Subspace partitioning in the human prefrontal cortex resolves cognitive interference

Jan Weber[a,b] (ID), Gabriela Iwama[a,b], Anne-Kristin Solbakk[c,d,e,f] (ID), Alejandro O. Blenkmann[c,d] (ID), Pal G. Larsson[e] (ID), Jugoslav Ivanovic[e] (ID), Robert T. Knight[g,h] (ID), Tor Endestad[c,d], and Randolph Helfrich[a,1]

The human prefrontal cortex (PFC) constitutes the structural basis underlying flexible cognitive control, where mixed-selective neural populations encode multiple task features to guide subsequent behavior. The mechanisms by which the brain simultaneously encodes multiple task–relevant variables while minimizing interference from task-irrelevant features remain unknown. Leveraging intracranial recordings from the human PFC, we first demonstrate that competition between coexisting representations of past and present task variables incurs a behavioral switch cost. Our results reveal that this interference between past and present states in the PFC is resolved through coding partitioning into distinct low-dimensional neural states; thereby strongly attenuating behavioral switch costs. In sum, these findings uncover a fundamental coding mechanism that constitutes a central building block of flexible cognitive control.

prefrontal cortex | intracranial EEG | cognitive control | population geometry

In a complex and ever-changing environment, humans need to dynamically adjust their actions according to immediate contextual demands. For example, switching between braking and accelerating in a traffic jam requires rapid shifts between competing actions. The ability to flexibly adjust in an ever-changing environment hallmarks flexible cognitive control. However, rapid shifts between competing actions often incur a behavioral cost, i.e., slower and more error-prone responses following a shift from one action to another.

Multiple lines of research in humans and animal models provide converging evidence that the prefrontal cortex (PFC) constitutes the key structure that enables cognitive flexibility to guide adaptive goal-directed behavior (1, 2). To support flexible operations, PFC neurons are not feature-specific, but instead exhibit mixed selectivity and context-dependent coding of task features [i.e., sensory stimuli, abstract rules or actions, (3, 4)]. Mixed-selective neurons contribute to multiple cognitive operations by participating in different transient coalitions of cell assemblies. Consequently, information encoded by mixed-selective neurons can only be understood at the level of population activity. This notion is now referred to as the population doctrine, which posits that neural populations reflect the fundamental unit of computation in the brain (5). The population doctrine further postulates that rapid switches between distinct cognitive operations can efficiently be implemented by adjusting the population geometry, i.e., the reconfiguration of cell assemblies. A key advantage of this model is that neural representations that evolve in parallel can be integrated into a unified, conjunctive representation (6). In support of this hypothesis, recent findings demonstrated that conjunctive coding schemes enable the flexible context-dependent remapping between sensory inputs and behavioral outputs; predicting behavioral performance on a trial-by-trial basis (6).

However, parallel encoding of different variables implies that not all encoded features are behaviorally relevant; hence, raising the question if and how encoding of task-irrelevant factors impacts subsequent behavior. Previously, it had been observed that past choices reflect latent factors (defined as model-inferred variables within the population activity, which are not directly observable), and might interfere with the currently active, task-relevant representations (7). At the behavioral level, previous choices have been shown to impact subsequent behavior [i.e., serial response bias (8)]. Integration of knowledge about past choices is oftentimes desirable to correct past mistakes (9), but might cause interference when consecutive choices are independent, thus, giving rise to behavioral switch costs (10). In this scenario, a conjunctive neural code that integrates information about the past is detrimental for behavioral performance (7). To date, little is known about how the human brain maintains multiple task-relevant representations, while minimizing interference from task-irrelevant factors, in order to guide subsequent behavior. Theoretical work has proposed that neural populations can reduce interference by orthogonalizing competing representations (11, 12). This change in population

geometry enables downstream regions to flexibly readout information about competing states from the same neural population (13). Evidence in support of this notion stems from a recent study in the mouse auditory cortex that demonstrated that past and present sensory representations reside in orthogonal subspaces (i.e. lower dimensional projections of the population activity, which reflect the activity that best predicted latent factors, such as past and present sensory representations, that otherwise cannot be directly measured) (14). However, it remained unaddressed if the identified neural representations had immediate behavioral relevance. Hence, to date it remains unknown if similar principles also apply to the human brain, especially in the higher order association cortex and whether orthogonalization constitutes a key mechanism to reduce interference between competing neural representations (15).

Here, we addressed these outstanding questions by directly recording intracranial electroencephalography (iEEG) from the human prefrontal (encompassing dorsolateral and ventrolateral and medial PFC; *SI Appendix,* Table S1) and the motor cortex while participants performed a modified stop-signal task that enabled disentangling how competing neural representations between past states and current goals influence human decision-making. We leveraged high-frequency activity (HFA), a surrogate marker of multiunit activity (MUA) in humans, to approximate local neural population activity (16, 17). Recent work has shown that HFA contains more behaviorally relevant information than single-unit/MUA, thus constituting a highly valuable aggregate metric of population activity in humans (18, 19). Here, we specifically tested if neural population activity in the PFC and motor cortex simultaneously encode information about the past and present. The key question was whether overlapping neural representations account for behavioral switch costs between movement inhibition and execution. We hypothesized that efficiently distributed computing at the level of population activity constituted a core mechanism to avoid interference between distinct latent factors. Collectively, we tested whether efficient cortical partitioning of distinct neural representations enables flexible cognitive control.

Results

We recorded iEEG from 19 pharmacoresistant patients with epilepsy (33.73 y ± 12.52, mean ± SD; 7 females) who performed a predictive motor task (Fig. 1*A*). Participants were instructed to closely track a vertically moving target and respond as soon as the moving target reached a predefined spatial location (*go-trials*). In a subset of trials, the moving target stopped prematurely and participants were instructed to withhold their response (*stop-trials*). A contextual cue indicated the likelihood of a premature stop (referred to as *predictive context*; *Methods*). Previously, we demonstrated that the human PFC integrates current contextual information to guide goal-directed behavior (20). Additionally, human behavior is also strongly modulated by a variety of latent factors, such as the tendency to systematically repeat choices, known as history-dependent serial biases (8). We here investigated the neural mechanisms underlying such history-dependent serial biases, quantified as the behavioral switch costs between distinct action demands—initiation or inhibition of a goal-directed movement. Thus, task-switching was defined based on the congruency between the trial type (*stop/go-trial*) at trial *n* and *n-1* (*congruent = go-trial followed by go-trial; incongruent = stop-trial followed by go-trial*) and directly reflects the switch in task demand for the participant from "withhold" to "execute" a response.
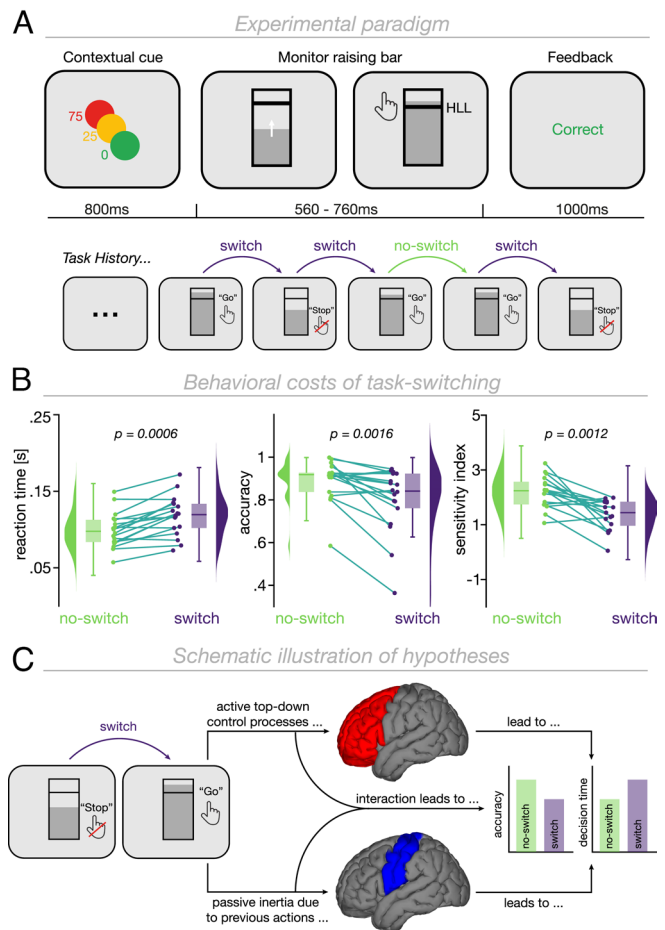
Fig. 1. Experimental design, behavioral performance, and schematic hypotheses. (*A, Top*) participants performed a predictive motor task where they had to track a vertically moving target and respond as soon as the target reached a predefined lower limit (go-trials; hit-lower limit, HLL; black horizontal line). At the start of every trial, participants received a contextual cue indicating the likelihood (0, 25 and 75% likelihood; green, orange, red circle, respectively) that the moving target would stop prematurely requiring participants to withhold their response (stop-trials). Feedback was provided at the end of each trial. (*Bottom*) Task-switching was defined as the trial-type (stop/go-trial) congruency between two successive trials (*Methods*). (*B*) Behavioral performance as a function of task-switching. (*Left*) Reaction time was significantly increased after switch trials (go-trial preceded by a stop-trial) as compared to no-switch trials (go-trial preceded by a go-trial). Accuracy (*Middle*) and d-prime (*Right*) significantly decreased after switch as compared to no-switch trials. Gray lines display individual participants, density-plots display the data distribution, and boxplots show the median (horizontal line), the first/third quartiles (upper/lower edge of box), and the minima/maxima (vertical lines). (*C*) Schematic illustration of hypothetical outcomes. In the first scenario, switch costs may reflect the time needed to engage active top-down control processes in the prefrontal cortex to reconfigure the cognitive system. In the second scenario, switch costs could reflect persistent inhibition (passive inertia) of motor areas after withholding a response. Note that both alternatives are not mutually exclusive and could also evolve in parallel.

