communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-026-10513-4

Dorsomedial prefrontal cortex links abstract planning to motor execution

Check for updates

Ricardo J. Alejandro, Iris Ikink, Emmanouela Foinikianaki & Clay B. Holroyd

Mental planning is essential for producing action sequences. Despite the contribution of planning to many everyday activities, the neurocognitive processes that map a selected plan into a concrete course of action are unclear. We asked whether planning and execution are linked by abstract task relationships that are independent of the specific perceptual and motor details of action implementation. Human participants underwent functional MRI while performing a task that required encoding and maintaining abstract representations in order to plan and later apply those representations to execute action sequences. Our analyses revealed that dorsomedial prefrontal cortex (DMPFC) encoded representations during the planning period that shifted progressively from information based on perceptual input to information associated with the abstract task context. These abstract representations were then used to translate between the action plan and its execution irrespective of the perceptual and motor details, indicating that DMPFC builds and sustains an abstract representational bridge that links planning to action.

Mental planning is a powerful neurocognitive mechanism that enables flexible production of action sequences. For instance, imagine a driver passing a traffic sign that indicates that a roundabout is close ahead, requiring them to examine the symbols on the sign quickly in order to determine and plan the appropriate exit (Fig. 1A, left). This complex planning process entails retrieving task-relevant information from memory, associating these concepts within the current task context, simulating possible behaviors, and predicting and evaluating their outcomes in terms of the task goals and subgoals, all of which often happen without the benefit of external guidance or cues (for example, because the traffic sign has already been passed). Moreover, once a plan is selected (such as taking the fifth exit; Fig. 1A, left), this information needs to be maintained until the actual roundabout is reached and the planned sequence of actions executed. The implementation of the plan requires mapping its abstract goal into a concrete course of action, all the while monitoring for performance errors and unforeseen constraints (such as encountering a roundabout with an unusual shape; Fig. 1A, right). Yet although planning is essential for many everyday behaviors[1], the cognitive and neural operations that link and translate planning to execution are largely unknown.

Several neuroimaging studies have examined neural activity occurring during a delay or planning period between an instruction cue and its designated response[2–5]. However, these experimental paradigms have mostly related the planning phase to a prospective motor action such as a button press, hand movement, or eye saccade. Consequently, many of these studies reported a mixture of signals related to cognitive control, action and motor preparation that are related but not specific to planning, across a distributed network of brain regions including dorsolateral-, ventrolateral-, and dorsomedial prefrontal cortex (DLPFC, VLPFC, DMPFC, respectively), anterior cingulate cortex (ACC), posterior parietal cortex (PPC), basal ganglia, cerebellum, insular cortex, hippocampus (HPC) and motor and pre-motor regions[6–21]. Conversely, the execution period in these studies often reveals a combination of brain areas involved in motor and cognitive control, including DLPFC, DMPFC, PPC, cerebellum, HPC, and basal ganglia (BG)[22–29].

The wide variety of brain regions highlights the fact that multiple cognitive processes related to perception, cognitive control, memory, and motor control work in parallel both when planning and implementing planned behaviors as specific actions[6]. This ubiquity of signals complicates identifying the specific contribution of each region to each of these processes[1]. Moreover, considering their reciprocal relationship, it has been noted that planning and execution are not truly separate processes[30–32]. Rather, it has been proposed that planning evokes a latent, coarse representation of the action plan that is maintained in an active state until the actions are completed, in order to monitor the execution of the plan and intervene as needed[33]. On this view, this representation would intermediate between developing a plan and implementing it, utilizing a language that is common to both task phases and that supersedes their differences. Such a high-level, abstract representation would provide the flexibility needed to map internal models to concrete courses of action while adapting the execution of the plan to external perturbations and behavioral errors (Fig. 1A).

Extending these ideas, we suggest that this superordinate process, which encodes task-relevant information at an abstract level as a bridge

Department of Experimental Psychology, Ghent University, Ghent, Belgium. e-mail: rjalejandrob@gmail.com

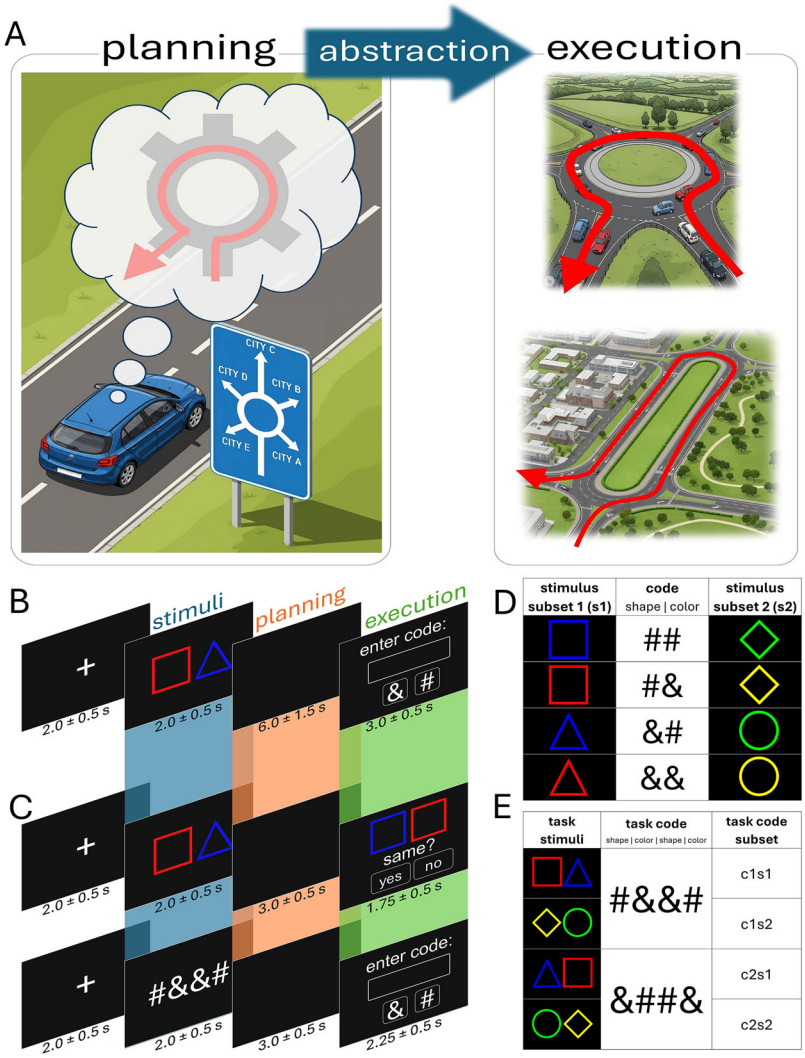


Fig. 1 | Study conceptualization and experimental paradigm(s). A Illustration of the abstract link between planning and execution. The goal of a driver is to go to City E; therefore when encountering the relevant traffic sign, they determine taking the fifth exit as their plan of action. This information about prospective action must be maintained until the plan is executed because the sign has been passed and will no longer be available as a cue. Once the roundabout is reached, its physical layout might be as expected (execution, top) or unusual (execution, bottom); in both cases the driver must map their plan of action from memory to the specifics of the current environment. Hence, the plan must be encoded in a flexible format that is independent of the specific details of its execution while still directing such execution. **B** Main task paradigm. On each trial a set of visual symbols ("stimuli", blue shading) instructed participants to map the stimuli to the corresponding codes (predefined sequences of the characters "#" and "&") during a planning phase ("planning", orange shading) and subsequently type the corresponding sequence into a password prompt ("execution", green shading). **C** Following the main task, participants completed two functional localizer tasks (shape and code localizer). Top: functional "shape" localizer. Participants were presented with colored geometric shapes (stimuli, blue shade; same combinations as in the main task), which were instructed to be remembered during a subsequent planning phase (planning, orange shade), after which another set of colored geometric shapes were presented that either were or were not the same as the initial set during the trial, together with the question "same?" (execution, green shade). Participants were required to respond by selecting "yes" to indicate that the second set was exactly the same across all stimulus dimensions (i.e., same shapes, same colors, and same positions) as the initial set, or "no" to indicate that they were not the same in one or more of the stimulus dimensions. This localizer was designed to capture perceptual processing elements related to the main task (see Methods). Bottom: functional "code" localizer. Participants were shown an abstract code (stimuli, blue shade; same combinations of "&" and "#" as in the main task) that they were required to remember during a planning phase (planning, orange shade), after which they were prompted with a screen to execute the abstract code (execution, green shade). This localizer was designed to capture abstract processing elements related to the main task (see Methods). **D** Visual stimuli and their corresponding abstract codes. Each geometric shape was represented by a 2-character code that was mapped to its shape (first character) and color (second character) following specific constraints. Squares and diamonds corresponded to "#", and triangles and circles corresponded to "&". Squares and triangles could be blue ("#") or red ("&"), and circles and diamonds could be green ("#") or yellow ("&"). **E** On each trial, participants were required to execute 2-stimulus sequences of 4 button presses. To increase the amount of observations per condition, the majority of codes presented during the tasks (~82% during the main task, ~88% during the localizer tasks) were c1: "# & & #" and c2: "& # # &" (standard trials, see Methods), which included two 2 shape-pairs per code (c1 represented by c1s1: "red square + blue triangle" or c1s2: "yellow diamond + green circle", c2 represented by c2s1: "blue triangle + red square" or c2s2: "green circle + yellow diamond"). Our analyses focused on these "standard" trials only. During the tasks, standard trials were intermixed with catch trials that were constructed from all of the remaining valid combinations of stimuli, comprising ~18% of the main task trials and ~12% of the localizer tasks trials (see Methods and Supplementary Table S1). The images in panel (**A**) were generated with Google Gemini and subsequently edited by the authors.

between plans and actions, does so specifically by harnessing learned representations that generalize over extensive repertoires of tasks and task requirements (i.e., goals and constraints). Notably, the behaviors that compose a task can be organized into hierarchical structures that facilitate goal-directed behavior[34–38], such that higher-order task goals and concepts modulate subordinate subgoals and explicit actions[39]. Such hierarchical representations allow for existing knowledge and task properties to be generalized to new tasks (or new instances of known tasks)[40–43]. For instance, when learning to play the piano, a skilled guitarist does not relearn all of musical theory from scratch, but instead utilizes task-related abstractions such as notes and musical chords to integrate the commonalities across instruments[44–46]. These higher-order representations enable mappings between action sequences across different domains of feature perception and motor execution (Fig. 1A). Consistent with this possibility, neuroimaging evidence has revealed that hierarchically-organized plans are represented in the brain at different levels of abstraction in orbitofrontal cortex (OFC), ACC, DLPFC, DMPFC, and HPC[47–50]. However, whether these areas link plans to action execution utilizing such a hierarchical encoding scheme is not clear.

Here, we identify brain areas that encode hierarchical representations for abstract planning and action execution. While undergoing functional magnetic resonance imaging (fMRI), participants were required to encode and maintain abstract representations of an upcoming action sequence and later apply those representations to execute said sequences. Crucially, this abstract context was independent of the visual stimuli and the motor actions used in the task. Following a pre-registered protocol, we trained a support vector machine (SVM) classifier –a machine learning algorithm designed to predict new data from training observations– on the multivariate fMRI activity during the planning period to discriminate different plans for sequence implementation, and then tested the classifier on the activity during the execution period. A searchlight application of this approach identified brain areas sensitive to the high-level task context that generalized over the low-level implementational details of the task. Additionally, to identify brain areas supporting abstract vs. perceptual aspects of planning, we assessed the representational similarity of brain activations with respect to two possible representational formats: abstract representations that relate different stimuli according to the task context, and perceptual representations that relate stimuli according to their shared visual features. These analyses converged in identifying a cluster of activity overlapping DMPFC and dorsal ACC (hereafter, simply DMPFC), which suggests that this brain region maintains abstract, higher-order task representations during planning that are then used to implement the plans during their execution. In contrast, low-level visual representations related to stimulus identity were identified in visual cortex, as expected. Lastly, an exploratory analysis revealed that as the planning interval progressed, the representation during the planning period shifted from a perceptually-based format to a more abstract format.

Results

Experimental task

Fifty participants took part in the experiment, which consisted of a main task (Fig. 1B) and two functional localizer tasks (Fig. 1C), while they underwent fMRI scanning. Participants were trained online beforehand during a separate session (see Methods) on task rules that required mapping colored geometric shapes into specific codes (sequence combinations of the characters "#" and "&"). On each trial of the fMRI main task (Fig. 1B), the participants were shown a set of two colored geometric shapes (red and blue squares and triangles, and green and yellow diamonds and circles; Fig. 1D) during a stimulus phase (blue highlight in Fig. 1B), followed by a planning phase (orange highlight in Fig. 1B) that required the participants to internally translate the observed sequence of colored shape stimuli into the corresponding sequential code (Fig. 1E). After the planning phase, participants were shown an input field, images of two buttons labeled with the characters "#" and "&" (the relative positions of which varied at random from trial to trial), and the instruction "enter code" (execution phase, green highlight in Fig. 1B). They were then required to type the corresponding (planned) 4-character code sequence as instructed on that trial (Fig. 1B, D).

