Mapping Causal Brain Interactions in fMRI Decoded Neurofeedback

Fahimeh Arab (farab002@ucr.edu)
Department of Electrical and Computer Engineering
University of California Riverside
Riverside, California, USA

AmirEmad Ghassami (ghassami@bu.edu)
Department of Mathematics and Statistics
Boston University
Boston, Massachusetts, USA

Megan A. K. Peters (megan.peters@uci.edu)
Department of Cognitive Sciences
University of California Irvine
Irvine, California, USA

Erfan Nozari (erfan.nozari@ucr.edu)
Department of Mechanical Engineering
Department of Electrical and Computer Engineering
Department of Bioengineering
University of California Riverside
Riverside, California, USA

Abstract

Neurofeedback (NF), including its specialized form Decoded Neurofeedback (DecNef), holds great promise for improving mental health and cognitive function by allowing individuals to voluntarily control their brain activity. However, there exists vast subject-to-subject and region-to-region variability in neurofeedback outcomes and the causal mechanisms involved in successful neurofeedback are largely unknown. In this paper, we investigate the neural mechanisms behind this variability using whole-brain causal connectomes derived from functional Magnetic Resonance Imaging (fMRI) data from a DecNef study aimed at reducing common fears via subconscious induction of feared images. During NF, we found strongest causal connections among regions of the attention and somatomotor subnetworks. Additionally, the net strength of causal effects between most pairs of subnetworks was significantly correlated with mean NF score, though with different signs. Specifically, we found most connections among visual, subcortical, default mode, and dorsal attention subnetworks to support NF success, while most connections among ventral attention, somatomotor, limbic, and control subnetworks correlated negatively with NF scores.

Keywords: decoded neurofeedback; causal discovery; brain networks

Introduction

Neurofeedback (NF) offers a unique avenue for influencing neural dynamics (Watanabe et al., 2017) and has become an attractive therapeutic option for conditions such as Attention-Deficit Hyperactivity Disorder (ADHD) (Enriquez-Geppert et al., 2019; Arns et al., 2020), anxiety (Pindi et al., 2022; Tolin et al., 2020), and substance use disorders (Russo et al., 2023; Trudeau, 2005; Fielenbach et al., 2019). Decoded NF (DecNef) is a particular form of NF that allows for the manipulation of precise neural activity patterns associated with specific cognitive processes or behaviors, without participants' explicit awareness of the experimental goals (Shibata et al., 2019). DecNef has been documented to induce changes in visual sensitivity and color perception (Amano et al., 2016), fear memory modulation (Koizumi et al., 2016; Taschereau-Dumouchel et al., 2018), enhancement of perceptual confidence (Cortese et al., 2016), and facial preference (Shibata et al., 2016). However, a notable challenge in neurofeedback stems from the variability in participants' ability to alter their target brain activity. Around 30% of NF participants struggle to self-regulate (Hammer et al., 2012; Sitaram et al., 2017) and some even experience diminished feedback scores over time (Sitaram et al., 2017). Understanding the neural mechanisms of self-regulation thus has significant potential for improving experimental and clinical protocols as well as advancing neurofeedback tools (Sitaram et al., 2017).

In this study we hypothesize that individual variations in neurofeedback outcomes stems from and can thus be attributed to differences in how different networks of brain regions causally interact with each other. Supporting the potential utility of this approach, a recent study employed Dynamical Causal Modeling (DCM) to find the brain processes associated with successful brain self-regulation of supplementary motor area (SMA) via NF (Vargas et al., 2023), and found significant distinctions in causal connectivity patterns between successful and non-successful learners. However, this study investigated causal connections only among SMA and three other regions of interest (ROIs). In contrast, the complexity of target brain activations in neurofeedback and the subjects' complete unawareness of them make *whole-brain* causal discovery essential for discovering the neural mechanisms underlying DecNef.

