Analytic variability in fMRI:
Multivariate and meta-analytic approaches to the problem and solution

Kendra Oudyk[1], Alexandre Pérez[1] , Peer Herholz[1] , Boris Bernhardt[1], Bratislav Misic[1], Jean-Baptiste Poline[1]
[1]Montréal Neurological Institute, McGill University, Montréal, Québec, Canada

Background

Simulations have shown that **analysis choices may influence results** in research using functional magnetic resonance imaging (fMRI)[1].

Until recently, it was unclear how much analytic variability there is in the field, between actual researchers.

Purpose

Here, we complement the work by Botvinik-Nezer and colleagues[2], demonstrating the use of a **multivariate analysis for investigating analytic variability.**

Further, we explore a novel meta-analysis-comparison method for evaluating the **meta-analytic result as a potential solution to analytic variability.**

Data

Neuroimaging Analysis Replication and Prediction Study (NARPS)[2] was the source of our data. In that study, 70 analysis teams analyzed the same task-fMRI dataset, performing whole-brain corrected analyses of task-related activation, using the raw or preprocessed data.

The original fMRI dataset had the following properties:

Participants: 108, with 54 in each condition

Task: Mixed-gambles task, which involves choosing whether to take a gamble(Tom, 2007)

Conditions: 1) Equal Range (range of possible gains = range of possible losses)
2) Equal Indifference (range of possible gains = 2 * range of possible losses)

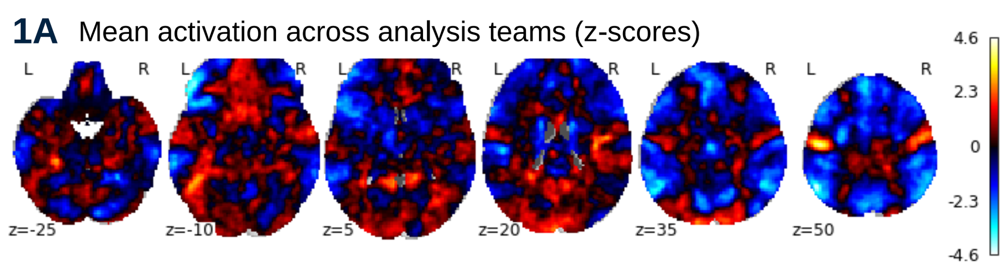
In the present study, we used 2 **group-level z-maps from 55 analysis teams**, both from a test of the positive effect of gain (winning a gamble), one for each condition. We used the Equal-Indifference maps for the main analysis, and the Equal-Range maps for replication (not shown here).

Descriptive statistics

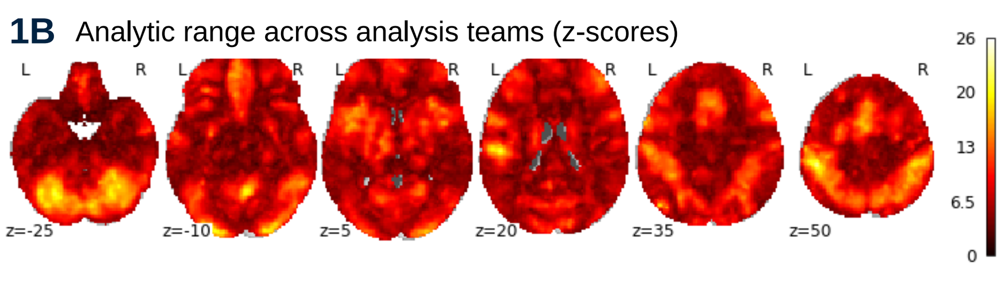
Figure **1A** and **1B** respectively show the mean activation and the analytic range of the group-level z-maps.

Figure **1C-1F** show the number of teams that made various analysis choices.

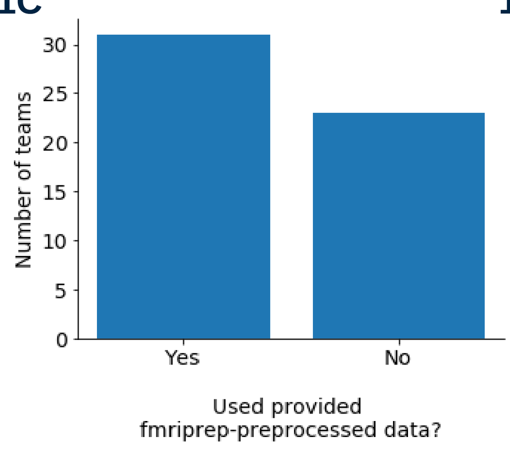
1A Mean activation across analysis teams (z-scores)