Behavioral results

RT signature of chunking and benefit of planning. Because each trial required executing a four-step sequence of button presses according to each shape's two-button sequence, we predicted that participants would encode this information hierarchically by consolidating the four-step code into two independent two-step subsequence codes (or "chunks"[51,52]). To test this prediction, reaction times (RTs) to the four individual actions were compared both within-subsequence, where we predicted that the first action within each subsequence would be slower than the second (RT1 > RT2, and RT3 > RT4), and between-subsequences, where we predicted that the transition between subsequences would result in the third action to be slower than the second action (RT3 > RT2)[53,54]. Additionally, behavioral evidence of planning can be seen in shorter RTs after longer planning periods[55,56]. We therefore predicted that the first step of the entire sequence (RT1) would be faster with increasing time for planning (namely, for planning periods of 4.5 s, 6 s, and 7.5 s; see Fig. 1B and Methods).

We fitted RT data with a linear mixed-effects model (LMM) that included sequence step (1, 2, 3, 4), and planning duration (short, medium, long, given by the jittered duration of the planning phase; see Methods) as fixed effects, as well as their interactions, and random intercepts per participant. The linear mixed-effects model revealed significant main effects of sequence step ($F(3, 32907) = 24256.56$, $p < 0.0001$, $\eta_p^2 = 0.69$) and planning duration ($F(2, 32908) = 37.64$, $p < 0.0001$, $\eta_p^2 = 0.002$), and a significant interaction between sequence step and planning duration ($F(6, 32907) = 55.69$, $p < 0.0001$, $\eta_p^2 = 0.01$). Subsequent post-hoc pairwise comparisons (using Tukey's HSD) estimated the direction of the main effects and their interaction. As shown in Fig. 2A, RT1 was significantly higher than RT2 ($p < 0.0001$), and RT3 was significantly higher than RT4 ($p < 0.0001$). However, the difference between RT2 and RT3 was not statistically significant ($p = 0.258$). We supposed that the lack of the expected effect of hierarchical chunking (i.e., a slow-down when transitioning subsequences) could have resulted from the participants' repeated exposure to the task both during their online behavioral training (see Methods) and during the fMRI session itself, thereby attenuating the effect. In fact, an analysis of the RT data from the online behavioral training session revealed the predicted signature of hierarchical chunking (RT2 < RT3; $p < 0.0001$) (Supplementary Fig. S1), in addition to replicating the significant effects observed in the main fMRI task (RT1 > RT2, $p < 0.0001$); and RT3 > RT4, $p < 0.0001$).

Further, the main effect of plan duration was driven by significant differences between long and medium (7.5 vs 6 s, $p = 0.0008$), long and short (7.5 vs 4.5 s, $p < 0.0001$), and medium and short (5 vs 4.5 s, $p < 0.0001$) planning durations, indicating that the speed of sequence execution was influenced by the length of preparation time. Regarding the interaction between sequence step and planning duration, the main effect of planning duration on RT was only significant for RT1 when contrasted against all planning durations (long vs medium, $p < 0.0001$; long vs short, $p < 0.0001$; medium vs short, $p < 0.0001$). All pairwise differences for RT2, RT3, and RT4 were not statistically significant (Supplementary Fig. S2). This result confirmed that participants used the planning period to prepare their responses, beginning sequence execution more quickly when provided with longer periods to plan their response.

Average accuracy across participants (Fig. 2B) was high (M = 0.94; SD = 0.08), with no significant differences across blocks (as revealed by a Kruskal–Wallis test: $\chi^2(3) = 3.99$, $p = 0.262$), as was expected due to their extensive training (see Methods).

Neuroimaging results

DMPFC, DLPFC, and basal ganglia encode abstract information during planning that is used during execution. We sought to identify brain areas that mediated between superordinate, abstract sequential codes during planning and execution. To this end, following our pre-registered

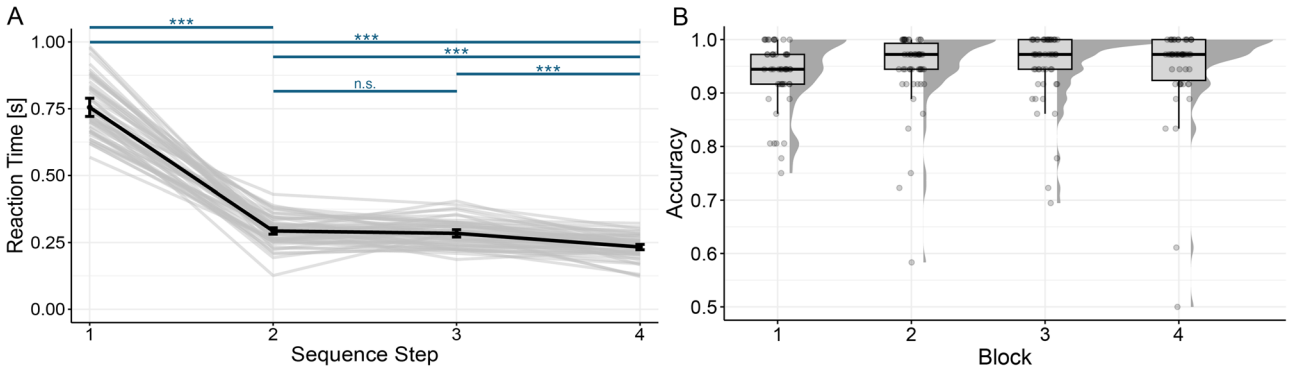


Fig. 2 | Behavioral results. A Reaction time (RT) per sequence step and post-hoc pairwise comparisons following the main effect of sequence step on RT. Bold black line represents the average RT across participants, gray lines represent individual participants. Asterisks denote statistical significance at $p < 0.0001$; n.s.: non-significant. **B** Task accuracy per block; there were no significant differences across blocks ($\chi^2(3) = 3.99$, $p = 0.262$). Error bars denote the standard error of the mean (SEM) across participants.

analysis plan we applied Multi-Voxel Pattern Analysis (MVPA)[57] to conduct a searchlight cross-decoding analysis on the multivariate fMRI data (see Supplementary Fig. S4 for results in pre-determined ROIs). The searchlight technique systematically inspects a spherical subset (i.e., a "searchlight") of data from the entire brain volume that is iteratively centered on every voxel in the volume. The voxel values within each searchlight are used by a classifier to discriminate between the conditions of interest, assigning the classification accuracy score to the voxel at the center of the searchlight. This process is repeated for every voxel location within the volume, resulting in a decoding accuracy map of the entire brain[58]. During decoding, part of the data serves to train the classifier, and the rest is used for testing, which gives an estimate of the predictability of new data points based on the training set. Here, we operationalized "abstraction" as the generalization performance of a multivariate decoder on conditions not used for training[59]. We trained the classifier to discriminate between abstract representations during the planning phase and tested it during the execution phase to predict the activity of the same representations. We hypothesized that if this abstract context connects both task phases (planning and execution), then the (abstract) information shared across phases would enable prediction of one phase from the other, regardless of the low-level (non-abstract) components of the action plan (i.e., the visual stimuli and motor responses).

Specifically, we trained a support vector machine (SVM) on the brain activity patterns (i.e., the beta value estimates, see Methods) during the planning phase of one of the stimulus subsets of each code (e.g., planning c1s1 and c2s1; Fig. 3A), and used it to classify activity patterns during execution of the left-out stimuli of each code (e.g., executing c1s2 and c2s2; Fig. 3A). Importantly, the task design and analysis incorporated multiple features that controlled for potential stimulus, motor, and collinearity confounds. First, the design controlled for stimulus confounds by 1) showing only a blank screen during the planning period, and 2) training the classifier on activity patterns of one set of visual stimuli (e.g., red square and blue triangle) and testing the classifier on the activity patterns of different visual stimuli that mapped to the same code (e.g., yellow diamond and green circle; Fig. 3A). Second, the design controlled for motor-related confounds by providing participants with the button-to-screen mappings (i.e., whether the left or the right button corresponded to the "#" and "&" symbols) only during the execution phase of the trial, which varied randomly from trial-to-trial. Third, the design incorporated several features that controlled for potential collinearity confounds between the training and testing data. In particular, because the training and testing occurred both on different trials and on different conditions (see above, and Methods), the hemodynamic response was temporally dissociated across adjacent events[60,61]; as a consequence, the hemodynamic response to planning could not contaminate the neural representations during execution. Finally, the jittered duration of the planning phase (4.5, 6, or 7.5 s, see Methods) additionally ensured that the temporally-adjacent planning and execution regressors were sufficiently decorrelated.

All together, this MVPA procedure trained and tested the data for different conditions (different visual stimuli belonging to the same abstract context), different trial phases (associated with planning versus execution), different trials (because the stimulus for training occurred on a different trial than the stimulus for testing), and different functional runs (via leave-one-run-out cross-validation). To confirm that these measures were effective at minimizing collinearity between our measures of the planning and execution phase, we assessed two parameters that test for collinearity: pairwise regressor correlations[62,63] and the variance inflation factor[64,65]. We found that the pairwise regressor correlations were low across the training and testing conditions (mean = −0.07, SD = 0.017; Supplementary Fig. S3, left), indicating that the hemodynamic responses associated with each condition were statistically separable. We further found that the variance inflation factor (VIF[62,64], see Methods) was minimal for all regressors of interest across all participants (mean = 1.14, SD = 0.03; Supplementary Fig. S3, right), indicating that the regressors of interest were free of dependences from each other. In sum, these experimental and statistical controls ensured that our findings reflect high-level task representations rather than sensorimotor overlap or temporal dependency of the hemodynamic response.

This searchlight cross-decoding analysis yielded no significant clusters after family-wise error (FWE) rate multiple comparison correction ($p_{FWE} = 0.05$). However, an exploratory analysis using a less restrictive, uncorrected threshold of ($p_{unc} < 0.001$) identified three clusters of activity in which the representations formed during planning could be decoded at an above chance level based on the representations applied during execution (Fig. 3B). One cluster encompassed right DMPFC (medial superior frontal gyrus, extending into medial frontal gyrus; Fig. 3B, left; peak at x = 10; y = 46; z = 40), another cluster was located in left DLPFC (middle frontal gyrus, extending into inferior frontal gyrus pars triangularis; Fig. 3B, middle; peak at x = -48; y = 30; z = 30) and the third cluster centered on a region of the BG (left lenticular nucleus: posterior putamen and pallidum; Fig. 3B, right; peak at x = −26; y = −14; z = −4).

Abstract decoding differs from perceptual decoding. To verify that these findings did not simply result from the visual similarities across different stimuli or codes, we conducted a control analysis (not pre-registered) by training an SVM classifier on the planning-related activations of the stimuli that mapped to the same abstract code (e.g., planning c1s1 and c1s2; Fig. 3C) and tested it on the execution-related brain activity of the stimuli that mapped to the other code (e.g., executing c2s1 and c2s2; Fig. 3C). Because each pair of stimuli denoted both codes, and thus shared a high degree of visual similarity, we hypothesized that application of the searchlight SVM cross-decoder would predominantly reveal activity clusters near visual areas. As expected, this analysis identified one significant cluster located in bilateral primary visual cortex (VC, calcarine fissure and surrounding cortex; Fig. 3D, peak at x = 4; y = −76; z = 10).