Shifting between Competing Task Demands Has a Behavioral Switch Cost. To assess whether task-switching modulated behavioral performance, we contrasted reaction time, accuracy, and d-prime (d') between *switch-* and *no-switch-trials*. Participants were significantly slower (Fig. 1*B*; $P = 0.0006$; Cohen's $d = 1.29$; +19.02 ± 14.73 ms; mean ± SD; Wilcoxon rank sum test) and made significantly more mistakes after *switch-* as compared to *no-switch-trials* (Fig. 1*B*; $P = 0.0016$; Cohen's $d = -0.9$; +8.99 ± 9.94%; mean ± SD). Note that our task design did not enable to fully separate accuracy from reaction times as errors were oftentimes the direct consequence of prolonged reaction

times. Thus, switch costs as indexed by reaction times were only computed on correct trials, i.e., trials in which participants released the button within the time interval between the lower and upper limits (cf. Fig. 1*A*). Furthermore, task-switching across successive trials reduced participants' sensitivity to correctly decide between two action alternatives (Fig. 1*B*; *d'*; *P* = 0.0012; Cohen's *d* = –1.04). We further confirmed that our behavioral analyses were not confounded by regressor collinearity (linear mixed-effect model; *SI Appendix,* Fig. S1; *Methods*). We observed a main effect of *predictive context* and *task-history* on reaction time [*task-history:* 95% CI = (0.016, 0.026), *P* < 0.0001; *predictive context:* 95% CI = (0.0005, 0.0006), *P* < 0.0001] and accuracy [*task-history:* 95% CI = (–0.99, –0.37), *P* < 0.0001; *predictive context:* 95% CI = (–0.03, –0.02), *P* < 0.0001]. Thus, *predictive context* and *task-history* independently modulated behavior. Critically, we did not find evidence for a consistent impact on both reaction time as well as accuracy of the outcome, choice, or context of the previous trials on behavior after accounting for collinearity between task factors [reaction time: *outcome:* 95% CI = (–0.011, –0.002), *P* = 0.005; *past choice:* 95% CI = (–0.004, 0.005), *P* = 0.91; *past cue:* 95% CI = (–0.0001, 0.00004), *P* = 0.41; accuracy: *outcome:* 95% CI = (–0.14, 0.42), *P* = 0.32; *past choice:* 95% CI = (–0.53, 0.036), *P* = 0.09; *past cue:* 95% CI = (–0.005, 0.005), *P* = 0.99]. Hence, the subsequent electrophysiological analyses were directly motivated by the behavioral switch cost and therefore, focused on the interaction between predictive context and task-history.

Overlapping Neural Populations Encode Past and Present States in Human PFC. To dissect the neural mechanisms underlying behavioral switch costs, we simultaneously recorded neural activity from electrodes located in the human PFC (26 ± 18.74 electrodes per participant; mean ± SD) and motor cortex (14.83 ± 15.23 electrodes). Previous studies had demonstrated the relevance of neural dynamics in the prefrontal as well as anterior cingulate cortex for behavioral switch costs (21). Furthermore, it had been shown that the motor cortex is involved in mediating history-dependent choice biases (22). To date, no study has directly contrasted the roles of the prefrontal and motor cortex in task-switching. Based on these studies as well as behavioral models (23), two possible scenarios—that are not mutually exclusive—were considered: 1) Switch costs could reflect the time needed to engage active top-down control processes in PFC (Fig. 1*C*) or alternatively, 2) switch costs might result from prolonged inhibition of the motor cortex between switching from movement-inhibition to movement-execution (Fig. 1*C*). Alternatively, switch costs might be driven by an interaction between these two scenarios (Fig. 1*C*). In order to differentiate between these models, we quantified the univariate neural information (unsigned, bias-corrected percent explained variance; *Methods*) about *task-history* and *predictive context* in the PFC and motor cortex. Therefore, we time-locked the HFA signal in every trial (*SI Appendix,* Fig. S2 for single subject examples) to the participants response (*t* = 0 reflects the moment when participants released the button). We orthogonalized the different factors to disentangle their unique behavioral relevance [unbalanced ANOVA that implicitly orthogonalized the different factors (24, 25); *Methods*]. We observed significant context- and history-dependent neural information in the PFC (Fig. 2 *A–D*; *task-history:* $P_{cluster}$ < 0.0001, Cohen's *d* = 1.81; *predictive context:* $P_{cluster}$ < 0.0001, Cohen's *d* = 2.07; cluster permutation based on dependent *t* tests against nonencoding electrodes) and motor cortex (*task-history:* $P_{cluster}$ < 0.0001, Cohen's *d* = 1.46; *predictive context:* $P_{cluster}$ = 0.005, Cohen's *d* = 2.41).

Having established a robust coding of *predictive context* and *task-history* in the prefrontal-motor network, we then investigated whether neural information coding predicted individual switch costs in behavior using a sliding window correlation analysis (*Methods*). Here, we correlated the behavioral switch costs against context- and history-dependent neural information over time (*SI Appendix,* Fig. S3 and *Methods*). We hypothesized anticorrelated effects of context- and history-dependent neural information on behavior, since an optimal agent should only rely on the currently relevant task context and disregard uninformative features such as *task-history* to efficiently guide decisions.

In line with our hypothesis, we found that increased neural coding of predictive context was associated with lower switch costs, whereas increased neural coding of task-history predicted higher switch costs as quantified by the overall error rate (Fig. 2*E* and *SI Appendix,* Fig. S4 for reaction times on correct trials). These effects were not sustained over the entire task period, but transient in nature and emerged prior to the decision. Importantly, a significant neurobehavioral correlation was only observed in the PFC (*predictive context:* *P* = 0.039, *r* = 0.59, cluster from –0.12 to –0.027 s; *task-history:* *P* = 0.013, *r* = –0.57; cluster from –0.2 to –0.078 s; cluster-corrected correlation), but not in motor cortex (*predictive context:* no cluster; *task-history:* *P* = 0.434). Furthermore, this association differed significantly between *predictive context* and *task-history* for the PFC [*P* = 0.028; cluster from –0.18 to –0.12 s; motor cortex: *P* = 0.854; FDR-corrected (26); black horizontal line in Fig. 2*E* and *Methods*].

Next, we quantified the link between information coding of *predictive context* and *task-history* using a sliding window correlation analysis (Fig. 2*F*, *SI Appendix,* Fig. S3, and *Methods*), which revealed a robust negative correlation between context- and history-dependent neural information in the PFC (*P* = 0.004, *r* = –0.64; cluster from –0.15 to –0.1 s; cluster-corrected correlation), but not in motor cortex (*P* = 0.227). Note, however, that the correlation between the PFC and motor cortex did not significantly differ (*P* = 0.183; *Methods*).

