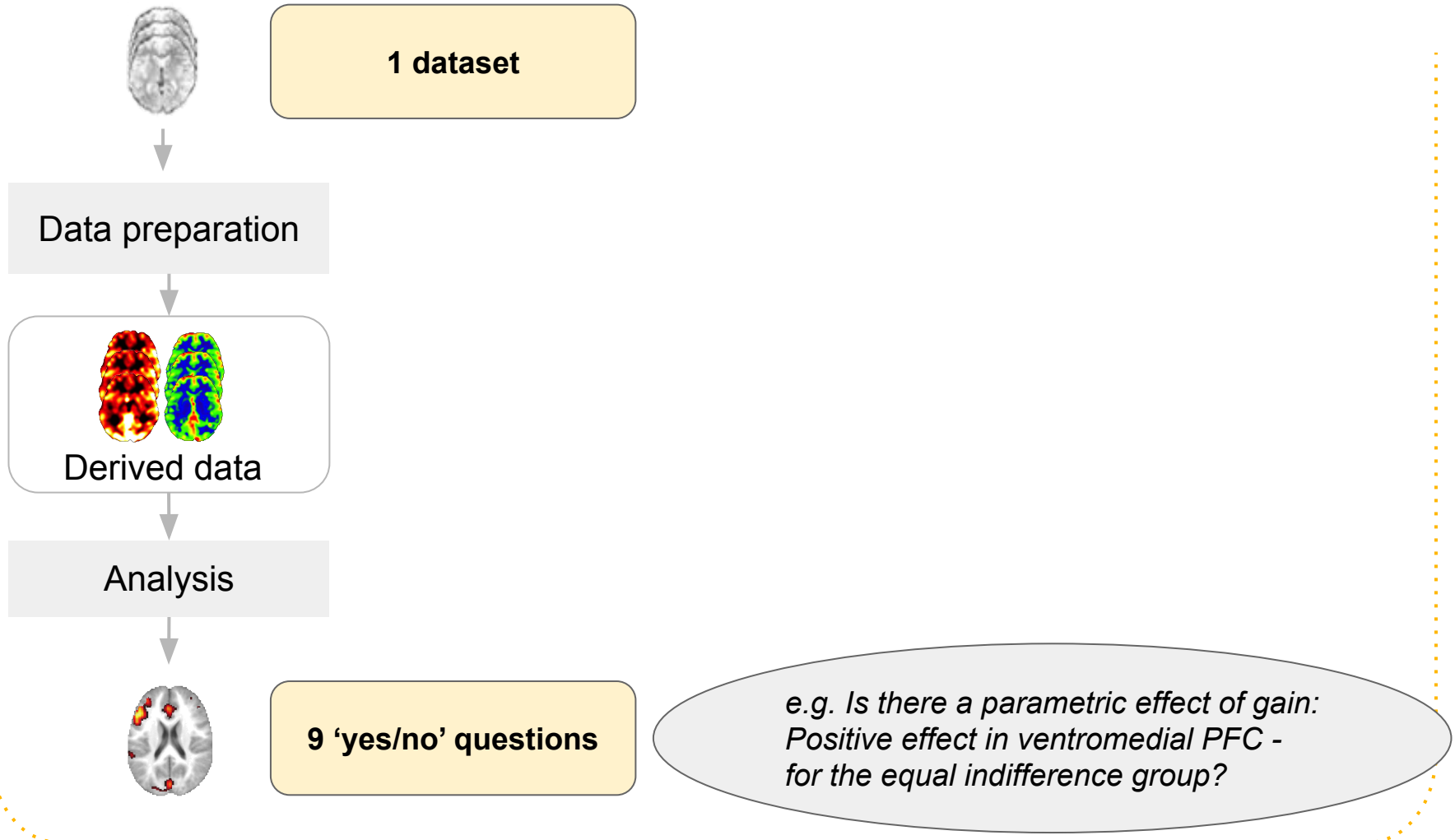


A family of acceptable pipelines
over 10^{30} combinations...

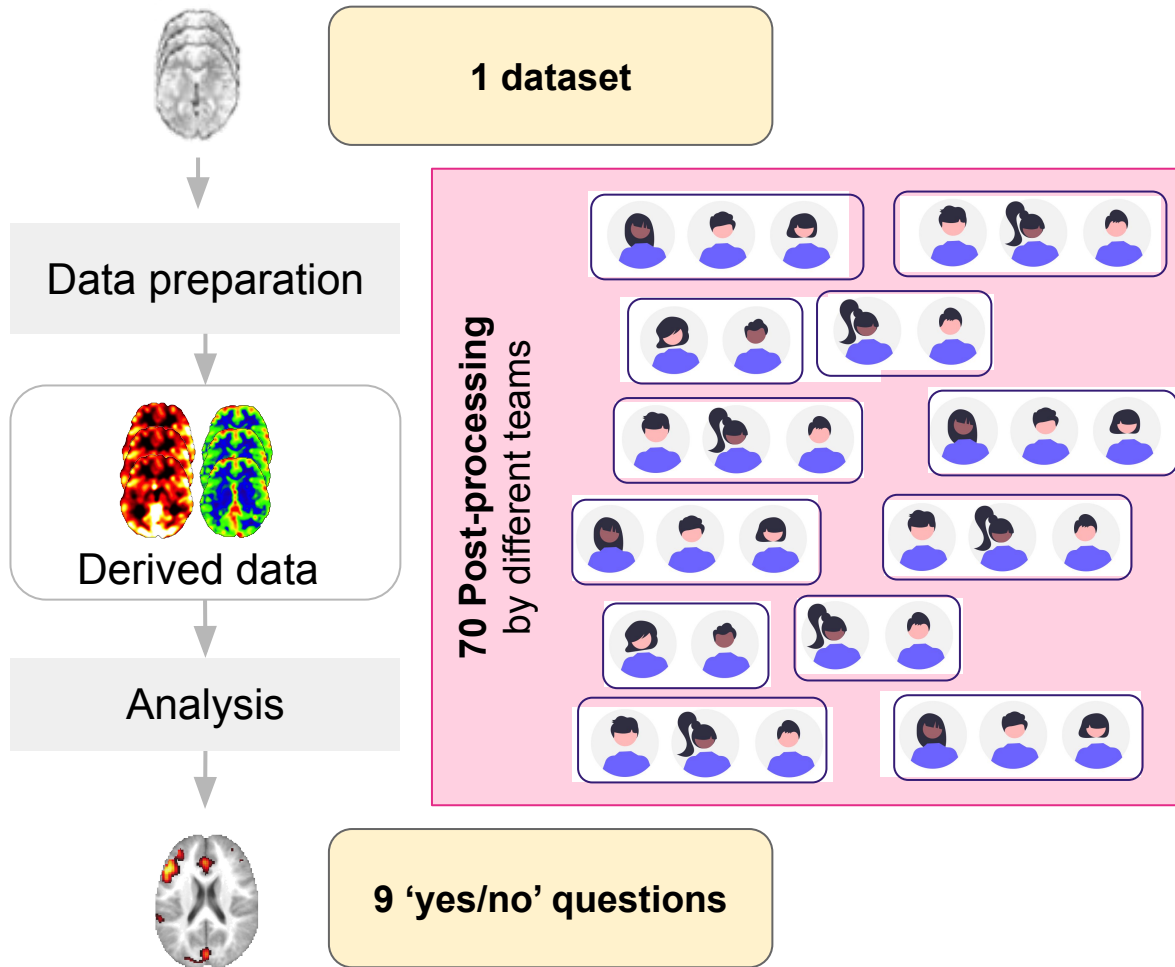
Many-analyst project : NARPS



Many-analyst project : NARPS

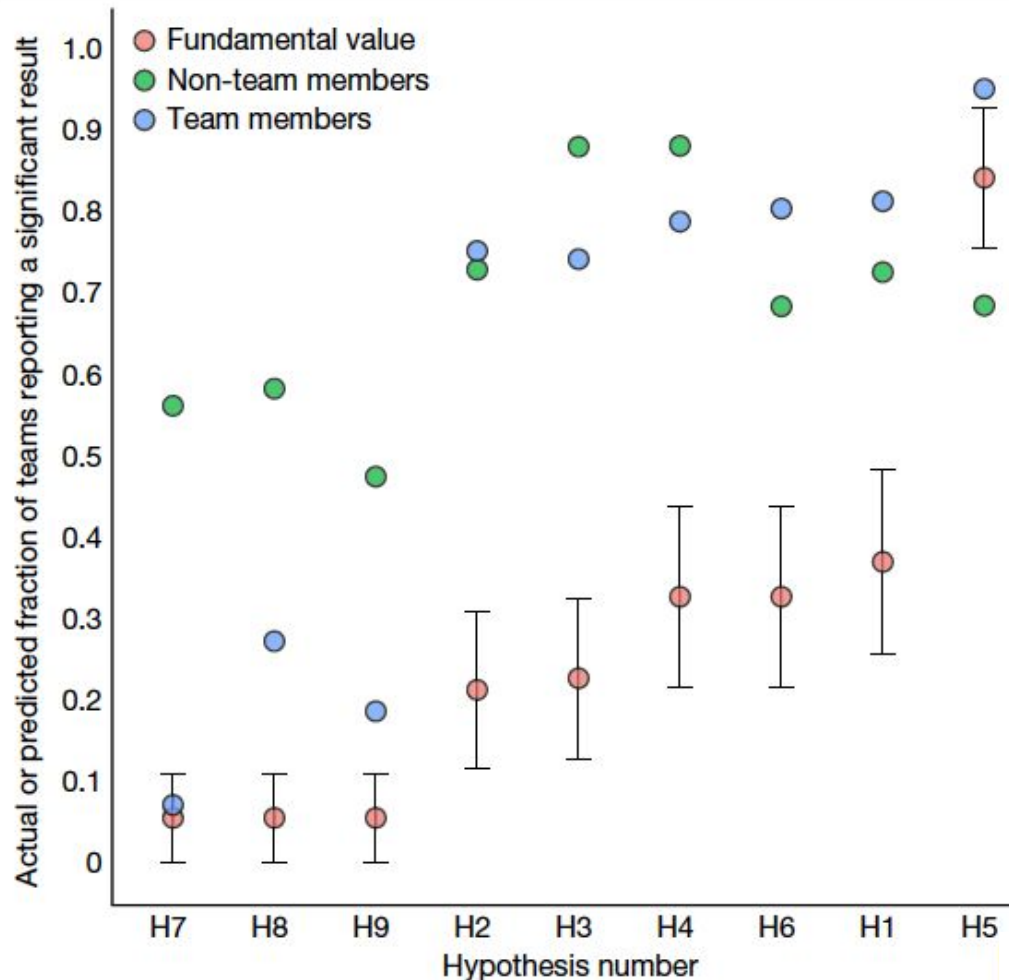


Many-analyst project : NARPS



Many-analyst project : NARPS

Results over the **70 post-processing pipelines**:

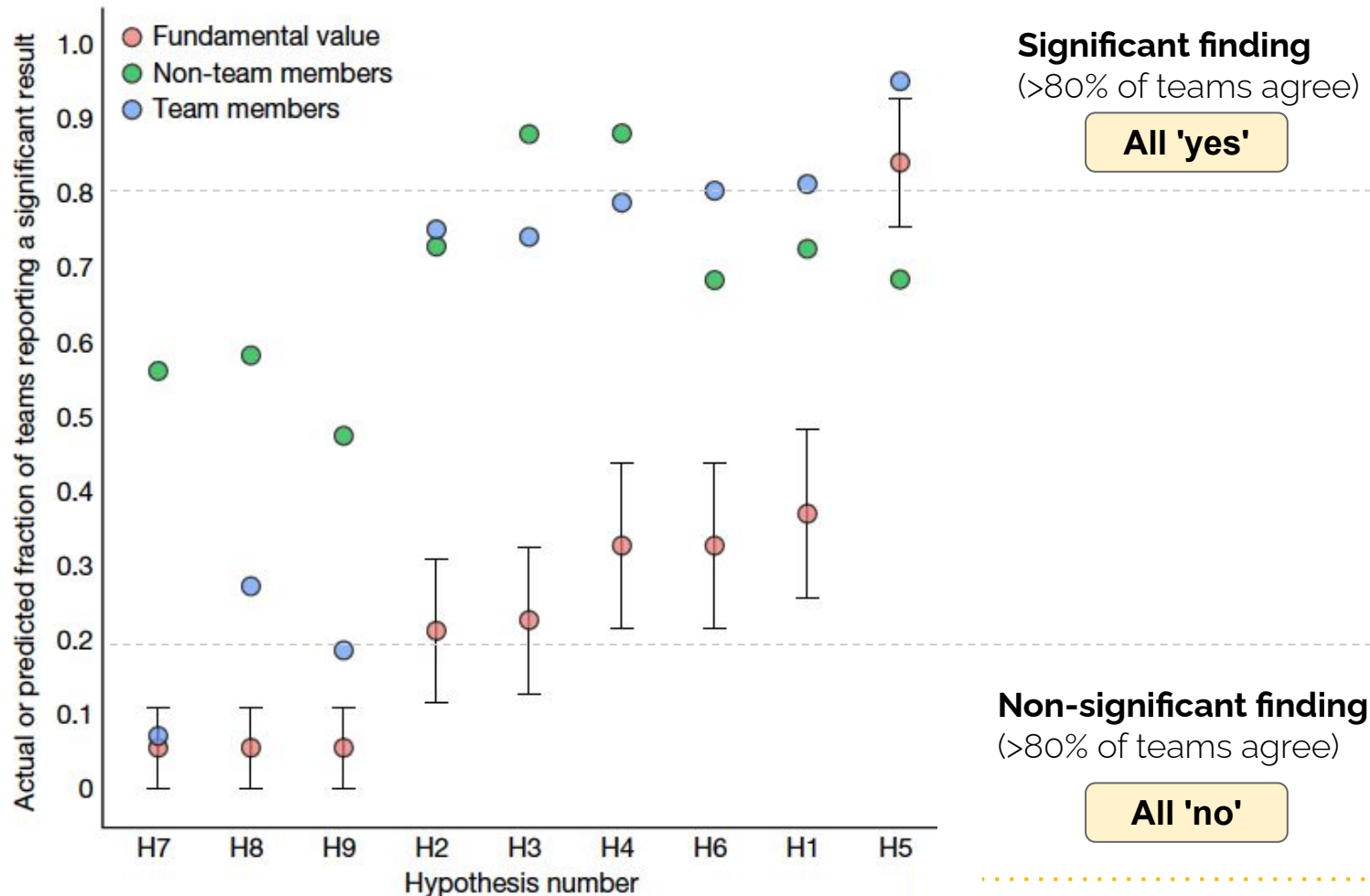


9 'yes/no' questions

[Botvinik-Nezer et. al, Nature 2020]

Many-analyst project : NARPS

Results over the **70 post-processing pipelines**:

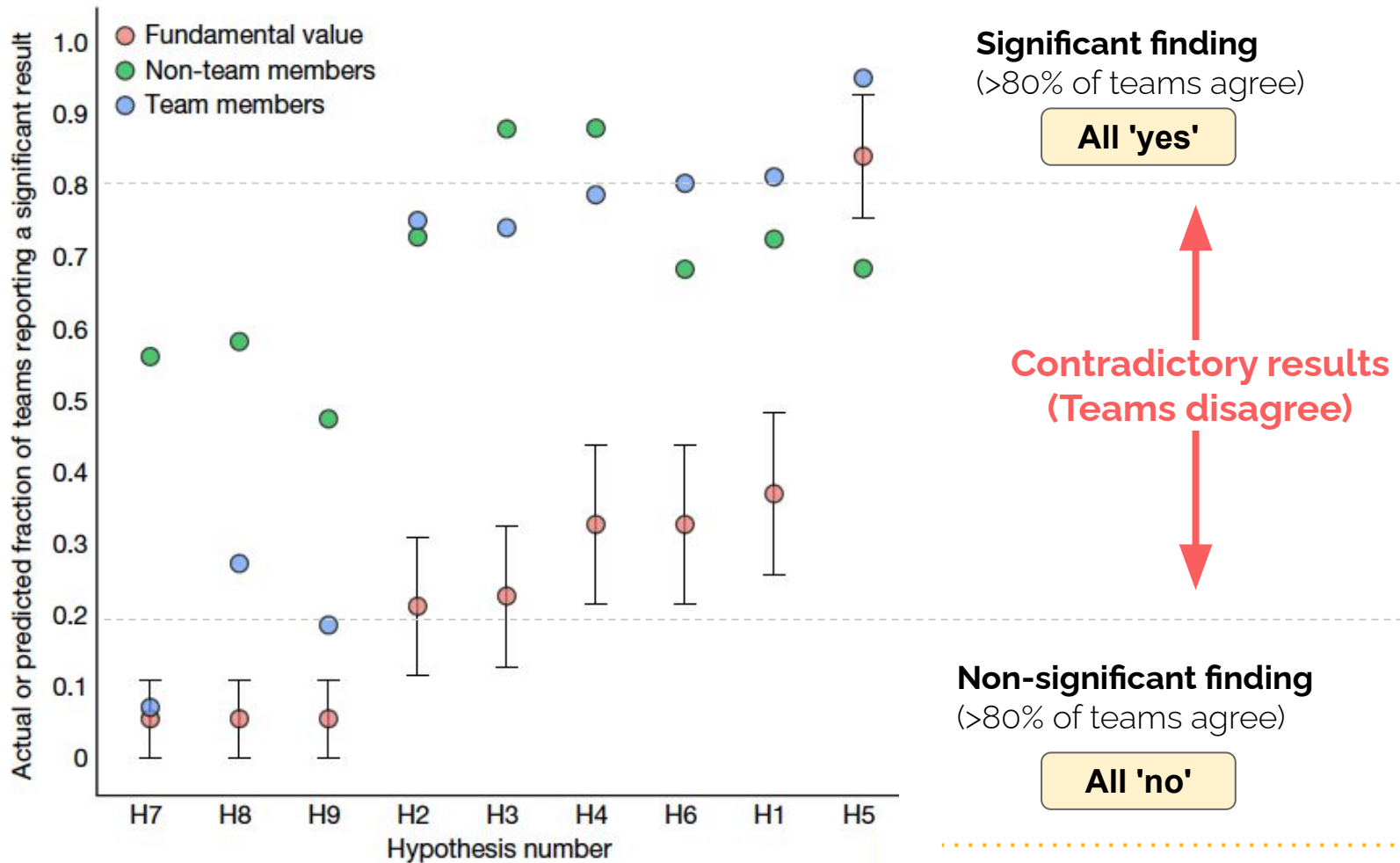


9 'yes/no' questions

[Botvinik-Nezer et. al, Nature 2020]

Many-analyst project : NARPS

Results over the **70 post-processing pipelines**:



9 'yes/no' questions

[Botvinik-Nezer et. al, Nature 2020]

Different levels of variability in post-processing pipelines

- Different end-to-end pipelines [Carp 2012]
- Different software version [Gronenschild 2012]
- Different operating system [Glatard 2015]