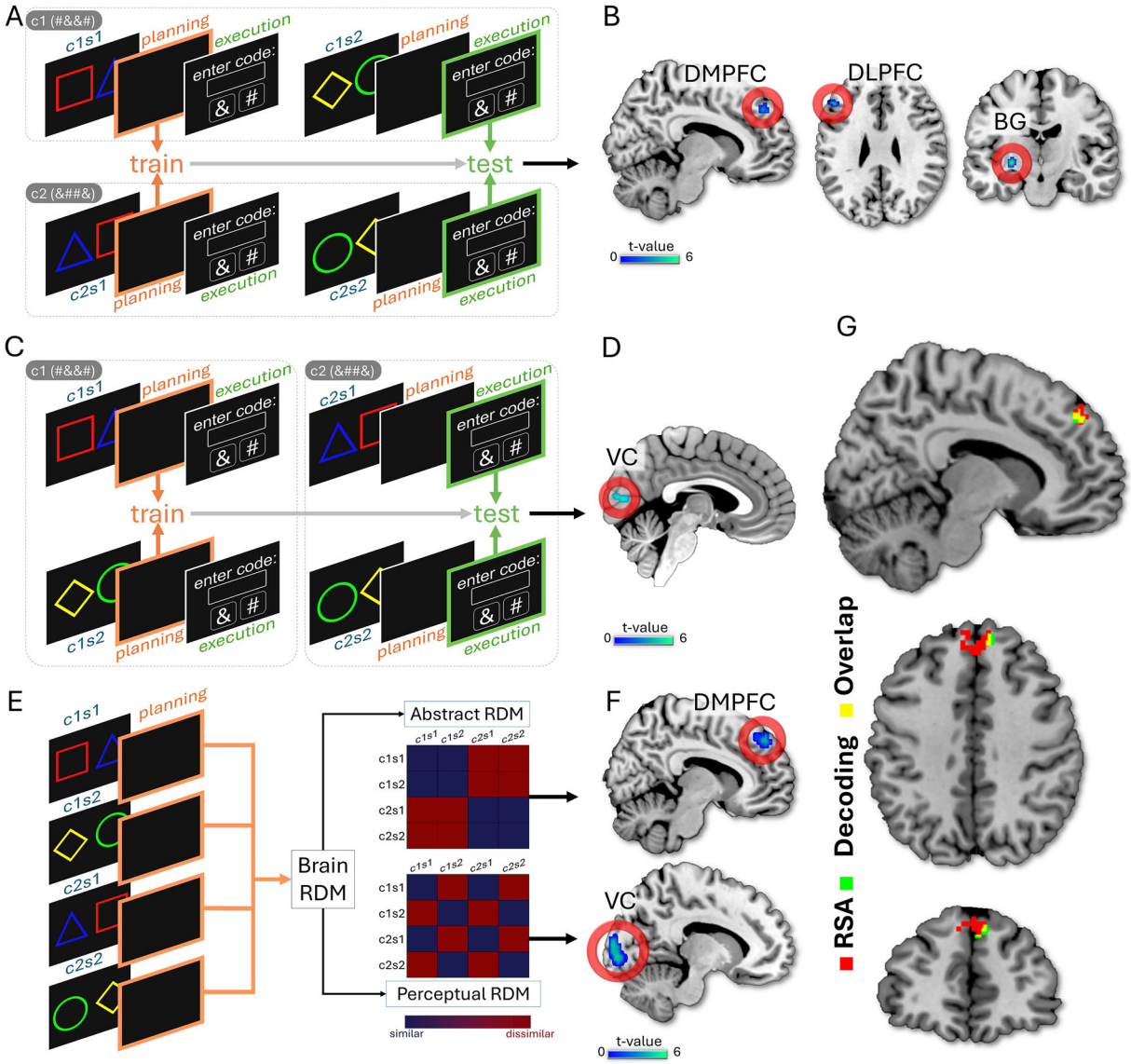


Fig. 3 | MVPA procedures and results. A Pre-registered searchlight cross-decoding analysis. We trained a SVM on the brain activity patterns during the planning phase of one of the stimulus subsets of each code (e.g., planning of c1s1 and c2s1, orange highlights), and tested this classifier on brain activity patterns during execution of the left-out stimulus subsets of each code (e.g., execution of c1s2 and c2s2, green highlights; note that the position of the buttons for "&" and "#" were pseudo-randomized across trials). Training and testing took place on data from different visual stimuli (which mapped to the same abstract context), on different trial phases (planning versus execution), on different trials (stimuli for training appeared on different trials than stimuli for testing), and on different functional runs (via leave-one-run-out cross-validation). **B** Searchlight cross-decoding results. Three clusters were identified where execution activity was decoded from planning activity at above-chance level (p_{unc} <0.001): right DMPFC (medial superior frontal gyrus, extending into medial frontal gyrus; peak at x = 10; y = 46; z = 40), left DLPFC (middle frontal gyrus, extending into inferior frontal gyrus pars triangularis; peak at x = −48; y = 30; z = 30), and BG (left lenticular nucleus: posterior putamen and pallidum, peak at x = −26; y = −14; z = −4). **C** Control searchlight cross-decoding analysis. An SVM was trained on the brain activity patterns during the planning phase of the visual sets of each code (e.g., planning of c1s1 and c1s2, orange highlights), and tested on brain activity patterns during execution of the visual sets of the other code (e.g., execution of c2s1 and c2s2, green highlights). Training and testing took place on data from different codes (which shared high visual similarities), on different trial phases (planning versus execution), on different trials (stimuli for training appeared on different trials than stimuli for testing), and on different functional runs (via leave-one-run-out cross-validation). **D** Control searchlight cross-decoding results. One cluster was identified where execution was decoded from planning at above-chance level (p_{unc} <0.001): bilateral primary VC (calcarine fissure and surrounding cortex; peak at x = 4; y = −76; z = 10). **E** Pre-registered searchlight representational similarity analysis (RSA). On brain activations only from the planning phase, we contrasted each participant's "brain-RDM" (all task conditions correlated against each other: c1s1 VS c1s2 VS c2s1 VS c2s2) against two candidate "hypothesis-RDMs": an abstract RDM was designed to capture abstract representations (representational distances depended on whether stimuli belong to within- or between code contexts), and a perceptual RDM which was designed to depict low-level visually-driven representations (distances depended on how visually similar the stimuli are). **F** Searchlight RSA results. Top: One cluster was identified with significant similarity (p_{FWE} = 0.014) between the brain RDM and the abstract RDM, located in bilateral DMPFC (medial superior frontal gyrus, extending into medial frontal gyrus; peak at x = 3; y = 38; z = 44). Bottom: One cluster was identified with significant similarity (p_{FWE} <0.001) between the brain RDM and the perceptual RDM, located in left occipito-VC (calcarine fissure, occipital gyri, lingual gyrus, and fusiform gyrus; peak at x = −18; y = −91; z = 1). **G** Overlay of the significant clusters found by the two multivariate approaches. The same DMPFC region was found significantly active by cross-decoding planning versus execution (neural data from different trial phases) and by comparing the representational similarity during planning (neural data only from the planning phase) against hypothesis-RDMs. SVM support vector machine, DMPFC dorsomedial prefrontal cortex, DLPFC dorsolateral prefrontal cortex, BG basal ganglia, VC visual cortex.

DMPFC represents abstract information during planning. Via the previous multivariate decoding analysis we tested whether abstract representations of plans formed during plan preparation could be subsequently decoded from the same brain areas during the execution of those plans. Next, we investigated the abstract representational profile expressed during planning that links plans to actions, independently of the actual execution of those plans. An abstracted plan could be expressed either as a representation that generalizes across different stimuli (such that different explicit action sequences are represented by the same pattern of activations), or as individual representations that map to each action sequence. Additionally, if the planning information is based on the perceptual features (instead of on the abstract links) of the task elements, then the task concepts should relate to each other by their visual similarities. Here we focused on the abstraction that subserves generalization. We hypothesized that if the planning information is indeed represented in an abstract format, then the task concepts should generalize across their higher-level relationships.

To test these predictions, we performed a pre-registered searchlight representational similarity analysis (RSA) on brain activations recorded during the planning phase (Fig. 3E). We constructed two hypothesis-driven RDMs that contrasted all the task conditions (standard trials) against each other (c1s1 VS c1s2 VS c2s1 VS c2s2): an Abstract RDM that tested for high-level, abstract relationships between plans (where the distances depended on whether the stimuli mapped to the same or different code contexts), and a Perceptual RDM that tested for low-level, visually-driven relationships between plans (where the distances depended on the visual similarity of the stimuli) (Fig. 3E). We expected this hypothesis-driven RSA-searchlight to identify clusters that overlapped with the areas identified from the decoding-searchlight analysis.

The searchlight-RSA analysis yielded a single significant ($p_{FWE} = 0.014$) cluster representing abstract planning information in bilateral DMPFC (medial superior frontal gyrus, extending into medial frontal gyrus; Fig. 3F, top; peak at x = 3; y = 38; z = 44). This cluster highly overlapped the DMPFC cluster identified by the decoding analysis (Fig. 3G). Likewise, the analysis yielded a single significant ($p_{FWE} < 0.001$) cluster representing perceptual information in left occipito-VC (calcarine fissure, occipital gyri, lingual gyrus, and fusiform gyrus; Fig. 3F, bottom; peak at x = −18; y = −91; z = 1), which also overlapped with the VC cluster identified by the decoding analysis.

Shifts of representational formats during planning. Next, we investigated the temporal evolution of the abstract representations during the planning period. Toward this end, we employed two functional localizer tasks that were performed after the main fMRI task during the same functional imaging session (see Methods). The "shape" localizer (Fig. 1C, top; Fig. 4A, top) captured variance in the main task explained by visual processing and visual memory (such as mentally recalling the visual shapes during the planning phase). By contrast, the "code" localizer (Fig. 1C, bottom; Fig. 4A, bottom) captured variance in the main task explained by abstract processing alone (i.e., the abstract codes independent of the properties of their stimulus representations). We expected

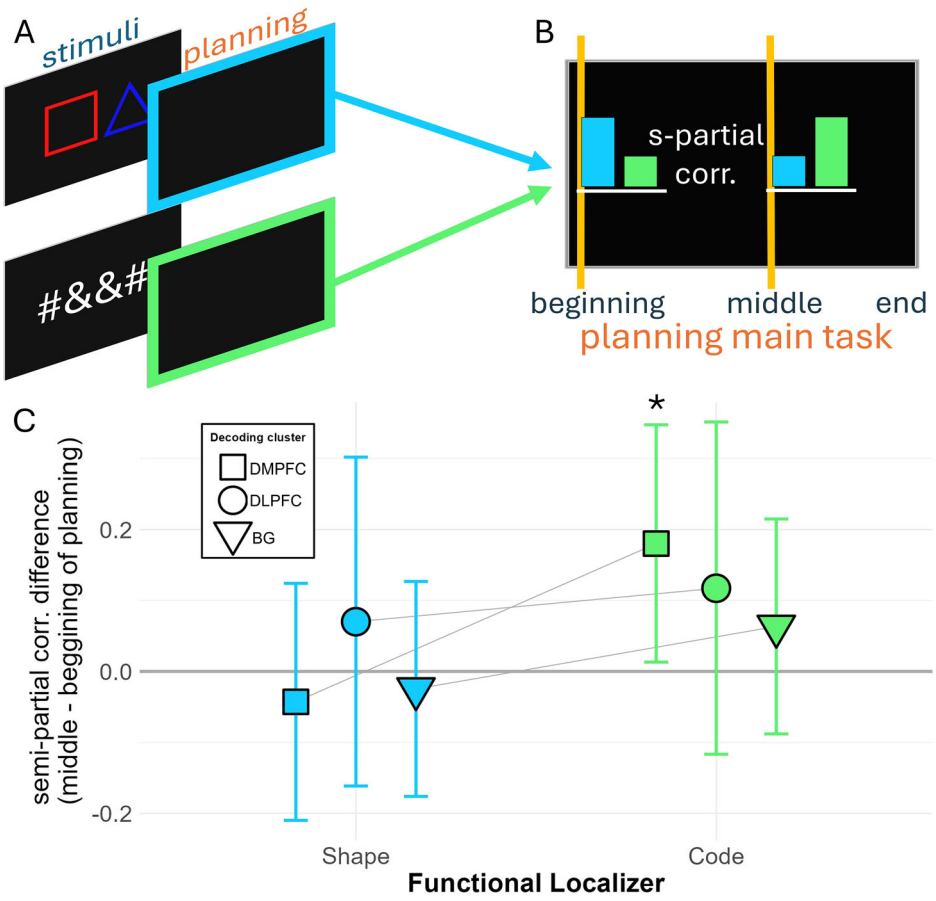


Fig. 4 | Functional localizer tasks and results. **A** Illustration of the semi-partial correlation analysis between the planning phase of the main fMRI task and the planning phase of each of the functional localizer tasks. We correlated the patterns of activity at the beginning and middle time points during the main task with the activity patterns during the planning phase of the shape localizer (blue frame, intended to capture the variance driven by visual processing and visual memory alone) and the code localizer (green frame, intended to capture the variance driven by abstract processing alone). **B** Predicted planning representational dynamics during the main task. We hypothesized that the beginning planning phase would be more correlated with the shape localizer, while the middle planning phase would be more correlated with the code localizer (see main text). **C** Semi-partial correlations of cluster activity with functional localizers. The x-axis indicates the representational format (shape/perceptual vs code/abstract) during the planning period. The y-axis indicates the difference between semi-partial correlations at the middle and the beginning of the planning phase, with positive or negative numbers signaling an increase or decrease in the respective representational format (shape/perceptual vs code/abstract), respectively, from the beginning to the middle of the planning phase. Activity in the DMPFC cluster (multivariate cross-decoding analysis, see Results) was significantly more correlated with the code localizer (green square) at the middle of the planning phase, indicating an enhanced abstract representational format late in the planning period, in contrast to the correlation with the shape localizer, which did not change during the planning phase. Activity in the other brain areas (DLPFC and BG) did not change significantly during planning for either of the representational formats. Asterisk denotes statistical significance at $p < 0.05$.

stronger perceptual contributions to neural activity at the beginning of the planning phase (due to visual reactivation of stimuli), resulting in greater correspondence with the shape localizer. By contrast, we expected stronger abstract contributions to neural activity at the end of the planning phase (following the translation of the instructive cues into their respective abstract codes), resulting in greater correspondence with the code localizer (Fig. 4B).

