

Toward real-time neurofeedback for the suppression of unwanted working memories

Caleb N. Jerinic-Brodeur ¹, Marie T. Banich ², Jarrod A. Lewis-Peacock ¹

¹ *Department of Psychology, University of Texas at Austin, Austin, TX, USA*

² *Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA*

Introduction

A cornerstone of human cognition is the ability to shape our conscious awareness—preserving goal-relevant information while pushing aside what is unwanted or intrusive. This ability is supported by working memory and allows individuals to curate the contents of their current thoughts. Previous research has demonstrated that individuals can exert inhibitory control over representations in working memory to reduce their accessibility in both the short and long term—a process we term maintenance suppression. Failures of maintenance suppression can give rise to rumination, characterized by repetitive negative or distressing thoughts without resolution. Dysfunctional memory control is a hallmark symptom of many psychiatric disorders (e.g., PTSD, OCD, depression), but challenges in controlling intrusive thoughts are also common in healthy individuals. Our recent work has shown that psychologically healthy individuals with higher trait-level difficulties in thought control show less distinct and less consistent fMRI activity patterns for working memory control processes in control-related functional networks. These findings suggest that, even in healthy individuals, variability in the neural substrates supporting maintenance suppression can have meaningful consequences. The goal of the current study was to test whether individuals could learn to more effectively and consistently engage the neural systems supporting maintenance suppression through decoded neurofeedback training. We hypothesized that neurofeedback scores would increase and become less variable across training, reflecting stronger and more consistent engagement of maintenance suppression, and that these improvements would transfer to behavior by reducing the accessibility of suppressed memories.

Methods

Participants (N = 16 healthy adults with no history of psychiatric or neurological disorders, aged 18–30 years) completed the experiment across three sessions, including a baseline session prior to training (Session 1), followed by two neurofeedback training sessions (Sessions 2–3). On each trial, participants encoded an emotional (negative or positive) image into working memory for 2 s before receiving a cue to either maintain or suppress the representation for 10 s while undergoing fMRI scanning. In Session 1, participants made a subjective performance rating at the end of each trial. In Sessions 2–3, participants received a visual neurofeedback score at the end of each trial reflecting their level of control engagement. A machine learning classifier was pre-trained on independent fMRI data to distinguish between working memory maintenance and suppression states, then applied to ongoing fMRI activity during the 10 s control period to generate this score. Classifier-derived engagement across sessions served as the primary neural outcome measure of learning. All sessions concluded with a behavioral transfer task to assess generalization of training effects on memory accessibility.

Results

Classifier-derived engagement scores remained stable across neurofeedback training sessions and did not show systematic increases relative to baseline. Similarly, neurofeedback scores did not exhibit reliable changes across runs within a session, providing little evidence for short-term modulation of maintenance suppression engagement. Variability in neurofeedback scores also remained stable across training, indicating that participants did not develop more consistent engagement of the target control strategy over time. Behavioral performance on the transfer task also showed no evidence of neurofeedback-related changes in memory accessibility.

Conclusions

The present study tested whether MVPA-based neurofeedback could serve as a targeted training intervention to strengthen the engagement of maintenance suppression and improve control over unwanted working memory representations. Although classifier-derived engagement and behavioral measures remained stable across training, these findings provide important constraints on the ability of short-term neurofeedback to modulate internally directed memory control processes. One possibility is that effective neurofeedback for memory control may require integrating signals that capture both the engagement of suppression-related control processes and their downstream effects on unwanted memory representations. Additionally, the relatively brief training duration and modest sample size may have limited the emergence and detection of learning-related effects. Ongoing offline analyses will further examine whether learning-related effects emerge through coordinated changes between suppression-related control activity and memory representations, as well as reorganization of large-scale functional networks. More broadly, these findings highlight challenges in using MVPA-based neurofeedback to modulate complex, internally directed cognitive processes and translating decoded neural states into behavioral change.