Multiverse Analysis in Functional Magnetic Resonance Imaging: An Attempt of Data Processing with 12 Pipelines

Note for second draft: I didn't add much to the written draft. I was mostly working on getting more participants' data as well as trying to figure out a analysis plan. With the one participant I mentioned last time, I was able to localize the voxels and compute the "percentage of match rate" across 12 pipelines with a thresholded t-value (I am aware this is invalid as the cutoff is doing inferential statistics, but I now have the code to implement it on more participants after I get their preprocessed data). The average "match" rate is computed across different regions of interests upon clustering the voxels of a specific brain region. I will definitely be continuing this project with all 1080 participants if I don't finish it by the end of the semester, but I for now will be focusing on a small subset of participants and conducting between-pipelines t-tests. I am also learning how to write .py and bash scripts to run parallel jobs on the lab's cluster.

Functional magnetic resonance imaging (fMRI) is a common method to measure neural activity in the brain in the field of cognitive neuroscience. As fMRI gains popularity in the research field, methodological concerns that of such technology arise, with reproducibility as one of the most recognized issues (Botvinik-Nezer & Wager, 2023; Niso et al., 2022). Unlike behavioral measures of human cognition, cognitive neuroscience techniques, specifically fMRI, requires numerous preprocessing steps from raw data prior to data analysis. Kristanto et al. (2024) reviewed 220 papers that utilize graph-based fMRI and discovered 220 unique analysis pipelines. Sixty-one steps of analysis pipelines were identified, 17 of which has frequently varied parameters, yielding in 102 total probable parameters that may vary. As such, it's common to identify that a single research question could be addressed in various ways solely dependent on the analysis pipeline of a specific research team (Dafflon et al., 2022). Multiverse analysis,

indeed, attempts to answer the question: how do different choices of preprocessing and analysis affect the results of fMRI studies?

The issue of variability in fMRI analysis pipelines has been identified by multiple studies, as choices including software environment (Glatard et al., 2015), segmentation (Palumbo et al., 2019), motion correction (Oakes et al., 2005), and registration (Klein et al., 2009) might contribute to various findings. Most recently, Botvinik-Nezer et al. (2020) effectively and empirically demonstrated the issue. Seventy independent research teams were provided with the same fMRI raw dataset and nine hypotheses to test, as they were instructed to conduct the analysis in a conventional way in their laboratories. Results found that conclusions for the hypothesis test of five out of the nine hypotheses yielded large variation from 21.4% to 37.1%. There is low consistency in the nine hypothesis, as significant finding of one hypothesis was discovered by 84.3% of the research teams, and the other three non-significant hypothesis reached 94.3% consensus. In fact, across all hypotheses, 20% teams reported a different conclusion from the majority of teams, indicating a level that almost falls midway between finding completely random results and completely consistency (i.e., the maximum possible variability is 50%). Several parameters in the preprocessing and analysis pipeline were discovered to result in significant findings at a higher chance. Size of the smoothing kernel, software package used, and method of multiple correction were all significant factors that could lead to variation in the analysis, as larger smoothing kernel (decreasing signal-to-noise ratio by taking the average value of nearby voxels), using FSL over SPM software, and parametric methods often yield in higher probability of significant outcomes.

Notably, while various studies have indicated the effect of choices in fMRI analysis that tend to lead to specific outcomes (Botvinik-Nezer et al., 2020; Bowring et al., 2022; Li et al.,

2021), Kiar et al. (2024) pointed out that there is no single stage in the pipeline that is solely responsible for accounting the differences; rather, the interaction complexity between steps also contributes to the variation. Such suggestion invites a conversation that discusses the higher-order mechanism of inter-pipeline variability. However, at the current stage, researchers' attempts continue to primarily involve parsing out the effect of step-specific differences.

The goal of this project is to investigate the variation in subject-level contrast maps processed with different pipelines. The methodology of the project is an adaptation of Germani et al. (2025), where they conducted a fMRI mega-analysis and processed the raw data with 24 unique pipelines in a motor task. Importantly, the present project focuses on a decision-making task that represents higher cognitive function. While the original study included analysis of the effect of software packages (i.e., FSL and SPM), this project attempts to analyze 12 pipelines using only the FSL software.

Method

This project analyzes the gambling task of the HCP Young Adult S1200 release (Van Essen et al., 2013). The experimental task was adapted from Delgado et al. (2000). Participants were instructed to guess the number on a card to win or lose money. They were told that the card number ranges from 1-9, and their task is to indicate if the number is less than 5. Participants were immediately given feedback after each trial as the correct number is shown, which included positive feedback (reward; "$1" with green arrow), negative feedback (loss; "-$0.50 with red arrow), and neutral feedback (number 5). There were two runs of the task. Within each run, there were 2 mostly reward blocks (6 reward trials paired with either 1 loss trial and 1 neutral trial, 2 loss trials, or 2 neutral trials) and 2 mostly loss blocks (6 loss trials paired with either 1 loss trial

and 1 neutral trial, 2 reward trials, or 2 neutral trials). Each block was consisted of six trials. Each trial starts with a question mark where participants were able to make a response (1500 ms) and is followed by 1000 ms feedback.