To do so, we inspected the differential contribution of the perceptual vs abstract representations at different points in time (beginning and middle) during the planning phase in the clusters obtained during the cross-decoding analysis (DMPFC, DLPFC, and BG; Fig. 3B). As preregistered, semi-partial correlations were estimated between the patterns of activity during planning in the main task, and the patterns of activity during each of the localizer tasks (Fig. 4A, B), for each of these clusters. We then tested effects of time point (beginning, middle) and localizer (shape, code) on the semi-partial correlation estimates, separately for each cluster using a two-way repeated-measures linear mixed-effects model with time point and localizer as fixed effects, as well as their interaction, and a random intercept for subject. For inference, Type III F-tests using Satterthwaite degrees of freedom were employed. This analysis returned a significant interaction between time point and localizer in DMPFC ($F(1,147) = 5.98$, $p = 0.015$, $\eta_p^2 = 0.03$). Pairwise post-hoc comparisons showed that the interaction was driven by a significant difference between the middle and beginning time points ($p = 0.028$) for the code localizer in DMPFC but not for the shape localizer (Fig. 4C). All of the other effects across all three clusters were not statistically significant (Fig. 4C).

Multi-dimensional scaling of planning representations. To visualize and compare the representational geometry associated with planning conditions across clusters of interest (DMPFC, DLPFC, BG), we applied multidimensional scaling (MDS[66]) to embed the planning-related RDMs (see above) into a low-dimensional space (Fig. 5; analysis not preregistered).

The MDS projections reveal that DMPFC (Fig. 5, left column) represents the conditions using a low-dimensional, abstract code that discriminates between the plans (c1 and c2) while generalizing over the stimuli (s1 and s2), with an excellent fit (e.g., stress = 1.5e−16) when the projection is constrained to two dimensions. By contrast, representations in the BG (Fig. 5, right column) do not discriminate between the plans (c1 and c2) in a two-dimensional space despite a similarly excellent fit (stress = 1.1e−16). Lastly, representations in DLPFC (Fig. 5, center column) are intermediate, fitting very well with only a single dimension (stress = 3.7e−7) despite not discriminating between conditions and requiring additional dimensions to generalize over the plans.

Discussion

Mental planning has been conceptualized as an iterative process of response selection and outcome prediction for goal-directed search, which is used to prepare or simulate upcoming sequential behaviors before their actual execution[67]. This process is understood to utilize internal representations (or models) derived from prior task experience and environmental constraints[68]. However, the variety of potential behaviors available for simulation often renders this process computationally intractable[69]. One strategy to make the process more efficient is to generalize over related sequences of actions in order to plan at a more abstract, hierarchically higher level, thereby simplifying the problem domain[70,71] while supporting the actual implementation of the plans. To investigate this possibility, we recorded brain activity of participants as they performed a hierarchical planning task expressly designed to test these predictions. On each trial, participants were required to map visual stimuli to abstract codes, maintain that abstract information throughout a delay period, and then implement the corresponding sequence of responses in an execution period. Behavioral analyses revealed that participants effectively used the planning period to prepare their upcoming responses (Supplementary Fig. S2). Further, participants leveraged the task structure during planning by encoding the task rules hierarchically (Supplementary Fig. S1), though this latter effect was attenuated with repeated exposure to the task (Fig. 2A).

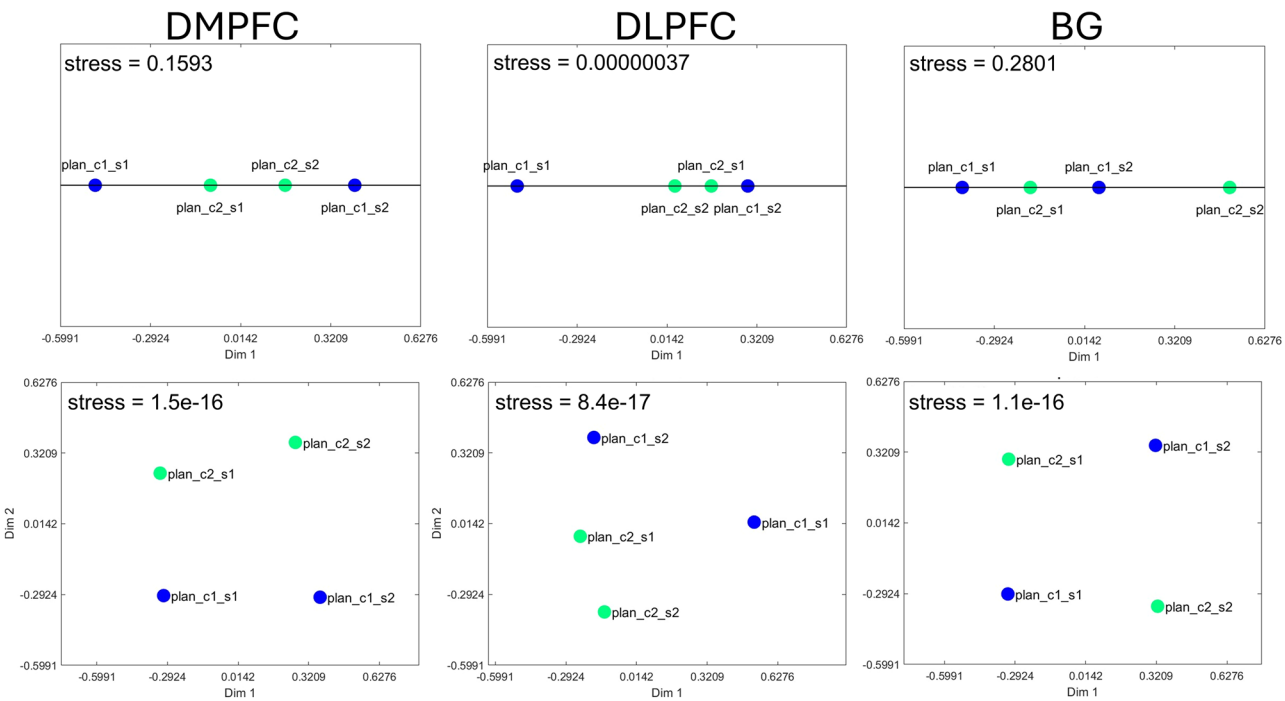


Fig. 5 | Multi-dimensional scaling (MDS) of the planning representational space. The DMPFC (left column) represents plans of upcoming actions with a two-dimensional code that distinguishes between the plans (c1 and c2) while generalizing over the stimuli (s1 and s2). By contrast, the BG (right column) is unable to separate the two plans (c1 and c2) in only two dimensions, whereas the DLPFC (middle column) is intermediate. These results suggest that all three brain areas encode plans that can be decoded during their execution, but DMPFC maintains the planning information in a relatively low-dimensional, abstract representational format.

Neuroimaging analyses revealed that the hypothesized abstract context was encoded by DMPFC (medial superior frontal gyrus, extending into medial frontal gyrus; Fig. 3B, left), DLPFC (middle frontal gyrus, extending into inferior frontal gyrus pars triangularis; Fig. 3B, center), and BG (left lenticular nucleus: posterior putamen and pallidum; Fig. 3B, right). Specifically, an MVPA classifier was trained to discriminate between fMRI data related to different abstract codes that mapped to the same stimuli during the planning phase. That classifier was then tested on fMRI data collected during the execution phase (Fig. 3A). This analysis revealed that these areas instantiated the same patterns of activity when planning and executing the behavioral sequences. These correspondences allowed for the identity of each hierarchical sequence to be predicted from the neural activity patterns occurring during its execution, based on the patterns of activity that occurred when planning an equivalent action sequence from a different, symbolic code. Hence these representations aligned according to their higher-order contexts (as specified by the higher-order relationships between the different stimuli) rather than to the stimulus and action specifics of task execution. Critically, these results were not due to the visual similarity of stimuli corresponding to the abstract codes, as confirmed by a control analysis that trained the classifier to discriminate between different stimuli that mapped to the same abstract code (Fig. 3C), which revealed activity in VC (Fig. 3D). Importantly, this control analysis also argues against hemodynamic carryover from planning into execution, as it did not reproduce the main decoding result associated with planning (namely, a cluster in DMPFC). These abstract representations were also unrelated to the implementational details of plan execution, because the current experimental design prevented participants from determining the specifics of sequence execution (i.e., code-to-button mappings) until the execution phase (see Methods). Furthermore, DMPFC in particular encoded planning information in an abstract format, as revealed by searchlight-RSA based on an RDM that grouped the planning phase data according to their higher-order similarity (Fig. 3F, top), whereas a searchlight-RSA based on an RDM that grouped the planning phase data according to the stimuli's visual similarity revealed activity in VC, as expected (Fig. 3F, bottom). Finally, a follow-up analysis employing functional localizers revealed that the abstract representations in DMPFC increased in strength as the delay period progressed, as would be expected of a region involved in planning the response in a higher-order space (Fig. 4C).

Despite the involvement of all of these neural systems in planning, our results point to a crucial distinction between the functions of DMPFC vs. BG and DLPFC. In particular, DMPFC was the only brain region that was sensitive both to the correspondence between planning and execution and to abstract processing (Fig. 3G). This property suggests that DMPFC encodes an abstract representation of a forthcoming behavioral sequence that is maintained until used, namely, during the execution of the sequence. By contrast, DLPFC and BG were identified by the searchlight cross-decoding analysis but not by the searchlight-RSA, which suggests that although these areas worked in tandem with the DMPFC to plan and produce the behavioral sequences, that information was most efficiently encoded using a low-dimensional, abstract representation in DMPFC. To gain insight into this relationship we applied multidimensional scaling (MDS[66]) to the planning representations (Fig. 5; not pre-registered). Importantly, this analysis revealed that DMPFC (Fig. 5, left column) clearly distinguished between the upcoming plans (c1 and c2) across stimuli using as few as two dimensions (Fig. 5, top left), which suggests that this region compresses planning information into a compact, low-dimensional format. By contrast, the BG (Fig. 5, right column) requires more than two dimensions to separate the planning representations. The DLPFC (Fig. 5, center column) failed to discriminate between the plans despite describing the data well along only a single dimension (i.e., with about half a million times less stress compared to DMPFC and BG), which suggests a representational format that is less aligned with abstract plan identity. This distinction suggests complementary functions: whereas DLPFC and BG may be more concerned with the implementational details of the plans by relying on specialized representations that facilitate decoding[38], the DMPFC appears to generalize across plans using a low-dimensional, abstract contextual representation that facilitates task maintenance during planning (Fig. 4C) and execution (Fig. 3G).

These results are supported by previous work that has highlighted DMPFC as a center for abstract, hierarchically-organized behavior and planning[72–78]. For example, one fMRI experiment found that during a variable planning interval DMPFC was involved in sustaining and "energizing" task-relevant processes[79]. In an fMRI study where participants planned their navigation in a virtual subway network, DMPFC activity was observed to encode the task context (i.e., current subway line), and was correlated with the cost of representing a hierarchical plan (i.e., the number of subway lines to take as opposed to the number of stations to visit)[47]. Similarly, when navigating a virtual city, information about upcoming paths was represented across a gradient of activity in anterior prefrontal cortex (APFC, including parts of DMPFC), with more anterior regions encoding places to be visited more distantly in the future[48]. Dorsal ACC activity was also observed to encode hierarchical representations in a sequential task that emulated making different beverages (coffee vs. tea), as these representations were predicted by the activation states of a recurrent neural network (RNN) model of the task[80]. More recently, a reanalysis of these data incorporating goal states into the RNN model demonstrated that more rostral regions of DMPFC (rostral ACC) encoded the goal-related representations in the task[35]. Further, our results align with a model-based hierarchical reinforcement learning (MB-HRL) framework that proposes that dorsal ACC is responsible for selecting and maintaining high-level task options (goals), where DLPFC follows by exerting top-down control over BG to facilitate the execution of the specific actions that implement the selected option[76]. Beyond maintaining such currently executed behaviors, we found that DMPFC prospectively maintains an abstract representation of the task context during planning and uses that representation to guide the execution of upcoming behavior. Relatedly, findings from primate electrophysiology show that neurons in medial frontal cortex represent upcoming action sequences rather than the current sensory input during planning periods[81,82]. This suggests that the perceptual-to-abstract shift we observed in DMPFC reflects an active constructive process, that progressively builds and maintains an abstract plan representation rather than simply signal maintenance across the delay.