Method

Data and Experimental Setup. In this study, we used whole-brain fMRI data from $n = 10$ subjects participating in a DecNef study aimed at reducing common fears via subconscious induction of feared images (Taschereau-Dumouchel et al., 2018). As is typical in DecNef, subjects were unaware of the target of NF and used trial and error to find mental strategies for maximizing end-of-trial scores, which measured how closely the subject's voxel-wise response in ventral temporal cortex matched a target pattern evoked during viewing of a feared animal. fMRI was collected with TR = 2s (see more details at (Cortese et al., 2021)). Each subject participated in 3 DecNef sessions with a minimum of 8 runs per session and 16 trials per run.

fMRI Preprocessing. We preprocessed the data using fMRIPrep (Esteban et al., 2019) with default parameters, followed by 9P confound regression (Ciric et al., 2017) and averaging into 100 cortical (Schaefer 100x7 atlas (Schaefer et al., 2018)) and 16 subcortical parcels (Melbourne Scale I atlas (Tian et al., 2020)).

Causal Discovery. We used the recently proposed CaLLTiF (Causal discovery for Large-scale Low-resolution Time-series with Feedback) algorithm (Arab et al., 2023) to extract whole-brain causal connectomes during DecNef. We truncated data to the minimum amount available for all subjects, resulting in 8 runs per subject-session and 175 fMRI volumes per run. To ensure a sufficient number of samples for CaLLTiF, we computed one causal graph for every pair of runs in each session. This resulted in $\binom{8}{2} = 28$ graphs per session and a total of 840 causal graphs across all subjects.

While CallTiF was originally designed for handling low-resolution time series (TR = 0.72s in the original study), we further modified it to handle the even slower fMRI data we had (TR = 2s). In CaLLTiF a causal link is established from a node (parcel) X_i to node X_j with a lag of $\tau \geq 0$ samples if $X_i(t-\tau)$ is significantly correlated with $X_j(t)$ after conditioning on all other nodes and their lagged values (ensuring correlation is not due to a common cause or mediation through other nodes). When $\tau = 0$, a bidirectional feedback connection is placed between X_i and X_j, unless when at least one

variable also causes the other with $\tau > 0$, in which case the direction of causality is determined based on the lagged effect(s). However, as noted in (Arab et al., 2023, Supp. Note 1), lagged effects become exponentially harder to detect with increasing TR and finite samples, even though the presence of a statistically significant contemporaneous effect ($\tau = 0$) is proof that a lagged effect must exist. Therefore, in this work we slightly modified CaLLTiF such that for pairs of nodes with a statistically significant contemporaneous effect (detected at the originally suggested strict level $\alpha = 0.0025$), we relaxed the threshold of statistical significance on their lagged effects from $\alpha = 0.0025$ to $\alpha = 0.05$.

Results and Discussion

Somatomotor and Attention Networks Exhibit Strongest Causal Connections During Neurofeedback. Both functional and causal graphs are highly consistent among subjects, sessions, and runs, while causal graphs are generally sparser than functional graphs due to their pruning of spurious correlations (Fig. 1a). Notably, in addition to strong internal connections within all 8 functional subnetworks (subcortical + 7 cortical (Yeo et al., 2011)), both functional and causal graphs show strong connections between attention (ventral and dorsal) and somatomotor subnetworks.

Furthermore, the lower density of causal graphs is *not* uniform across the whole network, but rather precisely structured along functional subnetworks. In particular, within-subnetwork connections have become relatively *denser* (relative to full graph density) in causal graphs compared to functional graphs (Fig. 1b). Even more notably, connections between attention and somatomotor subnetworks have also become relatively denser in the causal graphs, whereas the opposite is true for sets of randomly selected 3 subnetworks (Fig. 1c).

Strengths of Causal Connections Predict Neurofeedback Success. We summarized each parcel-level causal graph into a subnetwork-level causal graph by measuring the percentage of edges that go from parcels within one subnetwork to those within another (Fig. 2a). We then measured if the strength of each edge in this graph predicts how successful that subject was in inducing the desired brain activation pattern (DecNef score).

Interestingly, we found most subnetwork-level edges to correlate significantly with score (Fig. 2b). In particular, visual, subcortical, default mode, and dorsal attention subnetworks form a cluster where almost all edges positively and significantly correlate with scores. On the other hand, ventral attention, somatomotor, limbic, and control subnetworks form a cluster within which most causal effects *significantly negatively* correlate with the score. Most notably, the outstandingly-strong causal links within and between ventral attention and somatomotor networks largely *impede* neurofeedback success. In contrast, almost all edges associated with the visual and subcortical networks, despite being weak on average, support neurofeedback success.