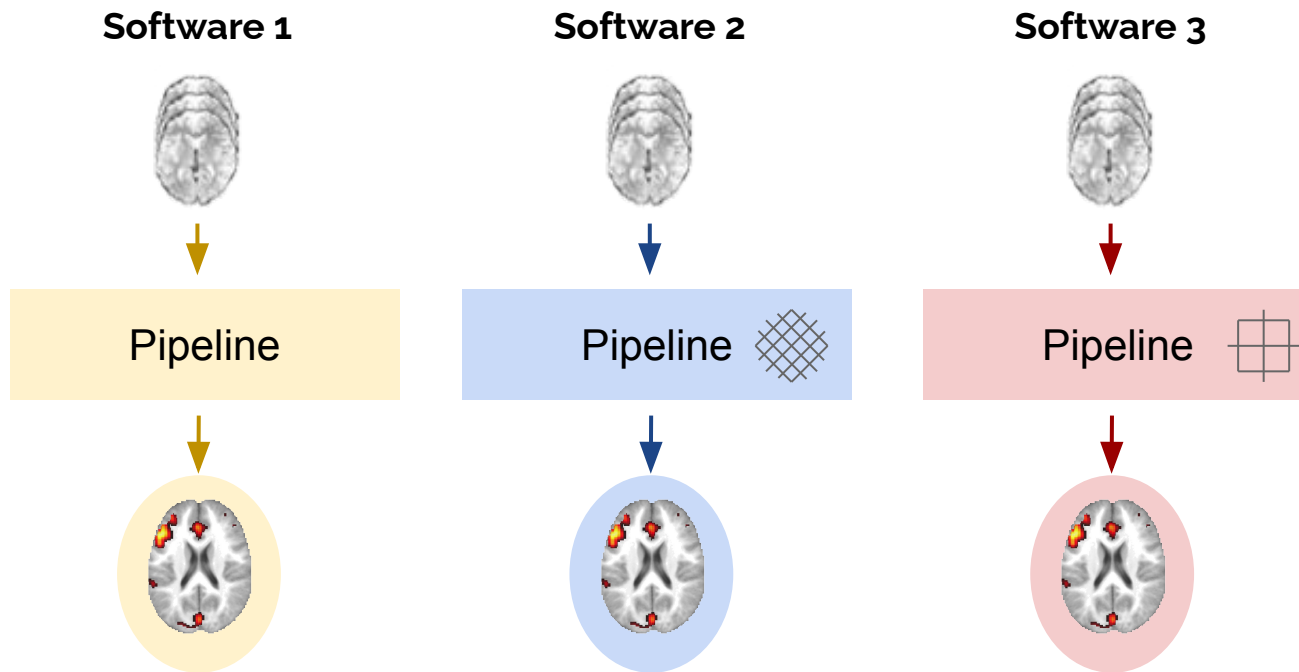
- Different **post-processing software packages** [Bowring 2019, 2021]

Variability across software

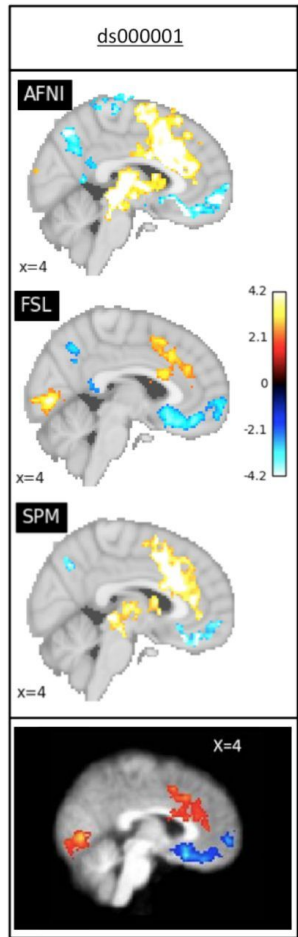
- Reproduced 3 published functional MRI studies
- Using 3 different software



Alex Bowring Tom Nichols



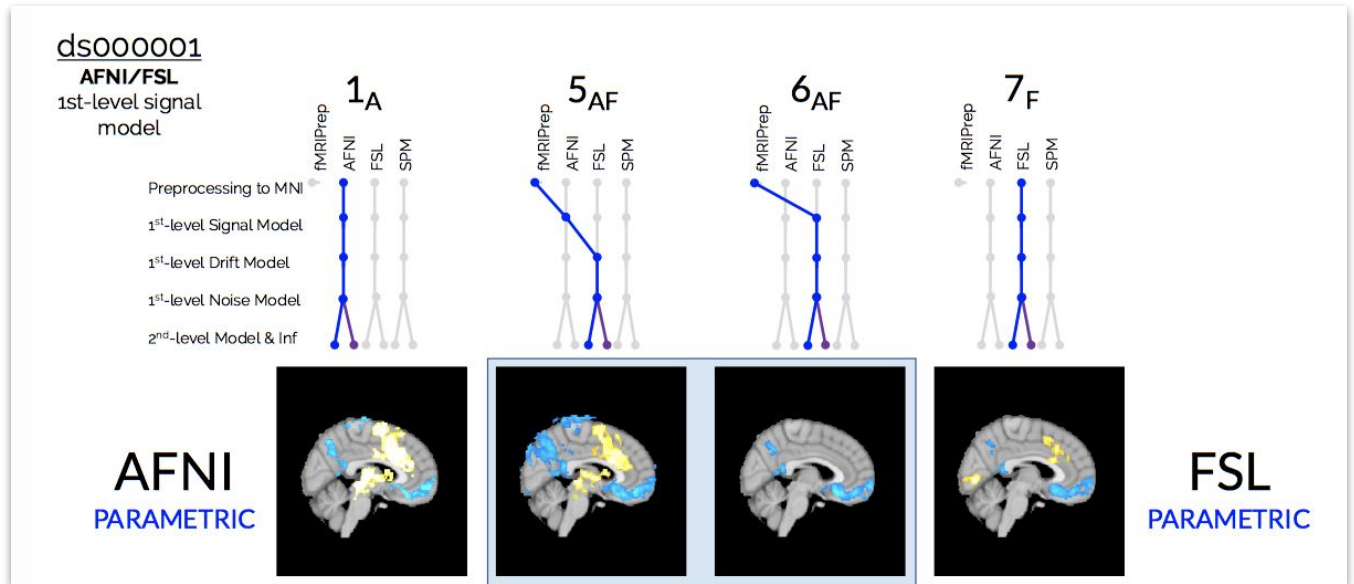
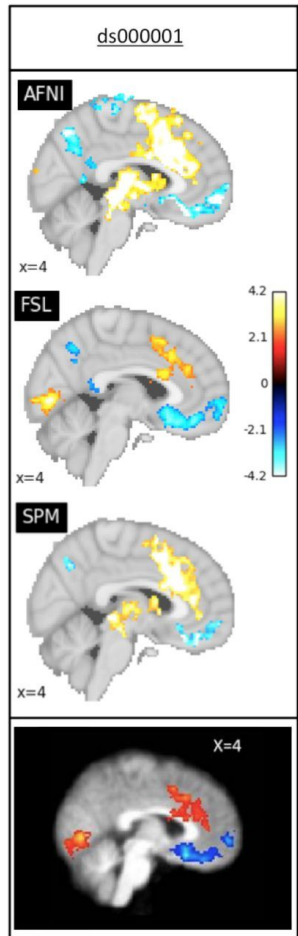
Variability across software



Thresholded maps in AFNI,
FSL, SPM + original

[Bowring et. al, HBM 2019]

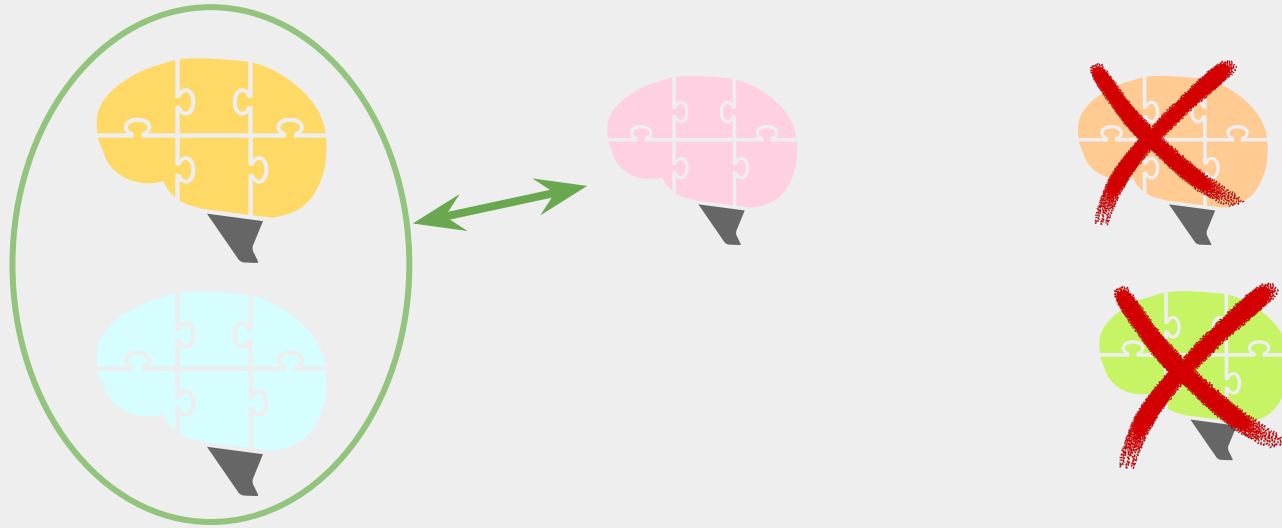
Variability across software



[Bowring et. al, HBM 2021]

Thresholded maps in AFNI,
FSL, SPM + original

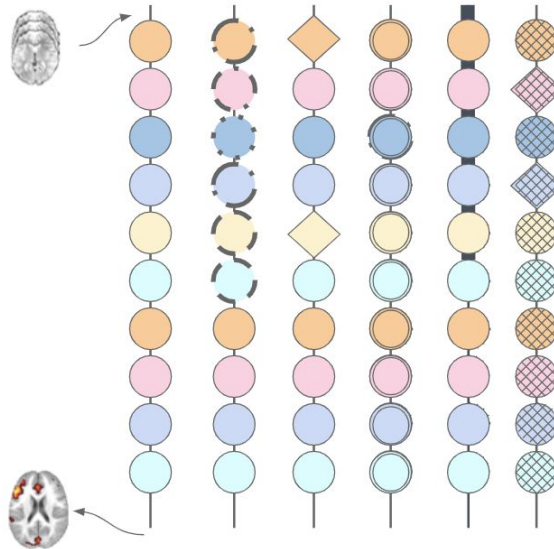
[Bowring et. al, HBM 2019]