To further illustrate the group-level correlation between information coding and behavioral switch costs, we performed a median split analysis (Fig. 2*G*). Participants with low switch costs revealed stronger neural coding for *predictive context*, but weaker neural coding for *task-history*. The opposite relationship was observed for participants with high switch costs (*predictive context:* $P_{cluster}$ = 0.026, Cohen's *d* = –1.6; *task-history:* $P_{cluster}$ = 0.084, Cohen's *d* = 1.49; cluster permutation based on independent *t* tests). Collectively, these findings indicate that limited neural resources impose a competition between feature-coding of past and present states and that overrepresentation of past states comes at substantial behavioral costs.

Competitive Coding of Past and Present States in the Human PFC Predicts Behavior. Having established that neural information coding about past and present states is anticorrelated and exerts dissociable effects on behavior, we tested whether a competitive coding scheme between past and present states in the human PFC could account for individual switch costs. Based on the highly distributed nature of feature-coding in the human PFC (cf. Fig. 2 *C* and *D*), a multivariate data analysis approach was used to test this prediction (Fig. 3*A* and *Methods*). We employed a variant of targeted-dimensionality reduction [TDR (27)]. In brief, linear regression was used to determine how *predictive context* and *task-history* modulate neural activity at every electrode. Subsequently, low-dimensional subspaces that capture *context-* and *history*-dependent variance in neural activity, were identified using principal component analysis (PCA). Consistent with previous findings demonstrating low-dimensional neural representations of task features (27, 28), we observed that neural coding of *predictive*
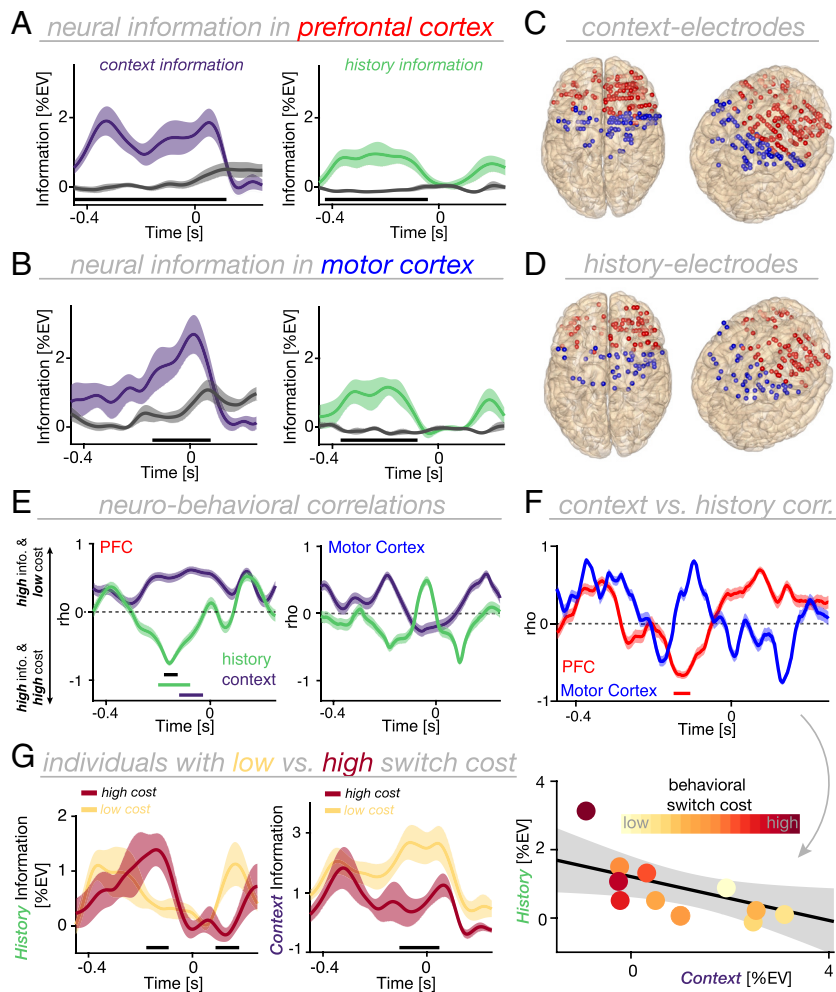
Fig. 2. Behavioral dissociation of neural dynamics encoding the past and present. (*A, Left*) Time course of context-dependent information averaged across all context-encoding electrodes in PFC. (*Right*) Time course of history-dependent neural information averaged across all history-encoding electrodes in the PFC. Lines and shaded regions show the mean and SEM. Gray traces indicate the time course of neural information across nonencoding electrodes. The lower horizontal black line shows the temporal extent of significant neural information. (*B*) Same as *A*, but for the motor cortex. (*C*) Context-encoding electrodes overlaid on a standardized brain in Montreal Neurological Institute and Hospital (MNI) coordinates for the PFC (red) and motor cortex (blue). Overall, 165 electrodes in the PFC and 89 electrodes in the motor cortex carried significant context-dependent information. (*D*) History-encoding electrodes, same conventions as in *C*. *Task-history* was significantly encoded in 96 electrodes in the PFC and 48 electrodes in the motor cortex. (*E, Left*) Temporally resolved correlation between neural information and individual switch costs (accuracy; see *SI Appendix,* Fig. S4 for reaction time) for the PFC (*Left*) and motor cortex (*Right*). Lines and shaded regions show the mean and 95% CIs of bootstrapped correlation coefficients (*Methods*). The lower colored horizontal lines show the temporal extent of significant correlation for context- (purple) and history-dependent (green) neural information. The black line shows the temporal extent of significant differences in neurobehavioral correlation between *predictive context* and -history. (*F, Top*) Temporally resolved correlation between context- and history-dependent neural information for the PFC (red) and motor cortex (blue). The red horizontal line shows the temporal extent of significant correlation between context- and history-dependent information in the PFC. Lines and shaded regions show the mean and 95% CIs of bootstrapped correlation coefficients (*Methods*). (*Bottom*) Correlation between context- and history-dependent information based on the significant temporal cluster shown in the top panel of *F*. Filled dots represent individual participants; color-coded by their individual switch costs. Gray-shaded area indicates 95% CIs of the trend line. (*G*) Median split analysis based on individual switch costs (shown for accuracy) for history- (*Left*) and context-dependent information (*Right*) in the PFC. The lower horizontal black line highlights the temporal extent of significant differences between individuals with a low (yellow) vs. high (red) switch cost. Lines and shaded regions show the mean and SEM.

context and *task-history* was restricted to a low-dimensional subspace (Fig. 3*B*). The activity subspaces in PFC spanned by the first five PCs captured 97.35 ± 0.76% (mean ± SD) of the variance for *predictive context* and 97.33 ± 0.83% of the variance for *task-history* without significant differences in dimensionality between the subspaces (Fig. 3*B*; $P = 0.952$; Wilcoxon rank sum test; comparable for different numbers of components). Critically, the number of PCs needed to explain at least 95% of the variance in the data was balanced across participants (5 PCs in 13/14 participants; $P = 0.002$; Binomial test).