The vast positive associations that we observed between

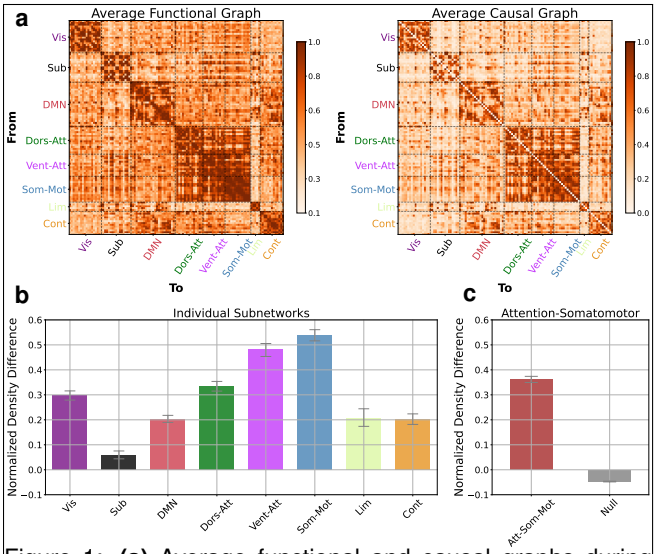

Figure 1: **(a)** Average functional and causal graphs during DecNef. **(b)** Distribution (mean $\pm$ 1 s.e.m.) of the difference of relative internal density of each functional subnetwork ($\frac{\rho^{\text{causal}}_{\text{subnet}}}{\rho^{\text{causal}}_{\text{full}}} - \frac{\rho^{\text{func}}_{\text{subnet}}}{\rho^{\text{func}}_{\text{full}}}$). **(c)** Same as in (b) but for density of connections *between* attention and somotomotor subnetworks (red) vs. between random subsets of 3 subnetworks (gray).

the strength of causal connections and neurofeedback success, particularly those involving dorsal attention, visual, default mode, and subcortical networks, underscore the active role of these networks in processing external stimuli, like visual feedback, and their impact on the ventral temporal area, a key focus of this neurofeedback experiment. The typically inactive default mode network during tasks requiring external attention, for instance, might heighten vigilance by monitoring personally relevant or internally generated stimuli. Conversely, negative correlations within limbic, ventral attention, somatomotor, and control networks suggest their interference with or irrelevance to the targeted processes required for NF success. Overall, our results provide initial insights into the causal mechanisms underlying DecNef and highlight the promise of causal discovery in understanding variability in NF success.

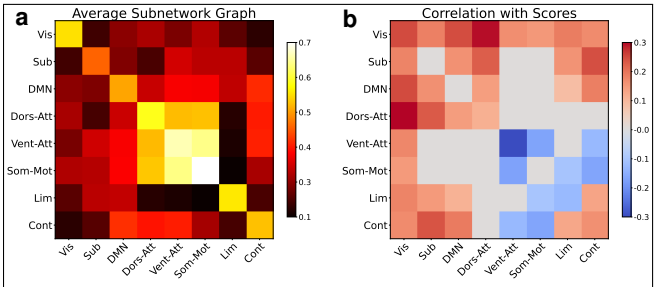

Figure 2: **(a)** Average subnetwork-level graph across subjects and sessions. **(b)** Spearman correlation, across subjects and sessions, between strength of each edge in the subnetwork-level graph and the average of neurofeedback score obtained during the two runs used to generate each graph. All non-gray blocks show a statistically significant correlation (t-test, $\alpha = 0.05$, FDR-corrected for multiple comparisons.

Acknowledgments

The research conducted in this study was partially supported by NSF Award #2239654 to EN and by the Canadian Institute for Advanced Research (fellowship awarded to MAKP).

References

Amano, K., Shibata, K., Kawato, M., Sasaki, Y., & Watanabe, T. (2016). Learning to associate orientation with color in early visual areas by associative decoded fmri neurofeedback. *Current Biology*, *26*(14), 1861–1866.

Arab, F., Ghassami, A., Jamalabadi, H., Peters, M. A., & Nozari, E. (2023). Whole-brain causal discovery using fmri. *bioRxiv*, 2023–08.

Arns, M., Clark, C. R., Trullinger, M., DeBeus, R., Mack, M., & Aniftos, M. (2020). Neurofeedback and attention-deficit/hyperactivity-disorder (adhd) in children: rating the evidence and proposed guidelines. *Applied psychophysiology and biofeedback*, *45*(2), 39–48.