Note that although the RSA and decoding analysis provide converging evidence on the role of DMPFC in planning, the decoding results (but not the RSA results) were determined using an uncorrected threshold and therefore should be considered exploratory. Nevertheless, the DLPFC and BG observations are consistent with a wide neuroimaging literature on this system. The prefrontal cortex (PFC) and BG are well-known to support model-based learning and planning[83,84], with the PFC constructing internal models of entire behaviors, simulating them, and directing their execution[85,86], and the BG learning action values specific to task states[87,88]. Further, the activity patterns observed in DLPFC are consistent with previous arguments that this region implements rule-based response selection[89,90]. In particular, DLPFC activity has been implicated when planning information is actively monitored[79] and processed[23] (as opposed to simply maintained) during a planning delay, and when rehearsing and preparing planned responses and updating evolving task goal states[91], for instance by selecting items and actions from working memory[92]. Likewise, the BG, which are highly interconnected with the frontal cortex, have been associated with rehearsal and implementation of appropriate motor and behavioral plans[28,29,93,94]. The DLPFC might therefore encode a set of task rules or instruction mappings[95] that are comparable during both planning and execution and are learned via BG associations. This interpretation is consistent with the role of BG in learning stimulus-association and response selection and of DLPFC in manipulating task information compliant with task rules[96–100], namely by matching shapes to codes during planning (first selecting actions internally) and from codes to button presses during execution (then executing the actions externally). In so doing, DLPFC prepares the upcoming sequence of actions[79,101] by relating this information between planning and execution[100,102,103]. This process enacts a necessary

counterpart to the abstract representations encoded by the DMPFC. Specifically, the DLPFC exhibits complementary dynamics where neurons actively transform stimulus-based information into action-based representations over the course of the delay[104]. This suggests that the DLPFC contributes by converting the evolving abstract context in DMPFC into the concrete action mappings required at execution.

The relationship presented here between planning and its execution might bear some relation to the Theory of Event Coding[32], which proposes that action plans are linked to their anticipated perceptual outcomes. However, the present results suggest that it cannot be fully explained in those terms. Because motor mappings were not predictable, and because DMPFC representations generalized across perceptually distinct stimulus sets, these findings are inconsistent with an account based solely on binding specific perceptual features to specific motor responses. Instead, our findings suggest that DMPFC operates at the level of task-relational structure, a dimension of abstraction that goes beyond, but is not incompatible with, feature-level binding. Furthermore, alternate accounts highlight the HPC at the forefront of sequence planning[28] and execution[29]. Crucially, HPC activity in those previous studies was observed when sequence planning involved knowledge not only of the behavioral sequences but also the specific temporal organization of their pre-assigned motor effectors (sequence-finger associations for execution), something that our task precluded by design. Considering the major role of HPC in the construction and maintenance of highly detailed memory episodes on the one hand, and the role of DMPFC in abstract and associative knowledge on the other, it is not surprising that DMPFC (and not HPC) was here responsible for leveraging statistical regularities between task representations in order to navigate between planning and execution. Moreover, note that some planning studies have utilized a go/no-go task design that contrasts go trials (planning followed by execution) against no-go trials (planning followed by suppression of execution)[21,105]. However, such contrasts aim to isolate motor preparatory signals during planning and then relate those activations to the motor activity during execution, whereas our study investigated how a plan is translated from instruction to implementation using an abstract format that is independent of any motor and perceptual representations. The no-go requirement also introduces a variety of confounds that are unrelated to planning, such as response inhibition, which encourages participants to prepare to not respond when they are uncertain about whether they will be prompted to respond[106]. To avoid this confound, and mindful our goal to identify representations that are shared across planning and execution, we designed the task to include go-trials only, while taking specific measures to prevent collinearity in our signals of interest (see Methods and Supplementary Fig. S3).

These conclusions are qualified by a few caveats. First, the pre-registered behavioral chunking effect[51–53] was attenuated during the main experiment (Fig. 2A). However, absence of chunking does not predict absence of planning, since we observed evidence of behavioral planning when analyzing the benefit of a longer planning interval (Results and Supplementary Fig. S2). Importantly, the participants' behavior showed the predicted signature of chunking during an online behavioral training session conducted prior to the fMRI experiment (see Methods, Results, and Supplementary Fig. S1). Although the interpretation is admittedly post-hoc, we suggest that the reduction of hierarchical chunking from behavioral training to fMRI testing reflects enhanced coding efficiency due to increased task exposure (given that the accuracy levels were comparable across the training and main experiment sessions). Second, the pre-registered cross-decoding analysis failed to reach statistical significance after FWE correction, though it was significant at a more liberal threshold. Nevertheless, this result is supported by an independent RSA that identified this same DMPFC region in abstract planning, and by an independent functional localizer that identified planning-related representations in this DMPFC region. Evaluated collectively, these results point to an important role for DMPFC in planning. Third, the limited time resolution inherent to fMRI prevented a finer analysis of the dynamics of the neural representations. However, by relating independent functional localizers to specific planning time points, our analyses provide insight into how these representations evolved during planning, namely that DMPFC representations became more abstract as the planning interval progressed. Taken together, these results provide evidence for how plans are transformed into action. Moving forward, prospective neuroimaging studies utilizing imaging modalities with better time resolution (such as electroencephalography or magnetoencephalography) can clarify how abstract representations for planning are acquired and developed, both within and across trials.

Conclusion

Planning and action operate along a continuum whereby planning allows for behaviors to be internally prepared and rehearsed, and action entails the rehearsed sequences of behavior to be implemented. These operations that translate planning to action require a common (abstract) representation that spans the transition from intention to execution. Crucially, real-world environments provide constraints that are sometimes revealed only during the execution of a behavior, and thus are difficult to plan for in advance. Hence, when driving on a roundabout, we might miss our turn if the roundabout has an unfamiliar layout despite having read the sign and planned our exit well in advance. Such abstract representations encode fundamental, high-level relational information that can be used to plan behavioral sequences and then adaptively translate the plans into specific details of implementation. Our results indicate that DMPFC plays a unique role in this hierarchical planning-to-action mechanism.

Methods

Experimental design, sample size, and statistical analyses were preregistered (https://doi.org/10.17605/OSF.IO/E58KT) and are described below. Deviations from the preregistration are noted as such.

Participants

Fifty-three participants (38 females, mean age = 24.5, range = 18.8–37.5) took part in an online behavioral training session. Before the subsequent fMRI scanning session, the data of two participants were removed due to failure to complete the training (online training timed out), and the data of one other participant were removed due to MRI-incompatibility. Our sample for analysis therefore included 50 participants (35 females, mean age = 24.4, range = 18.8–37.5) who took part in both the online behavioral training and in the fMRI experiment. All participants were screened for MRI-compatibility and normal or corrected-to-normal visual acuity, were compensated for their participation, and provided informed consent before each experimental session (online and fMRI). The Ethics Committee of Ghent University Hospital reviewed and approved this study. This study was conducted in accordance with the Declaration of Helsinki. All ethical regulations relevant to human research participants were followed.

Procedure

We conducted a two-part experiment. In the first part, participants underwent behavioral training online, which was intended for participants to learn the task stimuli and rules. For the second part, participants underwent fMRI scanning while performing a slightly modified version of the task (see Method Details) plus two additional functional localizer tasks inside an fMRI scanner. The two portions of the experiment were completed by the same set of participants.

Behavioral training (online). The online session comprised three increasingly difficult task training phases with two parts each: guided training (instructions + 10 practice trials) followed by task practice that required an accuracy criterion to be met (10 consecutive correct trials; same for all training stages) in order to proceed to the next training stage. The duration of the online training task depended on instruction reading time and task performance. However, the online training timed out after 120 min.

fMRI scanning session. Within 24 h after successfully completing the online session, participants came to the fMRI scanner for an imaging

session that included a main task and two functional localizer tasks. Due to scanner malfunctions, 3 of the participants were rescheduled for a later date.

Prior to entering the scanner, participants practiced the task (a shortened version of the online training) for about 10 min on a laptop. Inside the scanner they performed the main task, which was similar to the task they learned in the preceding session (see Method Details) for four blocks of about 10 min each. After the four main task blocks, they performed two functional localizer tasks for about 10 min each. In total, the scanning session took around 65 min.

Behavioral tasks

Online training task. Participants were instructed to decipher codes (combinations of characters "#" and "&") from abstract stimuli (colored geometric shapes), and that they would use the index and middle fingers of their right hand to press the "J" and "K" keys to select items (the characters "#" and "&") on the left and the right side of their screen, respectively. In general, the training started with an elementary task version of 1 stimulus with 1-character code (representing its shape), progressed to 1 stimulus with a 2-character code (representing shape + color), and finished with 2 stimuli with a 2-character code each (representing shape + color for each).

On the first training stage (training shape), participants were shown a single, randomly selected white shape (square, circle, diamond, or triangle) on a black background, and were instructed to select "#" if the shape was a square or a diamond, or "&" if the shape was a triangle or a circle. A trial ended when the correct response was chosen. They performed 10 practice trials irrespective of their accuracy. Then, in order to advance to the next training stage, they were required to correctly complete two 10-trial sequences of consecutive trials (thus meeting an accuracy criterion).

For the second training stage, a delay screen was introduced after stimulus presentation, during which the participants were required to plan the code that they would enter in the subsequent execution phase. As with the first training stage, they performed 10 practice trials, which they repeated until reaching the accuracy criterion.

For the next (third) training stage (training shape + color), the participants were presented with the same shapes as before (square, diamond, circle, triangle) but now displayed in specific colors (squares and triangles were either blue or red, circles and diamonds were either green or yellow; Fig. 1C), selected at random. Consequently, the codes to be input had two dimensions (i.e., character shape and character color). Initially, the participants were presented with each colored shape and were asked to input the 2-character code, and afterwards a planning phase was introduced (similar to stage 1). Finally, a time limit was introduced for the execution phase. The participants proceeded to the next stage once they have reached the accuracy criterion.

In the final (fourth) training stage (training shape + color with 2 stimuli), the stimuli at each trial consisted of 2 shapes (a shape pair) presented together. Participants were required to input two 2-character codes one after the other (equivalent to one 4-character code). Similar as the proceeding training stages, participants were trained in blocks without a planning phase, with a planning phase, and with a time limit for execution. They proceeded to the next stage once they had reached accuracy criterion.

After participants completed the training stages, they performed the "main training task" by performing 4 blocks of the task with two shape pairs (describing two 2-character codes). This stage was conducted without an accuracy criterion. Accuracy across blocks for the main training task was high (M = 0.93, SD = 0.007, range = 0.92–0.94) and comparable to the main fMRI task (M = 0.94, SD = 0.007, range = 0.93–0.95).

fMRI tasks. Each imaging session consisted of a main task and two functional localizer tasks.

For the main task, participants were placed inside the scanner and performed 4 blocks of around 10 minutes each of a task similar to the main training task. Each trial started with a fixation cross, followed by the presentation of the trial stimuli (colored geometric shapes). The durations of both the fixation cross and the stimuli were jittered by 2.0 ± 0.5 s. After stimulus presentation, a black screen was presented for 6.0 ± 1.5 s. (jittered), signaling the planning phase. During this period participants were required to map the stimuli (colored geometric shapes) presented on that trial to their corresponding codes (that were learned during the practice session). Finally, the execution phase was signaled by the presentation on screen of an input field, the instruction "enter code", and two buttons labeled with the characters "#" and "&" (the relative positions of these buttons varied at random from trial to trial). The execution phase lasted 3.0 ± 0.5 s. (jittered), and participants were required to enter the codes corresponding to the trial stimuli by pressing two spatially equivalent buttons (i.e., right button to select right item on the screen) on an MRI-compatible button box using the right-hand index and middle fingers. A feedback screen was presented only for incorrect trials with a fixed duration of 0.5 s.

The duration of each of the phases (fixation cross, stimuli presentation, planning, and execution) were assigned pseudo-randomly on each trial, such that all the trials in the experiment lasted for 13 s, and all possible durations were presented in equal proportion (e.g., the planning phase lasted an equal number of times for 4.5, 6.0, or 7.5 s).

The fMRI task differed from the (online) main training task (see above) according to the relative proportions of the presented stimuli. To increase the number of observations per condition, we restricted the number of codes to be analyzed to only two (c1: "# & & #", c2: "& # # &"), each of which was represented by two 2 shape-pairs -- i.e., c1 represented by c1s1 ("red square + blue triangle") and c1s2 ("yellow diamond + green circle"), and c2 represented by c2s1 ("blue triangle + red square") and c2s2 ("green circle + yellow diamond") (Fig. 1E). Each main task run contained 44 trials, about 82% of which consisted of the trials of interest or "standard trials" (c1 or c2; Fig. 1E), and the remaining 18% were "catch trials" (Supplementary Table S1) that were randomly selected from the allowed feature combinations (i.e., shapes in s1 combined only with shapes and colors within s1, and shapes in s2 combined only with shapes and colors within s2; Fig. 1D). Standard and catch trials were presented intermixed during each task run.