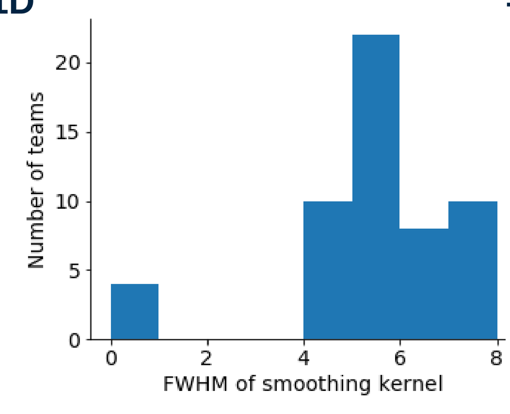
1B Analytic range across analysis teams (z-scores)



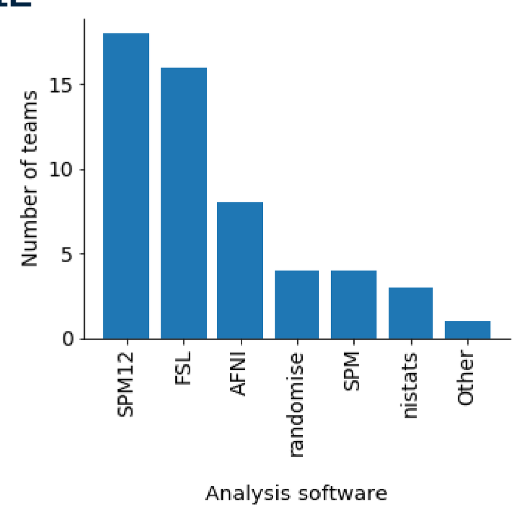
1C



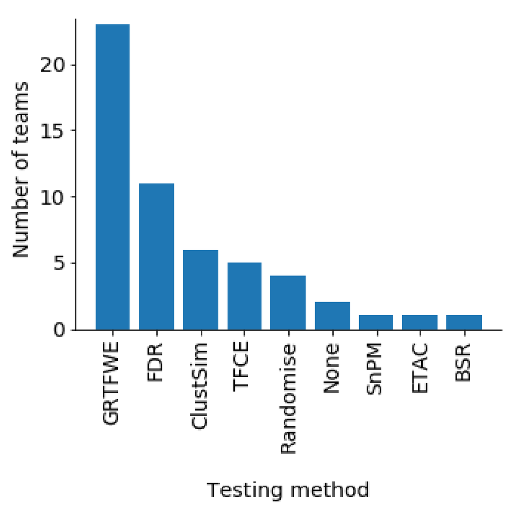
1D



1E



1F

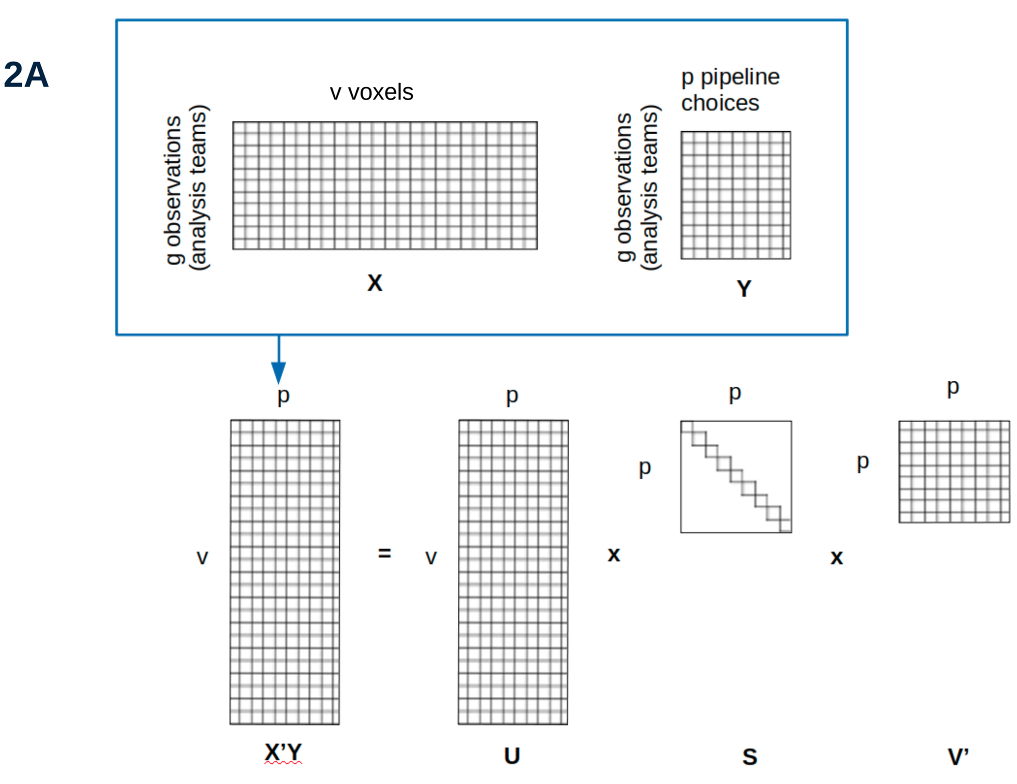


Assessing the problem

(2A) We used **partial least squares** to assess the association between methodological