Putamen Representation of Habit as Marker of Compulsivity

Amogh Johri (ajohri@caltech.edu), Vincent Man (vman@caltech.edu)
Rani Gera (sgera@caltech.edu), John P. O'Doherty (jdoherty@caltech.edu)
Division of the Humanities and Social Sciences, California Institute of Technology
1200 E. California Blvd., Pasadena, California, 91106, USA

Abstract

Compulsivity, marked by rigid and repetitive behaviors, is hypothesized to arise from overreliance on habitual control systems. We combined a habit-learning task with fMRI and multivariate pattern analysis to test whether the neural signatures of habitual control overlap with compulsive traits, as measured by the OCI-R. A cluster in the left posterior putamen reliably decoded Habitual versus Goal-Directed participants and classified high versus low OCI-R groups independently of behavioral habit expression. Cross-decoding revealed shared neural representations linking habitual control and compulsivity. Our results highlight overlapping neural substrates of habitual control and compulsivity, providing insight into the dimensional mechanisms of compulsive behavior.

Keywords: habit; goal-directed decision-making; compulsivity; fMRI; MVPA

Introduction

Compulsive traits are theorized to arise from an overreliance on habitual control systems, where actions become increasingly decoupled from goal-directed evaluation (Gillan, Robbins, Sahakian, van den Heuvel, & van Wingen, 2016). The posterior putamen, a key node in the sensorimotor corticostriatal circuit, has been implicated in both habit formation and compulsive behavior (Tricomi, Balleine, & O'Doherty, 2009), but direct evidence of overlapping neural representations remains limited. Although prior work shows behavioral links between compulsivity and habit bias, it is unclear whether shared neural substrates underpin these processes and whether such neural patterns can predict compulsivity at clinically relevant levels.

In this study, we combined a validated habit-learning task with fMRI and multivariate pattern analysis (MVPA) to examine whether the neural signatures of habitual control overlap with compulsive traits, as measured by the Obsessive-Compulsive Inventory-Revised (OCI-R) (Foa et al., 2002). Furthermore, we use the neural code from differentiating habitual vs. goal-directed control to decode high vs. low compulsivity status.

Experiment

The task required participants to use a trackball to swipe left or right and press buttons to collect gold and silver coins. The experiment comprised training, consumption, and choice-test phases (Figure 1.).

Training Phase: Participants received explicit instructions regarding the movements associated with each coin colour and cue. During training, a cue and the target coin appeared on the screen. Correct swiping movements (left or right) triggered visual confirmation via an orange triangle and highlighted rewarded coins. The training phase included 20 blocks, each lasting 37 - 60 seconds, during which participants aimed to collect as many coins as possible. We based the reward delivery on a random interval schedule (RR15) to mimic real-life reward unpredictability. Both coin colours held equivalent values. Each swipe carried a nominal cost of -1 point, while successful coin collection yielded +40 points, incentivizing frequent swiping—although we didn't inform the participants of this optimal strategy. The positions of cues and coin colours alternated across blocks. The mean swipe rate across all participants was 2.42 ± 0.89 swipes per second. We excluded the participants with a swiping rate of < 0.5 swipes per second.

Consumption Phase: After block 10, participants were told their piggy banks were nearly full and completed a dummy consumption test, selecting ten coins from a pool of ten gold and ten silver coins. Participants then proceeded with blocks $11 - 20$. Upon completing training, we informed participants that one piggy bank was full, effectively devaluing one coin type. In the subsequent consumption test, participants selected coins, and the number of valuable coins collected served as their consumption score. We excluded the participants with a consumption score of < 8 (for we cannot ascertain if they understood the devaluation).

Choice Test: Participants then performed the choice test, during which both cues appeared simultaneously, but the coins (hidden) behind a curtain. Participants swiped to collect coins without feedback, aiming for the still-valuable option. We calculated the devaluation ratio as the proportion of swipes directed toward the valuable coin out of the total swipes. We excluded participants with a devaluation ratio $< 1/3$ (overwhelming responses ($> 2/3$) in the wrong direction).