After completing the main task, participants performed two functional localizer tasks ("shape" and "code" localizers) that lasted ~10 min each. The shape localizer task included 68 task trials, including 15 trials per standard-trial shape-pair (or 30 trials per each of the two standard-trial codes), and 8 (~12%) catch trials. During each trial of the shape localizer task, participants first saw a fixation cross, followed by the presentation of a shape-pair (according to the same stimuli and constraints as in the main task). The durations of both the fixation cross and the stimulus presentations were jittered by 2.00 ± 0.50 s. Next, a black screen appeared for 3.00 ± 0.50 s (jittered across trials). During this planning phase participants were required to maintain the stimulus information in working memory. Then, a shape-pair with the same stimulus properties and constraints as in the main task (including both standard and catch trials) was presented for a duration of 1.75 ± 0.50 s. During this execution phase, participants were required to indicate whether the stimuli were identical to those presented at the start of the trial by pressing one of two buttons on a response box corresponding to the answers "yes" and "no". The correspondence between the button location (left or right) and the response options ("yes" and "no") was cued on the screen and was pseudo-randomized across trials such that each button corresponded to "yes" and "no" an equal proportion of times. The duration of each of the phases (fixation cross, stimulus presentation, planning, and execution) were assigned pseudo-randomly on each trial, such that all the trials in the shape localizer task lasted for 8.75 s, and all possible durations were presented in equal proportion (e.g., the planning phase lasted an equal number of times for 2.50, 3.00, or 3.50 s).

The code localizer task included 68 task trials (30 trials per each of the two standard-trial codes) and 8 (~12%) catch trials. During each trial, participants were presented with a fixation cross (2.00 ± 0.50 s) and then an abstract code consisting of the same combinations of the symbols "&" and "#" as they appeared in the main task (2.00 ± 0.50 s), the duration of which was jittered across trials. Then a black screen was displayed for 3.00 ± 0.50 s jittered across trials, during which the participants were required to

remember the observed code. Finally, a screen that was identical to the execution phase of the main task appeared for 2.25 ± 0.50 s, during which participants were required to execute a response (see above and Fig. 1). The duration of each of the phases (fixation cross, stimulus presentation, planning, and execution) were assigned pseudo-randomly for each trial such that all the trials in the code localizer task lasted for 9.25 s, with all possible durations presented in equal proportions (e.g., the planning phase lasted an equal number of times for 2.50, 3.00, or 3.50 s).

For both the main task and the functional localizer tasks, the standard and catch trials were presented intermixed. The standard trial codes were constructed such that 1) each character appeared an equal number of times in each code, 2) the codes were composed of more than one symbol (excluding, for example, the stimulus "####"), and 3) the dimensions were orthogonal across codes (e.g., red in code 1 is blue in code 2). During the scanning session all of the participants performed the main task first, followed by the two localizer tasks; the order of the two localizer tasks was counterbalanced across participants.

Participants pressed the left and right buttons of the MRI-compatible button box to select the code characters ("#" and "&") that were presented on the left and the right side of the screen. Importantly, the location of the characters on the screen was assigned pseudorandomly across trials and participants, such that either code character appeared at the same (left or right) screen location for no more than 2 consecutive trials.

MRI data collection

MRI data were collected in a Siemens MAGNETOM PrismaFit 3 T MRI scanner using a 64-channel head coil. For each participant we collected a high-resolution T1-weighted anatomical scan using an MPRAGE sequence (176 slices; voxel size = 1 mm isotropic; TR = 2250 ms; TE = 4.18 ms; FoV = 256 mm), and whole-brain, T2*-weighted multi-band echo-planar imaging (EPI) functional scans (54 slices per volume; TR = 1780 ms; TE = 27 ms; voxel size = 2.5 mm isotropic; FoV = 210 mm; interleaved slice acquisition).

Statistics and reproducibility

R (version 4.4.0) running under RStudio (version 2024.09.1 + 394) was used for all behavioral analyses. Per participant, we estimated mean reaction times (RTs) per sequence step, and accuracies with the Rmisc package. We performed a Kruskal–Wallis test to assess overall differences in accuracy across blocks.

Neuroimaging analyses were conducted using SPM12 software (Statistical Parametric Mapping; Wellcome Department of Imaging Neuroscience, London, UK) running on Matlab (version R2023a) with customized scripts and in tandem with purpose-specific toolboxes: The Decoding Toolbox (version 3.99 F[107]) for decoding analysis, and the rsatoolbox[108] for representational similarity analysis (RSA).

To estimate voxel activity patterns via MVPA, we used multivariate decoding[57], and Representational Similarity Analysis[109]. We first ran a voxel-wise general linear model (GLM) on unsmoothed, unnormalized individual data. This GLM included regressors of interest for planning and execution phases, modeled as boxcars. Error and late trials were modeled as regressors that comprised the duration of the entire trial. These task-related regressors were convolved with a canonical hemodynamic response function plus time and dispersion derivatives as implemented in SPM12. Confound regressors (estimated during preprocessing) included six head motion parameters, framewise displacement, and global signal (within cerebrospinal fluid, white matter, and brain mask); these were entered into the model without convolution. The resulting beta maps from the canonical HRF were used to perform our searchlight MVPA analyses. Additionally, region-of-interest analyses were performed for multivariate decoding-, representational similarity-, and semi-partial correlation analyses (Supplementary Fig. S4). As preregistered, we used a p-value of 0.05 to establish statistical significance, and to control for multiple comparisons where applicable we used a familywise error (FWE) cluster correction of $p = 0.05$ and a primary voxel-wise threshold of $p < 0.001$.

Considering that the activity of interest for the main cross-decoding analysis took place in sequential temporal periods, we took several measures to mitigate potential contamination of neural activity occurring during the execution phase by neural activity occurring during the planning phase, as well as other confounds inherent to fMRI research. First, to prevent activation related to motor preparation from occurring during the planning phase, a blank screen was shown during the planning period, followed by an image of the button-to-screen mappings during the execution period. Second, in order to mitigate against the lag of the hemodynamic response from contaminating the execution period with stimulus-related neural activity occurring during planning, the data for training (on the planning representations) and testing (on the execution representations) were taken from different trials and from different conditions (for example, training on the plan for the red square and blue triangle while testing on the execution of the yellow diamond and green circle; Fig. 3A). In addition, the duration of the planning phase was jittered across trials (Fig. 1B), as randomized timing of events has been shown to ameliorate temporal overlap due to the hemodynamic lag[110]. Altogether, these measures were expected to minimize collinearity among the regressors of interest.

We verified the validity of our experimental design and analysis by assessing (not pre-registered) the 1) pairwise correlations[62,63] and 2) the variance inflation factor (VIF)[62,111] between the regressors of interest (Supplementary Fig. S3, left and right, respectively). The value of the VIF for regressor X_i is estimated by $VIF_i = 1/\left(1 - R_i^2\right)$, where R_i^2 is the coefficient of determination obtained by modeling X_i as a function of all other regressors of interest ($X_i = \beta_1 X_1 + \cdots + \beta_{i-1} X_{i-1} + \beta_{i+1} X_{i+1} + \cdots + \beta_N X_N + \epsilon$). Since the magnitude of the VIF indicates the extent to which each regressor is expressed as a linear combination of the other regressors, it is effectively a direct measure of collinearity. Values of VIF closer to 1 reflect greater independence between regressors and hence no inflation of variance, while a typical threshold of VIF = 5 or 10 is considered to signal a collinearity problem.

As expected, the VIF was low for all regressors of interest (mean = 1.14, SD = 0.03; Supplementary Fig. S3, right). Further, although the pairwise correlations between planning and execution regressors from the same condition/trial were moderate (mean = 0.29, SD = 0.04; Supplementary Fig. S3, left), the pairwise correlations between planning and execution regressors from different conditions/trials were low (mean = -0.07, SD = 0.017; Supplementary Fig. S3, left). Note that the decoding analysis compared representations only across conditions and trials, never within a condition or trial. Therefore, this analysis confirms that the decoded activations between planning and execution result from shared components between those stages rather than from any lingering effects of the hemodynamic response.

Behavioral data analysis

One RT was estimated for each of the 4 steps in the sequence: RT for step 1 was measured from the onset of execution until the first button press, and RTs for steps 2–4 were measured as the intervals between successive button presses (e.g., RT2 was measured from the first to the second button press). RTs were analyzed fitting a linear mixed-effects model (LMM) using the lme4 package. The model included step (1, 2, 3, 4) and planning duration (short, medium, long) as fixed effects, as well as their interaction. To account for individual differences, we included random intercepts for subjects. We assessed the significance of the fixed effects using Type III ANOVA with Satterthwaite's degrees of freedom approximation. For significant main effects or interactions, we conducted post-hoc pairwise comparisons of estimated marginal means with Tukey's Honestly Significant Difference (HSD) correction implemented in the emmeans R package.

MRI data preprocessing

Before preprocessing, raw (DICOM) MRI data were converted into BIDS-format[112] using BIDScoin 3.0.8. Results included in this manuscript come from preprocessing performed using fMRIPrep 23.2.3, a Nipype 1.8.6[113] based tool. As recommended by the fMRIPrep developers, we report below

the automatically generated outputs from preprocessing both the anatomical and the functional MRI data.

Anatomical data preprocessing. A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1w image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection[114], distributed with ANTs 2.5.0[115], and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast[116] (FSL). Brain surfaces were reconstructed using recon-all (FreeSurfer 7.3.2), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.5.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization and accessed with TemplateFlow (23.1.0): ICBM 152 Nonlinear Asymmetrical template version 2009c (TemplateFlow ID: MNI152NLin2009cAsym).

Functional data preprocessing. For each of the 6 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume was generated, using a custom methodology of fMRIPrep, for use in head motion correction. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt[117] (FSL). The estimated fieldmap was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration[118]. Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions)[119], and Jenkinson (relative root mean square displacement between affines)[117]. FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al[119,120]). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128 s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF + WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's aseg segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each[121]. Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance timeseries are calculated by means of principal components analysis of the signal found within a thin band (crown) of voxels around the edge of the brain, as proposed by Patriat et al.[122]. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using nitransforms, configured with cubic B-spline interpolation.

Multi-Voxel Pattern Analysis (MVPA): searchlight cross-decoding of action execution from action planning. To answer our primary question of whether the brain maintains an abstract, higher-order context during planning and uses this representation when executing the plan, we conducted a one-vs-one multi-class cross-validated (leave-one-run-out) searchlight cross-classification of the executed action sequences, with classification accuracy minus chance values per voxel (accuracy maps) as the output. Beta maps from the planning phase of one shape-pair from each condition (e.g., c1s1 and c2s1) were used for training, and beta maps from the execution phase of the left-out conditions (e.g., c1s2 and c2s2) were used for testing (Fig. 3A). This procedure was repeated by switching the conditions (i.e., training on the planning phases of c1s2 and c2s2, and testing on the execution phase of c1s1 and c2s1, respectively; Fig. 3A). Crucially, because this procedure used one set of conditions for training and a different set of conditions for testing, the training and testing split was carried out on data taken from different trials (because different conditions took place on different trials), different trial phases (training on planning vs testing on execution), different planning conditions (because the classifier was trained on one set stimuli and tested on a different set of stimuli), and different functional runs (via leave-one-run-out cross-validation), per subject. The output of this analysis consisted of an accuracy map per subject (from planning to execution), indicating the degree to which the abstract codes generalized across task phases regardless of their low-level stimulus features and motor demands.

The analysis was run using a linear support vector machine (SVM) implemented within The Decoding Toolbox[107] software package. Decoding was performed in native space by scanning a spherical searchlight (6 mm of radius) across the condition beta values for the entire brain to obtain an accuracy map (cross-decoding accuracy score assigned to each searchlight-centered voxel) per subject. We normalized each subject's z-scored accuracy map into MNI space and applied spatial smoothing using a 3 mm FWHM gaussian kernel. Statistical significance was assessed at the group level by performing a one-sample t-test against the null hypothesis (chance performance level of 0.25) across participants followed by cluster-wise inference. Specifically, on the resulting map we applied a recommended voxel-level cluster-defining threshold of $p < 0.001$ (uncorrected)[123]. Then, cluster significance was determined using familywise error (FWE) correction at $p < 0.05$ at the cluster level, based on Gaussian Random Field Theory as implemented in SPM. Finally, for visualization purposes, the resulting statistical map was thresholded at a cluster size of 20 voxels.

Representational Similarity Analysis (RSA): searchlight RSA of planned sequences and hypothesis RDMs. To investigate the format in which information is represented while planning the execution of a sequence, a cross-validated (leave-one-run-out) searchlight representational similarity analysis (RSA) of planned sequences was applied to the data of the planning phase of the experiment. For the data of each participant, a 6 mm radius sphere was slid across the brain image in native space, obtaining for each voxel a vector of neural activations per

condition (c1s1, c1s2, c2s1, c2s2). Then, on each voxel we constructed a 4×4 brain-RDM by estimating the cross-validated Pearson (1 - Pearson) pair-wise correlation across all conditions (c1s1 VS c1s2 VS c2s1 VS c2s2). On each fold of the cross-validation process the dissimilarity estimates were averaged across runs of the training set, and this averaged value was contrasted against the left-out set. The dissimilarity values were averaged across cross-validation folds to obtain the RDM estimate per voxel. We contrasted this "brain-RDM" against two candidate "hypothesis-RDMs" (Fig. 3E): an "abstract" RDM that encoded whether the stimuli mapped to within- or between code contexts, and a perceptual RDM that encoded the visual similarity of the stimuli. To contrast the brain-RDM against each of the hypothesis-RDMs, from each voxel's RDM we extracted the elements above the matrix diagonal, producing a 6-element vector that was compared using Spearman's rank correlation to an equivalent 6-element vector extracted from each hypothesis-RDM. At the group level, for each hypothesis-RDM we performed a one-sample *t*-test. On the resulting statistical map we applied the recommended voxel-level cluster-defining threshold of $p < 0.001$ (uncorrected)[123]. Then, cluster significance was determined using familywise error (FWE) correction at $p < 0.05$ at the cluster level, based on Gaussian Random Field Theory as implemented in SPM. Finally, we report clusters with activity patterns significantly correlated with each hypothesis-RDM; for clarity, clusters smaller than 20 voxels were not considered.