Variation between the pipelines include smoothing kernel Full-Width Half-Maximum (FWHM), the number of motion regressors included in the general linear model (GLM) of the first-level analysis, and the presence of the derivatives of the hemodynamic response function (HRF) in the first-level GLM. There are 2 levels of the smoothing kernel FWHM (5 mm, 8 mm), 3 levels of the number of motion regressors (0, 6: 3 rotations + 3 translations, 24: 3 rotations + 3 translations + 6 derivatives and the 12 corresponding squares), and 2 levels of the derivatives of the HRF (present, absent).

Beyond calculating preprocessing steps as outlined above, first-level t-maps of each participant were obtained for each pipeline, such that there are 12 t-maps for each participant. The Harvard Oxford atlas is applied to categorize all voxels into 48 regions of interest (ROIs) in a three-dimensional space. The t-distributions of the contrast parameter were examined using a independent 2-samples t-test on a voxel-level, with alpha level of 0.05 and FWE correction for multiple comparison.

References

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., … Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84–88. https://doi.org/10.1038/s41586-020-2314-9

Botvinik-Nezer, R., & Wager, T. D. (2023). Reproducibility in Neuroimaging Analysis: Challenges and Solutions. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *8*(8), 780–788. https://doi.org/10.1016/j.bpsc.2022.12.006

Bowring, A., Nichols, T. E., & Maumet, C. (2022). Isolating the sources of pipeline-variability in group-level task-fMRI results. *Human Brain Mapping*, *43*(3), 1112–1128. https://doi.org/10.1002/hbm.25713

Dafflon, J., F. Da Costa, P., Váša, F., Monti, R. P., Bzdok, D., Hellyer, P. J., Turkheimer, F., Smallwood, J., Jones, E., & Leech, R. (2022). A guided multiverse study of neuroimaging analyses. *Nature Communications*, *13*(1), 3758. https://doi.org/10.1038/s41467-022-31347-8

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of Neurophysiology*, *84*(6), 3072–3077. https://doi.org/10.1152/jn.2000.84.6.3072

Germani, E., Rolland, X., Maurel, P., & Maumet, C. (2025). On the validity of fMRI mega-analyses using data processed with different pipelines. *Imaging Neuroscience*, *3*, imag_a_00522. https://doi.org/10.1162/imag_a_00522

Glatard, T., Lewis, L. B., Ferreira da Silva, R., Adalat, R., Beck, N., Lepage, C., Rioux, P., Rousseau, M.-E., Sherif, T., Deelman, E., Khalili-Mahani, N., & Evans, A. C. (2015). Reproducibility of neuroimaging analyses across operating systems. *Frontiers in Neuroinformatics*, *9*, 12. https://doi.org/10.3389/fninf.2015.00012

Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., Song, J. H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R. P., Mann, J. J., & Parsey, R. V. (2009).

Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, *46*(3), 786–802. https://doi.org/10.1016/j.neuroimage.2008.12.037

Kristanto, D., Burkhardt, M., Thiel, C., Debener, S., Gießing, C., & Hildebrandt, A. (2024). The multiverse of data preprocessing and analysis in graph-based fMRI: A systematic literature review of analytical choices fed into a decision support tool for informed analysis. *Neuroscience & Biobehavioral Reviews*, *165*, 105846. https://doi.org/10.1016/j.neubiorev.2024.105846

Li, X., Esper, N. B., Ai, L., Giavasis, S., Jin, H., Feczko, E., Xu, T., Clucas, J., Franco, A., Heinsfeld, A. S., Adebimpe, A., Vogelstein, J. T., Yan, C.-G., Esteban, O., Poldrack, R. A., Craddock, C., Fair, D., Satterthwaite, T., Kiar, G., & Milham, M. P. (2024). *Moving Beyond Processing and Analysis-Related Variation in Neuroscience* (p. 2021.12.01.470790). bioRxiv. https://doi.org/10.1101/2021.12.01.470790

Niso, G., Botvinik-Nezer, R., Appelhoff, S., De La Vega, A., Esteban, O., Etzel, J. A., Finc, K., Ganz, M., Gau, R., Halchenko, Y. O., Herholz, P., Karakuzu, A., Keator, D. B., Markiewicz, C. J., Maumet, C., Pernet, C. R., Pestilli, F., Queder, N., Schmitt, T., … Rieger, J. W. (2022). Open and reproducible neuroimaging: From study inception to publication. *NeuroImage*, *263*, 119623. https://doi.org/10.1016/j.neuroimage.2022.119623

Oakes, T. R., Johnstone, T., Ores Walsh, K. S., Greischar, L. L., Alexander, A. L., Fox, A. S., & Davidson, R. J. (2005). Comparison of fMRI motion correction software tools. *NeuroImage*, *28*(3), 529–543. https://doi.org/10.1016/j.neuroimage.2005.05.058

Palumbo, L., Bosco, P., Fantacci, M. E., Ferrari, E., Oliva, P., Spera, G., & Retico, A. (2019). Evaluation of the intra- and inter-method agreement of brain MRI segmentation software

packages: A comparison between SPM12 and FreeSurfer v6.0. *Physica Medica: PM: An International Journal Devoted to the Applications of Physics to Medicine and Biology: Official Journal of the Italian Association of Biomedical Physics (AIFB)*, *64*, 261–272. https://doi.org/10.1016/j.ejmp.2019.07.016

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., & WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041