Distributed activity in the human posterior putamen distinguishes goal-directed from habitual control in humans

Amogh Johri[1], Lisa Kluen[1], Rani Gera[1], Vincent Man[1], Omar Perez[1,8], Julia Pia Simon[1], Weilun Ding[1], Aniek Fransen[1], Scarlet Cho[1], Sarah Oh[1], Jeff Cockburn[1,2], Jamie D. Feusner[3,4,5], Reza Tadayonnejad[1,6,7], John P. O'Doherty[1]

[1]Division of Humanities & Social Sciences, California Institute of Technology.
[2]Psychological & Brain Sciences, University of Iowa, Iowa City, IA, USA .
[3]Centre for Addiction and Mental Health, 250 College St., Toronto, ON, M5T 1R8, Canada.
[4]Department of Psychiatry, University of Toronto, Toronto, Canada.
[5]Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden.
[6]Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA.
[7]TMS Clinical and Research Service, Neuromodulation Division, Semel Institute for Neuroscience and Human Behavior at UCLA, Los Angeles, CA, USA.
[8]Department of Industrial Engineering, University of Chile, Santiago, Chile.

There is considerable variability across individuals in the behavioral expression of goal-directed and habitual control, and yet little is known about the underlying functional neuroanatomical basis of such differences. Here, we aimed to determine whether distinct patterns of brain activity during initial learning of instrumental actions can indicate whether an individual subsequently utilizes a goal-directed or habitual behavioral strategy. We applied multivariate pattern analyses to a large fMRI dataset composed of 108 participants performing an instrumental action-learning task. Region-of-interest (ROI) and whole-brain searchlight analyses with cross-validation indicate that distributed activation patterns in the left posterior putamen during training distinguish goal-directed from habitual participants, ascertained via subsequent administration of an outcome devaluation test. This finding was further replicated and generalized to two fully held-out samples: an additional cohort of healthy participants (n = 38) and a separate group of psychiatric patients with variable diagnoses (n = 55). These findings reveal that it is possible to predict an individual's overall behavioral strategy from circumscribed activity in the human brain, while specifically implicating the posterior putamen in the implementation of distinct behavioral control strategies.

Introduction

Extensive research suggests that reward-driven action selection in both humans and non-human animals relies on at least two distinct modes of behavioral control: goal-directed and habitual strategies [1–3]. Goal-directed control involves the deliberate selection of actions based on the current value of expected outcomes, allowing for flexible adaptation when goals or circumstances change. In contrast, habitual control is reflexive and insensitive to the value of the outcome produced by the action.

Considerable research has characterized the neural mechanisms of goal-directed and habitual control in both humans and other animals. Brain regions, including the ventromedial prefrontal cortex, dorsolateral prefrontal cortex, anterior dorsal striatum, and putative homologues in the rodent brain, such as the prelimbic cortex and the dorsomedial striatum, have been implicated in goal-directed control [4–6]. Conversely, the posterior putamen and its rodent equivalent, the dorsolateral striatum, have been implicated in habitual control [7–11].

In humans, activity in the posterior putamen has been found to increase as a function of training [9, 12]. The posterior putamen has also been found to contain multivariate representations of stimuli and actions but not outcomes, in contrast to dorsolateral and ventromedial prefrontal cortex, which have been found to contain representations of actions, outcomes, and goals [13].

However, there is also considerable individual variability in the behavioral expression of goal-directed and habitual control [14]. Some individuals can quickly form particular habits, while others take much longer to do so or may not even do so at all in a given timeframe [15]. A natural question that arises in this context is whether there are differences in underlying neuronal activity patterns during the performance of a task that relate systematically to whether or not a given individual adopts a habitual or goal-directed strategy. In other words, is it possible to predict from a given individual's brain activity whether or not they are behaving habitually or in a goal-directed manner?

To address this question, human participants performed a novel free-operant action selection task [Figure-1] designed to assess whether behavior is habitual or goal-directed using an outcome devaluation procedure, while being scanned with fMRI. We used a multivariate pattern analysis approach to decode each individual's behavioral strategy from their brain activity. We were able to recruit a relatively large sample of participants for an fMRI study, which gave us a sufficient sample size to perform whole-brain and region-specific decoding of behavioral strategy across participants. We could therefore ask the question, where in the brain do distributed activity patterns during initial training predict the extent to which a given individual is going to express goal-directed or habitual behavior on the task when subsequently tested during a devaluation phase?

The fact that we had a large sample to draw from also enabled us to implement a fully held-out sampling approach, in which we trained our classifier on an initial cohort using cross-validation while retaining a separate cohort of healthy participants for subsequent out-of-sample testing of the classifier's performance. While the use of cross-validation on the initial cohort should protect against overfitting, the

use of a completely held-out sample helps strengthen claims about the generalizability of the classification approach. Furthermore, we utilized yet another population of participants who were also scanned on the same task: a cohort of patients with variable psychiatric disorder diagnoses. We could thus also test for the robustness and generalizability of our classification findings across different participants and populations.

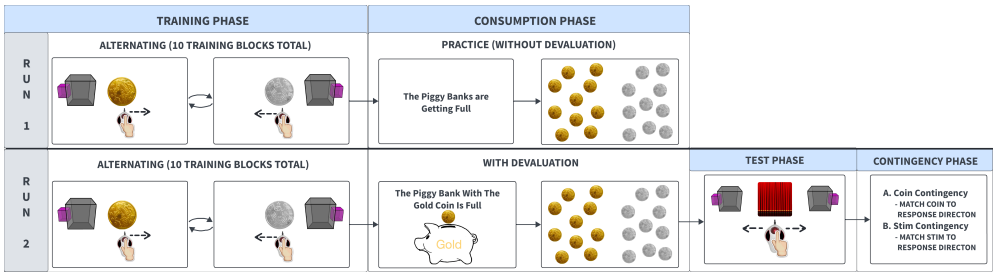

Fig. 1: **Illustration of task design**. During training, participants were presented with two different stimuli across different training blocks (one stimulus per block, the blocks were 37-60 s in duration). For one of these stimuli, if a participant performed a leftward response movement with a trackball (using their dominant hand), they would, according to a random ratio schedule (with rewards delivered after 8 to 15 swipes), receive (in one particular participant) a gold coin. The other stimulus, a rightward response movement, would (based on a similar reinforcement schedule) lead to (in that same participant) a silver coin. The specific mapping between stimuli, movements, and coins was fully permuted across participants, but remained consistent within each participant. Through this training procedure, participants could learn to associate stimuli, actions, and outcomes for both the silver and gold coins separately. Participants were instructed that the coins being collected were placed in virtual piggy banks and that these piggy banks could later be exchanged for monetary outcomes. After 10 blocks of training on each action-coin (total blocks = 20) association, a devaluation procedure was implemented whereby one of the coin types (silver or gold) was devalued. Participants were instructed that the piggy bank holding that coin was full, implying that any further collection of that coin would yield no additional reward. A consumption test was used to test the efficacy of the devaluation procedure, as discussed below. Participants then underwent a testing procedure in which both stimuli were presented at the same time, and they could choose which action to perform in order to obtain either a gold or silver coin (this was conducted using a pseudo-extinction procedure in which any coin that was won on each trial was hidden). Goal-directed control would be associated with predominant selection of the action associated with the non-devalued coin, whereas habitual control was associated with a tendency to continue choosing (at least some of the time) the action associated with the now devalued coin. Assignment of coin color (gold vs. silver) to the devalued outcome, as well as cue–response mappings, was counterbalanced across participants to control for stimulus or response biases. Further details are in Methods: Task Description.

Results

We analyzed data from 144 healthy participants who completed the task in the MRI scanner (post exclusions; Methods: Exclusions). We divided the sample into an Initial Cohort (N = 108) used for training and cross-validated testing of the classifier, and a fully independent Held-Out Cohort (N = 36) reserved for validation. To further assess generalizability, we applied the trained classifier to an additional Held-Out Patient Cohort (N = 55) with mixed clinical diagnoses.

Behavioral Results

Devaluation Ratio

We computed a devaluation ratio (DR) for each participant, i.e., the proportion of swipes directed toward the still-valued coin relative to the total number of actions during the choice test (see Methods: Choice Test Phase for details). A DR of 1.0 indicates an exclusive response associated with the still-valued outcome, consistent with perfectly goal-directed behavior. In contrast, a DR of 0.5 reflects an equal preference for response in both the devalued and still-valued outcomes, consistent with habitual responding. As illustrated in Figure 2A, participants exhibited a broad spectrum of behavior, with peaks at goal-directed (DR = 1.0) and habitual (DR = 0.5).

To facilitate subsequent analyses, we categorized participants into goal-directed and habitual subgroups based on their devaluation ratios using a data-driven threshold derived from a two-component Beta mixture model (see Group Classification: Beta Mixture Modeling in Methods). This model identified two latent subpopulations corresponding to distinct behavioral strategies and estimated a posterior decision boundary at a devaluation ratio of 0.648. Participants with devaluation ratios greater than or equal to this threshold were classified as goal-directed, reflecting a predominant preference for the still-valued outcome. Those with ratios below the threshold were classified as habitual, reflecting outcome-insensitive behavior. This threshold separates the Initial Cohort (N = 108) into two almost equal groups (N = 55 goal-directed, N = 53 habitual), providing a principled basis for binary subgroup analyses of neural and behavioral differences.

From this point onward, we refer to the devaluation ratio (DR) in terms of its binary classification (goal-directed vs. habitual) unless otherwise specified.

Learning Related Changes in Response Rates

Participants completed 10 training blocks in each of two runs. For each run, we fit a linear mixed-effects regression model with block number, group, and their interaction as fixed effects, and subject as a random intercept (see Methods: Response Rates), using data from the Initial Cohort.