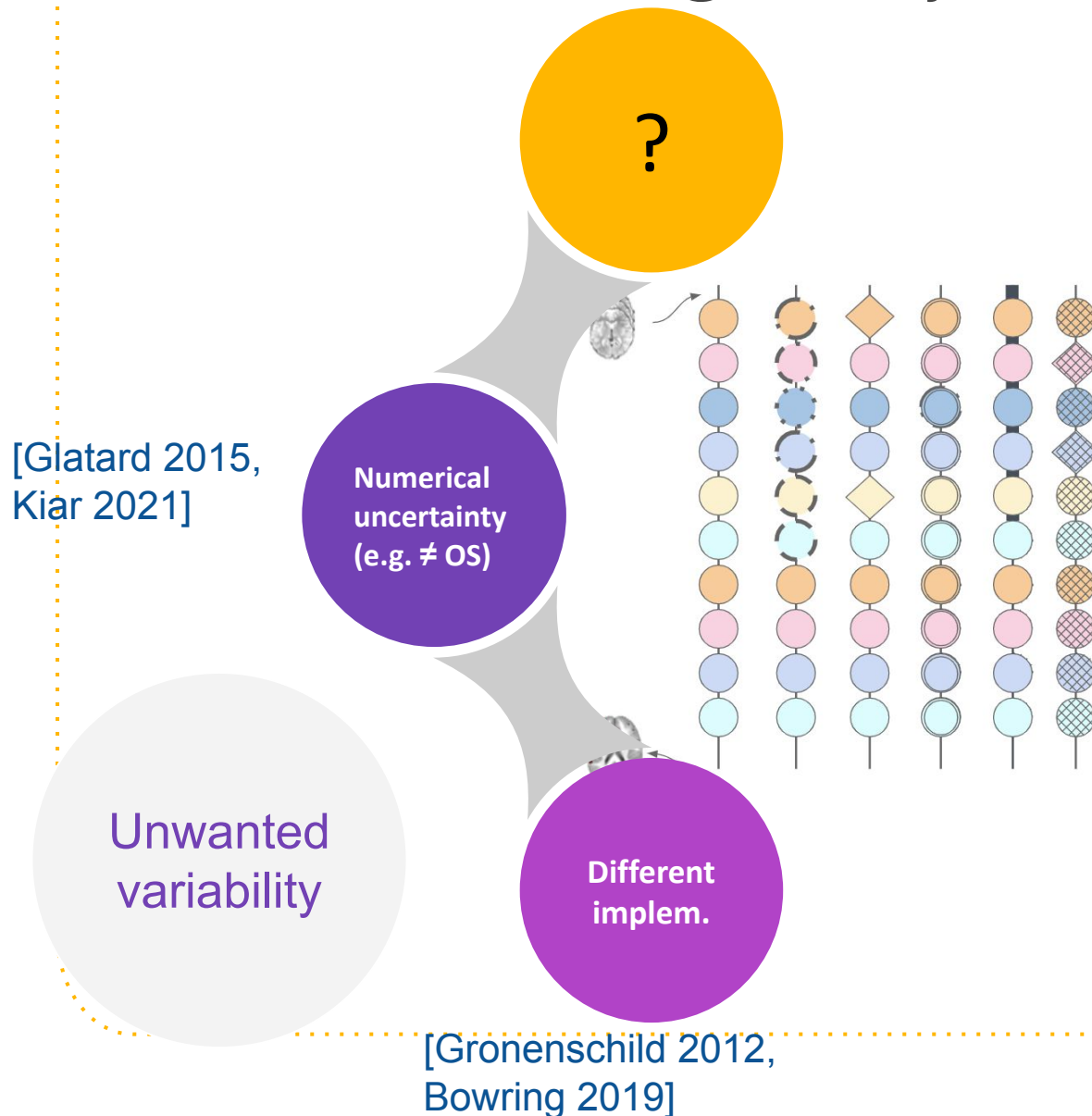
Understanding
analytical variability

Understanding analytical variability...

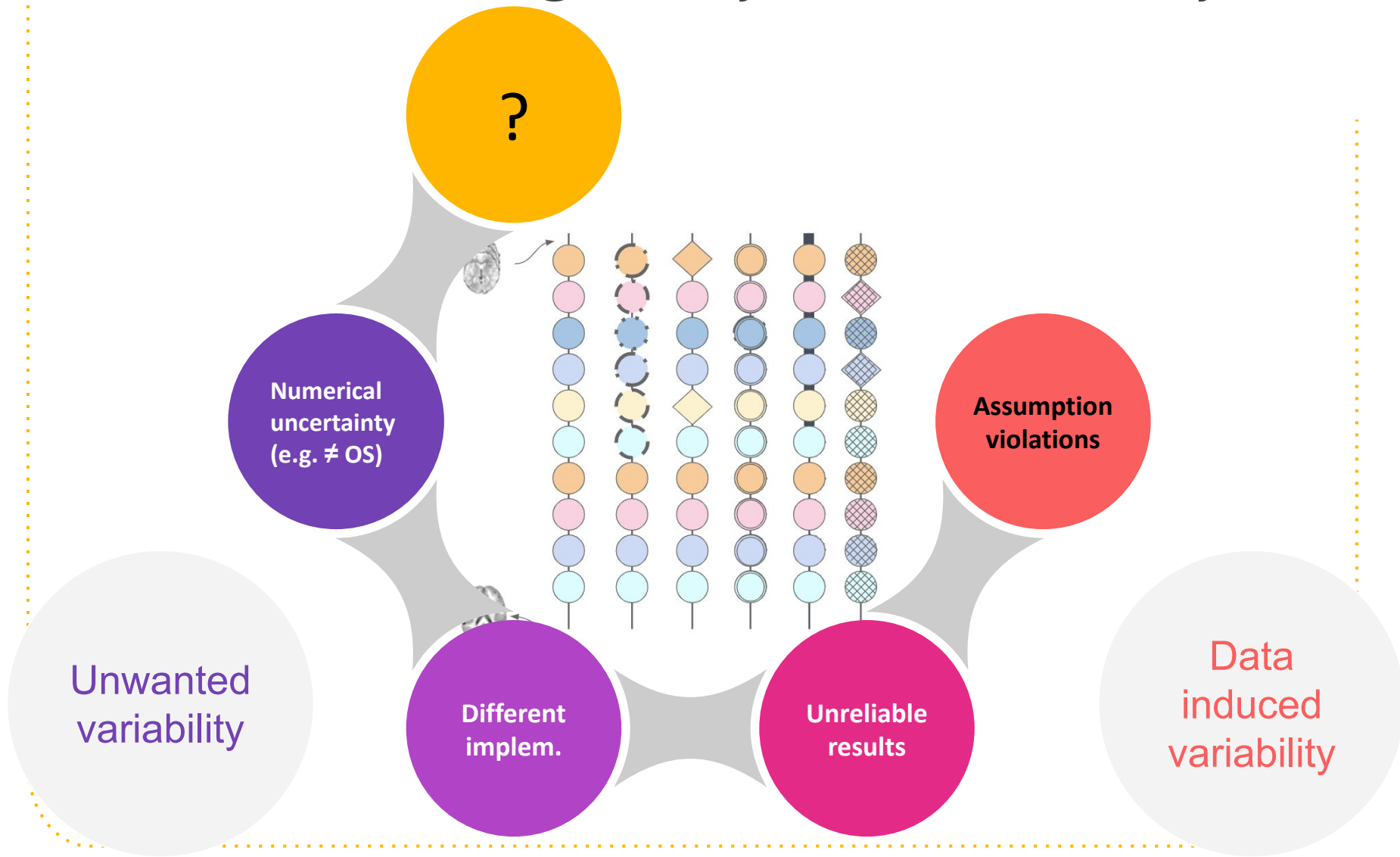
?



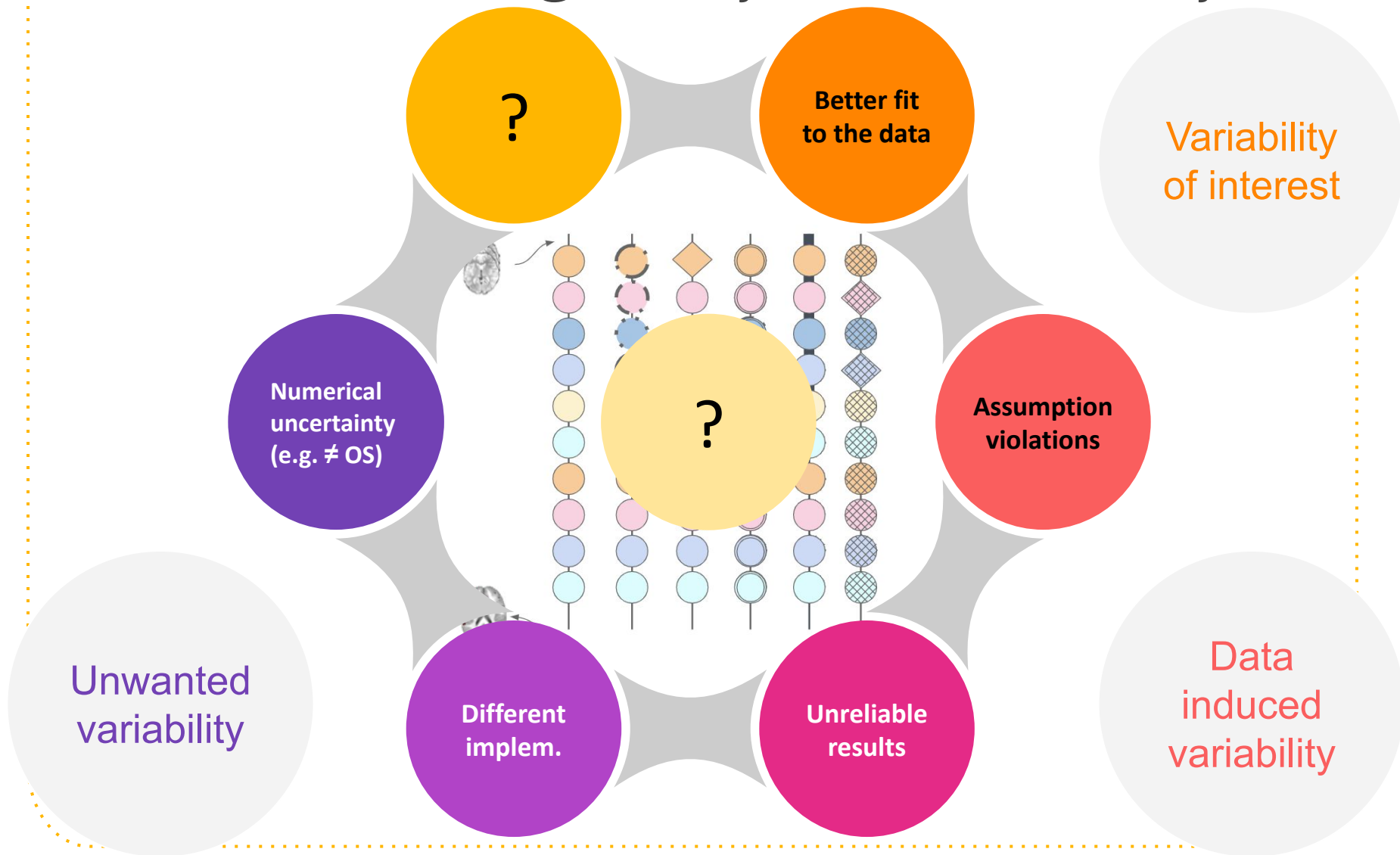
Understanding analytical variability...