Semi-partial correlation: main task and functional localizers. We hypothesized that during the planning phase of the main task, participants initially recalled the geometric shapes and progressively mapped the presented shapes to their specific codes, establishing the correct code sequence for that trial. Accordingly, we designed two functional localizer tasks to capture each of these processes: a "shape" functional localizer (Fig. 1C, top) that was sensitive to the visual properties of the stimuli, and a "code" functional localizer (Fig. 1C, bottom) that required implementing the abstract codes as action sequences.

The beta maps for this analysis were estimated from voxel-wise GLMs performed independently on the main task brain data and on each of the functional localizers. For the main task, we included regressors for planning at the beginning and middle time points (the middle time point varied due to the jittered duration of the planning period, see above), modeled as delta functions. For the functional localizers, we included regressors for planning at the beginning of the planning phase. For all GLMs, error and late trials were modeled as regressors comprising the duration of the entire trial. Confound regressors (estimated during preprocessing) included six head motion parameters, framewise displacement, and global signal (within cerebrospinal fluid, white matter, and brain mask); these were entered into the model without convolution. All task-related regressors were convolved with a canonical HRF plus time and dispersion derivatives as implemented in SPM12.

To inspect how perceptual and abstract representational formats evolved across the planning phase, we analyzed the brain activity patterns at the beginning and middle time points of the planning phase focused on the clusters obtained from the cross-decoding analysis (see Results and Fig. 3) as regions of interest (ROIs), as well as additional pre-registered ROIs (see Supplementary Fig. S4). Following a template tracking procedure[124], we estimated semi-partial correlations between the patterns of activity (beta maps) of the main task in each condition during the planning period, and the patterns of activity of each of the functional localizers during the planning period (Fig. 4A). We used semi-partial correlations to assess the individual/unique contribution of each representational format (e.g., abstract) with respect to the main task while partializing out the contribution of the other representational format (e.g., perceptual).

We predicted that the planning-related activity in the regions of interest would be more correlated with the shape localizer than with the code localizer at the beginning of the planning phase, whereas the opposite (higher correlation with the code localizer) would be true during the middle of the planning phase. To test these predictions we first estimated the semi-partial correlation between each localizer and the main task while controlling for the influence of the other localizer on the main task, that is, we correlated the activity patterns of one localizer (e.g., code) with the residuals obtained from regressing the activity patterns of the other localizer (e.g., shape) onto the activity patterns of the main task. We then repeated this procedure at each time point (beginning, middle) for each ROI (DMPFC, DLPFC, and BG clusters). Then, for each ROI, the semi-partial correlation estimates were analyzed via a two-way repeated measures linear mixed-effects model with fixed effects of time point (beginning, middle), localizer (shape, code), and their interaction, and subject included as a random effect (random intercept). Fixed effects were evaluated with Type III F-tests using Satterthwaite degrees of freedom. Post-hoc comparisons were tested using estimated marginal means (emmeans), with pairwise contrasts corrected for multiple comparisons using the Tukey method.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data and code are available at https://osf.io/jw3xy/overview?view_only=8020ee1405274d39b2abf5ee9b830a02.

Code availability

Data and code are available at https://osf.io/jw3xy/overview?view_only=8020ee1405274d39b2abf5ee9b830a02.

Received: 17 February 2026; Accepted: 9 June 2026;
Published online: 23 June 2026

References

1. Rowe, J. B., Owen, A. M., Johnsrude, I. S. & Passingham, R. E. Imaging the mental components of a planning task. *Neuropsychologia* **39**, 315–327 (2001).
2. Cohen, J. D. et al. Temporal dynamics of brain activation during a working memory task. *Nature* **386**, 604–608 (1997).
3. Curtis, C. E. & D'Esposito, M. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* **7**, 415–423 (2003).
4. Fuster, J. M. & Alexander, G. E. Neuron Activity Related to Short-Term Memory. *Science* **173**, 652–654 (1971).
5. Kubota, K. & Niki, H. Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J. Neurophysiol.* **34**, 337–347 (1971).
6. Elsinger, C. L., Harrington, D. L. & Rao, S. M. From preparation to online control: Reappraisal of neural circuitry mediating internally generated and externally guided actions. *NeuroImage* **31**, 1177–1187 (2006).
7. Gallivan, J. P., McLean, D. A., Valyear, K. F., Pettypiece, C. E. & Culham, J. C. Decoding Action Intentions from Preparatory Brain Activity in Human Parieto-Frontal Networks. *J. Neurosci.* **31**, 9599–9610 (2011).
8. Henderson, M. M., Rademaker, R. L. & Serences, J. T. Flexible utilization of spatial- and motor-based codes for the storage of visuo-spatial information. *eLife* **11**, e75688 (2022).
9. Hoshi, E. & Tanji, J. Area-Selective Neuronal Activity in the Dorsolateral Prefrontal Cortex for Information Retrieval and Action Planning. *J. Neurophysiol.* **91**, 2707–2722 (2004).
10. Ito, M. Control of mental activities by internal models in the cerebellum. *Nat. Rev. Neurosci.* **9**, 304–313 (2008).
11. Koechlin, E., Corrado, G., Pietrini, P. & Grafman, J. Dissociating the role of the medial and lateral anterior prefrontal cortex in human planning. *Proc. Natl. Acad. Sci. USA* **97**, 7651–7656 (2000).
12. Lazeron, R. H. C. et al. Visualizing Brain Activation during Planning: The Tower of London Test Adapted for Functional MR Imaging. *Am. J. Neuroradiol.* **21**, 1407 (2000).

13. Momennejad, I. & Haynes, J.-D. Human anterior prefrontal cortex encodes the 'what' and 'when' of future intentions. *NeuroImage* **61**, 139–148 (2012).
14. Momennejad, I. & Haynes, J.-D. Encoding of Prospective Tasks in the Human Prefrontal Cortex under Varying Task Loads. *J. Neurosci.* **33**, 17342–17349 (2013).
15. Rousseau, C., Barbiero, M., Pozzo, T., Papaxanthis, C. & White, O. Actual and Imagined Movements Reveal a Dual Role of the Insular Cortex for Motor Control. *Cereb. Cortex* **31**, 2586–2594 (2021).
16. Ruge, H., Müller, S. C. & Braver, T. S. Anticipating the consequences of action: An fMRI study of intention-based task preparation: Anticipating the consequences of action. *Psychophysiology* https://doi.org/10.1111/j.1469-8986.2010.01027.x (2010).
17. Tanaka, M. et al. Roles of the Cerebellum in Motor Preparation and Prediction of Timing. *Neuroscience* **462**, 220–234 (2021).
18. Tosoni, A. et al. Decision and action planning signals in human posterior parietal cortex during delayed perceptual choices. *Eur. J. Neurosci.* **39**, 1370–1383 (2014).
19. Van Den Heuvel, O. A. et al. Frontostriatal system in planning complexity: a parametric functional magnetic resonance version of tower of london task. *NeuroImage* **18**, 367–374 (2003).
20. Wagner, G., Koch, K., Reichenbach, J. R., Sauer, H. & Schlösser, R. G. M. The special involvement of the rostrolateral prefrontal cortex in planning abilities: An event-related fMRI study with the Tower of London paradigm. *Neuropsychologia* **44**, 2337–2347 (2006).
21. Yewbrey, R., Mantziara, M. & Kornysheva, K. Cortical Patterns Shift from Sequence Feature Separation during Planning to Integration during Motor Execution. *J. Neurosci.* **43**, 1742–1756 (2023).
22. Brandi, M.-L., Wohlschläger, A., Sorg, C. & Hermsdörfer, J. The Neural Correlates of Planning and Executing Actual Tool Use. *J. Neurosci.* **34**, 13183–13194 (2014).
23. Crescentini, C., Seyed-Allaei, S., Vallesi, A. & Shallice, T. Two networks involved in producing and realizing plans. *Neuropsychologia* **50**, 1521–1535 (2012).
24. Hanakawa, T., Dimyan, M. A. & Hallett, M. Motor Planning, Imagery, and Execution in the Distributed Motor Network: A Time-Course Study with Functional MRI. *Cereb. Cortex* **18**, 2775–2788 (2008).
25. Papitto, G., Friederici, A. D. & Zaccarella, E. Distinct neural mechanisms for action access and execution in the human brain: insights from an fMRI study. *Cereb. Cortex* **34**, bhae163 (2024).
26. Turella, L., Rumiati, R. & Lingnau, A. Hierarchical Action Encoding Within the Human Brain. *Cereb. Cortex* **30**, 2924–2938 (2020).
27. Unterrainer, J. M. & Owen, A. M. Planning and problem solving: From neuropsychology to functional neuroimaging. *J. Physiol.* **99**, 308–317 (2006).
28. Yewbrey, R. & Kornysheva, K. The Hippocampus Preorders Movements for Skilled Action Sequences. *J. Neurosci.* **44**, e0832242024 (2024).
29. Dolfen, N., Reverberi, S., Op De Beeck, H., King, B. R. & Albouy, G. The Hippocampus Represents Information about Movements in Their Temporal Position in a Learned Motor Sequence. *J. Neurosci.* **44**, e0584242024 (2024).
30. Cisek, P. & Kalaska, J. F. Neural Mechanisms for Interacting with a World Full of Action Choices. *Annu. Rev. Neurosci.* **33**, 269–298 (2010).
31. Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W. & Donchin, E. A psychophysiological investigation of the continuous flow model of human information processing. *J. Exp. Psychol. Hum. Percept. Perform.* **11**, 529–553 (1985).
32. Hommel, B., Müsseler, J., Aschersleben, G. & Prinz, W. The Theory of Event Coding (TEC): A framework for perception and action planning. *Behav. Brain Sci.* **24**, 849–878 (2001).
33. Konidaris, G., Kaelbling, L. P. & Lozano-Perez, T. From Skills to Symbols: Learning Symbolic Representations for Abstract High-Level Planning. *J. Artif. Intell. Res.* **61**, 215–289 (2018).
34. Botvinick, M. M. Hierarchical models of behavior and prefrontal function. *Trends Cogn. Sci.* **12**, 201–208 (2008).
35. Colin, T. R., Ikink, I. & Holroyd, C. B. Distributed Representations for Cognitive Control in Frontal Medial Cortex. *J. Cogn. Neurosci.* **37**, 941–969 (2025).
36. Tomov, M. S., Yagati, S., Kumar, A., Yang, W. & Gershman, S. J. Discovery of hierarchical representations for efficient planning. *PLOS Comput. Biol.* **16**, e1007594 (2020).
37. Wientjes, S. & Holroyd, C. B. The successor representation subserves hierarchical abstraction for goal-directed behavior. *PLOS Comput. Biol.* **20**, e1011312 (2024).
38. Badre, D. Cognitive Control. *Annu. Rev. Psychol.* **76**, 167–195 (2025).
39. Miller, G. A., Galanter, E. & Pribram, K. H. *Plans and the Structure of Behavior* (Henry Holt and Co, New York, 1960). https://doi.org/10.1037/10039-000.
40. Hall-McMaster, S., Tomov, M. S., Gershman, S. J. & Schuck, N. W. Neural evidence that humans reuse strategies to solve new tasks. *PLOS Biol.* **23**, e3003174 (2025).
41. Lee, W.-T., Hazeltine, E. & Jiang, J. Decoding task representations that support generalization in hierarchical task. Preprint at https://doi.org/10.1101/2024.12.02.626403 (2024).
42. Skorstad, J., Genter, D. & Medin, D. Abstraction Processes during Concept Learning: A Structural view. *Proc. Annu. Meet. Cogn. Sci. Soc*. **10**, 419–425 (1988).
43. Theves, S., Neville, D. A., Fernández, G. & Doeller, C. F. Learning and Representation of Hierarchical Concepts in Hippocampus and Prefrontal Cortex. *J. Neurosci.* **41**, 7675–7686 (2021).
44. Collins, A. G. E. & Frank, M. J. Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychol. Rev.* **120**, 190–229 (2013).
45. Reverberi, C., Görgen, K. & Haynes, J.-D. Compositionality of Rule Representations in Human Prefrontal Cortex. *Cereb. Cortex* **22**, 1237–1246 (2012).
46. Xia, L. & Collins, A. G. E. Temporal and state abstractions for efficient learning, transfer, and composition in humans. *Psychol. Rev.* **128**, 643–666 (2021).
47. Balaguer, J., Spiers, H., Hassabis, D. & Summerfield, C. Neural Mechanisms of Hierarchical Planning in a Virtual Subway Network. *Neuron* **90**, 893–903 (2016).
48. Brunec, I. K. & Momennejad, I. Predictive Representations in Hippocampal and Prefrontal Hierarchies. *J. Neurosci.* **42**, 299–312 (2022).
49. Grossman, C. D., Man, V. & O'Doherty, J. P. The representation and valuation of subgoals in the human brain during model-based hierarchical behavior. *Neuron* **114**, 1306–1320.e7 (2026).
50. Liang, Q., Li, J., Zheng, S., Liao, J. & Huang, R. Dynamic Causal Modelling of Hierarchical Planning. *NeuroImage* **258**, 119384 (2022).
51. Verwey, W. B. Concatenating familiar movement sequences: the versatile cognitive processor. *Acta Psychol.* **106**, 69–95 (2001).
52. Sakai, K., Kitaguchi, K. & Hikosaka, O. Chunking during human visuomotor sequence learning. *Exp. Brain Res.* **152**, 229–242 (2003).
53. Eckstein, M. K. & Collins, A. G. How the Mind Creates Structure: Hierarchical Learning of Action Sequences. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Cognitive Science Society, 2021).
54. Rosenbaum, D. A., Kenny, S. B. & Derr, M. A. Hierarchical control of rapid movement sequences. *J. Exp. Psychol. Hum. Percept. Perform.* **9**, 86–102 (1983).
55. Afshar, A. et al. Single-Trial Neural Correlates of Arm Movement Preparation. *Neuron* **71**, 555–564 (2011).
56. Rosenbaum, D. A. Human movement initiation: Specification of arm, direction, and extent. *J. Exp. Psychol. Gen.* **109**, 444–474 (1980).