In Run 1, participants exhibited significant increases in response rate across blocks, consistent with within-session learning (Figure 2C). The model revealed a robust main effect of block number, indicating that response rates increased over time. Critically, there was a significant group × block interaction: goal-directed participants

showed a steeper slope (0.112 swipes/sec/block, $z = 15.75$, $p < 0.001$) than habitual participants (0.077 swipes/sec/block, $z = 6.16$, $p < 0.001$). The slope difference was statistically significant ($\beta = -0.036$, $z = -3.52$, $p < 0.001$), confirming that goal-directed participants increased their responding more strongly across blocks.

In Run 2, response rates showed no significant within-group change across blocks (goal-directed: -0.0096 swipes/sec/block, $z = -1.62$, $p = 0.105$; habitual: 0.0072 swipes/sec/block, $z = 0.70$, $p = 0.484$). However, there was a small but significant group $\times$ block interaction ($\beta = 0.017$, $z = 1.99$, $p = 0.047$), indicating that slopes differed between groups (habitual slightly increasing vs. goal-directed slightly decreasing). There was also a significant group difference in intercept ($\beta = -0.564$, $z = -2.74$, $p = 0.006$), with goal-directed participants maintaining higher overall response rates across Run 2.

Together, these results suggest that learning occurs primarily during the first session, after which performance stabilizes. Goal-directed participants exhibit a steeper learning-related increase in Run 1 and maintain an overall advantage in response vigor across both runs compared to habitual participants.

Behavioral and Trait-Related Predictors of Devaluation Sensitivity

Next, we identified the relationship between task behavior during training and devaluation ratio, focusing on the Initial Cohort. We fit a logistic regression predicting DR (binarized at a threshold 0.648; 1 = goal-directed) using standardized predictors with KNN-imputed missing values ($k = 5$), and evaluated Type-III (Wald) effects. The model converged (pseudo-$R^2 = 0.245$; LLR $p = 0.048$). Participants with greater response rates were significantly more likely to be sensitive to devaluation (Wald $\chi^2(1) = 4.82$, $p = 0.028$; $\beta = 0.65$). We also asked participants to report on their knowledge of the contingencies after the experiment. Specifically, we presented the silver coin outcome and asked participants to report on the action needed to obtain it (and the same for the gold coin; i.e., a left or right swipe). This is a test of response–outcome knowledge, which can index explicit knowledge of such associations. We found a significant relationship between self-reported knowledge of response–outcome associations and goal-directed behavior on the task (Wald $\chi^2(1) = 9.31$, $p = 0.002$; $\beta = 0.90$). Furthermore, we probed stimulus–response knowledge at the end of the experiment by presenting the initial cues present at trial onset and asking participants to select the appropriate response (left or right swipe). We did not observe a statistically significant relationship between stimulus–response knowledge and the tendency to express habits (Wald $\chi^2(1) = 0.04$, $p = 0.837$; $\beta = -0.05$).

Furthermore, when extending the analysis to the full sample (N = 199; healthy controls and patients), the predictive effect of response rate was attenuated to a non-significant trend. We also found that in the full sample, total OCI-R scores (a scale measuring obsessive-compulsive traits) exhibited a significant association with increased habitual behavior, consistent with prior reports linking compulsivity to diminished goal-directed control (Supplementary: Figure S2).

Training-Related Changes in fMRI Activity

Next we examined changes in fMRI activity associated with training by modeling a linear increase across blocks during the training phase. We fit a general linear model (GLM) to the fMRI data from the Initial Cohort ($N = 108$), incorporating a parametric modulator that linearly increased with block number to capture learning-related trends. We interpret the brain regions positively correlating with this regressor as exhibiting increased activity across training. This analysis revealed a significant cluster in the left posterior putamen (Figure 2D), where the BOLD signal increased progressively over training blocks (small-volume FDR corrected within the anatomically defined RoIs (Methods: RoI); $\alpha = 0.05$, cluster-size threshold = 10 voxels).

Subject-wise Neural Decoding of Goal-Directed Decision Making vs. Habitual Responding

A central aim of this study was to identify distributed neural activity patterns that distinguish between individuals exhibiting predominantly goal-directed versus habitual decision-making strategies. To this end, we implemented a subject-wise decoding approach using MVPA to classify participants based on their neural responses during the training phase.

We evaluated neural decoding of goal-directed versus habitual control within the Initial Cohort ($N = 108$) using a four-fold cross-validation framework. To obtain neural features, we fit a block-wise general linear model (GLM) for each participant (see Methods: First Level). Training blocks (10 per run, i.e., 20 in total) were modelled with block regressors, contrasted against the implicit baseline (all other time points, including inter-trial intervals, consumption phase, etc.). This yielded a single beta coefficient map per subject, reflecting the mean activation pattern across the brain during instrumental training.

To mitigate the confounding effects of individual differences in motor activity such as response rates, we regressed each participant's average response rate and response rate slopes (separately for each run) from the neuroimaging data. We used the resulting residuals in all subsequent analyses (see Methods, Equation 6).

We trained a Linear Discriminant Analysis (LDA) classifier to predict whether a participant belonged to the goal-directed or habitual group, as defined by their devaluation ratio. LDA serves as a well-established linear classifier that provides a conservative test of whether multivoxel patterns carry discriminative information, without overfitting or requiring strong distributional assumptions. We assessed classification performance using mean accuracy across four cross-validation folds, as the two groups were nearly equal in size (53 vs. 55), and accuracy provides an appropriate and interpretable measure of separability.

We employed non-parametric permutation testing with 5,000 random label shuffles to generate null distributions of decoding accuracy for each voxel-centered searchlight. We assess statistical significance using small-volume family-wise error correction, with additional Bonferroni adjustment across 11 predefined regions of interest (RoIs) (see Methods: MVPA Statistical Analysis).

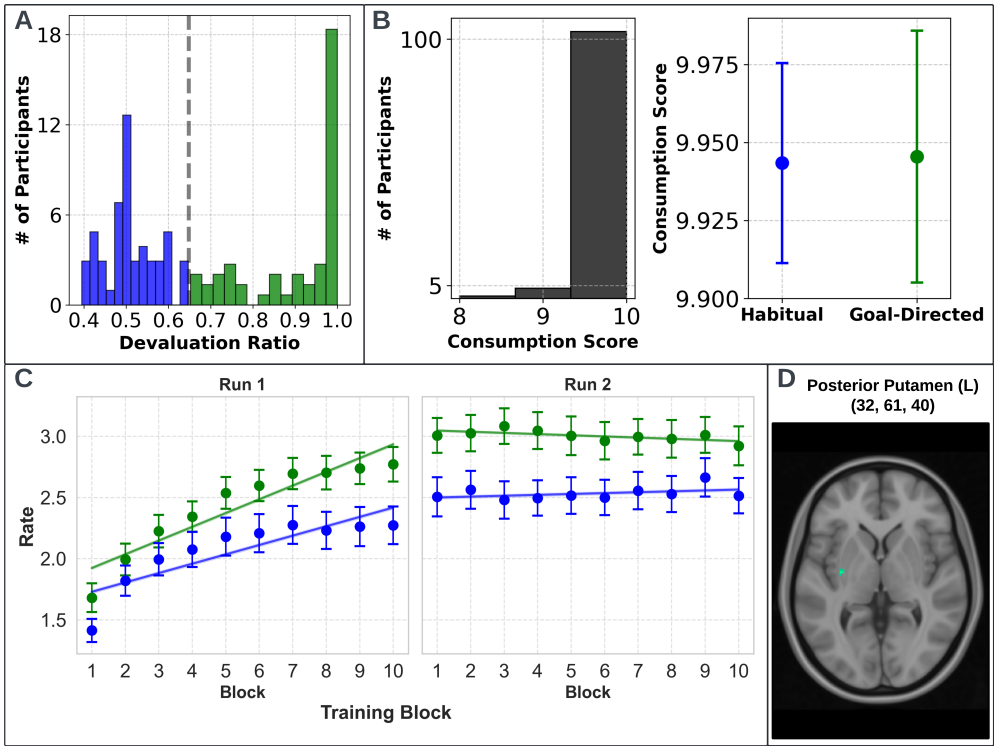

Fig. 2: (A) Distribution of continuous devaluation ratios (DR) across participants. Values near 1.0 indicate exclusive selection of the still-valued outcome (goal-directed behavior), while values near 0.5 indicate indifference (habitual behavior). We divide the participants in the *Initial Cohort* ($N = 108$) into Goal-Directed ($N = 55$) and Habitual groups based on the DR threshold of 0.648. (B) Distribution of Consumption scores. It reflects the successful selection of the still-valued coin following devaluation (and avoidance of the now devalued coin). Comparable scores between groups confirm an equivalent understanding of task structure and of the devaluation procedure. (C) Response rates (responses per second) during the Training Phase increased across Run 1 and plateaued in Run 2 for both groups. The goal-directed participants exhibited higher mean response rates. The lines represent the best linear fit. (D) fMRI results showing increased activity in the left posterior putamen across training blocks, identified using a parametric regressor modeling linear learning-related change. Results are small-volume corrected (FDR, $\alpha = 0.05$) within the left posterior putamen (14 voxels). Markers denote group means; error bars indicate ± 1 standard error of the mean (SEM). All plots reflect participants who passed exclusion criteria (see Methods: Exclusions).