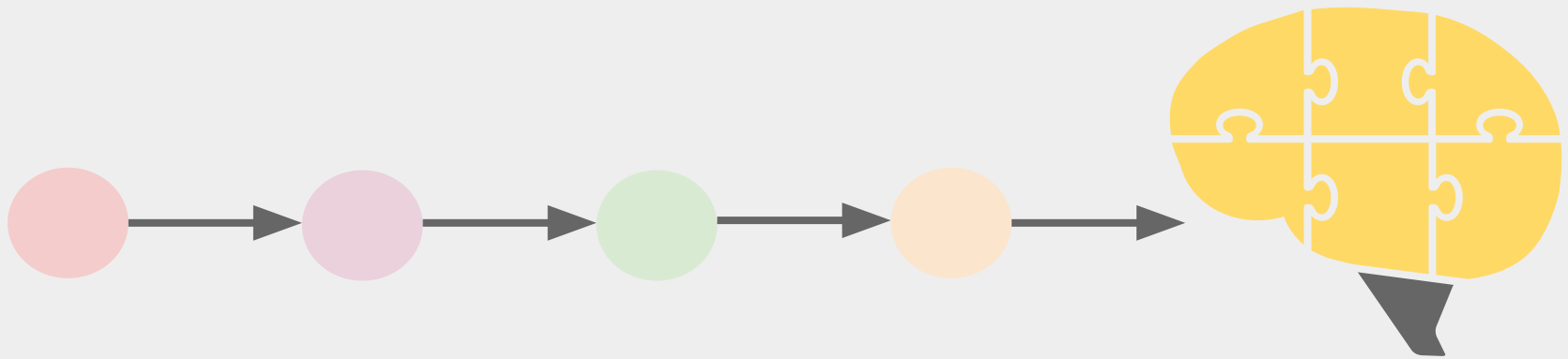


Understanding analytical variability...



Understanding analytical variability...





**Describing the
neuroimaging
pipeline space**



Boris
Clenet

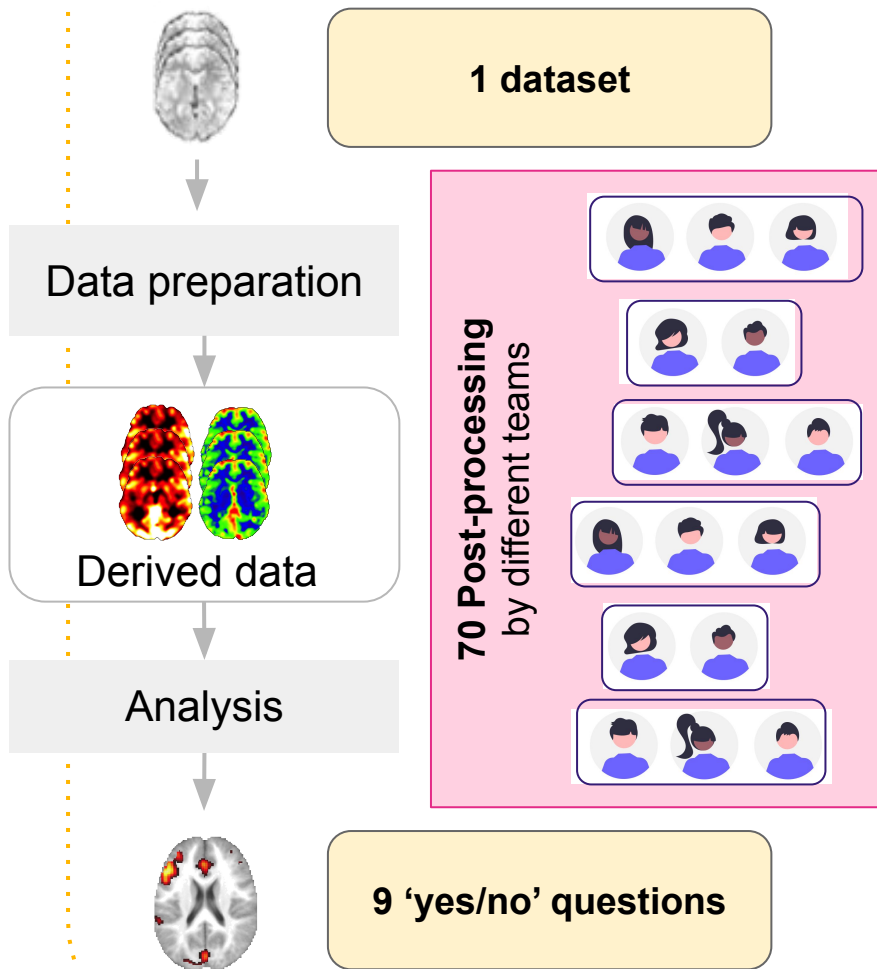


Elodie
Germani



Jérémy
Lefort-Besnard

NARPS open pipelines



NARPS open pipelines



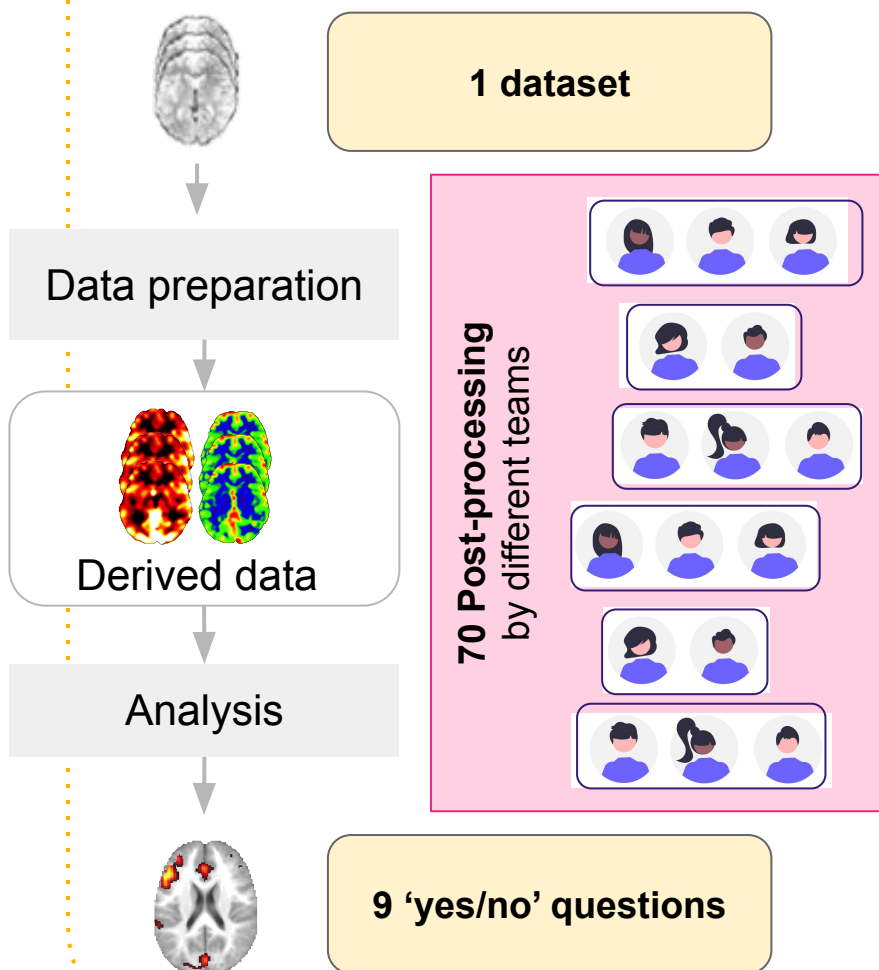
Boris
Clenet



Elodie
Germani



Jérémy
Lefort-Besnard



NARPS open pipelines

- (Re)-coding the 70 pipelines
- Collaborative effort



Project NARPS open pipelines

Join us at the Brainhack Global!

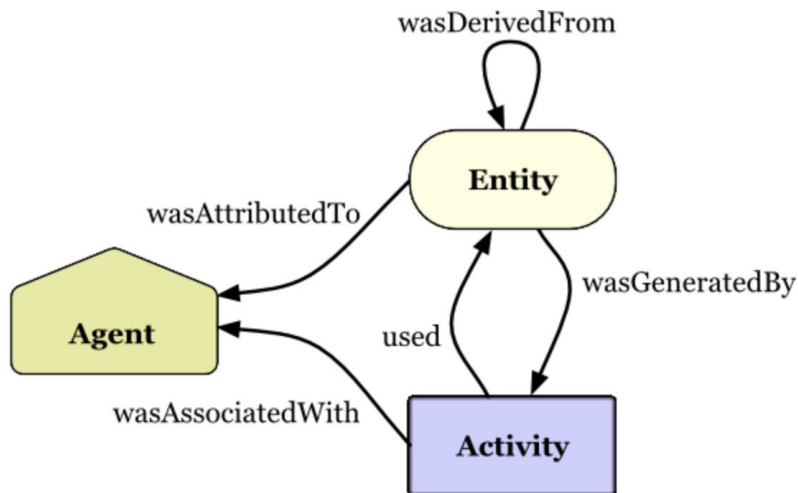


https://github.com/Inria-Empenn/narps_open_pipelines/

Provenance

“**Provenance** describe how an artifact or set of artifacts was produced” [Wikipedia](#)

“**Provenance** is information about entities, activities, and people involved in producing a piece of data [...]” [W3C PROV](#).



