

Unconscious modulation of emotional valence using decoded neurofeedback (DecNef)

Nirmitee Nitin Mulay^{1,2}, Vincent Taschereau-Dumouchel^{3,4}, Cesar Caballero- Gaudes^{1,5}, David Soto^{1,5}

¹ Basque Center on Cognition, Brain and Language (BCBL), San Sebastián, Spain

² University of Basque Country, San Sebastián, Spain

³ University of Montreal, Montreal, Canada

⁴ Research Center of the Institut Universitaire en Santé Mentale de Montréal, Montreal, Canada

⁵ Ikerbasque - Basque Foundation for Science, San Sebastián, Spain

Introduction

Subjective emotional experiences refer to the phenomenological aspect of our emotions, namely, how it feels to be angry, joyful, anxious, or calm. While emotions have long been studied from physiological and behavioral perspectives, their subjective qualities have remained elusive, in part because of the inherent challenges of measuring internal states. The study of consciousness and emotion is now converging on the need to better characterise and regulate these inner experiences.

Effective emotion regulation is critical for mental health and adaptive functioning. However, current protocols for studying and training emotion regulation are limited. Traditional neurofeedback approaches, particularly using fMRI, have shown promise in enabling individuals to modulate activity in key emotion-related brain regions such as the amygdala and insula. Yet, these methods often use univariate feedback without any specific emotional content being modulated. As such, they lack precision in shaping the subjective and neural representations of distinct emotional states.

Decoded neurofeedback (DecNef) (see Watanabe et al., 2017, Taschereau-Dumouchel, V. et al., 2018, Shibata, K. et al., 2011, Cortese, A., 2016. , Watanabe, T. et al., 2016.) represents a significant advancement. By targeting distributed, multivariate patterns of brain activity associated with specific emotional states, DecNef enables the modulation of neural representations linked to those states without the need for explicit emotion induction or reporting. This opens new possibilities for understanding the neural basis of subjective emotion and developing fine-grained tools for training emotional regulation.

According to several theoretical positions, emotions are intimately tied to conscious awareness: what makes an episode genuinely emotional is not just bodily or behavioural change, but the way it is consciously experienced (e.g., Siewert, 1998). The aim of this study is to investigate whether DecNef can be used to unconsciously modulate the neural representations of specific emotional valence and in doing so address concomitant changes in subjective experience. The work can help overcome the limitations of conventional neurofeedback and move toward more personalized and effective interventions for emotional awareness and regulation. One promising application of this work lies in alexithymia- a cognitive-affective deficit marked by difficulty identifying and describing feelings, distinguishing emotions from bodily sensations, reduced imaginative activity, and a tendency toward externally oriented thinking. Individuals with alexithymia often experience anhedonia, heightened vulnerability to negative affect, and impaired emotional regulation and stress management. By targeting the neural underpinnings of specific emotional states, DecNef may offer a novel intervention pathway for individuals with alexithymia, allowing them to develop richer emotional awareness

Methods

We are recruiting 24 healthy young adults (18-35 years). Sample size estimation is consistent with prior DecNef studies and powered to detect medium effect sizes. We acquired neuroimaging data on a Siemens 3T Magnetom Prismafit scanner using T2*-weighted EPI (TR=2000ms, TE=28ms, 3×3×3.5mm³ voxels, 34 slices) and high-resolution T1-weighted MP-RAGE (1mm isotropic). Physiological signals of Photoplethysmography (PPG), Electrodermal activity (EDA) and pneumatic Respiration (RSP) recorded during all fMRI scans are acquired using Biopac MP160 system.

In the decoder construction session, participants viewed 12-second blocks of emotional images (from the International Affective Picture System (IAPS), 4 images per block of particular valence). We trained whole-brain logistic regression

classifiers to discriminate positive versus negative valence, using leave-one-run-out cross-validation. We used searchlight analyses (9mm radius) to identify regions contributing to decoding performance. We also collected psychological assessment surveys. Scores of anxiety and alexithymia, along decoder accuracies(>75) were the inclusion criteria.

In the DecNef Training, participants completed three DecNef sessions with monetary reinforcement. We gave the visual feedback (disc size) that reflects decoder-derived likelihood that brain patterns match the target valence state. The study uses a double-blind, between-subjects design where neither participants nor experimenters running sessions know whether positive or negative valence is being reinforced. Participants are instructed to use any mental strategy to increase disc size but receive no explicit emotion-related instructions. We implemented the opensource decnef package PyDecNef for our realtime pipeline.

To check the effects of decnef training, we have some Pre-post behavioural measures of subjective valence/arousal ratings (9-point Likert scales), affective priming task, and neural measures include univariate pattern analysis of emotion appraisal task, and physiological monitoring (electrodermal activity, photoplethysmography, respiration) during all fMRI sessions.

Results

Individual whole-brain logistic regression classifiers successfully discriminated positive versus negative emotional valence across participants tested to date (n=25). Cross-validated performance using leave-one-run-out validation yielded strong classification metrics: mean AUC = 0.85 (range: 0.75-0.95) and mean accuracy = 0.90 (range: 0.75- 0.95), demonstrating robust individual-level decoding of emotional valence from distributed brain patterns.

We identified prefrontal cortex (PFC) regions as primary contributors to decoding performance in the majority of participants using searchlight analyses. The distributed nature of informative regions supports targeting prefrontal representations, which are implicated in emotion regulation and higher-order emotional processing as DecNef targets.

These results demonstrate the feasibility of whole-brain valence decoding at the individual level and motivate use of distributed valence representations as DecNef targets. Given the double-blind nature of the experiment, and as the data collection is still ongoing, DecNef training results cannot yet be disclosed in time for the abstract submission. Planned analyses will assess whether DecNef induces shifts in prefrontal-limbic neural patterns, subjective emotional ratings, affective priming performance, physiological responses, and prefrontal engagement during emotional tasks. Post-training awareness assessments will evaluate whether learning occurred implicitly.

Conclusion

Our preliminary results demonstrate that emotional valence can be decoded with high accuracy using whole-brain fMRI patterns, with dominant contributions from prefrontal regions associated with emotional regulation. By leveraging these representations, decoded neurofeedback offers a method for implicitly modulating emotional brain states without explicit emotion induction or conscious strategy use. This approach may overcome key limitations of traditional neurofeedback by enabling fine-grained, content-specific modulation of emotional experiences. If successful, this work will advance our understanding of the neural basis of subjective emotions and open new translational avenues for disorders characterized by impaired emotional awareness and regulation, such as alexithymia. More broadly, the findings contribute to ongoing debates in affective neuroscience and consciousness research regarding how emotions are represented, accessed, and regulated in the human brain.

References

- Watanabe, T., Sasaki, Y., Shibata, K., & Kawato, M. (2017). Advances in fMRI Real-Time Neurofeedback. *Trends in Cognitive Sciences*, 21(12), 997–1010. <https://doi.org/10.1016/j.tics.2017.09.010>
- Taschereau-Dumouchel, V., Cortese, A., Chiba, T., Knotts, J. D., 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. <https://doi.org/10.1073/pnas.1721572115>
- Shibata, K., Watanabe, T., Sasaki, Y., & Kawato, M. (2011). Perceptual Learning Incepted by Decoded fMRI Neurofeedback Without Stimulus Presentation. *Science*, 334(6061), 1413–1415. <https://doi.org/10.1126/science.1212003>
- 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.
- Siewert, C. (1998). The significance of consciousness.
- Cortese, A. (2016). Online fMRI Decoded Neurofeedback to Explore Metacognition and Awareness.
- LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, 73(4), 653–676.