Initial Cohort

Our primary decoding analysis focused on predefined regions of interest (RoIs) previously implicated in goal-directed and habitual behavior [13]. These included the posterior putamen, anterior caudate, dorsolateral prefrontal cortex (dlPFC), and ventromedial prefrontal cortex (vmPFC). Within this anatomically constrained search space, we identified a single significant cluster in the left posterior putamen where decoding accuracy exceeded chance (mean accuracy score $= 0.627 \pm 0.023$, number of voxels $= 60$; small-volume FWE-corrected at $p = 0.0198$ (after Bonferroni Correction across 11 RoIs; Figure 3A).

The left-lateralized selectivity we found in the posterior putamen may reflect the strong right-handed bias in the MVPA training sample (97% right-handed), consistent with contralateral motor representation in the basal ganglia.

To test for effects outside of our specific RoIs, we also conducted a whole-brain searchlight analysis. No additional regions outside the a priori RoIs exhibited statistically significant classification accuracy, further highlighting the specificity of posterior putamen involvement in distinguishing goal-directed from habitual individuals.

Generalization: Held-Out Cohort

Next, we tested whether the posterior putamen decoder would generalize to unseen participants using the Held-Out Cohort ($N = 36$, 25% of healthy participants). Participants were classified as goal-directed or habitual using the same devaluation-ratio threshold ($DR \geq 0.648$) as in the Initial Cohort, yielding 13 habitual and 23 goal-directed individuals (Figure4A).

We applied a region-based decoder restricted to the 60-voxel cluster in the left posterior putamen identified by the searchlight analysis. To approximate the effective resolution of the searchlight spheres and enhance generalizability, the decoder was implemented as a bagging ensemble of linear discriminant classifiers (see Methods: Generalization Analysis), each trained on random feature subsets. This design reduces variance across folds and mirrors the feature subsampling inherent to the searchlight procedure. Because the Held-Out Cohort was somewhat imbalanced (13 habitual vs. 23 goal-directed participants), we assessed decoder performance using ROC AUC, which provides a threshold-independent measure of separability and is more appropriate than accuracy under class imbalance. Re-evaluating performance in the Initial Cohort with this 60-voxel RoI yielded a mean ROC AUC of 0.656 ± 0.018 (Figure3C).

Crucially, the Held-Out Cohort was completely untouched during model training: the decoder was trained exclusively on the Initial Cohort and then applied directly to the Held-Out Cohort without any retraining or parameter tuning. Applying it to the Held-Out Cohort yielded a mean ROC AUC of 0.666 ± 0.020, significantly above chance based on 5,000 label permutations ($p = 0.044$; Figure4B). These results confirm that the posterior putamen signal reliably distinguishes habitual from goal-directed participants in an independent sample. As detailed in the Supplementary: Figure S6, we further verified that performance is robust across different decoder implementations, underscoring that the observed effect is not dependent on the specific classifier choice.

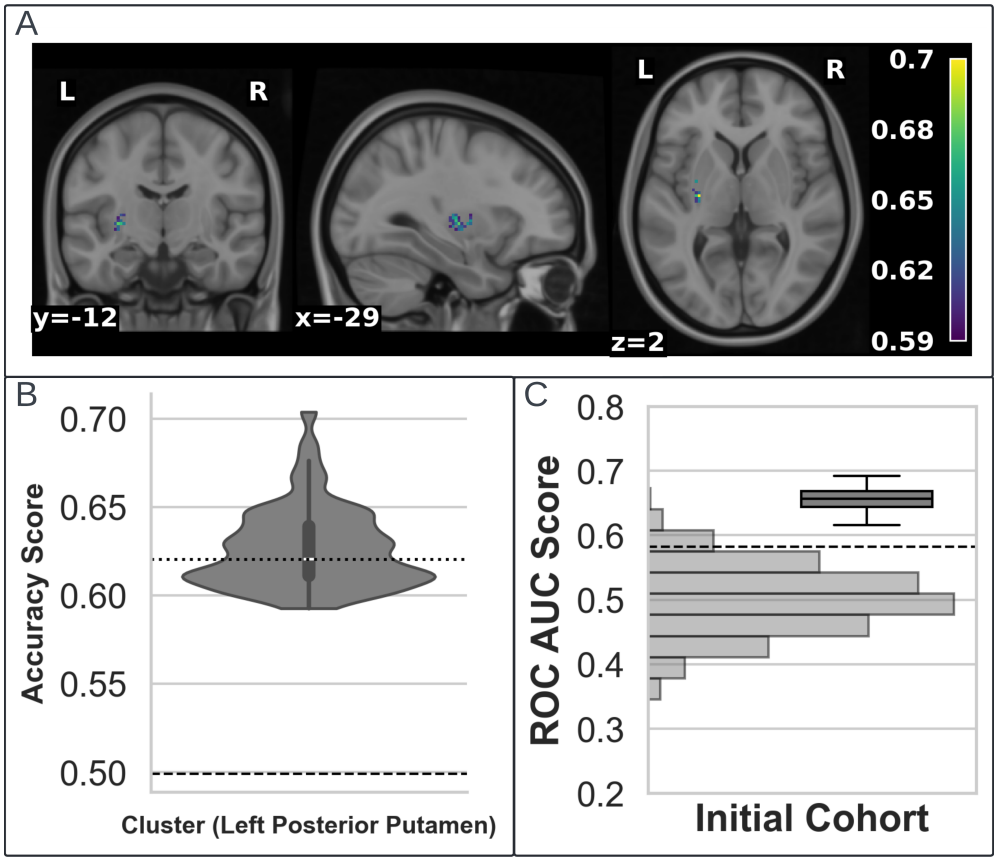

Fig. 3: **Decoding Evidence within the Significant Posterior Putamen Cluster.** (A) Spatial map of accuracy scores within the significant posterior putamen cluster (FWER corrected $p = 0.0018$, Bonferroni corrected $p < 0.0198$ for 11 RoIs). Each voxel value reflects a weighted average of nearby decoding scores within a 4 mm radius, normalized across the cluster. (B) Voxel-wise accuracy scores from searchlight decoding within the cluster as a violin plot. The dotted line shows the mean (0.627), and the dashed line shows the permutation mean (0.49). (C) Boxplot of decoding accuracy using the cluster as a single RoI. Mean performance ROC AUC = 0.656, SEM = 0.02. We present the null distribution for reference, with a dashed line at the 95th percentile. We don't report inference here due to RoI selection bias.

Cross-Population Generalizability

Our study sample also included an additional cohort of psychiatric patients with a range of different diagnoses (N = 55 after exclusions; see TableS1 and TableS2 for details on the patient sample diagnostic composition). Although the inclusion of these patients in our sample was for purposes beyond the scope of the present study, we nevertheless elected to establish whether behavioral strategy could also be decoded

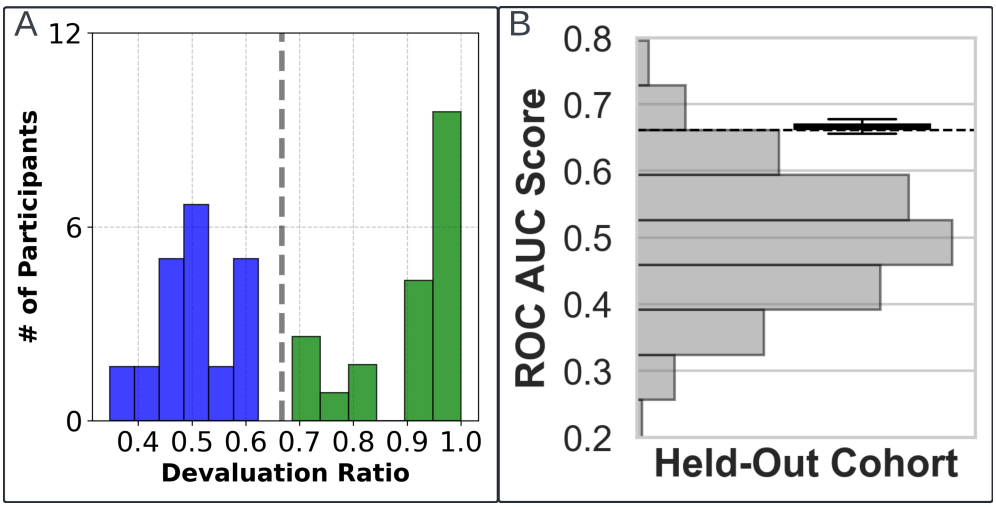

Fig. 4: **Held-Out Validation of Behavioral Classification and Neural Decoding.** (A) Distribution of devaluation ratios in the Generalization Set (Held-Out Sample, $N = 36$), using the same threshold (0.648) as the Initial Cohort to classify participants into Goal-Directed (green) and Habitual (blue). (B) Decoding accuracy for the held-out sample using the RoI identified from the Initial Cohort. Mean decoding accuracy was 0.666 ± 0.02, significantly above chance ($p = 0.0442$), as assessed against a permutation-based null distribution. Error bars reflect ± 1 SEM.

in this second held-out sample, when applying the same devaluation ratio threshold as used for the healthy participants. Applying this threshold separated the patient cohort into 26 goal-directed and 29 habitual participants. Using the posterior putamen RoI from the Initial Cohort, the decoder once again achieved significant classification accuracy in this patient sample (ROC AUC $= 0.739 \pm 0.019$, $p = 0.001$), confirming that both the behavioral classification and its neural signature generalize across diverse populations. This result thus constitutes a further replication and generalization of the decoding capacity of posterior putamen (Supplementary: Figure S5).

Nature of Representations (Multivariate vs Univariate)

We asked whether the posterior putamen signal reflects multivariate structure, univariate differences, or both. First, we regressed out the mean (univariate) signal from each subject's voxel pattern within the MVPA cluster. Decoding remained significant ROC AUC ($= 0.609$, $p < 0.022$), showing that multivariate pattern information alone carries discriminative value. We then examined mean activation by pooling all voxels within the same cluster. Habitual participants showed a higher average signal ($t(199) = 3.773$, $p < 0.0005$). This univariate effect did not survive voxel-wise small-volume correction across the broader posterior putamen RoI but did reach significance when aggregated across the MVPA cluster.