W3C PROV model

<https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/>

Provenance Use Case 1

An unambiguous the methods



Provenance Use Case 1 **An unambiguous description of the methods**



Which release of
SPM was used?

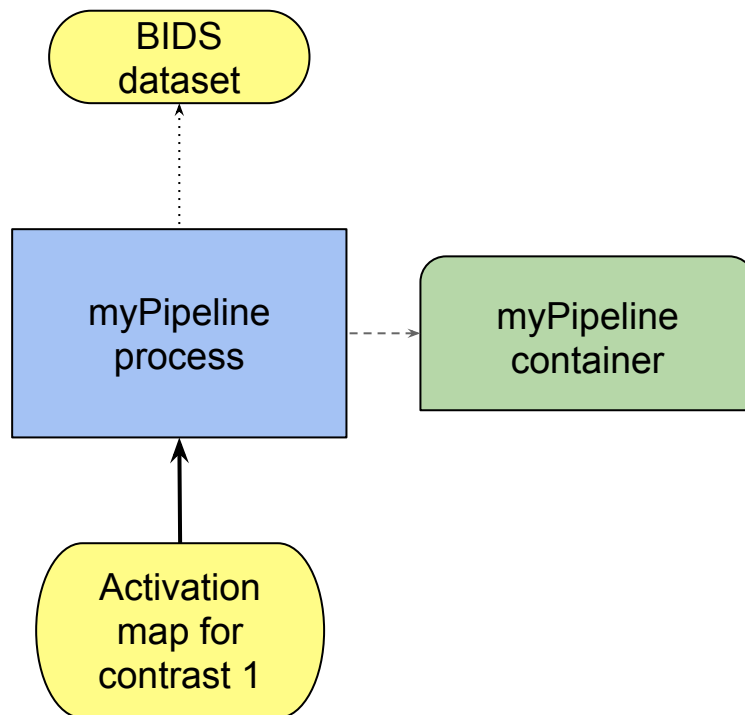
Which registration
algorithm was
selected?

How was the brain
mask computed?

Provenance Use Case 1 **An unambiguous description of the methods**

Can we
reproduce
exactly the same
results
(same data, same
methods)?

Provenance Use Case 1 **An unambiguous description of the methods**



Can we
**reproduce
exactly the same
results**
(same data, same
methods)?

Provenance Use Case 2 **Dealing with unwanted variability**

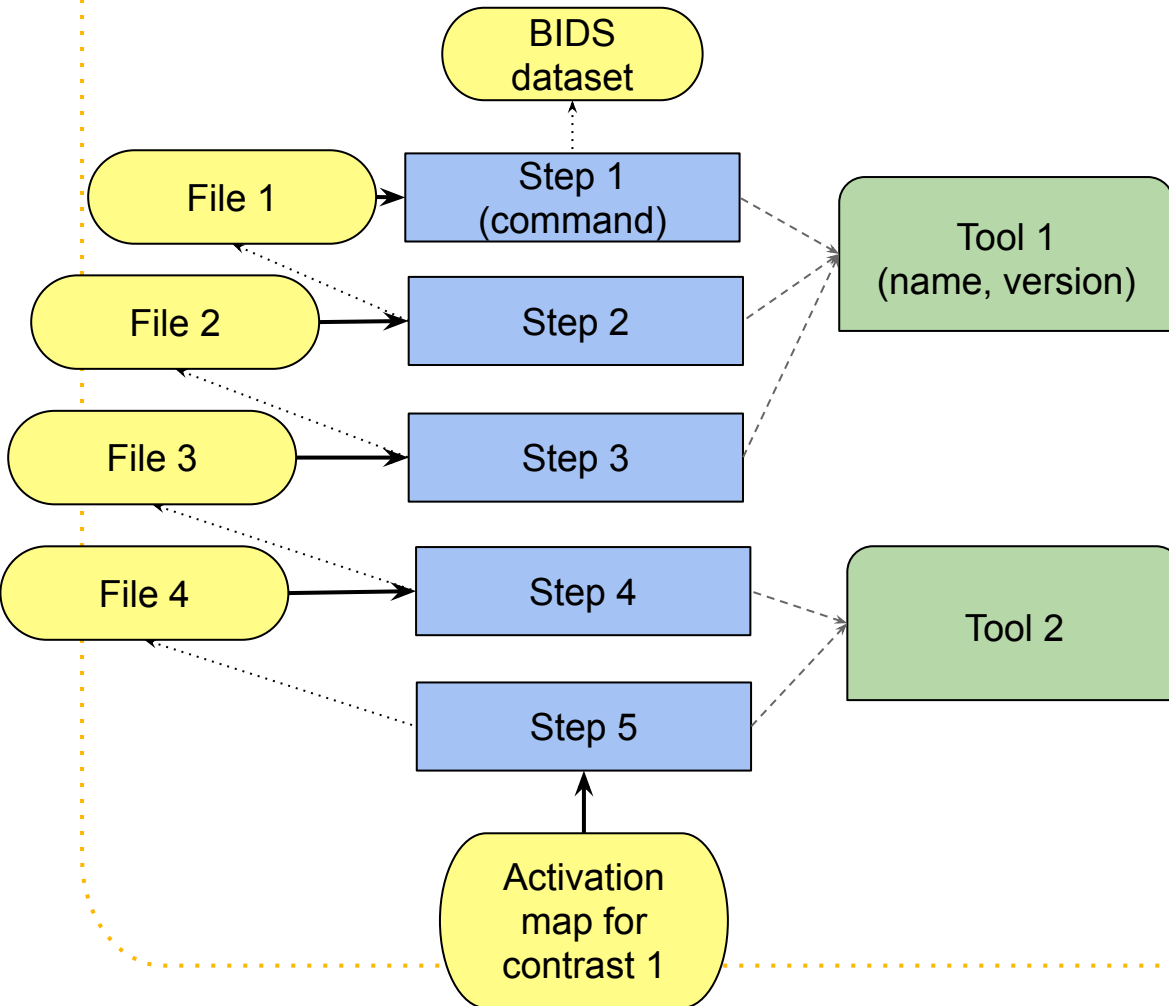
A previously widely-used **parameter is no longer considered as best practice**

Provenance Use Case 2 **Dealing with unwanted variability**

A previously widely-used **parameter is no longer considered as best practice**

A bug was found in a given tool version, can we retrieve all corresponding results?

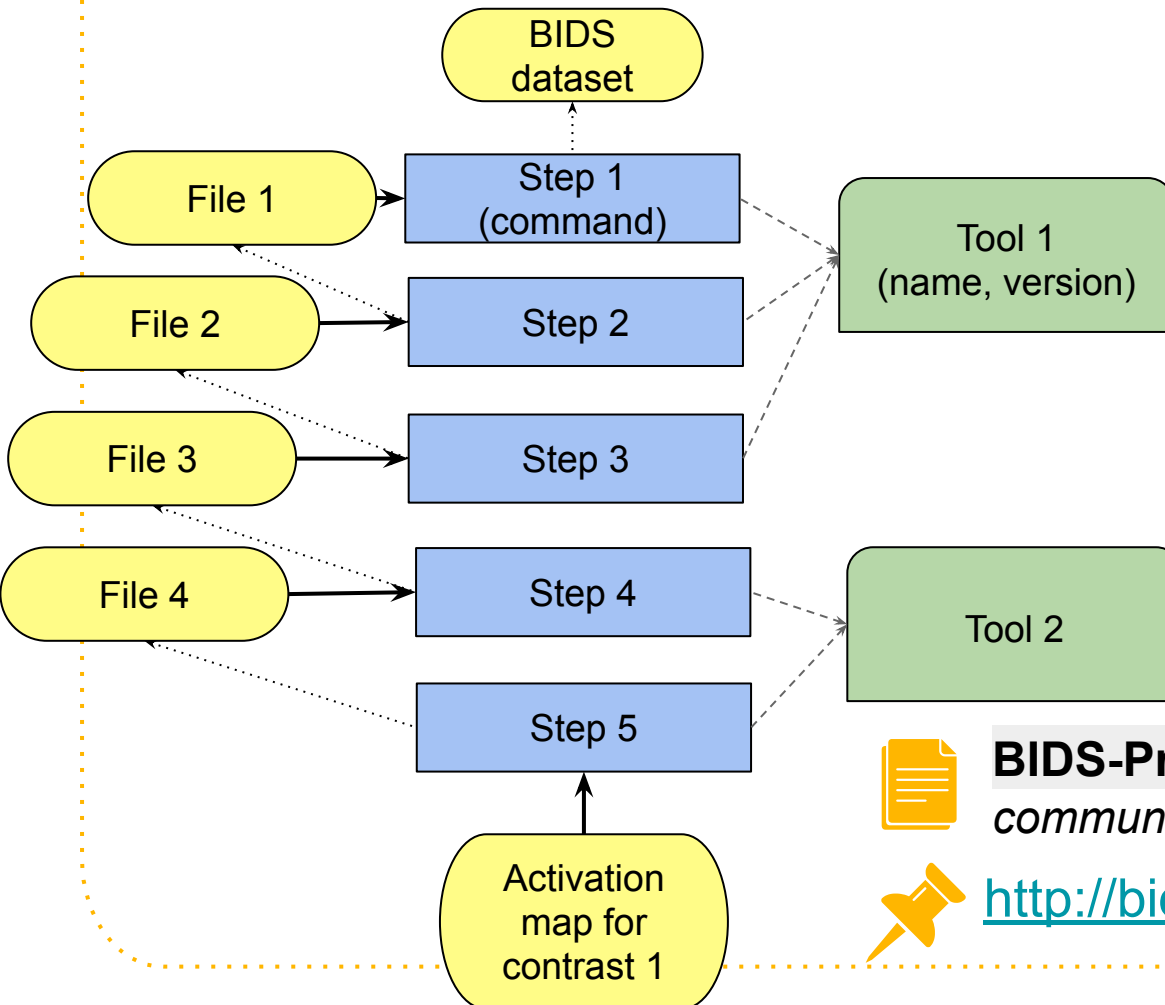
Provenance Use Case 2 **Dealing with unwanted variability**



A previously widely-used **parameter is no longer considered as best practice**

A bug was found in a given tool **version**, can we retrieve all corresponding results?