choices and the whole-brain group-level results.

This figure illustrates the use of partial least squares on this data.

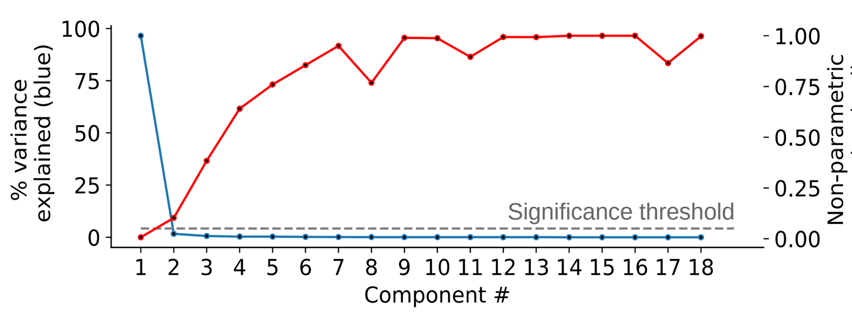
2A



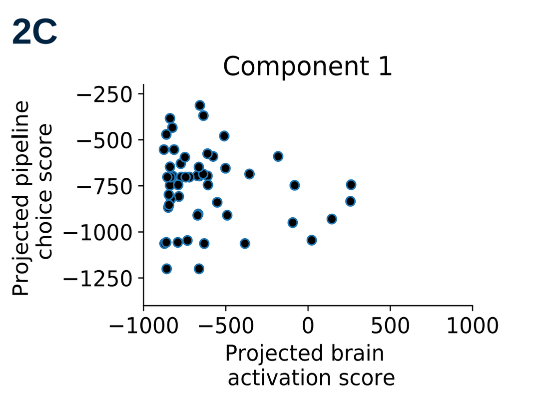
(2B-C) We used **permutation testing** to assess the significance of each component.

Only the first component was significant.