We emphasize that these analyses are post hoc and exploratory, and the statistics are not independent inferential tests. Instead, they serve to characterize the nature of

the signal, suggesting that the decoding effect reflects both distributed multivariate structure and modest, directionally consistent univariate differences within the cluster.

Additional robustness checks

We conducted supplementary analyses to confirm that decoding reflects stable neural structure rather than noise or analytic artifacts. Accuracy declined in a graded, interpretable manner when we reduced participants (Figure S9A), subsampled voxels (Figure S9B), increased smoothing (Figure S10A), or filtered voxels by decoding-weight SNR (Figure S10B). Performance remained stable across cross-validation strategies and scoring metrics (Figure S8), and voxel-level weights showed consistent signs with low variability across repetitions (Figure S14). We obtained consistent results across alternative group-definition schemes, including the mixture-model boundary (DR = 0.648; AUC = 0.676 ± 0.010), a median split (AUC = 0.679 ± 0.009), and a top-vs-bottom quartile split (reduced N; AUC = 0.648 ± 0.020; Table **??**). We also replicated the effect when we decoded from the entire left posterior putamen RoI, rather than the data-driven cluster, and when we residualized all task, psychological, and demographic confounds (AUC = 0.629 ± 0.016, $p < .05$). Finally, even when we adopted a stereotypical whole-brain searchlight analysis without our targeted constraints, we still identified clusters in the left posterior putamen with high overlap to the main result (Figure S4). Together, these results elucidate that decoding robustness does not depend on arbitrary analytic choices, cohort sizing, or residual confounds.

Discussion

We find that distributed neural activity in the left posterior putamen during initial training of instrumental associations can distinguish between whether a given individual deploys a goal-directed or habitual strategy following a subsequent outcome devaluation procedure, a canonical test of goal-directed and habitual control in the animal learning literature [1, 16].

Critically, because the decoding procedure is occurring during training and not during the devaluation test itself, the decoder we used is not relying on features of the devaluation procedure test that would be trivially different between groups (such as neural correlates of reduced responding on the devalued action) to facilitate successful strategy decoding. Moreover, we also carefully controlled for confounding aspects of behavior during training that could potentially be used to distinguish goal-directed from habitual participants, such as the response rate and/or the slope of the learning curve. Importantly, the participants in our study who were classed as habitual did not simply exhibit a lack of understanding of the devaluation procedure in particular. The consumption test is a crucial element of our experimental design that protects against this possibility – in this test, participants freely selected between the two token outcomes after the devaluation had occurred but before the action-outcome devaluation test. If participants do not understand the devaluation procedure, then they would not show a substantial preference for the still valued coin. We excluded those participants (a relatively small proportion) who performed poorly on the consumption test. Thus, the participants included in our sample showed clear evidence that

they understood the devaluation procedure because they exhibited coin preferences consistent with optimal choice. This eliminates a confounding explanation that the classifier is picking up on those participants appearing to exhibit habit-like behavior solely because they did not understand or pay attention to the devaluation procedure.

The decoder's ability to generalize to an independent held-out sample is strong evidence that neural patterns in the left posterior putamen reflect robust and behaviorally meaningful representations of each individual's behavioral control strategy. Unlike most subject-wise MVPA studies that typically report only cross-validated performance within the same sample, our use of an ensemble classifier tested on unseen participants demonstrates true out-of-sample generalizability. It suggests that the identified multivoxel patterns don't overfit the training set but instead capture a consistent neural signature distinguishing goal-directed from habitual individuals across datasets. It's also important to note that our classification approach also generalized to a different cohort of individuals: patients with one or more psychiatric diagnoses, thereby showing that the generalization applies even to a different population of individuals beyond our original healthy participant sample.

The fact that decoded activity in posterior putamen allows us to distinguish between the form of behavioral control governing an individual's behavior on a given task suggests that neural activity in this region differs when individuals are engaged in either goal-directed or habitual behavior. This finding extends our understanding of the functional contributions of the posterior putamen, while inviting several possible interpretations.

A longstanding proposal in the literature is that this region is involved in encoding the stimulus-response associations that underpin learned habits [17–19]. It is possible that the goal-directed participants in the present study were less successful at learning stimulus-response associations, which was reflected in different patterns of activity in the putamen during training compared to individuals who did learn robust stimulus-response associations and subsequently deployed habits, and that these differences are being exploited by the decoder in the present study. It is relevant to note that we didn't find a significant relationship between stimulus-response knowledge (probed after the experiment) and goal-directedness, which would appear to indicate against this possibility. However, explicit knowledge of stimulus-response contingencies may not be synonymous with learned stimulus-response associations that are likely implicit and not easily accessible to meta-cognitive knowledge.

Alternatively, the posterior putamen may be doing more than merely encoding stimulus-response associations, in that activity in this region could be playing an active role as part of the mechanism by which the specific mode of behavioral control is implemented. This possibility resonates with emerging evidence in the rodent literature indicating a relationship between neuronal activity in the rodent homologue of the posterior putamen, the dorsolateral striatum (DLS), and the behavioral expression of habits [20]. Neurons in DLS have been found to decrease engagement during goal-directed actions [8], while DLS neuronal activity during the onset of a behavioral sequence correlates with the degree to which animals subsequently pursue devalued outcomes [21]. Stimulation of the DLS during behavioral onset elicits increased habit-like behavior [22], as well as inducing reduced goal-directed control [23]. Thus, it is

plausible that in the present study, the multivariate distributed signal being leveraged to predict behavioral strategy reflects neuronal populations directly involved in driving the behavioral onset of habitual actions.

Surprisingly, we did not find robust evidence for differential signatures of goal-directed and habitual control anywhere outside of the posterior putamen, such as in the anterior caudate or prefrontal cortex. The specificity of our findings is unexpected, given that other brain regions would also be expected to participate in the allocation of behavioral control over these strategies. These include areas implicated in arbitration between model-based and model-free reinforcement-learning, such as the ventrolateral prefrontal cortex and frontopolar cortex [24], as well as areas involved in outcome monitoring and feedback-related behavioral adjustment, [25, 26] such as the anterior cingulate cortex. There could be a number of reasons why these areas were not implicated in the present study. The first is that we are focusing on a particular time frame of initial learning, and if those other areas are involved in active top-down control of behavioral strategy, they might not be differentially engaged during initial learning, but rather become more prominently involved when the two behavioral strategies conflict markedly in their proposed action selection policy, which would occur during the devaluation phase. This possibility could be tested, provided trivial action selection differences are taken into account, as noted earlier. However, in the present study, it is not feasible to reliably decode neural activity patterns during the devaluation probe because it is a brief one-shot manipulation, severely limiting the amount of data available for a decoder to utilize. To overcome these limitations, a future study could employ repeated devaluation and revaluation so as to yield sufficient data for robust decoding of activity associated with behavioral control during such scenarios [27]. Another possible reason for the absence of robust neural signatures of behavioral strategy in these other regions is that patterns of neural activity associated with the engagement of habitual and goal-directed control might not be sufficiently similar across participants to enable between-participant decoding to be successful. Perhaps the posterior putamen is relatively unique in possessing sufficiently similar neural coding patterns associated with behavioral strategy across participants so as to enable between-participant decoding.

The tendency to express habits behaviorally (and/or exhibit reduced goal-directed control) has been found to correlate with a number of psychological traits and dispositions, including increased stress, anxiety, and compulsivity [14, 28–31]. In the present study, because we examine behavioral performance on a single task with a one-shot devaluation procedure, we cannot ascertain whether the putamen signal we detect reflects trait-related dispositional effects of an individual's general propensity to engage in habits and goal-directed control, or whether the putamen activity is instead indexing transient neuronal correlates of a habitual or goal-directed response tendency. Consequently, neuronal correlates of the behavioral control strategy observed here could pertain to either stable between-participant differences in neuronal encoding within the putamen, or else a contextually-specific snapshot of current task performance. However, none of the measures we obtained of individual traits, including of obsessive compulsive traits, anxiety or mood, could explain the observed effects of

devaluation sensitivity in the putamen. Thus, we can at least rule out an explanation of putamen activity in terms of the specific trait measures we obtained.

The finding that it is possible to decode an individual's overall behavioral strategy on a task from activity measured even within even a single brain area, opens up the possibility that brain activity patterns could be used to predict subsequent behavioral strategies and ultimately performance, in the context of brain machine interfaces and even potentially in real-world settings. Such a capability could have utility in assigning individuals to particular tasks or roles to which they are most suited, to provide guidance as to the most appropriate learning strategy to adopt in a particular individual, or even indicating whether and when to intervene in order to promote the adoption of alternative learning strategies. Although the present decoding accuracy in the range of 62% is robust and significantly above chance, it is likely to be qualitatively below the accuracies that would be needed in a practical application. Furthermore, it is obviously infeasible to use MRI for such an application due to its lack of portability and expense. Nevertheless, the present findings provide a proof of principle that activity patterns can be used to predict even high-level behavioral strategies across individuals, opening the door to possible applications in the future if out-of-sample accuracies can be enhanced with improved methodology alongside less noisy and more portable brain measures.

To conclude, the present study shows that distributed activity in the human posterior putamen contains information about which behavioral control strategy an individual utilizes during performance of instrumental action selection. These findings provide evidence that it is possible to ascertain an individual's overall behavioral control strategy from circumscribed and localized brain activity. Furthermore, the findings provide novel insights into the potential functions of the human posterior putamen, consistent with the possibility that the posterior putamen is involved in promoting the onset of different behavioral strategies at the time of action initiation.