57. Norman, K. A., Polyn, S. M., Detre, G. J. & Haxby, J. V. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* **10**, 424–430 (2006).
58. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc. Natl. Acad. Sci.* **103**, 3863–3868 (2006).
59. Bernardi, S. et al. The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. *Cell* **183**, 954–967.e21 (2020).
60. Huettel, S. A. Event-related fMRI in cognition. *NeuroImage* **62**, 1152–1156 (2012).
61. Kruggel, F. & Von Cramon, D. Y. Temporal properties of the hemodynamic response in functional MRI. *Hum. Brain Mapp.* **8**, 259–271 (1999).
62. Mumford, J. A., Poline, J.-B. & Poldrack, R. A. Orthogonalization of Regressors in fMRI Models. *PLOS ONE* **10**, e0126255 (2015).
63. Josephs, O., Turner, R. & Friston, K. Event-related f MRI. *Hum. Brain Mapp.* **5**, 243–248 (1997).
64. Mumford, J. A. et al. Unintended bias in the pursuit of collinearity solutions in fMRI analysis. *Imaging Neurosci.* **3**, IMAG.a.958 (2025).
65. O'brien, R. M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual. Quant.* **41**, 673–690 (2007).
66. Kruskal, J. B. Nonmetric Multidimensional Scaling: A Numerical Method. *Psychometrika* **29**, 115–129 (1964).
67. Dehaene, S. & Changeux, J.-P. A hierarchical neuronal network for planning behavior. *Proc. Natl. Acad. Sci. USA* **94**, 13293–13298 (1997).
68. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208 (1948).
69. Mattar, M. G. & Lengyel, M. Planning in the brain. *Neuron* **110**, 914–934 (2022).
70. Ho, M. K. et al. People construct simplified mental representations to plan. *Nature* **606**, 129–136 (2022).
71. Ying, L. et al. Assessing Adaptive World Models in Machines with Novel Games. Preprint at https://doi.org/10.48550/ARXIV.2507.12821 (2025).
72. Alejandro, R. J. & Holroyd, C. B. Hierarchical control over foraging behavior by anterior cingulate cortex. *Neurosci. Biobehav. Rev.* **160**, 105623 (2024).
73. Alexander, W. H. & Brown, J. W. Hierarchical Error Representation: A Computational Model of Anterior Cingulate and Dorsolateral Prefrontal Cortex. *Neural Comput.* **27**, 2354–2410 (2015).
74. Duverne, S. & Koechlin, E. Hierarchical Control of Behaviour in Human Prefrontal Cortex. In *The Wiley Handbook of Cognitive Control* (ed. Egner, T.) 207–220 (Wiley, 2017). https://doi.org/10.1002/9781118920497.ch12.
75. Foinikianaki, E., Ikink, I., Colin, T. R., Alejandro, R. J. & Holroyd, C. B. ACC representations of reward-driven motivation over hierarchically-organized behavior. *NeuroImage* **317**, 121380 (2025).
76. Holroyd, C. B. & Verguts, T. The Best Laid Plans: Computational Principles of Anterior Cingulate Cortex. *Trends Cogn. Sci.* **25**, 316–329 (2021).
77. Holroyd, C. B. & Yeung, N. Motivation of extended behaviors by anterior cingulate cortex. *Trends Cogn. Sci.* **16**, 122–128 (2012).
78. Zarr, N. & Brown, J. W. Hierarchical error representation in medial prefrontal cortex. *NeuroImage* **124**, 238–247 (2016).
79. Vallesi, A., McIntosh, A. R., Shallice, T. & Stuss, D. T. When Time Shapes Behavior: fMRI Evidence of Brain Correlates of Temporal Monitoring. *J. Cogn. Neurosci.* **21**, 1116–1126 (2009).
80. Holroyd, C. B., Ribas-Fernandes, J. J. F., Shahnazian, D., Silvetti, M. & Verguts, T. Human midcingulate cortex encodes distributed representations of task progress. *Proc. Natl. Acad. Sci. USA* **115**, 6398–6403 (2018).
81. Saito, N., Mushiake, H., Sakamoto, K., Itoyama, Y. & Tanji, J. Representation of Immediate and Final Behavioral Goals in the Monkey Prefrontal Cortex during an Instructed Delay Period. *Cereb. Cortex* **15**, 1535–1546 (2005).
82. Constantinidis, C. et al. Persistent Spiking Activity Underlies Working Memory. *J. Neurosci.* **38**, 7020–7028 (2018).
83. Fermin, A. S. R. et al. Model-based action planning involves cortico-cerebellar and basal ganglia networks. *Sci. Rep.* **6**, 31378 (2016).
84. Miller, E. K. & Buschman, T. J. Rules through recursion: how interactions between the frontal cortex and basal ganglia may build abstract, complex rules from concrete, simple ones. *Neurosci. Rule Guid. Behav.* **419**, 440 (2008).
85. Simon, D. A. & Daw, N. D. Neural Correlates of Forward Planning in a Spatial Decision Task in Humans. *J. Neurosci.* **31**, 5526–5539 (2011).
86. Tricomi, E., Balleine, B. W. & O'Doherty, J. P. A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* **29**, 2225–2232 (2009).
87. Kawagoe, R., Takikawa, Y. & Hikosaka, O. Expectation of reward modulates cognitive signals in the basal ganglia. *Nat. Neurosci.* **1**, 411–416 (1998).
88. Samejima, K., Ueda, Y., Doya, K. & Kimura, M. Representation of Action-Specific Reward Values in the Striatum. *Science* **310**, 1337–1340 (2005).
89. Bunge, S. A., Hazeltine, E., Scanlon, M. D., Rosen, A. C. & Gabrieli, J. D. E. Dissociable Contributions of Prefrontal and Parietal Cortices to Response Selection. *NeuroImage* **17**, 1562–1571 (2002).
90. Cole, M. W., Bagic, A., Kass, R. & Schneider, W. Prefrontal Dynamics Underlying Rapid Instructed Task Learning Reverse with Practice. *J. Neurosci.* **30**, 14245–14254 (2010).
91. Fincham, J. M., Carter, C. S., Van Veen, V., Stenger, V. A. & Anderson, J. R. Neural mechanisms of planning: A computational analysis using event-related fMRI. *Proc. Natl. Acad. Sci. USA* **99**, 3346–3351 (2002).
92. Sakai, K. & Passingham, R. E. Prefrontal interactions reflect future task operations. *Nat. Neurosci.* **6**, 75–81 (2003).
93. Graybiel, A. M. The basal ganglia. *Curr. Biol.* **10**, R509–R511 (2000).
94. Groenewegen, H. J. The Basal Ganglia and Motor Control. *Neural Plast.* **10**, 107–120 (2003).
95. Mian, M. K. et al. Encoding of rules by neurons in the human dorsolateral prefrontal cortex. *Cereb. Cortex* **24**, 807–816 (2014).
96. Sakai, K. & Passingham, R. E. Prefrontal Set Activity Predicts Rule-Specific Neural Processing during Subsequent Cognitive Performance. *J. Neurosci.* **26**, 1211–1218 (2006).
97. Schumacher, E. H. & D'Esposito, M. Neural implementation of response selection in humans as revealed by localized effects of stimulus–response compatibility on brain activation. *Hum. Brain Mapp.* **17**, 193–201 (2002).
98. Abe, M. et al. Functional coupling of human prefrontal and premotor areas during cognitive manipulation. *J. Neurosci.* **27**, 3429–3438 (2007).
99. Cohen, M. X. & Frank, M. J. Neurocomputational models of basal ganglia function in learning, memory and choice. *Behav. Brain Res.* **199**, 141–156 (2009).
100. Chatham, C. H. & Badre, D. Multiple gates on working memory. *Curr. Opin. Behav. Sci.* **1**, 23–31 (2015).
101. Pochon, J.-B. The Role of Dorsolateral Prefrontal Cortex in the Preparation of Forthcoming Actions: an fMRI Study. *Cereb. Cortex* **11**, 260–266 (2001).
102. Sakai, K., Rowe, J. B. & Passingham, R. E. Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nat. Neurosci.* **5**, 479–484 (2002).
103. Vaidya, A. R. & Badre, D. Abstract task representations for inference and control. *Trends Cogn. Sci*. S1364661322000675, https://doi.org/10.1016/j.tics.2022.03.009 (2022).
104. Tanji, J. & Hoshi, E. Role of the Lateral Prefrontal Cortex in Executive Behavioral Control. *Physiol. Rev.* **88**, 37–57 (2008).
105. Ariani, G., Oosterhof, N. N. & Lingnau, A. Time-resolved decoding of planned delayed and immediate prehension movements. *Cortex* **99**, 330–345 (2018).

106. Chikazoe, J. et al. Preparation to Inhibit a Response Complements Response Inhibition during Performance of a Stop-Signal Task. *J. Neurosci.* **29**, 15870–15877 (2009).
107. Hebart, M. N., Görgen, K. & Haynes, J.-D. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Front. Neuroinformatics* **8**, 88 (2015).
108. Nili, H. et al. A Toolbox for Representational Similarity Analysis. *PLoS Comput. Biol.* **10**, e1003553 (2014).
109. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci*. https://doi.org/10.3389/neuro.06.004.2008 (2008).
110. Dale, A. M. Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.* **8**, 109–114 (1999).
111. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning*, vol. 103 (Springer, 2013).
112. Gorgolewski, K. J. et al. BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLOS Comput. Biol.* **13**, e1005209 (2017).
113. Gorgolewski, K. et al. Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Front. Neuroinformatics* **5**, 13 (2011).
114. Tustison, N. J. et al. N4ITK: Improved N3 Bias Correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).
115. Avants, B., Epstein, C., Grossman, M. & Gee, J. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).
116. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).
117. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage* **17**, 825–841 (2002).
118. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* **48**, 63–72 (2009).
119. Power, J. D. et al. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* **84**, 320–341 (2014).
120. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage* **37**, 90–101 (2007).
121. Satterthwaite, T. D. et al. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage* **64**, 240–256 (2013).
122. Patriat, R., Reynolds, R. C. & Birn, R. M. An improved model of motion-related signal changes in fMRI. *NeuroImage* **144**, 74–82 (2017).
123. Woo, C.-W., Krishnan, A. & Wager, T. D. Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage* **91**, 412–419 (2014).
124. Palenciano, A. F., Senoussi, M., Formica, S. & González-García, C. Canonical template tracking: Measuring the activation state of specific neural representations. *Front. Neuroimaging* **1**, 974927 (2023).

Acknowledgements

The authors thank the members of the Learning and Cognitive Laboratory (Ghent University) for valuable discussions during the development of this project.

Author contributions

R.J.A.: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. II: Writing – review & editing, Software, Investigation. E.F.: Writing – review & editing, Investigation. C.B.H.: Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Funding

This project was supported by funding from the European Research Council (ERC) under the EU's Horizon 2020 Research and Innovation Programme (grant agreement no. 787307).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-026-10513-4.

Correspondence and requests for materials should be addressed to Ricardo J. Alejandro.

Peer review information *Communications Biology* thanks Ivan Toni and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Benjamin Bessieres.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026