2B



2C

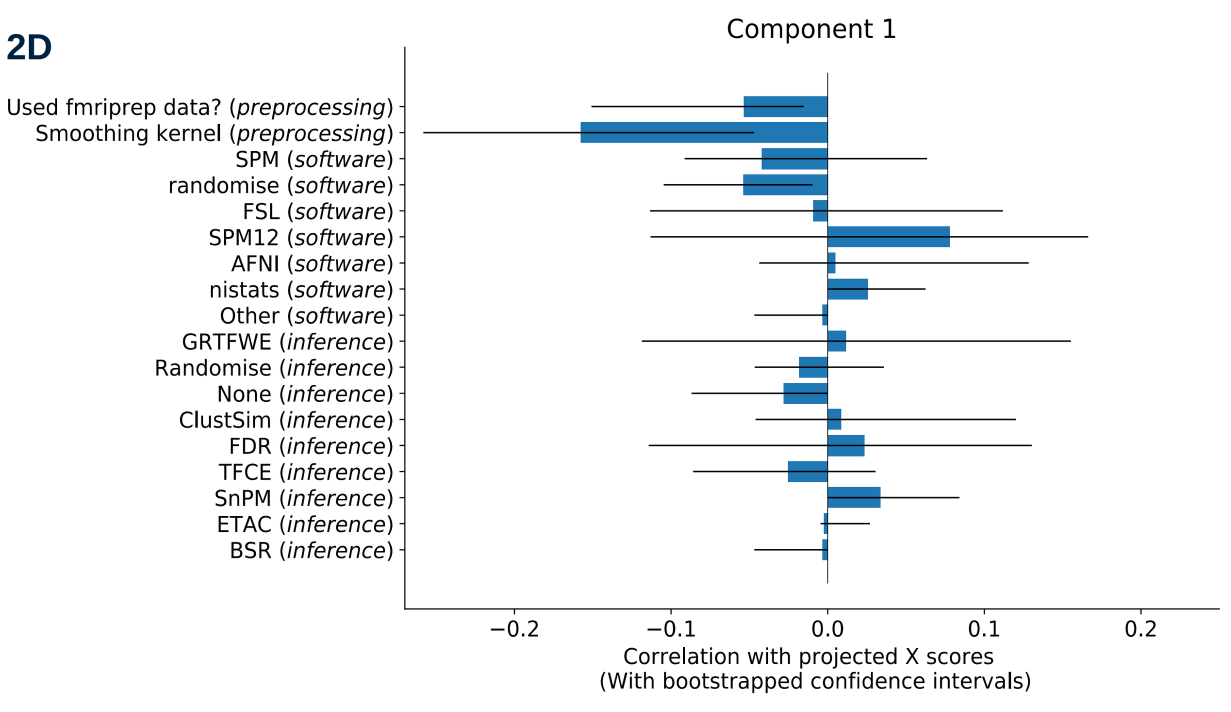


(2D) Within the first component, we used **bootstrap resampling** with replacement to test the stability of each analysis variable's contribution to the component.

The variables with a stable contribution were

- Whether they used the provided preprocessed data or did their own preprocessing,
- The FWHM of their smoothing kernel
- Whether they used randomise as their analysis software.

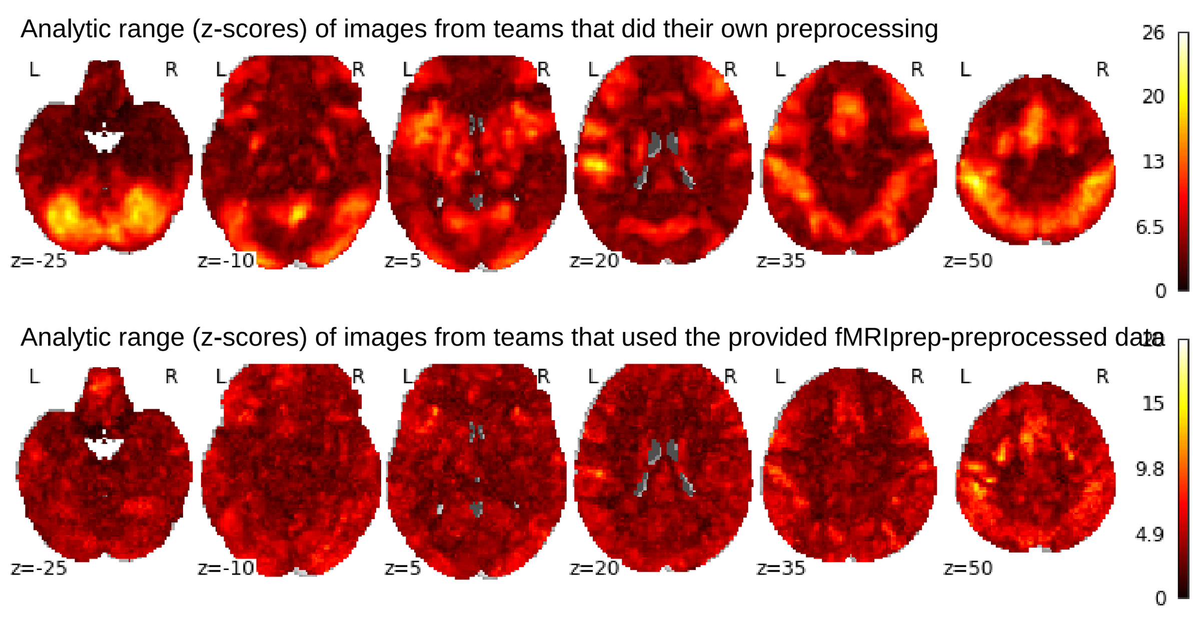
2D



(2E) We followed up by comparing the analytic range of maps that did and did not use the provided preprocessed data.

It appears that the analytic range is more spatially uniform across the teams that used the provided preprocessed data.