Acknowledgements

We thank Julian Michael Tyszka from the Caltech Brain Imaging Center for assistance with the functional neuroimaging. We thank Emmily Hovhannisyan from UCLA for help in participant recruitment. This study was supported by a grant from the National Institute of Mental Health (R01MH121089) to J.P.O.D.

Methods

Task Description

Participants completed a reinforcement learning task using a trackball to swipe left or right and button presses to collect gold and silver coins. The task comprised three sequential phases: Training, Consumption, and a Choice Test (Figure 1). Outside the scanner, a trained research assistant guided participants through detailed task instructions to ensure comprehension.

Training Phase

Participants first received explicit instructions mapping coin colors (gold or silver) to specific swiping directions (left or right). Each trial presented a cue and a target coin. Correct swiping triggered feedback in the form of a visual confirmation (an orange triangle) and highlighted the collected coin.

The training phase consisted of 20 blocks (10 in each run), each lasting 37–60 seconds. We instructed the participants to collect as many coins as possible. Rewards were delivered on a random ratio schedule (RR15), simulating the unpredictability of real-world reinforcement [32]. Both coin types had equal value. Each swipe incurred a cost of -1 point, while successful coin collection yielded +40 points, encouraging frequent responding. However, we did not explicitly inform them of this optimal strategy. The mean swipe rate across all participants was 2.42 ± 0.89 swipes per second (Healthy Controls $= 2.46 \pm 0.91$, Patients $= 2.35 \pm 0.87$).

Consumption Phase

After the first 10 training blocks, participants completed a pre-devaluation consumption test in which they selected 10 coins from a grid containing 10 gold and 10 silver coins within 10 seconds. It is included to familiarize the subjects with the consumption procedure before the actual test and minimize surprise. Here, the correct response was to collect any 10 coins. Participants then completed training blocks 11–20 using the same task structure (Methods: Training Phase). Upon completing training, we devalued one coin type by informing participants that the corresponding piggy bank was full. In the post-devaluation consumption test, participants again selected among 10 gold and 10 silver coins. We defined **consumption score** as the number of valuable (non-devalued) coins chosen.

Choice Test Phase

In the final phase, participants completed a choice test. Each trial presented both cues simultaneously, with the associated coins hidden behind a curtain. Participants swiped for 90 seconds without feedback, aiming to maximize total points. We computed the **devaluation ratio (DR)** as the proportion of swipes directed toward the valuable (non-devalued) coin ($DR \in [0, 1]$).

Contingency Memory Assessment

Following the main task, participants completed a brief memory test to assess explicit knowledge of task contingencies. Specifically, participants answered four binary, mutually exclusive questions: two assessed which cue symbols were associated with which swiping direction (left/right), and two assessed which coin colors corresponded to swiping direction. We didn't provide any feedback. To ensure a stringent criterion, we only considered responses accurate if both answers within a category were correct. We defined these binary outcome measures as the **stimulus contingency score** and the **coin contingency score**.

Participants

We recruited 203 healthy participants (ages 18–65 years) from the Greater Los Angeles area through flyers, the Caltech Chen Participation Center [33], word of mouth, and Craigslist [34]. We additionally screened them to ensure they did not exhibit symptoms of respiratory infection on the day of scanning, in accordance with Caltech policy during the data acquisition phase. We instructed the participants to abstain from recreational drug and alcohol use on the day before and the day of scanning, and to limit caffeine intake to their typical daily amount. All participants were fluent in English, had normal or corrected-to-normal vision, met MRI safety criteria, and reported no history of neurological disorders or drug abuse. We compensated the participants at a rate of $30 per hour for MRI scanning sessions and $20 per hour for remotely completed pre-scan tasks. California Institute of Technology Institutional Review Board (IRB) [35] approved the study, and all participants provided written informed consent.

Psychiatric Patient Sample

In addition to healthy participants, we recruited an undifferentiated sample of psychiatric patients with a diverse array of psychiatric disorders. Patients initially self-reported their diagnoses and were screened over the phone. Those who passed this screening completed an extended in-person evaluation with a clinical psychologist or psychiatrist before MRI scanning to confirm their current diagnosis. Thus all patients underwent a structured DSM-5 based structured clinical interview. In total, we recruited 68 patients spanning 12 different diagnoses (Supplementary: Table S2); diagnostic distributions in Supplementary: Table S1.

Exclusions

Participants completed both runs of the habit task during fMRI scanning. We excluded eighteen participants due to incomplete imaging data. We predefined additional exclusion criteria as:

1. **Low consumption score (< 8):** participants who may not have understood the devaluation manipulation (Healthy Controls = 25, Patients = 9).

2. **Devaluation ratio (DR < 0.333):** responses opposite to task contingencies, making it unclear whether behavior reflected habit or misunderstanding of action-outcome mappings (Healthy Controls = 12, Patients = 3).
3. **Slow response rate (< 0.5 swipes/s) during training:** indicating difficulty acquiring the task (Healthy Controls = 4, Patients = 1).

After applying exclusions, the final sample consisted of 144 Healthy Controls (78 female, 66 male; mean age = 30.00 ± 9.82 years) and 55 Patients (45 female, 10 male; mean age = 28.22 ± 9.49 years).

Data Analysis

We performed all data analysis in R and Python. We present the results as mean $\pm$ standard deviation unless reported otherwise.

Group Classification: Beta Mixture Modeling

We modeled the distribution of devaluation ratios (DR) using a mixture of Beta distributions fit with an Iterated Method of Moments Expectation–Maximization (IMM–EM) algorithm [36].

Equations 1–2 define the generative model, which assumes each observed DR value x_i is drawn from a convex combination of K Beta distributions. The mixing proportions π_k capture the prevalence of latent subgroups, while the Beta parameters (α_k, β_k) characterize their response profiles.

We assessed model adequacy using the Kolmogorov–Smirnov (KS) statistic (Eq. 3), which quantifies the maximum discrepancy between the empirical distribution of DR values and the fitted mixture distribution.

We selected this approach because (i) the IMM–EM method supports Beta-distribution fitting even in the presence of boundary values (e.g., 0 and 1), which occur in our data, and (ii) the KS statistic has been previously argued to provide a reliable measure of fit for mixture models [36].

$$x_i \sim \sum_{k=1}^{K} \pi_k \, \mathrm{Beta}(\alpha_k, \beta_k), \quad \text{for } i = 1, \ldots, N \tag{1}$$

$$F_\theta(x) = \sum_{k=1}^{K} \pi_k \, F_{\mathrm{Beta}}(x \mid \alpha_k, \beta_k), \tag{2}$$

$$D_N = \sup_{x \in [0,1]} |F_N(x) - F_\theta(x)| \tag{3}$$

The IMM algorithm estimates both the component distributions and their mixing proportions via a responsibility-weighted Expectation step and a moment-matching Maximization step, with explicit handling of boundary values (0 and 1). Initialization was performed with 1000 random seeds using a distance-aware strategy [37]. To identify the optimal number of mixture components, we estimated models with one to five components on the Initial Cohort only and compared fits using the Kolmogorov–Smirnov (KS) distance. A two-component solution provided the best trade-off between

fit quality and parsimony (Supplementary: Figure S1). The model parameters were $\alpha = 29.636$, $\beta = 28.435$ for the goal-directed subgroup and $\alpha = 3.375$, $\beta = 0.460$ for the habitual subgroup, with nearly equal mixing proportions ($\pi = 0.459$, $\pi = 0.541$). The posterior decision boundary between components was 0.648, which we used to classify participants.

To evaluate the robustness of this fit, we implemented a cross-validated parametric bootstrap (5000 replicates) based on the KS statistic. For each replicate, we simulated train–test splits under the fitted model, re-estimated mixture parameters on bootstrap training sets, and computed the KS distance on held-out test sets. This procedure accounts for both sampling variability and refitting uncertainty.

In the Initial Cohort, the mean KS distance was 0.1639 with a bootstrap $p = 0.41$. In the full set of healthy controls ($N = 144$), the mean KS distance was 0.1544 with a bootstrap $p = 0.28$. Across datasets, bootstrap p-values indicated no evidence of systematic misfit, supporting the robustness and generalizability of the two-component solution.

Response Rates

To evaluate participants' engagement and learning during the training phase, we analyzed their response rate across task blocks. Response rate served two methodological purposes. First, it provided a behavioral index of task acquisition: although participants were not explicitly instructed to swipe as frequently as possible, this was the optimal strategy for maximizing outcomes, and successful learning should manifest as an increase in swiping rate over time. Second, it allowed us to test whether goal-directed and habitual participants differed in engagement or learning dynamics, which could otherwise confound later group comparisons.

We analyzed response rates using a linear mixed-effects model fit separately for each run of the training phase on the Initial Cohort. Participants were binarized into Goal-Directed or Habitual groups using the mixture-model threshold ($DR = 0.648$). Block number was treated as a continuous predictor (centered within run), group as a categorical fixed factor (Goal-Directed = reference), and subject as a random intercept. The model specification was:

$$\begin{aligned} \text{rate}_{isr} = \beta_0 + \beta_1 \,\text{blockid}_{isr} + \beta_2 \,\mathbb{I}[g_s = \text{HB}] + \beta_3 \,\text{blockid}_{isr} \cdot \mathbb{I}[g_s = \text{HB}] \\ + b_{0s} + \varepsilon_{isr}, \end{aligned} \tag{4}$$

$$b_{0s} \sim \mathcal{N}(0, \sigma_b^2), \qquad \varepsilon_{isr} \sim \mathcal{N}(0, \sigma^2). \tag{5}$$

where β_0 is the intercept for the Goal-Directed group, β_1 the slope of block number for the Goal-Directed group, β_2 the intercept difference for the Habitual group, and β_3 the group $\times$ block interaction. b_{0s} captures subject-level random intercepts, and ε_{isr} is the residual error.