Provenance Use Case 3 **Exploring factors impacting analytical variability**



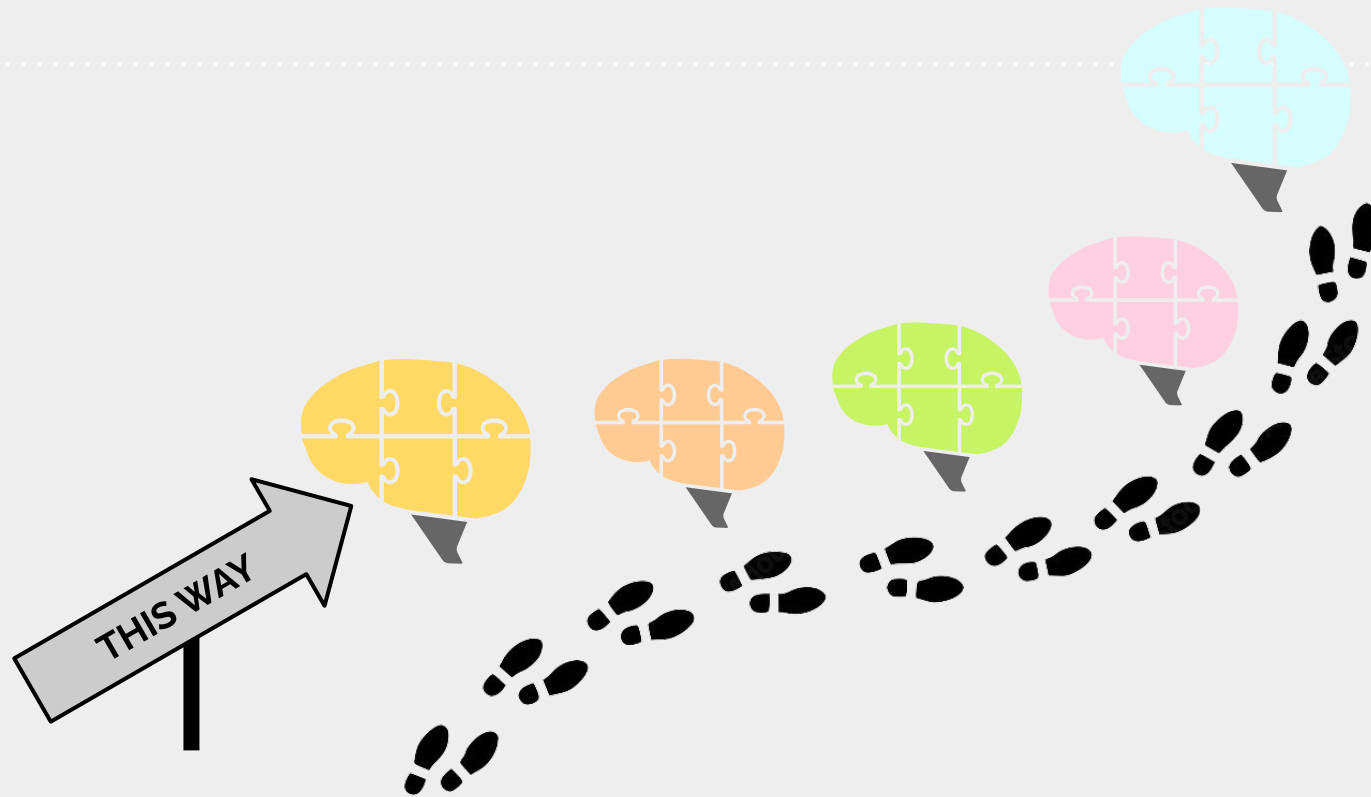
Are my conclusions dependant on the **size of the smoothing kernel** used at Step 5?



BIDS-Prov will soon be open for community feedback!



<http://bids.neuroimaging.io/bep028>



Which solutions **for** **practitioners?**

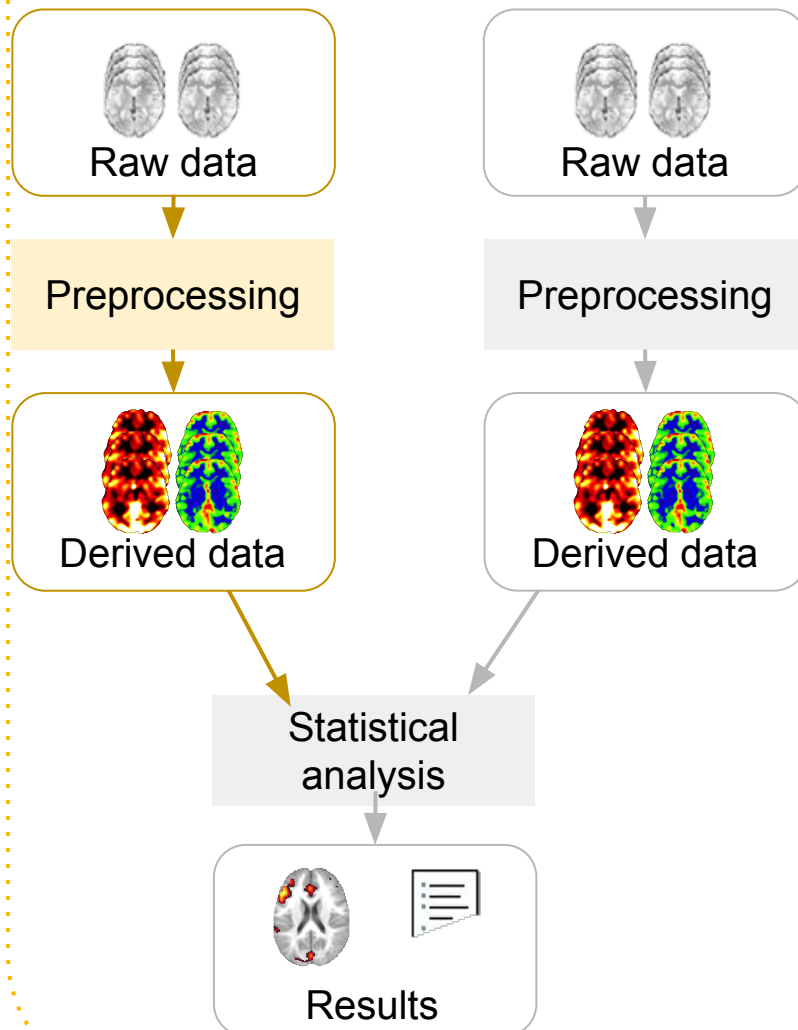
Data re-use



Xavier
Rolland



Pierre
Maurel

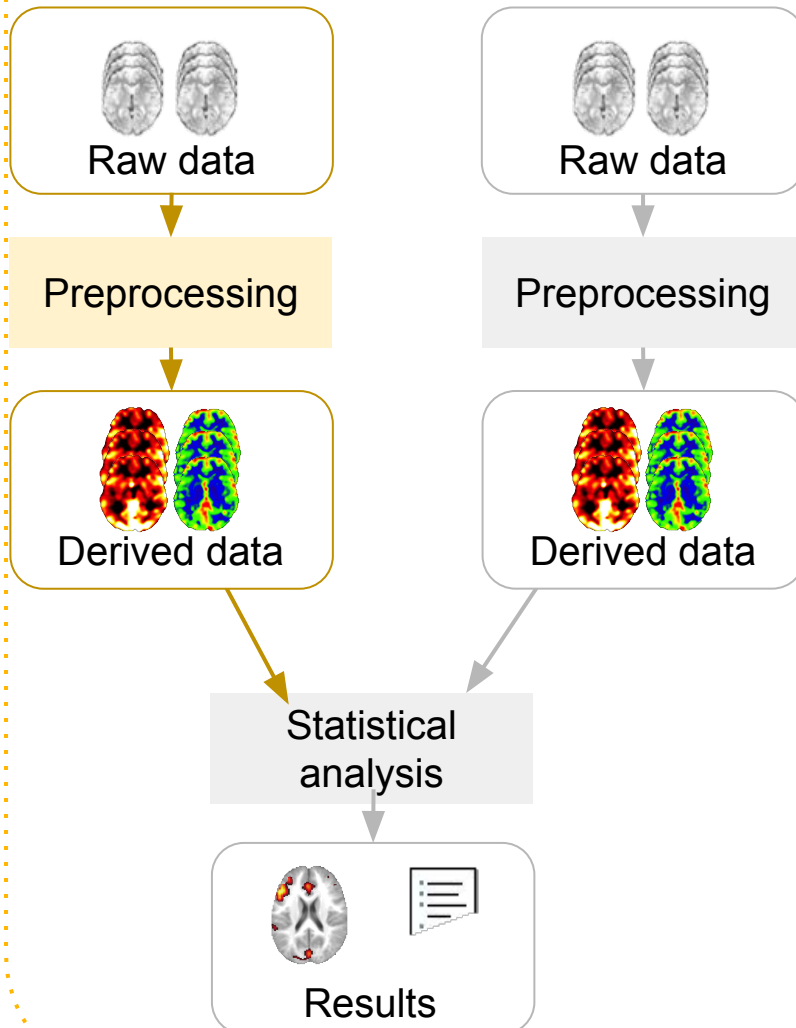


Data re-use

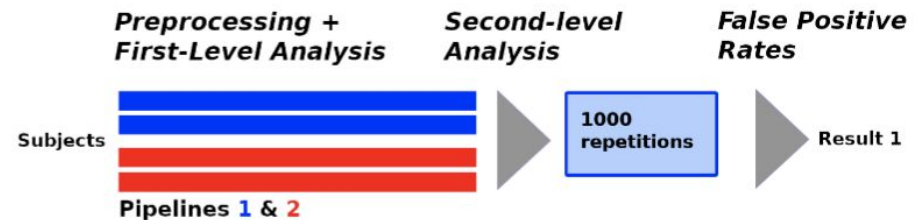


Xavier
Rolland

Pierre
Maurel



- Using HCP data under the null
- Comparisons of pipelines 2-by-2
- 3 variations
 - Smoothing: 5 or 8
 - HRF derivative: yes or no
 - Motion regressors: 0, 6, 24
- “Default” : 5mm, 24 regressors, no temp. deriv.