2E

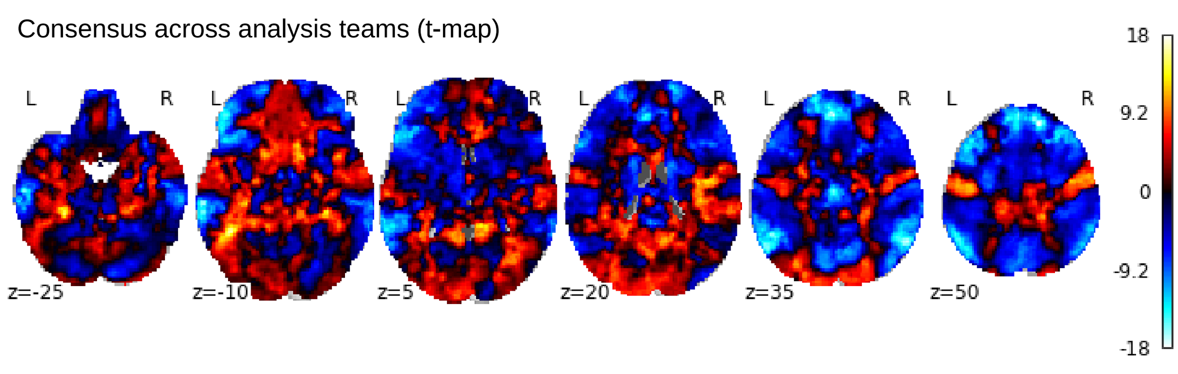


Exploring a solution

Here, we explore whether the consensus result across teams was more accurate than the individual group-level maps, with 'accuracy' defined as similarity (correlation) to the consensus map from the literature.

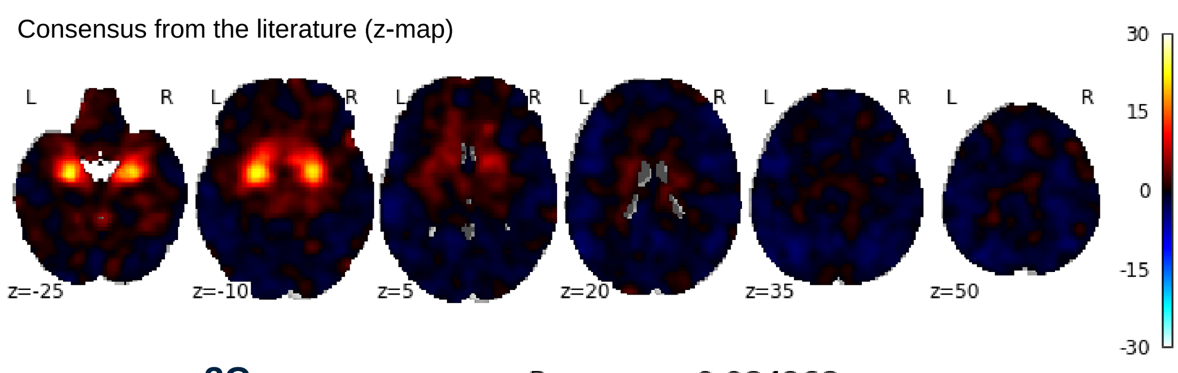
(3A) We used a **fixed-effects general linear model** to find the consensus across teams.

3A Consensus across analysis teams (t-map)



(3B) We used **NeuroQuery** (Dockès, 2018 to find the consensus from the literature, searching for "Loss aversion in decision-making under risk", which is part of the title of the original paper using the task with fMRI[3].

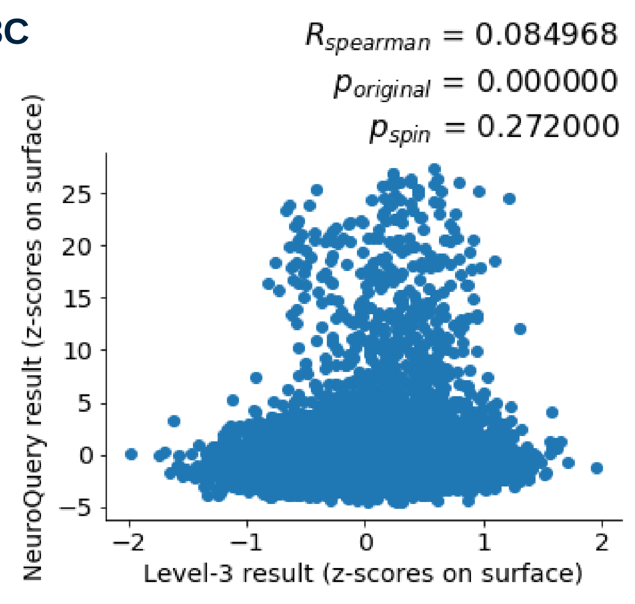
3B Consensus from the literature (z-map)



(3C) We used **Spearman correlation** to measure the similarity between the two consensus maps, and **spin permutation**[cite] to test for significance

There was almost no correlation, and it was not significant.