Models were implemented in `statsmodels` (Python) using restricted maximum likelihood and the L-BFGS optimizer (1000 iterations). Group-specific slopes were derived by combining the main effect of block number with the interaction term. Wald

tests provided z-statistics and p-values. The fitted models revealed reliable between-subject differences in baseline response rates (Run 1: $\sigma_b^2 = 0.771$; Run 2: $\sigma_b^2 = 0.983$). Fixed-effect estimates are reported in the Results section.

Questionnaire Self-Reports

Participants completed three batteries of psychological questionnaires, which we aimed to administer approximately one week before the fMRI scan. The STAI-State was collected immediately before scanning to capture transient anxiety levels, and participants who missed questionnaires completed them after the scan (within the next week). To avoid unnecessary data loss, we do not exclude participants with incomplete responses. Instead, missing values in the psychological confound matrix were imputed using a K-nearest neighbors algorithm [38]. These psychological measures were used as confounding variables in the fMRI analyses to control for individual differences in affective and cognitive state. Although we included the NEO Personality Inventory in the assessment battery, we excluded it from analysis due to a coding error that occurred during administration. A complete list of psychological measures, along with response distributions, is provided in Supplementary: Figure S3.

Logistic Regression Predicting Devaluation Sensitivity

To identify behavioral and trait-level predictors of sensitivity to outcome devaluation, we fit logistic regression models with binarized devaluation ratio (DR) as the dependent variable (goal-directed: $DR \geq 0.648$; habitual: $DR < 0.648$). Analyses were conducted in `statsmodels` (Python) using maximum likelihood estimation. Predictors included (i) task-related variables (e.g., response rate, task order, devalued coin identity, devalued response direction, contingency knowledge), (ii) psychological questionnaire scores (OCI-R, COHS, SRS, OLIFE, STAI, LSPS, AUDIT, EAT, PDSS, PSWQ, BDI, Apathy Scale, BIS/BAS), and (iii) demographic covariates (age, sex, handedness). We standardized all predictors before model fitting. We handle missing data via the KNN imputation described above.

We assessed model fit using log-likelihood ratio tests and pseudo R^2 statistics. We used Type III Wald χ^2 tests to evaluate the contribution of individual predictors. For the Initial Cohort ($N = 108$), the logistic regression converged successfully (pseudo-$R^2 = 0.245$; LLR $p = 0.048$). Response rate ($\beta = 0.65$, Wald $\chi^2(1) = 4.82$, $p = 0.028$) and contingency knowledge ($\beta = 0.90$, Wald $\chi^2(1) = 9.31$, $p = 0.002$) emerged as significant predictors of goal-directed behavior, while stimulus–response knowledge was non-significant ($\beta = -0.05$, $p = 0.837$).

In the full sample ($N = 199$; healthy controls and patients), the model also converged (pseudo-$R^2 = 0.127$; LLR $p = 0.068$). Coin–contingency knowledge again emerged as the strongest predictor of goal-directed behavior ($\beta = 0.74$, Wald $\chi^2(1) = 14.67$, $p < 0.001$). OCI-R total score was negatively associated with goal-directed behavior ($\beta = -0.45$, Wald $\chi^2(1) = 4.04$, $p = 0.044$), indicating greater compulsivity predicted habitual responding. Response rate showed a trend-level effect ($\beta = 0.33$, Wald $\chi^2(1) = 3.67$, $p = 0.055$). All other psychometric, task, and demographic predictors were non-significant (all $p > 0.10$).

Task administration technical issues

A subset of participants were affected by two minor technical issues during task administration:

One of these issues (in n=25 participants after exclusions) concerned a reset in the trackerball position during the choice phase after a response had been recorded such that if the trackerball had sufficient momentum, multiple responses could be recorded where only one was intended. Because this issue only affected the choice trials, it did not impact on the training data used in the fMRI analysis, thus only potentially affecting the behavioral results. To correct for this, we removed trailing responses with intertrial intervals $< 0.084s$ when the preceding response was in the same direction, thereby eliminating unintended responses from the behavioral choice data. In any event, our main behavioral measure of interest (devaluation ratio) would in principle not have been affected by this issue (as both swiping directions were equally affected). Thus, this issue did not impact the results in any way.

Another issue concerned the default "mouse sensitivity" setting in the stimulus PC that was set to a different faster speed setting than intended in a subset of participants (n = 52; post-exclusions). This also resulted in some trailing responses being registered as multiple swipes. In this case, we also removed trailing responses $< 0.034s$ under the same criterion. We further checked for differences in key behavioral variables in individuals with and without the fast mouse sensitivity. There were no significant differences in either overall response rates during training (t=1.24; p=0.22) or in the devaluation sensitivity measure during the choice phase (t=-0.95;p=0.35) when comparing between individuals with or without fast mouse sensitivity. Thus, we included these participants in the analysis. For completeness, we also confirmed that our fMRI findings are not affected by either of these issues, described in more detail below.

MRI methodology

Acquisition

fMRI data were acquired at the Caltech Brain Imaging Center (Pasadena, CA), using a Siemens Prisma 3T scanner with a 32-channel radio frequency coil. The functional scans were conducted using a multi-band echo-planar imaging (EPI) sequence with 72 slices, $-30°$ slice tilt from the AC-PC line, 192 mm $\times$ 192 mm field of view, 2 mm isotropic resolution, repetition time (TR) of 1.12 s, echo time (TE) of 30 ms, multi-band acceleration of 4, 54° flip angle, in-plane acceleration factor 2, echo spacing of 0.56 ms, and EPI factor of 96. Following each run, both positive and negative polarity EPI-based field maps were collected using similar parameters to the functional sequence, but with a single band, TR of 5.13 s, TE of 41.40 ms, and 90° flip angle.

T1-weighted and T2-weighted structural images were also acquired for each participant with 0.9 mm isotropic resolution and 230 mm $\times$ 230 mm field of view. For the T1-weighted scan, TR of 2.55 s, TE of 1.63 ms, inversion time (TI) of 1.15 s, flip angle of 8°, and in-plane acceleration factor 2 were used. The T2-weighted scan was acquired with TR of 3.2 s, TE of 564 ms, and in-plane acceleration factor of 2.

Preprocessing

All fMRI data underwent a two-stage preprocessing pipeline. First, we use fMRIPrep (version 23.1.3) [Reference] for anatomical and functional preprocessing, which includes standard procedures such as motion correction, slice-time correction, anatomical normalization, and confound estimation. In the second stage, we utilized Nilearn (version 0.10.1: manually corrected for the z-score bug [39]) to further denoise and standardize the BOLD time series. It involved high-pass filtering, temporal detrending, and regression of nuisance signals, including motion parameters, CompCor components, and global signals. The resulting time series for each run were standardized, optionally reduced using PCA (retaining 95% of the variance), and concatenated to form participant-level inputs for multivariate analyses. Supplementary Supplementary: Figure S12. provides a visual summary of this preprocessing workflow.

Regions of Interest

We selected eleven regions of interest (RoIs) based on prior research that has linked these regions to both goal-directed and habitual behavior. These included bilateral anterior and posterior caudate and putamen, defined anatomically using the striatal parcellation from Pauli et al. (2018) [40], as well as functionally defined (6mm spherical searchlight) prefrontal RoIs. Specifically, we used a dorsolateral prefrontal cortex (dlPFC) RoI centered at $(48, 9, 36)$, derived from Gläscher et al. (2010) [41], and mirrored it to the left hemisphere. A ventromedial prefrontal cortex (vmPFC) ROI centered at $(-3, 41, -11)$, based on McNamee et al. (2013) [13]. We use all RoIs as masks for both searchlight-based and region-wise decoding analyses.

fMRI - General Linear Modelling

We used Nilearn (version 0.10.1) [42] for all fMRI-GLM analysis.

First Level

We constructed first-level general linear models for each participant using a design matrix that spans both concatenated fMRI runs, but models only the training blocks. Task-related regressors did not capture variance from the subsequent choice-test block, preserving those segments from confounding downstream prediction analyses. Each run included 10 training blocks, and two task-related regressors were defined: a block-mean regressor capturing sustained activation during each block, and a block-linear regressor indexing linear progression across blocks. We use a canonical SPM hemodynamic response function [43] with temporal and dispersion derivatives. The design matrix included the intercept, along with orthogonalized and standardized versions of each regressor and their derivatives, resulting in a total of five regressors. We use a 5mm Gaussian kernel for spatial smoothing. Supplementary: Figure S11 presents the visualizations of the design matrix and contrast.

Second Level (Parametric Analysis)

To examine how individual differences in psychological traits and task behavior relate to neural activity, we performed a second-level (parametric) analysis using participant-wise confound variables as predictors. The design matrix included a broad set of standardized confounds (e.g., task-related confounds, psychological evaluations, and demographic variables), with missing values imputed using the K-nearest neighbors' (N = 5) method [44, 45]. The resulting design matrix was standardized and used to model voxel-wise brain responses, with a specified contrast vector isolating the predictor of interest. Supplementary Supplementary: Figure S13 provides a visual summary of the analysis.

Statistical Inference

For first-level statistical inference, we identified task-related activation using a parametric cluster-level correction based on Gaussian Random Field Theory (RFT), as implemented in AFNI. Results reported in the main text reflect clusters that survived this correction approach. To ensure robustness, we also conducted supplementary non-parametric testing using Threshold-Free Cluster Enhancement (TFCE) with permutation-based inference, as well as small-volume corrections using the False Discovery Rate (FDR) within anatomically defined masks. These additional tests served as confirmatory checks to validate the reliability of activation patterns identified through the primary parametric analysis.