Methods

We acquired fMRI data during the Habit Task, two functional runs of 646 seconds and 735 seconds. 11 regions of interest were identified (bilateral anterior and posterior caudate, bilateral anterior and posterior putamen, bilateral dlPFC and vmPFC). We performed pre-processing using fMRIPrep $23.1.3$ and post-processing and GLM modelling with Nilearn $0.10.1$. For GLM modelling, we use a mean block regressor across all the training blocks after concatenating both runs and obtaining subject-wise beta maps. All participants completed the Obsessive-Compulsive Inventory-Revised (OCI-R) questionnaire.

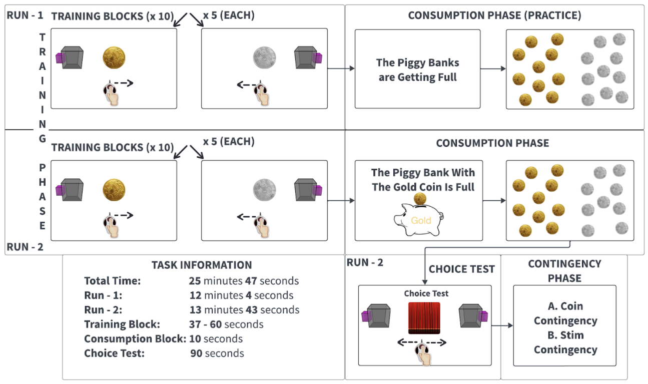

Figure 1: Habit Task. Devalued Coin: Gold.

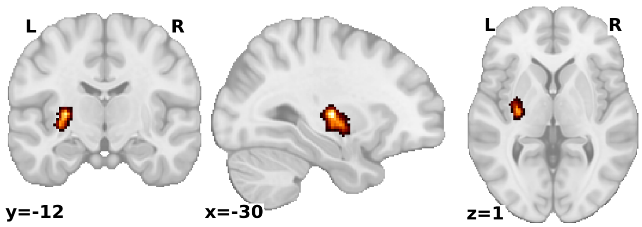

Figure 2: MVPA (Goal Directed vs. Habit) Cluster: Left Posterior Putamen.

MVPA: Posterior Putamen Decodes Goal Directed vs. Habit Participants

The final number of participants was 198 (post-exclusion). We labelled the subjects with a devaluation ratio $< 2/3$ as Habitual (N $= 93$) and the rest as Goal-Directed. For decoding (subject-wise) Habit vs. Goal-Directed, we applied Linear Discriminant Analysis with 4-fold stratified cross-validation. In each ROI, we used a $4mm$ radius searchlight. We performed permutation tests (N $= 5000$) to obtain statistics. Given the possible confounding effect of the swiping rate (i.e., Goal-Directed participants had a higher swiping rate during training as compared to Habitual participants), we applied cross-validated confound regression (CVCR) before decoding (Snoek, Miletić, & Scholte, 2019) to regress out the subject-wise mean swiping rate. We identified a cluster in the left posterior putamen (number of voxels $= 44$, FWER $p < 0.05$ after applying Bonferroni correction for ROIs (i.e., $x11$)) to significantly decode Habitual vs. Goal-Directed subjects (mean accuracy $= 0.65 \pm 0.01$) (Figure 2.). The complete analysis will be reported separately in a forthcoming manuscript.