Activity within these low-dimensional subspaces reflects time-varying neural population dynamics predictive of past and present states. Based on the principle of communication subspaces (29), we tested whether the magnitude of partitioned information [quantified by the Euclidean distance analogous to

(30, 31)] between the feature-subspaces predicts individual switch costs. Note that we quantified the Euclidean distance within low-dimensional subspaces accounting for at least 95% of the variance in the data. The dimensionality of these subspaces was balanced across participants (5 PCs in 13/14 participants; $P = 0.002$; Binomial test). In this context, distance can be interpreted as the discriminability of neural representations between past and present states. Although both factors are encoded in the same population-wide activity, our decomposition approach allows to directly infer the discriminability between neural representation of distinct states and link it to behavior. We employed the maximum distance to quantify the discriminability between past and present states (for details, see *Methods* and *SI Appendix,* Fig. S4 for a different approach). This approach directly reflects the rationale behind commonly
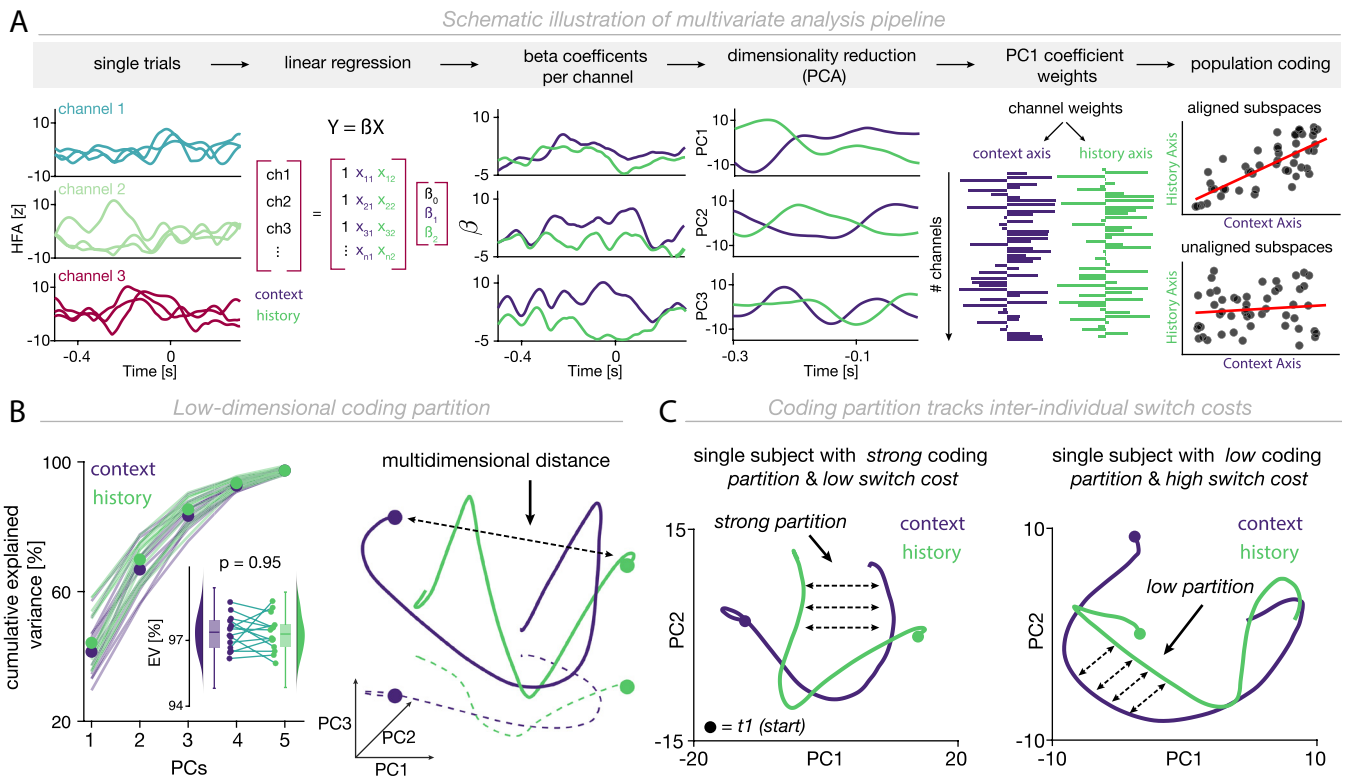
Fig. 3. Multivariate data analysis approach to identify low dimensional coding subspaces. (*A*) Schematic illustration of the multivariate analysis applied to estimate population coding subspaces (*Methods*). (*B*, *Left*) Cumulative explained variance estimated using the first five PCs for history- and context-coding subspaces. Small *Inset* illustrates that the dimensionality of context- and history-coding subspaces does not differ. (*Right*) Schematic of two coding trajectories through a three-dimensional state space. The black dotted line reflects the multidimensional distance between the two trajectories at time *t* = 1. The colored dotted lines denote the projection line onto PC1 and PC2. (*C*, *Left*) Single subject example with a low switch cost and antagonistically evolving context- and history-dependent coding trajectories projected onto the first two PCs. (*Right*) Single subject example with a high switch cost and strongly resembling coding trajectories projected onto the first two PCs (see *SI Appendix,* Fig. S5 for group-level correlation).

utilized decoding algorithms (i.e., linear discriminant analysis) that quantify the discriminability (or separability) between two categories by maximizing the distance between component axes for class separation. Therefore, we computed the time-resolved multidimensional distance between the two coding trajectories (projected into a common subspace; *Methods*) and extracted the magnitude of maximal divergence between the population coding trajectories (Fig. 3 *B*, *Right*, *SI Appendix,* Fig. S5, and *Methods*). In line with our hypothesis, a stronger coding subspace partition (less overlap between coding subspaces) predicted reduced switch costs [Fig. 3*C*; $P = 0.015$, Spearman $r = −0.64$, 95% CI = (−0.88, −0.17), $N = 14$; *SI Appendix,* Fig. S5]. We performed additional control analyses to rule out that the dimensionality of the underlying data drove the relationship between multidimensional distance and behavioral switch costs. First, we repeated the correlation analyses by fully equating the number of PCs (3 to 5 PCs) used across participants. This revealed similarly strong correlation coefficients (3 PCs: $P = 0.02$, Spearman $r = −0.62$; 4 PCs: $P = 0.011$, Spearman $r = −0.67$; 5PCs: $P = 0.015$, Spearman $r = −0.64$). Additionally, to rule out dimensionality as a confounding factor, we computed partial rank correlations between the multidimensional distance and behavioral switch cost while controlling for the number of channels. When controlling for the number of channels, our results remained unchanged ($P = 0.017$, partial Spearman $r = −0.64$). These additional control analyses demonstrate that the relationship between multidimensional distance and behavioral switch cost is in fact robust and cannot be fully explained by the electrode coverage.