Multivariate Pattern Analysis

Searchlight Analysis

First, we performed multivoxel pattern analysis (MVPA) on the Initial cohort, using a searchlight decoding approach to identify neural patterns predictive of individual differences in behavior. For each participant, we used first-level contrast maps (block-mean, Supplementary: Figure S11) as input features and classified individuals based on their high versus low devaluation ratios (threshold = 0.648). We conducted separate analyses within eleven anatomically defined regions of interest (RoIs) implicated in goal-directed and habitual control, using a spherical searchlight with a 4 mm radius. We trained a linear discriminant analysis (LDA) classifier with stratified 4-fold cross-validation and evaluated performance using the accuracy score. To verify that results were not specific to predefined regions, we repeated the analysis using a whole-brain searchlight.

Confound Regression

To control for nuisance effects, we performed *train-on-train* confound residualization with an intercept inside each cross-validation fold. Let $\mathbf{C} \in \mathbb{R}^{N \times k}$ denote the full confound matrix for all participants. For a given fold, let $\mathcal{I}_{\text{train}}$ and $\mathcal{I}_{\text{test}}$ index the training and test rows, respectively. We z-scored confounds using training-set statistics only and added a column of ones (intercept); the resulting parameters (means, standard

deviations, and $\boldsymbol{\beta}$) were fit on the training split and then applied to the held-out split (cf. [46]).

Let

$$\mathbf{X}_{\text{train}} \in \mathbb{R}^{n\times v}, \quad \mathbf{X}_{\text{test}} \in \mathbb{R}^{m\times v},$$
$$\mathbf{C}_{\text{train}} = \mathbf{C}[\mathcal{I}_{\text{train}}, :] \in \mathbb{R}^{n\times k}, \quad \mathbf{C}_{\text{test}} = \mathbf{C}[\mathcal{I}_{\text{test}}, :] \in \mathbb{R}^{m\times k}.$$

Compute train-only standardization of confounds and add an intercept:

$$\boldsymbol{\mu} = \tfrac{1}{n}\mathbf{1}_n^\top \mathbf{C}_{\text{train}}, \quad \boldsymbol{\sigma} = \mathrm{sd}(\mathbf{C}_{\text{train}}) \in \mathbb{R}^k,$$

$$\widetilde{\mathbf{C}}_{\text{train}} = (\mathbf{C}_{\text{train}} - \mathbf{1}_n\boldsymbol{\mu}^\top) \oslash \boldsymbol{\sigma}, \quad \widetilde{\mathbf{C}}_{\text{test}} = (\mathbf{C}_{\text{test}} - \mathbf{1}_m\boldsymbol{\mu}^\top) \oslash \boldsymbol{\sigma},$$
$$\widehat{\mathbf{C}}_{\text{train}} = [\mathbf{1}_n \;\; \widetilde{\mathbf{C}}_{\text{train}}] \in \mathbb{R}^{n\times(k+1)}, \quad \widehat{\mathbf{C}}_{\text{test}} = [\mathbf{1}_m \;\; \widetilde{\mathbf{C}}_{\text{test}}] \in \mathbb{R}^{m\times(k+1)}.$$

Estimate coefficients $\boldsymbol{\beta} \in \mathbb{R}^{(k+1)\times v}$ on the training split:

$$\boldsymbol{\beta} = (\widehat{\mathbf{C}}_{\text{train}}^\top \widehat{\mathbf{C}}_{\text{train}})^{-1} \widehat{\mathbf{C}}_{\text{train}}^\top \mathbf{X}_{\text{train}}. \tag{6}$$

Residualize features in train and test:

$$\mathbf{X}_{\text{train}}^{\text{resid}} = \mathbf{X}_{\text{train}} - \widehat{\mathbf{C}}_{\text{train}}\boldsymbol{\beta}, \quad \mathbf{X}_{\text{test}}^{\text{resid}} = \mathbf{X}_{\text{test}} - \widehat{\mathbf{C}}_{\text{test}}\boldsymbol{\beta}.$$

Generalization Analysis

To evaluate the out-of-sample generalizability of our decoder, we trained on the Initial cohort and tested on an independent Held-out cohort. Whereas our primary (Initial cohort) analyses relied on searchlight decoding with a single linear discriminant analysis (LDA) classifier, we conducted the generalization analysis using the entire significant cluster as the feature space.

For classification, we employed an ensemble-based *bagging classifier* with LDA as the base estimator. Bagging (bootstrap aggregating) improves stability by averaging across multiple resampled training sets, thereby reducing variance and mitigating the risk of overfitting. Specifically, we constructed an ensemble of 100 LDA classifiers, each trained on a bootstrap sample comprising 75% of participants and 50% of features, with both participants and features resampled with replacement. We chose the 75% subsampling of participants to mirror the effective training proportion in the stratified 4-fold cross-validation used in the MVPA analyses, ensuring consistency in the inductive bias across approaches. The final prediction for a held-out participant was obtained by averaging the predicted probabilities across the ensemble. We use feature subsampling to approximate the reduced and variable feature spaces encountered in the 4 mm searchlight analyses, thereby aligning the generalization analysis with the representational constraints of the searchlight setting. Such an ensemble approach is particularly valuable in multivariate decoding of high-dimensional voxel patterns, where single decision boundaries may otherwise be unstable.

As in the initial analysis, confound regression was performed in a leakage-safe manner using train-only confound statistics, and we standardize features within each training fold. We evaluate model performance in the generalization analysis using the mean area under the ROC curve (AUC) across repetitions, with significance assessed

by permutation testing of group labels in the held-out set. Unlike the discovery cohort, where the groups were balanced (55 goal-directed vs. 53 habitual) and accuracy provided a suitable measure, the generalization cohort was moderately imbalanced (23 goal-directed vs. 13 habitual). In this context, we use ROC AUC as it is insensitive to class imbalance and provides a more reliable estimate of predictive performance.

Statistical Inference

To statistically evaluate decoding performance, we applied a non-parametric cluster-level correction procedure based on permutation testing. We computed voxel-wise accuracy maps from the MVPA and compared them to a null distribution generated from 5,000 permutations.

In the searchlight analysis, for each voxel, we calculated empirical p-values by comparing the accurate decoding accuracy to the permutation-derived distribution. To identify significant clusters, we applied a cluster-forming threshold based on the 95th percentile of the voxel-wise null distribution and identified contiguous supra-threshold regions within predefined anatomical masks. We calculated cluster-level statistics by summing the difference between observed and thresholded values across each cluster. To estimate family-wise error (FWE)–corrected significance, we repeated this procedure across all permutations and computed the null distribution of maximum cluster scores. We then compared observed clusters to this distribution to obtain corrected p-values. This approach was applied both within individual ROIs and across a whole-brain grey matter mask to ensure the robustness and anatomical specificity of decoding effects. We present the detailed algorithm in the Supplement (Algorithm **??**).

Robustness Analyses

Unless otherwise specified, all robustness analyses followed a typical multivariate decoding pipeline designed for consistency and comparability across tests. Specifically, we used all data ($N = 199$) and trained an LDA classifier within a 5-fold stratified cross-validation framework, and 100 repeated random resamplings of the data. Within each fold, features were residualized with respect to confounds using train-only statistics, z-scored using training-set parameters, and then applied to the held-out set. Unless otherwise noted, the confound set included mean response rate and response slopes from both training runs. Model performance was quantified using the area under the ROC curve (AUC), with scores averaged across folds and repetitions. This framework was applied uniformly across robustness checks, ensuring that differences in performance reflect the tested variation rather than methodological inconsistencies.

Confirming fMRI results are unimpacted by technical issues during task administration

Using the procedure described for the other robustness checks, to rule out any residual impact of task administration issues described earlier on the fMRI results, we included separate categorical nuisance regressors in the confound regression procedure for both issues noted (see Equation 6), and we found that decoding performance was still

robust (ROC AUC = 0.69). We also verified that the decoding effect holds even when training the decoder on unimpacted participants (n = 122) and cross-generalizing to the impacted group (n = 77; AUC = 0.65 +/- 0.02).

References

[1] Dickinson, A.: Actions and habits: the development of behavioural autonomy. Philosophical Transactions of the Royal Society of London. B, Biological Sciences **308**(1135), 67–78 (1985) https://doi.org/10.1098/rstb.1985.0010

[2] Balleine, B.W., O'Doherty, J.P.: Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology **35**(1), 48–69 (2010) https://doi.org/10.1038/npp.2009.131

[3] Dolan, R.J., Dayan, P.: Goals and habits in the brain. Neuron **80**(2), 312–325 (2013) https://doi.org/10.1016/j.neuron.2013.09.007

[4] Yin, H.H., Ostlund, S.B., Knowlton, B.J., Balleine, B.W.: The role of the dorsomedial striatum in instrumental conditioning. European Journal of Neuroscience **22**(2), 513–523 (2005) https://doi.org/10.1111/j.1460-9568.2005.04218.x

[5] Hart, G., Balleine, B.W.: Consolidation of goal-directed action depends on mapk/erk signaling in rodent prelimbic cortex. Journal of Neuroscience **36**(47), 11974–11986 (2016) https://doi.org/10.1523/JNEUROSCI.1760-16.2016

[6] Shipman, M.L., Hart, G., Killcross, S., Balleine, B.W., Corbit, L.H.: Inactivation of prelimbic and infralimbic cortex respectively affects minimally-trained and extensively-trained goal-directed actions. Neurobiology of Learning and Memory **155**, 164–172 (2018) https://doi.org/10.1016/j.nlm.2018.07.002