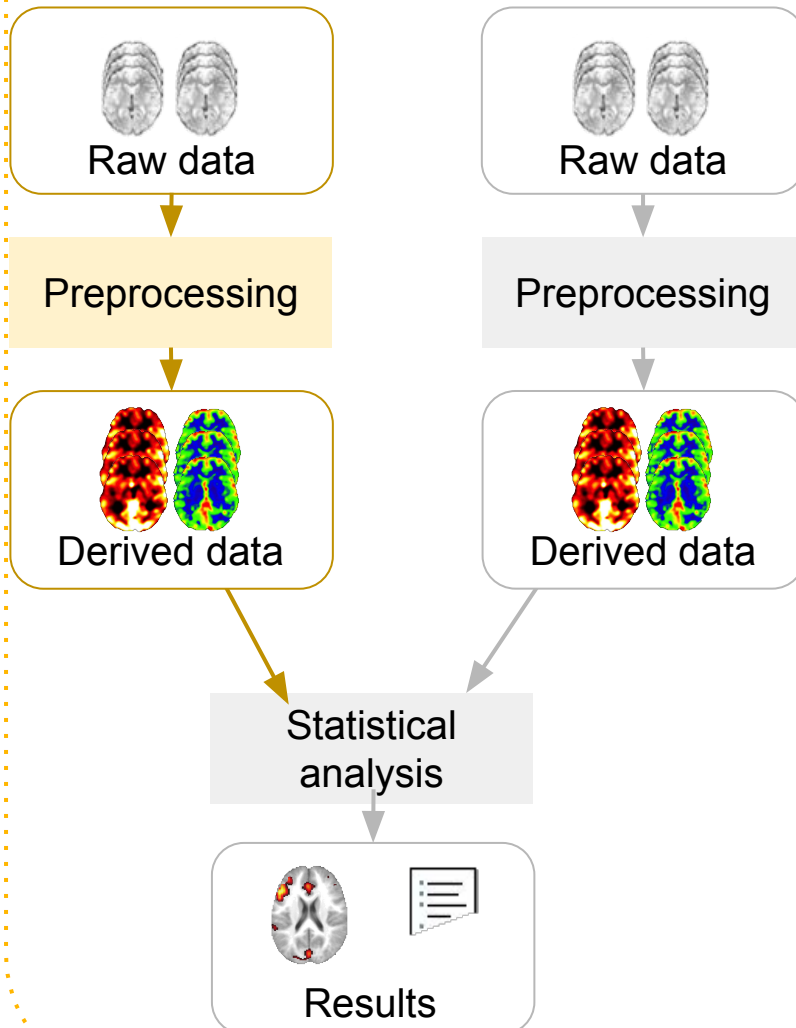


Data re-use



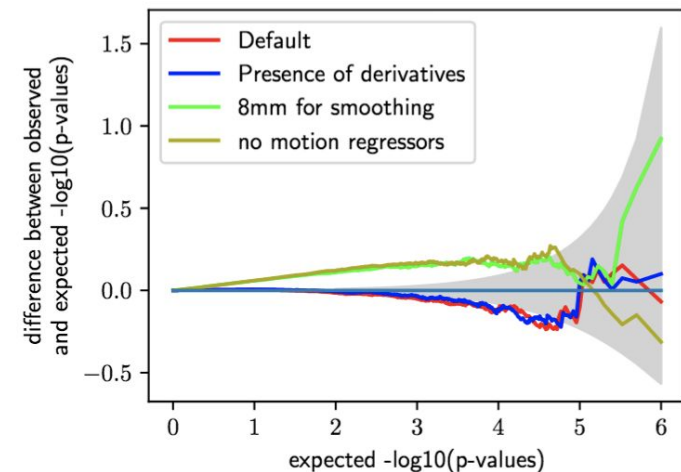
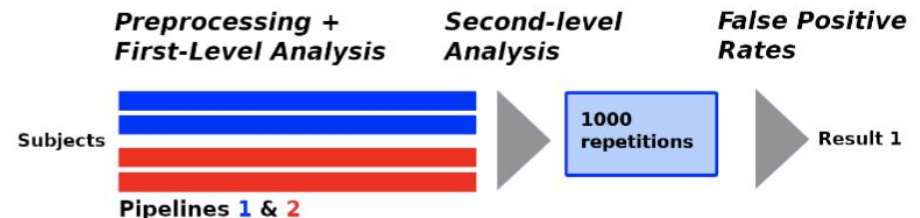
Xavier
Rolland

Pierre
Maurel



[Rolland et. al, ISBI 2022]

- Using HCP data under the null
- Comparisons of pipelines 2-by-2
- 3 variations
 - Smoothing: 5 or 8
 - HRF derivative: yes or no
 - Motion regressors: 0, 6, 24
- “Default” : 5mm, 24 regressors, no temp. deriv.



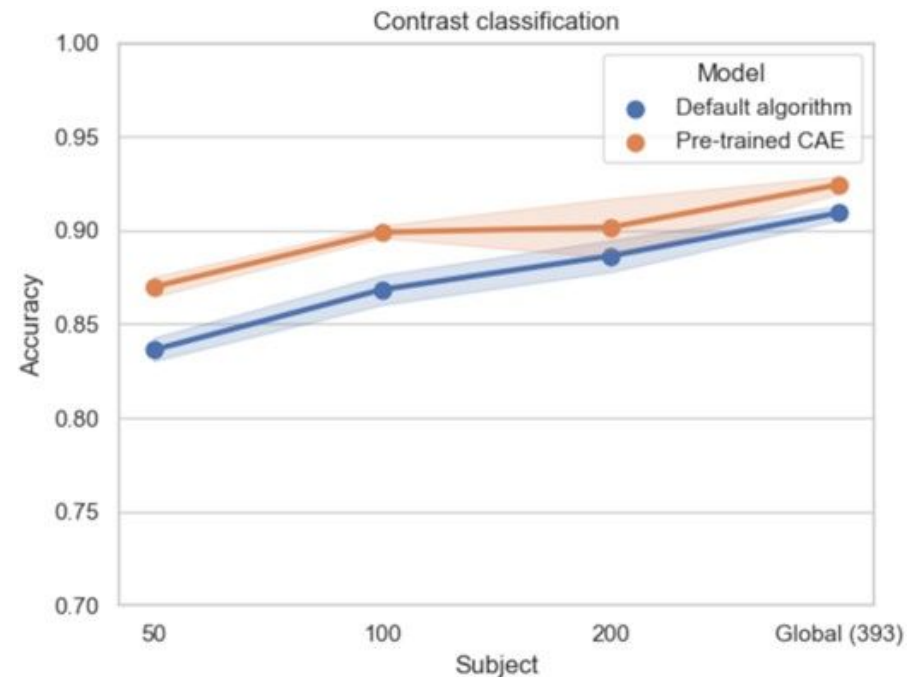
Building on large and more diverse samples



Elodie
Germani

Elisa
Fromont

- Latent representation of fMRI statistic maps based on NeuroVault
- Apply to a **decoding task**



[Germani et. al, Gigascience 2023]

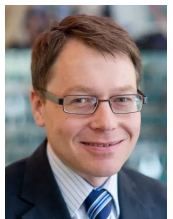


<https://inria.hal.science/hal-03769993>

Same data meta-analysis



Jérémie Lefort-Besnard

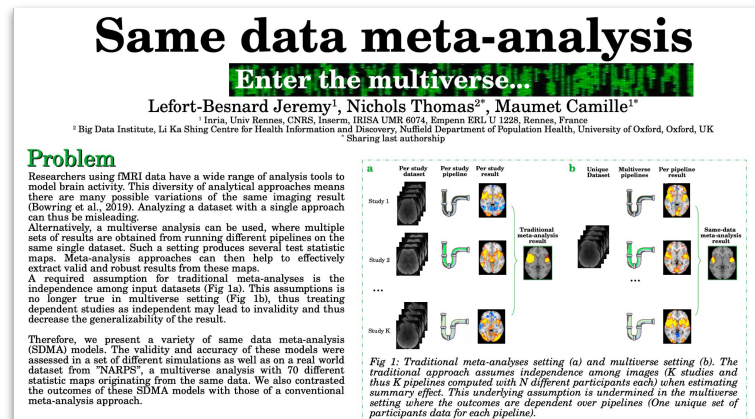


Tom Nichols

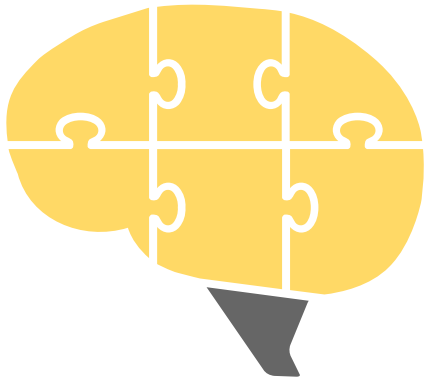
- A statistical approach to **combine the results of different pipelines** and obtain **a single (consensus) summary**



Poster *Systematic review and evaluation of meta-analysis methods for same data meta-analyses* Jérémie Lefort-Besnard



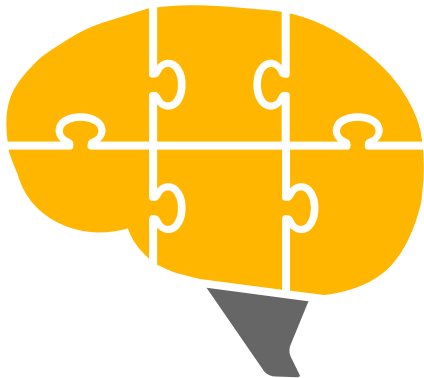
<https://inserm.hal.science/inserm-04222730>



Towards reproducible neuroimaging across different analysis pipelines

Camille Maumet

Thank you!



We are hiring

<https://team.inria.fr/empenn/job-offers>

Credit: Presentation template and icons by
[SlidesCarnival](https://slidescarnival.com/), adapted



@cmaumet@fediscience.org



@cmaumet