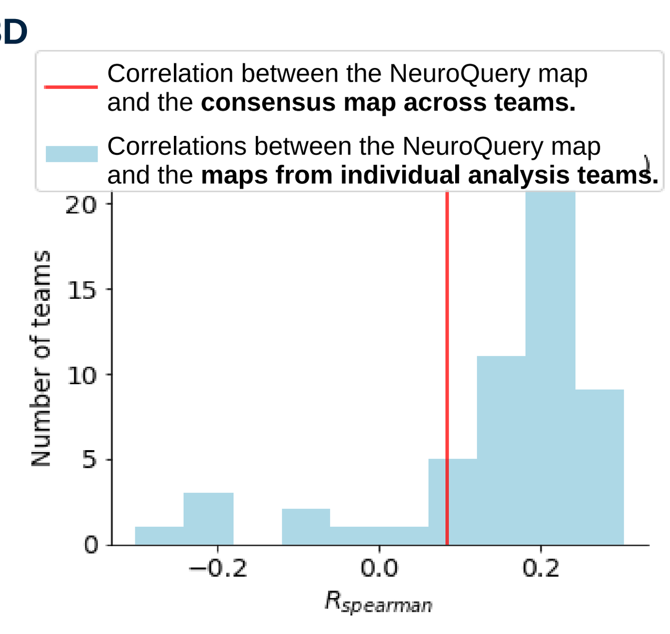
3C



(3D) We also performed **Spearman correlations** between the NeuroQuery result and the result map from each analysis team.

The consensus result across analysis teams was not more highly correlated with the consensus from the literature than were the maps from individual teams.

3D



Discussion

Our **multivariate analysis for investigating analytic variability** suggests that the more important analysis choices were

- Whether the team used the fmriprep-preprocessed data,
- The size of their smoothing kernel, and
- Whether they used randomise to produce their statistical images (in FSL).

Note that in the replication of these analyses in the second task condition (with separate participants), the smoothing kernel did not make a stable contribution to the first component.

Notably, while the **original publication using this dataset**[2] also found that variability in results was related to the smoothing kernel and some choices of analysis software, they reported **no significant effect for the use of the provided preprocessed data** versus performing custom preprocessing.

Our evaluation of the **meta-analytic result as a potential solution to analytic variability** did not show that the consensus across teams was more 'accurate' than the results from individual teams. However, our use of the consensus result from the literature as the ground-truth for evaluating accuracy may not be ideal.

Code

Code for this project is available at **github.com/koudyk/narps_meta**, and draws extensively from code by Ross Markello (https://netneurotools.readthedocs.io/en/latest/auto_examples/plot_mirchi_2018.html) and BrainSpace (https://brainspace.readthedocs.io/en/development/generated/brainspace.null_models.spin.SpinPermutations.html)

References

[1]Carp, J. (2012). On the plurality of (methodological) worlds: estimating the analytic flexibility of FMRI experiments. Frontiers in neuroscience, 6, 149.

[2]Botvinik-Nezer, R., Holzmeister, F., Camerer, C.F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J.A., ..., Nichols, T.E., Poldrack, R.A., Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. Nature. https://doi.org/10.1038/s41586-020-2314-9.

[3]Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007, January). The Neural Basis of Loss Aversion in Decision-Making Under Risk. Science, 315 (5811), 515–518. Retrieved 2019-10- 26, from http://www.sciencemag.org/cgi/doi/10.1126/science.1134239 doi: 10.1126/science.1134239

Acknowledgements

This work was partially funded by the National Institutes of Health (NIH) NIH-NIBIB P41 EB019936 (ReproNim) NIH-NIMH R01 MH083320 (CANDIShare) and NIH RF1 MH120021 (NIDM), the National Institute Of Mental Health under Award Number R01MH096906 (Neurosynth), as well as the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative and the Brain Canada Foundation with support from Health Canada. Author KO is funded through the Fonds de la recherche en santé du Québec.