Flexible Change in Population Geometry Reduces Switch Costs.
What might be a computational mechanism that enables neural ensembles to partition distinct coding features and thereby allows downstream readers to robustly decode feature-specific information (cf. Fig. 2*G*)? A viable mechanism would be a flexible, feature-dependent set of weights between upstream and downstream neural populations supporting robust read-out of feature-specific information. To test this prediction, we quantified the *subspace alignment* [Fig. 3*A* and *Methods*] between the coding subspaces accounting for variance of *past* and *present* task-features. A positively aligned subspace indicates that neural ensembles share the same low-dimensional population code (geometric orientation) for *past* and *present* states (e.g., similar local populations contributing to the respective coding). Consequently, downstream populations would not be able to untangle information about past and present states, ultimately leading to information interference between them. In line with this proposed mechanism, we observed that a stronger alignment between subspaces representing past and present states comes with a stronger individual switch cost [Fig. 4 *A–C*; $P = 0.002$, Spearman $r = 0.78$, 95% CI = (0.43, 0.93), $N = 14$]. We performed additional control analyses that supported the robustness of the observed correlation (*SI Appendix,* Figs. S6 and S7). This result based on multivariate analysis approaches provides complementary insights into the observation that competition for information-coding resources comes at a behavioral switch cost (cf. Fig. 2*G* for univariate results). Importantly, this set of results substantially extends the univariate findings by demonstrating that the transformation of population geometry might constitute a central mechanism of flexible cognitive control. Taken together,
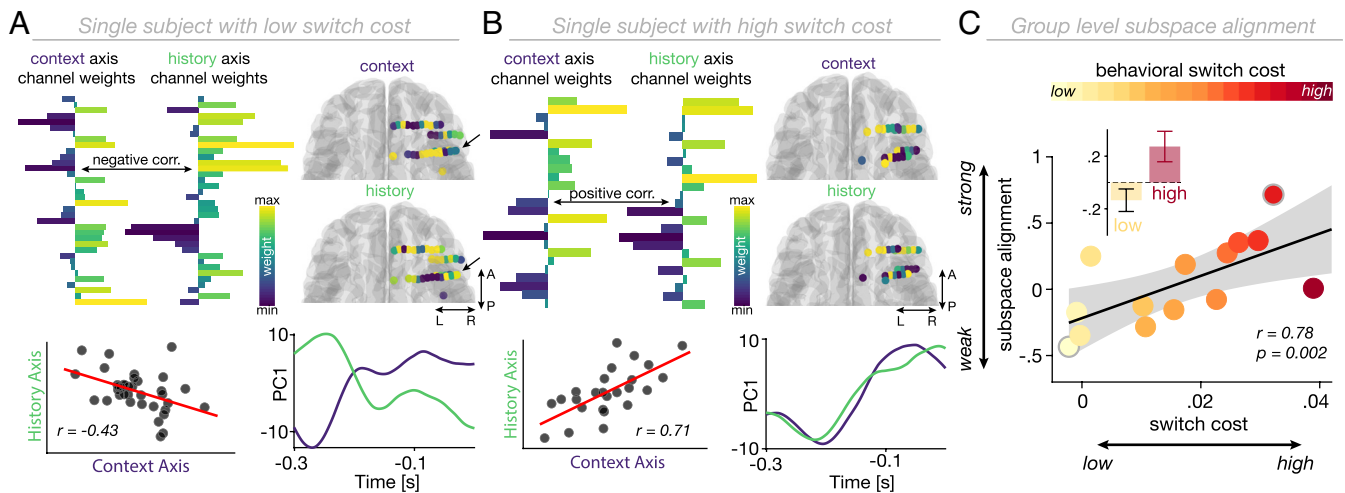
Fig. 4. Low-dimensional coding partition predicts individual switch costs. (*A*) Example of subspace alignment in the PFC for a participant with a low-behavioral switch cost. (*Top Left*) Channel weights obtained for the dominant mode (PC1) across context (*Left*) and history axes (*Right*). (*Top Right*) Channel weights obtained for the dominant mode across context (*Top*) and history axes (*Bottom*) overlaid on an individual brain. (*Bottom Left*) Strong negative correlation between channel weights projecting onto the dominant mode across context and history axes. (*Bottom Right*) Temporal evolution of low-dimensional context- and history-coding subspaces, shown for the dominant mode. Note the anticorrelated traces. (*B*) Example of subspace alignment in the PFC for a participant with low-behavioral switch costs. Same conventions as in *A*. (*C*) Group-level correlation reveals a positive relationship between subspace alignment strength and individual switch cost. Strong subspace alignment is associated with a high switch cost, whereas weak subspace alignment is associated with a low switch cost. Outlined circles in gray indicate the example subjects in *A* and *B*. The inset highlights the strength of subspace alignment based on a median split for participants with a low (yellow) and high (red) switch cost. Gray-shaded area indicates 95% CIs of the trend line.

this set of observations suggests that conjunctive coding of the past and present leads to mutual interference between the past and present in the human PFC, ultimately deteriorating behavioral performance; whereas, efficient coding in distinct population subspaces benefits behavior and enables rapid switching and might therefore, constitute a central mechanism underlying cognitive flexibility.

Discussion

Humans rapidly switch between actions to meet shifting task demands or internal goals (32). Flexibly shifting between competing task demands incurs a *switch cost* reflected in slower and more error-prone behavioral responses (10). Despite robust behavioral evidence of switch costs (33), their underlying neural mechanisms and neurocomputational principles to overcome interference and diminish switch costs remain elusive. Prior studies have implicated various regions in switching, including prefrontal (1, 21) and sensorimotor regions (34–36). Yet, the majority of evidence was obtained from single-unit recordings or local field potentials in rodents (37) or nonhuman primates (NHP) (1, 21, 38). Thus, it remains unclear whether similar principles apply to the human brain. Here, we bridged this gap using intracranial HFA recordings as a MUA surrogate from the human prefrontal and motor cortex. Our findings demonstrate that (I) neural coding of past and present states coexists in the human prefrontal and motor cortex (Fig. 2 *A–D*), and that (II) coding competition between past and present states manifests in switch costs (Fig. 2 *E–G*). Crucially, (III) the human PFC solves the interference problem by efficiently partitioning information about the past and present into distinct low-dimensional neural states (Figs. 3*C* and 4 *A* and *B*). Finally, (IV) interindividual variability of low-dimensional coding partition predicts interindividual variability in switch costs (Fig. 4*C*). In sum, these findings reveal a fundamental coding principle that might constitute a central building block of human cognitive flexibility.

Neural Coding of Past and Present Behavioral States in the Human PFC and Motor Cortex. Prior studies have demonstrated that task-relevant features are encoded in a distributed network of brain regions, including frontoparietal (25, 27, 39) and sensorimotor regions (40–42). Recently, it had been shown that the conjunctive neural representation of goal-relevant features is critical for action selection in humans (6). However, despite mounting evidence emphasizing the strong influence that past states can exert over goal-relevant present states (*serial biases*), little is known about the intricate interplay between past and present state representations in the human brain. Using direct brain recordings in humans, we demonstrate that neural information about the past and present coexists in space and time in the human prefrontal and motor cortex. Moreover, univariate information coding revealed similar temporal profiles distributed across both regions. These results further expand recent evidence obtained in NHP (15) and show that neural information about the past and present is not unique to the PFC, but is equally represented in the human motor cortex. Importantly, neural activity in NHP in the anterior-cingulate cortex (ACC) displays significantly higher as well as earlier task-selectivity after a rule-switch as compared to the lateral PFC (21). Hence, we cannot preclude that the ACC in humans also encodes information about the past state earlier or stronger than the PFC. In sum, these results demonstrate a widely distributed coding of past and present states across the prefrontal-motor hierarchy.

Competition among Past and Present States in the Human PFC Manifest in Switch Costs. When the past does not predict the future, an optimal agent should in principle discard irrelevant information about the past in order to efficiently guide currently relevant decisions. Prior studies across different species have identified several neural signatures that correlated with past events and choice history during perceptual decision-making, highlighting the key roles of the prefrontal (43, 44), parietal (8, 45) or motor (8, 46) cortex in shaping subsequent behavior. Most previous studies focused on neural signatures of past states without accounting

for coemerging neural signatures of present states. This precluded strong inference about how coexisting neural dynamics of past and present states jointly shape behavior. Here, we overcome this limitation by combining human iEEG recordings with information theoretical approaches to quantify the neural information about past and present states directly from HFA as a proxy of MUA (16). We demonstrate a behaviorally relevant dissociation between neural coding of past and present states. Specifically, the strength of neural coding of past and present information predicts individual switch costs. Strong neural coding of the present and weak neural coding of the past predicts low switch costs, whereas the opposite pattern increases switch costs. We observed that this link between neural activity and behavior was only sustained over a brief period of time, in line with previous studies that showed a short-lived representation of past states (47, 48). These results suggest that limited neural resources give rise to a short-lived biased competition between coexisting representations of past and present states: An overrepresentation of the past interferes with goal-directed behavior in the present and is detrimental for behavior. Given that our task design did not enable to fully separate reaction times from the overall error rate, a conceivable approach for future studies would involve disentangling and quantifying the impact of coexisting representations of past and present states on lapses and decision timing.

Efficient Low-Dimensional Coding Partition of Past and Present Mitigates Switch Costs.

Is there a neural mechanism that reduces mutual interference between competing representations? Theoretical work has proposed that neural populations might resolve interference by orthogonalizing competing representations (12). In this way, the same neural population encodes a competing set of stimuli, but keeps their representations separable in neural state space. This enables downstream populations to optimally decode information about a particular state. Evidence supporting this notion has been obtained in two recent animal model studies (14, 15). These studies demonstrated that both sensory [auditory cortex (14)] and association regions [medial PFC (15)] can maintain competing internal or external inputs by generating orthogonal representations. Yet, whether these representations have immediate behavioral relevance and apply to the human brain remained unaddressed. Given that large-scale single-unit recordings are not yet feasible in humans, we used HFA as a direct proxy of MUA. Recent work has demonstrated that HFA is robustly associated with population dynamics as tracked using single units (17). Thus, despite the fact that HFA reflects an aggregate metric of the underlying population activity, it constitutes a suitable level of abstraction to study low-dimensional coding representations (18, 19). Here, our results reveal that the human PFC efficiently resolves mutual interference by partitioning information about past and present states into distinct low-dimensional subspaces. Importantly, the magnitude of overlap between past and present states predicts interindividual switch costs. This demonstrates that the population activity contains multiple, behaviorally relevant latent factors that dynamically evolve over time. In sum, these results provide evidence that the segregation of competing representations into distinct population subspaces delineates an efficient coding mechanism that has immediate behavioral relevance constituting an integral component of flexible cognitive control. In the current study, we used between-subject comparisons to identify viable neural mechanisms that might explain interindividual variance in behavioral switch costs. Hence, a viable approach for future work is to employ within-subject designs to infer if these mechanisms remain stable within the individual (i.e., trait effect) or if these mechanisms can be flexibly altered (i.e., state-dependent). One

testable hypothesis directly emerging from the present results is that the interaction between subspace alignment and cognitive flexibility might be determined by the predictive value of the past. Hence, if the past is *relevant* for the present (as often in nature), integrating past and present states via subspace alignment should benefit behavior. In contrast, if the past is *irrelevant* or might even interfere with the present (as in the present experiment), segregating coding subspaces should benefit behavior.