[7] Yin, H.H., Knowlton, B.J., Balleine, B.W.: Inactivation of dorsolateral striatum enhances sensitivity to changes in the action–outcome contingency in instrumental conditioning. Behavioural Brain Research **166**(2), 189–196 (2006) https://doi.org/10.1016/j.bbr.2005.07.012

[8] Gremel, C.M., Costa, R.M.: Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. Nature Communications **4**(1), 2264 (2013) https://doi.org/10.1038/ncomms3264

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

[10] Balleine, B.W., Yin, H.H., Hart, G., Corbit, L.H.: The dorsomedial striatum: an optimal cellular environment for encoding and updating goal-directed learning. Current Opinion in Behavioral Sciences **41**, 38–44 (2021) https://doi.org/10.1016/j.cobeha.2021.03.007

[11] Knowlton, B.J., Patterson, T.K.: Habit formation and the striatum. In: Current Topics in Behavioral Neurosciences vol. 37, pp. 275–295. Springer, Cham (2018). https://doi.org/10.1007/7854_2016_451

[12] Gera, R., Or, M.B., Tavor, I., Roll, D., Cockburn, J., Barak, S., Schonberg, T.: Characterizing habit learning in the human brain at the individual and group levels: A multi-modal mri study. NeuroImage **272**, 120002 (2023) https://doi.org/10.1016/j.neuroimage.2023.120002

[13] McNamee, D., Liljeholm, M., Zika, O., O'Doherty, J.P.: Characterizing the associative content of brain structures involved in habitual and goal-directed actions in humans: a multivariate fmri study. The Journal of Neuroscience **35**(9), 3764–3771 (2015) https://doi.org/10.1523/JNEUROSCI.4677-14.2015

[14] Pool, E.R., Gera, R., Fransen, A., Perez, O.D., Cremer, A., Aleksic, M., Tanwisuth, S., Quail, S., Ceceli, A.O., Manfredi, D.A., Nave, G., Tricomi, E., Balleine, B.W., Schonberg, T., Schwabe, L., O'Doherty, J.P.: Determining the effects of training duration on the behavioral expression of habitual control in humans: a multilaboratory investigation. Learning & Memory **29**(1), 16–28 (2022) https://doi.org/10.1101/lm.053413.121

[15] Lally, P., Jaarsveld, C.H.M.V., Potts, H.W.W., Wardle, J.: How are habits formed: Modelling habit formation in the real world. European Journal of Social Psychology **40**(6), 998–1009 (2010) https://doi.org/10.1002/ejsp.674

[16] Adams, C.D., Dickinson, A.: Instrumental responding following reinforcer devaluation. The Quarterly Journal of Experimental Psychology Section B **33**(2b), 109–121 (1981)

[17] Horvitz, J.C.: Stimulus–response and response–outcome learning mechanisms in the striatum. Behavioural Brain Research **199**(1), 129–140 (2009) https://doi.org/10.1016/j.bbr.2008.10.049

[18] Devan, B.D., Hong, N.S., McDonald, R.J.: Parallel associative processing in the dorsal striatum: Segregation of stimulus–response and cognitive control subregions. Neurobiology of Learning and Memory **96**(2), 95–120 (2011) https://doi.org/10.1016/j.nlm.2011.06.002

[19] Hart, G., Leung, B.K., Balleine, B.W.: Dorsal and ventral streams: the distinct role of striatal subregions in the acquisition and performance of goal-directed actions. Neurobiology of Learning and Memory **108**, 104–118 (2014) https://doi.org/10.1016/j.nlm.2013.11.003

[20] Malvaez, M., Wassum, K.M.: Regulation of habit formation in the dorsal striatum. Current Opinion in Behavioral Sciences **20**, 67–74 (2018) https://doi.org/10.1016/j.cobeha.2017.10.004

[21] Smith, K.S., Graybiel, A.M.: A dual operator view of habitual behavior reflecting cortical and striatal dynamics. Neuron **79**(2), 361–374 (2013) https://doi.org/10.1016/j.neuron.2013.05.038

[22] Crego, A.C., Štoček, F., Marchuk, A.G., Carmichael, J.E., Meer, M.A.A., Smith, K.S.: Complementary control over habits and behavioral vigor by phasic activity in the dorsolateral striatum. Journal of Neuroscience **40**(10), 2139–2153 (2020) https://doi.org/10.1523/JNEUROSCI.2080-19.2020

[23] Hart, G., Chieng, B., Balleine, B.W.: Unilateral stimulation of the dorsolateral striatum attenuates goal-directed action. European Journal of Neuroscience **62**(3), 70175 (2025) https://doi.org/10.1111/ejn.70175

[24] Lee, S.W., Shimojo, S., O'Doherty, J.P.: Neural computations underlying arbitration between model-based and model-free learning. Neuron **81**(3), 687–699 (2014) https://doi.org/10.1016/j.neuron.2013.11.028

[25] Watson, P., Van Wingen, G., De Wit, S.: Conflicted between goal-directed and habitual control, an fmri investigation. eNeuro **5**(4), 0240–182018 (2018) https://doi.org/10.1523/ENEURO.0240-18.2018

[26] Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D.: Conflict monitoring and cognitive control. Psychological Review **108**(3), 624–652 (2001) https://doi.org/10.1037/0033-295X.108.3.624

[27] Gera, R., Barak, S., Schonberg, T.: A novel free-operant framework enables experimental habit induction in humans. Behavior Research Methods **56**(4), 3937–3958 (2024) https://doi.org/10.3758/s13428-023-02277-6

[28] Voon, V., Derbyshire, K., Rück, C., Irvine, M.A., Worbe, Y., Enander, J., Bullmore, E.T.: Disorders of compulsivity: a common bias towards learning habits. Molecular Psychiatry **20**(3), 345–352 (2015) https://doi.org/10.1038/mp.2014.44

[29] Gillan, C.M., Morein-Zamir, S., Urcelay, G.P., Sule, A., Voon, V., Apergis-Schoute, A.M., Robbins, T.W.: Enhanced avoidance habits in obsessive-compulsive disorder. Biological Psychiatry **75**(8), 631–638 (2014) https://doi.org/10.1016/j.biopsych.2013.02.002

[30] Schwabe, L., Wolf, O.T.: Stress prompts habit behavior in humans. Journal of Neuroscience **29**(22), 7191–7198 (2009) https://doi.org/10.1523/JNEUROSCI.0979-09.2009

[31] Vogel, S., Klumpers, F., Schröder, T.N., Oplaat, K.T., Krugers, H.J., Oitzl, M.S., Fernández, G.: Stress induces a shift towards striatum-dependent stimulus-response learning via the mineralocorticoid receptor. Neuropsychopharmacology **42**(6), 1262–1271 (2017) https://doi.org/10.1038/npp.2016.262

[32] Haw, J.: Random-ratio schedules of reinforcement: The role of early wins and unreinforced trials. Journal of Gambling Issues **21**, 56–67 (2008) https://doi.org/10.4309/jgi.2008.21.6

[33] Chen Institute for Neuroscience: Chen Studies. https://chenstudies.caltech.edu/. Accessed: 2025-07-23 (n.d.)

[34] Craigslist: Craigslist Los Angeles. https://losangeles.craigslist.org/. Accessed: 2025-07-23 (n.d.)

[35] California Institute of Technology: Institutional Review Board (IRB). https://researchcompliance.caltech.edu/research-committees/institutional-review-board. Accessed: 2025-07-23 (n.d.)

[36] Li, Y., Han, R., Bi, C., Li, M., Wang, S., Gao, X., Li, Y.: Deepsimulator: a deep simulator for nanopore sequencing. Algorithms for Molecular Biology **12**(1), 1–12 (2017) https://doi.org/10.1186/s13015-017-0112-1

[37] Arthur, D., Vassilvitskii, S.: K-means++: The advantages of careful seeding. In: Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 1027–1035. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (2007)

[38] Murti, D.M.P., Pujianto, U., Wibawa, A.P., Akbar, M.I.: K-nearest neighbor (k-nn) based missing data imputation. In: 2019 5th International Conference on Science in Information Technology (ICSITech), pp. 83–88 (2019). https://doi.org/10.1109/ICSITech46713.2019.8987530

[39] Nilearn contributors: Pull Request in Nilearn. https://github.com/nilearn/nilearn/pull/3474. GitHub pull request (2025)

[40] Pauli, W.M., Nili, A.N., Tyszka, J.M.: A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. Scientific Data **5**, 180063 (2018) https://doi.org/10.1038/sdata.2018.63

[41] Gläscher, J., Daw, N., Dayan, P., O'Doherty, J.P.: States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron **66**(4), 585–595 (2010) https://doi.org/10.1016/j.neuron.2010.04.016

[42] Nilearn contributors, Chamma, A., Frau-Pascual, A., Rothberg, A., Abadie, A., Abraham, A., Gramfort, A., Savio, A., Cionca, A., Sayal, A., Thual, A., Kodibagkar, A., Kanaan, A., Pinho, A.L., Joshi, A., Idrobo, A.H., Kieslinger, A.-S., Kumari, A., Rokem, A., Najera, O.: nilearn (0.12.0). Zenodo (2025). https://doi.org/10.5281/zenodo.15709579

[43] Wikibooks contributors: SPM/Haemodynamic Response Function. https://en.wikibooks.org/wiki/SPM/Haemodynamic_Response_Function. Accessed: 2025-07-24 (2024)

[44] scikit-learn developers: sklearn.impute.KNNImputer — Scikit-learn Documentation. scikit-learn, (2024). scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.impute.KNNImputer.html

[45] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in python. Journal of Machine Learning Research **12**, 2825–2830 (2011)

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