Results

To assess the specificity of this neural signature, we further regressed out OCI-R total scores as a potential confound and confirmed that classification performance remained robust (mean accuracy $= 0.64 \pm 0.02, p < 0.001$). **It suggests that the posterior putamen encodes habitual tendencies independent of compulsive traits.**

Next, we examined whether the same posterior putamen cluster carried information about compulsivity, operationalized via high (≥ 21) versus low (< 21) OCI-R total scores, with 21 being the recommended threshold for clinically significant obsessive-compulsive symptoms (Wootton et al., 2015). Addressing class imbalance with borderline-SMOTE resampling, decoding accuracy for OCI-R group classification was significantly above chance (mean accuracy $= 0.62 \pm 0.02, p < 0.01$). Importantly, this decoding remained significant after regressing out Habitual vs. Goal-Directed labels, **indicating that compulsivity-related signals in the posterior putamen are not reducible to behavioural habit expression alone.**

We formed cross-decoding to analyze shared neural representations between habitual control and compulsivity. A classifier trained to distinguish Habitual from Goal-Directed individuals generalized successfully to classify participants with high versus low OCI-R total scores when aligning Habitual labels with high OCI-R status (mean cross-decoding accuracy $= 0.61 \pm 0.02, p < 0.01$). **This result implies a shared representational space in the posterior putamen linking normative variations in habitual control with compulsivity traits.**

Finally, to probe the translational potential of this shared neural signature, we trained the classifier exclusively on sub-threshold participants (OCI-R < 21) and tested its ability to classify high OCI-R individuals (≥ 21). The classifier generalized successfully (mean accuracy $= 0.72 \pm 0.04, p < 0.01$), **demonstrating the potential use of neural patterns associated with habitual control in a sub-threshold population to predict clinically-relevant compulsivity.**

As a robustness check, we evaluated additional performance metrics beyond accuracy, including precision, recall, F1-score, and specificity, to ensure that the classification of High vs. Low OCI-R groups isn't a trivial byproduct of class imbalance or metric inflation. Importantly, all these metrics yielded statistically significant results, reinforcing the reliability and robustness of the observed decoding effects.

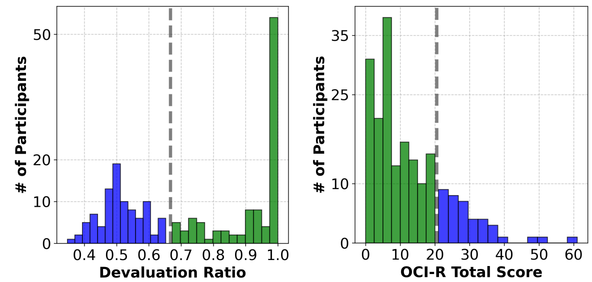

Figure 3: Histograms (Devaluation Ratio and OCI-R Total Scores). Green: Goal Directed and Low Compulsivity. Blue: Habit and High Compulsivity.

Acknowledgments

This work was supported by the National Institute of Mental Health under grant R01MH121089-01. We thank the Caltech

Brain Imaging Center for MRI data acquisition and all participants for their involvement in the study.

References

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The obsessive-compulsive inventory: Development and validation of a short version. *Psychological Assessment*, *14*(4), 485–496. doi: 10.1037/1040-3590.14.4.485

Gillan, C. M., Robbins, T. W., Sahakian, B. J., van den Heuvel, O. A., & van Wingen, G. (2016). The role of habit in compulsivity. *European Neuropsychopharmacology*, *26*(5), 828-840. doi: https://doi.org/10.1016/j.euroneuro.2015.12.033

Snoek, L., Miletić, S., & Scholte, H. S. (2019). How to control for confounds in decoding analyses of neuroimaging data. *NeuroImage*, *184*, 741-760. doi: https://doi.org/10.1016/j.neuroimage.2018.09.074

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience*, *29*(11), 2225-2232. doi: https://doi.org/10.1111/j.1460-9568.2009.06796.x

Wootton, B. M., Diefenbach, G. J., Bragdon, L. B., Steketee, G., Frost, R. O., & Tolin, D. F. (2015). A contemporary psychometric evaluation of the obsessive compulsive inventory—revised (oci-r). *Psychological Assessment*, *27*(3), 874–882. doi: 10.1037/pas0000075