Conclusion

Collectively, these findings uncover a fundamental computational principle how the human PFC resolves interference between competing past and present information. The results establish that competition between overlapping representations of past and present states can be reduced by partitioning state-specific information into nonoverlapping, low-dimensional coding subspaces.

Materials and Methods

Participants. We obtained intracranial recordings from a total of 19 pharma-coresistant epilepsy patients (33.73 y ± 12.52, mean ± SD; 7 females) who underwent presurgical monitoring and were implanted with intracranial depth electrodes (DIXI Medical, France). Data from one patient were excluded from neural analyses because a low-pass filter was applied at 50 Hz during data export from the clinical system, thus, precluding analyses focusing on HFA. All patients were recruited from the Department of Neurosurgery, Oslo University Hospital. Electrode implantation site was solely determined by clinical considerations and all patients provided written informed consent to participate in the study. All procedures were approved by the Regional Committees for Medical and Health Research Ethics, Region North Norway (#2015/175) and the Data Protection Officer at Oslo University Hospital as well as the University Medical Center Tuebingen (049/2020BO2) and conducted in accordance with the Declaration of Helsinki.

iEEG Data Acquisition. Intracranial EEG data were acquired at Oslo University Hospital at a sampling frequency of 512 Hz using the NicoletOne (Nicolet, Natus Neurology Inc., USA) or at a sampling frequency of 16 KHz using the ATLAS (Neuralynx) recording system.

CT and MRI Data Acquisition. We obtained anonymized postoperative computed tomography scans and presurgical MRI scans, which were routinely acquired during clinical care.

Electrode Localization. Two independent neurologists visually determined all electrode positions based on individual scans in native space. For further visualization, we reconstructed the electrode positions as outlined recently (49). In brief, the preimplant MRI and the postimplant CT were transformed into Talairach space. Then, we segmented the MRI using Freesurfer 5.3.0 (50) and coregistered the T1 to the CT. 3D electrode coordinates were determined using the Fieldtrip toolbox (51) on the CT scan. Then, we warped the aligned electrodes onto a template brain in the Montreal Neurological Institute and Hospital coordinate system for group-level analyses.

Experimental Procedure. Participants performed a predictive motor task where they had to continuously track a vertically moving target and respond as soon as the target hits or withhold their response if the target stops prior to a predefined spatial position using their dominant hand (20). Each trial started with a baseline period of 500 ms followed by a cue (presented for 800 ms centered) that informed participants about the likelihood that the moving target would stop prior to the lower limit (hit lower limit; HLL; Fig. 1A). Thus, the predictive cue could be directly translated into the probability that participants had to release the space button (BR) or withhold their response. Participants were instructed to either release the button as soon as the target hits (*go-trials*) or withhold their response if the target stops prior to the HLL (*stop-trials*). We parametrically modulated the likelihood of stopping. A green circle indicated a 0% likelihood, an orange circle indicated a 25% likelihood, and a red circle indicated a 75% likelihood that the moving target would stop prior to the HLL. Upon receiving the predictive cue, participants

were able to start the trial in a self-paced manner by pressing the space bar on the keyboard. By pressing the space bar, the target would start moving upward with constant velocity and reach the HLL after 560 to 580 ms. The upper boundary was reached after 740 to 760 ms, thus, leaving a response window of 160 ms between the HLL and the upper boundary. If participants released the button within this 160-ms interval, the trial was considered as correct. Trials in which the button was released either before or after this interval were considered as incorrect. Feedback was provided after every trial for 1,000 ms.

Behavioral Analysis. The effect of *predictive context* on behavioral performance has been extensively reported in a previous study (20). Here, we defined task-switching based on the congruency between the trial type (*stop/go-trial*) at trial *N* and at trial *N-1* (*congruent = go-trial followed by go-trial; incongruent = stop-trial followed by go-trial*). Note that, given the experimental setup, the reverse step (*go-trial followed by stop-trial*) was not feasible since our analyses required time-locking relative to the motor response, which was withheld in case of stop-trials. To account for collinearity between various factors (*SI Appendix,* Fig. S1) on behavioral performance, we employed a generalized linear mixed effect model including *current predictive context (likelihood of stop), past predictive context (likelihood of stop in past trial), task-history (stop/go-trial), past choice (button release/withhold response)* and *past feedback (correct/incorrect)* as fixed-effect predictors, *participants* as random-effects, and *response time/accuracy* as response variables.

iEEG Preprocessing. Intracranial EEG data were demeaned, linearly detrended, locally rereferenced (bipolar derivations to the next adjacent lateral contact) and if necessary down-sampled to 512 Hz. To remove line noise, data were notch-filtered at 50 Hz and all harmonics. Subsequently, a neurologist visually inspected the raw data for epileptic activity. Channels or epochs with interictal epileptic discharges and other artifacts were removed. We segmented the cleaned data into 10 s long, partially overlapping trials to prevent edge artifacts due to subsequent filtering. Unless stated otherwise, trials were event locked to the participants' response.

Extraction of HFA. HFA is broadband activity in the range from 70 to 150 Hz (also sometimes referred to as high gamma or high-frequency broadband activity) that can only be recorded intracranially, given that the scull acts as a low-pass filter. HFA has been shown to track human behavior with high spatiotemporal resolution (52–57). HFA reflects a surrogate of MUA in humans (58). Recent work demonstrated that HFA contains more behaviorally relevant information than MUA [quantified as decoding/classification accuracy of behavior; (18, 19)], since it also captures additional afferent inputs (16). In sum, HFA constitutes a highly valuable aggregate metric of population activity in humans, which enables recordings from large-scale cortical networks, which are not yet feasible at the level of single-unit and MUA in humans. The extraction of the HFA time series was conducted in a three-step process. First, we bandpass-filtered the raw data epochs (10 s) between 70 and 150 Hz into eight, nonoverlapping 10-Hz-wide bins. We then applied the Hilbert transform to obtain the instantaneous amplitude of the filtered time series. In a last step, we normalized the high-frequency traces using a bootstrapped baseline distribution (59). This involved randomly resampling baseline values (from −0.2 to −0.01s relative to cue onset) 1,000 times with replacement and normalizing single high-frequency traces by subtracting the mean and dividing by the SD of the bootstrap distribution. The eight individual high-frequency traces were then averaged to yield a single time series of high-frequency band activity.

Regions of Interest. We classified electrodes into discrete prefrontal or (pre) motor electrodes based on anatomical and functional characteristics using the human Brainnetome Atlas (60) (see *SI Appendix,* Table S1 for details about electrode classification).