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., ... others (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage*, *154*, 174–187.

Cortese, A., Amano, K., Koizumi, A., Kawato, M., & Lau, H. (2016). Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nature communications*, *7*(1), 13669.

Cortese, A., Tanaka, S. C., Amano, K., Koizumi, A., Lau, H., Sasaki, Y., ... Kawato, M. (2021). The decnef collection, fmri data from closed-loop decoded neurofeedback experiments. *Scientific data*, *8*(1), 65.

Enriquez-Geppert, S., Smit, D., Pimenta, M. G., & Arns, M. (2019). Neurofeedback as a treatment intervention in adhd: current evidence and practice. *Current psychiatry reports*, *21*, 1–7.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., ... others (2019). fmriprep: a robust preprocessing pipeline for functional mri. *Nature methods*, *16*(1), 111–116.

Fielenbach, S., Donkers, F. C., Spreen, M., & Bogaerts, S. (2019). The ability of forensic psychiatric patients with substance use disorder to learn neurofeedback. *International Journal of Forensic Mental Health*, *18*(3), 187–199.

Hammer, E. M., Halder, S., Blankertz, B., Sannelli, C., Dickhaus, T., Kleih, S., ... Kübler, A. (2012). Psychological predictors of smr-bci performance. *Biological psychology*, *89*(1), 80–86.

Koizumi, A., Amano, K., Cortese, A., Shibata, K., Yoshida, W., Seymour, B., ... Lau, H. (2016). Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. *Nature human behaviour*, *1*(1), 0006.

Pindi, P., Houenou, J., Piguet, C., & Favre, P. (2022). Real-time fmri neurofeedback as a new treatment for psychiatric disorders: a meta-analysis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *119*, 110605.

Russo, G. M., Smith, S., & Sperandio, K. R. (2023). A meta-analysis of neurofeedback for treating substance use disorders. *Journal of Counseling & Development*, *101*(2), 143–156.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., ... Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, *28*(9), 3095–3114. doi: 10.1093/cercor/bhx179

Shibata, K., Lisi, G., Cortese, A., Watanabe, T., Sasaki, Y., & Kawato, M. (2019). Toward a comprehensive understanding of the neural mechanisms of decoded neurofeedback. *NeuroImage*, *188*, 539–556.

Shibata, K., Watanabe, T., Kawato, M., & Sasaki, Y. (2016). Differential activation patterns in the same brain region led to opposite emotional states. *PLoS biology*, *14*(9), e1002546.

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., ... others (2017). Closed-loop brain training: the science of neurofeedback. *Nature Reviews Neuroscience*, *18*(2), 86–100.

Taschereau-Dumouchel, V., Cortese, A., Chiba, T., Knotts, J., Kawato, M., & Lau, H. (2018). Towards an unconscious neural reinforcement intervention for common fears. *Proceedings of the National Academy of Sciences*, *115*(13), 3470–3475.

Tian, Y., Margulies, D. S., Breakspear, M., & Zalesky, A. (2020). Hierarchical organization of the human subcortex unveiled with functional connectivity gradients. *bioRxiv*. doi: 10.1101/2020.01.22.916357

Tolin, D. F., Davies, C. D., Moskow, D. M., & Hofmann, S. G. (2020). Biofeedback and neurofeedback for anxiety disorders: a quantitative and qualitative systematic review. *Anxiety disorders: Rethinking and understanding recent discoveries*, 265–289.

Trudeau, D. L. (2005). Applicability of brain wave biofeedback to substance use disorder in adolescents. *Child and Adolescent Psychiatric Clinics*, *14*(1), 125–136.

Vargas, G., Araya, D., Sepulveda, P., Rodriguez-Fernandez, M., Friston, K. J., Sitaram, R., & El-Deredy, W. (2023). Self-regulation learning as active inference: dynamic causal modeling of an fmri neurofeedback task. *Frontiers in Neuroscience*, *17*, 1212549.

Watanabe, T., Sasaki, Y., Shibata, K., & Kawato, M. (2017). Advances in fmri real-time neurofeedback. *Trends in cognitive sciences*, *21*(12), 997–1010.

Yeo, T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, *106*(3), 1125–1165.