Univariate Information Dynamics. We quantified the information encoded in a neural population about two main factors of interest, *predictive context* and *task-history*, using a well-established information theoretical approach (61, 62). We employed a six-way ANOVA to quantify the percentage of HFA variance explained by the following task factors: *predictive context (likelihood of stop), choice, past predictive context (likelihood of stop in last trial), task-history (stop or go trial), past choice,* and *past feedback*. Importantly, due to collinearity between the task factors, an unbalanced ANOVA (25), that implicitly orthogonalized the different factors, was employed. Thus, variance explained by *task-history* or *predictive*

context could not be explained by any other residual regressor. The amount of percent explained variance was quantified using the debiased effect size ω^2, which is defined as

$$\omega^2 = \frac{SS_{between-groups} - (df \times MSE)}{SS_{total+MSE}},$$

where SS_{total} reflects the total sum of squares across *n* trials,

$$SS_{total} = \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

$SS_{between-groups}$ the sum of squares between *G* groups (e.g. factor levels),

$$SS_{between-groups} = \sum_{group}^{G} n_{group}(\overline{x}_{group} - \overline{x})^2.$$

MSE the mean square error,

$$MSE = \sum_{i=1}^{n} (x_i - \overline{x}_{group})^2,$$

and *df* the degrees of freedom specified as $df = G - 1$. We estimated ω^2 using a sliding window of 50 ms that was shifted in steps of 2 ms to obtain a time course of neural information. This approach is insensitive with respect to the time of task-related activation and to the direction of encoding (i.e., HFA increases or decreases). Electrodes that revealed a significant main effect of *task-history* and/or *predictive context* for at least 10% of the trial length were considered as information-encoding electrodes (54). Finally, to minimize interindividual variance and maximize the sensitivity to identify a temporally consistent pattern that accounts for most of the variance across participants, we used PCA, (61). PCA was applied to the *F* value time series concatenated across participants [channel × time matrix, (20, 61)]. In order to define PCs that explain a significant proportion of variance in the data, we employed nonparametric permutation testing (1,000 iterations) to determine the proportion of variance explained by chance. Electrodes that exhibited a strong weight (75th percentile) on any of the high variance-explaining PCs were used for further analyses as outlined recently (61).

The orthogonalized computation of neural information allowed us to further quantify the similarities of information dynamics linked to *predictive context* and *task-history* in a time-resolved manner using spearman correlation. To further link the strength of regressor information to the individual switch costs, we first computed the individual switch cost by subtracting reaction time/accuracy on *no-switch-trials* from reaction time/accuracy on *switch-trials*. In a second step, we then correlated the individual switch cost with the strength of neural information related to *predictive context* or *task-history* in a time-resolved manner (see *SI Appendix,* Fig. S3 for schematic illustration of the analysis). Significance of correlation across time was assessed using cluster-based permutation testing to correct for multiple comparisons (1,000 iterations; randomly shuffling participant indices without replacement). To compare two distributions of correlation coefficients, we used Fisher's z-transformation to convert the correlation coefficients to the normally distributed variable *z*, based on which the *P*-value was derived.

Multivariate Population Dynamics. We characterized low-dimensional coding dynamics with respect to *predictive context* and *task-history* using a variant of TDR, (27). A multivariable linear regression was employed to determine how *predictive context* and *task-history* contribute to the response of every electrode in a temporally resolved manner. Here, we included all electrodes with no prior constraint on information strength to estimate latent coding dynamics across the entire sampled population. Subsequently, PCA was employed to identify low-dimensional subspaces capturing variance due to *predictive context* and *task-history*. The subsequent population analysis focused on the time window prior to the participants' response (300 ms to 0 ms; cf. Fig. 2 *E* and *F*) to maximize the temporal sensitivity of population coding analysis.

Neural coding trajectories. To verify that coding trajectories are low-dimensional, we computed the cumulative explained variance of the first 5 principal components (PCs). We then quantified the Euclidean distance *D* between coding

trajectories for *predictive context* and *task-history* in a temporally resolved manner (window size = 50 ms; shift size = 10 ms) as

$$D(Pcontext_t^n, \; Phistory_t^n),$$

where P represents a point in the context- or history-trajectory embedded in an n-dimensional space where n reflects the number of PCs required to cumulatively explain at least 95% of the total variance. The number of dimensions required to explain at least 95% of the total variance was balanced across participants (5 PCs in 13/14 participants). The Euclidean distance allows inference on how neural states dynamically evolve over time (5). Importantly, we first projected the history-subspace into the context-subspace by multiplying the beta coefficient time series (channel × time) for task-history with the PC weights obtained for the low-dimensional context-subspace (63). We then determined the maximal divergence between the two coding trajectories across time to further infer the behavioral relevance of coding separation between predictive context and task-history on a group level. We excluded participants with less than 10 electrodes per region of interest from this analysis to ensure robust PCA estimates. We used the maximum distance approach given that our univariate analyses revealed only transient time periods of competition between past and present states that were not sustained in nature (cf. Fig. 2E). Hence, averaging would likely obscure subtle, time-sensitive changes in neural trajectories. However, we repeated this analysis by using multiple time-windows around the maximum distance (±50 ms and ±100 ms) to ensure that our maximum distance approach did not capture random changes in activation while still being sensitive to time-periods in which the neural representation between past and present states diverged (*SI Appendix,* Fig. S4).

Subspace alignment analysis. In a final step, we computed the alignment of low-dimensional coding subspaces. We used the coefficient matrix based on the PCA approach outlined above, which contains the individual channel weights per PC. We quantified alignment A between low-dimensional context- and history-coding subspaces as the correlation coefficient (Spearman's Rho) between vector C containing the weights of all electrodes n contributing to the subspace defined by *predictive context* and vector H containing the weights of all electrodes n contributing to the subspace defined by *task-history* $A = corr(C_n, H_n)$. We excluded participants with less than 10 electrodes per region of interest from this analysis.

Statistics. Unless stated otherwise, we used nonparametric cluster–based permutation testing (64) to analyze data in the time domain (Fig. 2 A, B, E, and F).

Neural information. Clusters for neural information time series (Fig. 2 A, B, and G) were formed by thresholding a dependent (Fig. 2 A and B) or independent (Fig. 2G) t test at a critical alpha of 0.05. We generated a permutation distribution by randomly shuffling encoding vs. nonencoding electrode labels (Fig. 2 A and B) or condition labels (Fig. 2G) and recomputing the cluster statistic. The permutation P-value was obtained by comparing the cluster statistic to the random permutation distribution. Clusters were considered to be significant at $P < 0.05$.

Correlation analyses. Clusters for correlation time series (Fig. 2 E and F) were formed by thresholding the resulting correlation P-values at 0.05. We generated a permutation distribution by randomly shuffling participant labels and recomputing the cluster statistic. The permutation P-value was obtained by comparing the cluster statistic to the random permutation distribution. Clusters were considered to be significant at $P < 0.05$.

Author affiliations: [a]Hertie Institute for Clinical Brain Research, Center for Neurology, University Medical Center Tübingen, 72076 Tübingen, Germany; [b]International Max Planck Research School for the Mechanisms of Mental Function and Dysfunction, University of Tübingen, 72076 Tübingen, Germany; [c]Department of Psychology, University of Oslo, 0373 Oslo, Norway; [d]RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, 0373 Oslo, Norway; [e]Department of Neurosurgery, Oslo University Hospital, 0372 Oslo, Norway; [f]Department of Neuropsychology, Helgeland Hospital, 8657 Mosjøen, Norway; [g]Helen Wills Neuroscience Institute, UC Berkeley, Berkeley, CA 94720; and [h]Department of Psychology, UC Berkeley, Berkeley, CA 94720

1. T. J. Buschman, E. L. Denovellis, C. Diogo, D. Bullock, E. K. Miller, Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron* **76**, 838–846 (2012).
2. R. F. Helfrich, R. T. Knight, Oscillatory dynamics of prefrontal cognitive control. *Trends Cogn. Sci.* **20**, 916–930 (2016).
3. M. Rigotti *et al.*, The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
4. A. Parthasarathy *et al.*, Mixed selectivity morphs population codes in prefrontal cortex. *Nat. Neurosci.* **20**, 1770–1779 (2017).
5. R. B. Ebitz, B. Y. Hayden, The population doctrine in cognitive neuroscience. *Neuron* **109**, 3055–3068 (2021), 10.1016/j.neuron.2021.07.011.
6. A. Kikumoto, U. Mayr, Conjunctive representations that integrate stimuli, responses, and rules are critical for action selection. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10603–10608 (2020).
7. K. G. Guise, M. L. Shapiro, Medial prefrontal cortex reduces memory interference by modifying hippocampal encoding. *Neuron* **94**, 183–192.e8 (2017).
8. A. E. Urai, T. H. Donner, Persistent activity in human parietal cortex mediates perceptual choice repetition bias. *Nat. Commun.* **13**, 6015 (2022).
9. K. Desender, A. Boldt, T. Verguts, T. H. Donner, Confidence predicts speed-accuracy tradeoff for subsequent decisions. *Elife* **8**, e43499 (2019).
10. S. Monsell, Task switching. *Trends Cogn. Sci.* **7**, 134–140 (2003).
11. O. L. White, D. D. Lee, H. Sompolinsky, Short-term memory in orthogonal neural networks. *Phys. Rev. Lett.* **92**, 148102 (2004).
12. F. Bouchacourt, T. J. Buschman, A flexible model of working memory. *Neuron* **103**, 147–160.e8 (2019).
13. D. Badre, A. Bhandari, H. Keglovits, A. Kikumoto, The dimensionality of neural representations for control. *Curr. Opin. Behav. Sci.* **38**, 20–28 (2021).
14. A. Libby, T. J. Buschman, Rotational dynamics reduce interference between sensory and memory representations. *Nat. Neurosci.* **24**, 715–726 (2021).
15. S. Maggi, M. D. Humphries, Activity subspaces in medial prefrontal cortex distinguish states of the world. *J. Neurosci.* **42**, 4131–4146 (2022).
16. M. Leszczynski *et al.*, Dissociation of broadband high-frequency activity and neuronal firing in the neocortex. *Sci. Adv.* **6**, eabb0977 (2020).
17. C. Gallego-Carracedo, M. G. Perich, R. H. Chowdhury, L. E. Miller, J. A. Gallego, Local field potentials reflect cortical population dynamics in a region-specific and frequency-dependent manner. *Elife* **11**, e73155 (2022).
18. S. T. Kanth, S. Ray, Electrocorticogram (ECoG) is highly informative in primate visual cortex. *J. Neurosci.* **40**, 2430–2444 (2020).
19. E. L. Rich, J. D. Wallis, Spatiotemporal dynamics of information encoding revealed in orbitofrontal high-gamma. *Nat. Commun.* **8**, 1139 (2017).
20. J. Weber *et al.*, Population coding and oscillatory subspace synchronization integrate context into actions. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2021.12.17.473118 (Accessed 10 January 2022).
21. K. Johnston, H. M. Levin, M. J. Koval, S. Everling, Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. *Neuron* **53**, 453–462 (2007).
22. A. A. Pape, M. Siegel, Motor cortex activity predicts response alternation during sensorimotor decisions. *Nat. Commun.* **7**, 13098 (2016).
23. M. J. Imburgio, J. M. Orr, Component processes underlying voluntary task selection: Separable contributions of task-set inertia and reconfiguration. *Cognition* **212**, 104685 (2021).
24. T. J. Buschman, M. Siegel, J. E. Roy, E. K. Miller, Neural substrates of cognitive capacity limitations. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11252–11255 (2011).
25. M. Siegel, T. J. Buschman, E. K. Miller, Cortical information flow during flexible sensorimotor decisions. *Science* **348**, 1352–1355 (2015).
26. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate–A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
27. V. Mante, D. Sussillo, K. V. Shenoy, W. T. Newsome, Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
28. Z. Fu *et al.*, The geometry of domain-general performance monitoring in the human medial frontal cortex. *Science* **376**, eabm9922 (2022).
29. A. Kohn *et al.*, Principles of corticocortical communication: Proposed schemes and design considerations. *Trends Neurosci.* **43**, 725–737 (2020).
30. R. B. Ebitz, E. Albarran, T. Moore, Exploration disrupts choice-predictive signals and alters dynamics in prefrontal cortex. *Neuron* **97**, 475 (2018).
31. J. A. Michaels, B. Dann, R. W. Intveld, H. Scherberger, Neural dynamics of variable grasp-movement preparation in the macaque frontoparietal network. *J. Neurosci.* **38**, 5759–5773 (2018).
32. E. K. Miller, J. D. Cohen, An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
33. L. Qiao, L. Zhang, A. Chen, T. Egner, Dynamic trial-by-trial recoding of task-set representations in the frontoparietal cortex mediates behavioral flexibility. *J. Neurosci.* **37**, 11037–11050 (2017).

34. S. A. Bunge, I. Kahn, J. D. Wallis, E. K. Miller, A. D. Wagner, Neural circuits subserving the retrieval and maintenance of abstract rules. *J. Neurophysiol.* **90**, 3419–3428 (2003).

35. A. B. Smith, E. Taylor, M. Brammer, K. Rubia, Neural correlates of switching set as measured in fast, event-related functional magnetic resonance imaging. *Hum. Brain Mapp.* **21**, 247–256 (2004).

36. F. M. Korb, J. Jiang, J. A. King, T. Egner, Hierarchically organized medial frontal cortex-basal ganglia loops selectively control task- and response-selection. *J. Neurosci.* **37**, 7893–7905 (2017).

37. C. A. Duan, J. C. Erlich, C. D. Brody, Requirement of prefrontal and midbrain regions for rapid executive control of behavior in the rat. *Neuron* **86**, 1491–1503 (2015).

38. C. R. Hussar, T. Pasternak, Flexibility of sensory representations in prefrontal cortex depends on cell type. *Neuron* **64**, 730–743 (2009).

39. D. A. Crowe et al., Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nat. Neurosci.* **16**, 1484–1491 (2013).

40. T. H. Donner, M. Siegel, P. Fries, A. K. Engel, Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Curr. Biol.* **19**, 1581–1585 (2009).

41. H. Nienborg, B. G. Cumming, Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature* **459**, 89–92 (2009).

42. D. Thura, P. Cisek, Deliberation and commitment in the premotor and primary motor cortex during dynamic decision making. *Neuron* **81**, 1401–1416 (2014).

43. A. Genovesio, S. Tsujimoto, G. Navarra, R. Falcone, S. P. Wise, Autonomous encoding of irrelevant goals and outcomes by prefrontal cortex neurons. *J. Neurosci.* **34**, 1970–1978 (2014).

44. G. Mochol, R. Kiani, R. Moreno-Bote, Prefrontal cortex represents heuristics that shape choice bias and its integration into future behavior. *Curr. Biol.* **31**, 1234–1244.e6 (2021).

45. E. J. Hwang, J. E. Dahlen, M. Mukundan, T. Komiyama, History-based action selection bias in posterior parietal cortex. *Nat. Commun.* **8**, 1242 (2017).

46. F. P. de Lange, D. A. Rahnev, T. H. Donner, H. Lau, Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *J. Neurosci.* **33**, 1400–1410 (2013).

47. G. Ranieri, A. Benedetto, H. T. Ho, D. C. Burr, M. C. Morrone, Evidence of serial dependence from decoding of visual evoked potentials. *J. Neurosci.* **42**, 8817–8825 (2022).

48. B. O. Rangel, E. Hazeltine, J. R. Wessel, Lingering neural representations of past task features adversely affect future behavior. *J. Neurosci.* **43**, 282–292 (2023).

49. A. Stolk et al., Integrated analysis of anatomical and electrophysiological human intracranial data. *Nat. Protoc.* **13**, 1699–1723 (2018).

50. A. M. Dale, B. Fischl, M. I. Sereno, Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).

51. R. Oostenveld, P. Fries, E. Maris, J. M. Schoffelen, FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).

52. E. Edwards, M. Soltani, L. Y. Deouell, M. S. Berger, R. T. Knight, High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *J. Neurophysiol.* **94**, 4269–4280 (2005).

53. R. T. Canolty et al., High gamma power is phase-locked to theta oscillations in human neocortex. *Science* **313**, 1626–1628 (2006).

54. B. Voytek et al., Oscillatory dynamics coordinating human frontal networks in support of goal maintenance. *Nat. Neurosci.* **18**, 1318–1324 (2015).

55. M. Haller et al., Persistent neuronal activity in human prefrontal cortex links perception and action. *Nat. Hum. Behav.* **2**, 80–91 (2018).

56. Y. Oganian, E. F. Chang, A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Sci. Adv.* **5**, eaay6279 (2019).

57. N. Natraj, D. B. Silversmith, E. F. Chang, K. Ganguly, Compartmentalized dynamics within a common multi-area mesoscale manifold represent a repertoire of human hand movements. *Neuron* **110**, 154–174.e12 (2022).

58. S. Ray, J. H. Maunsell, Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* **9**, e1000610 (2011).

59. R. F. Helfrich et al., Neural mechanisms of sustained attention are rhythmic. *Neuron* **99**, 854–865.e5 (2018).

60. L. Fan et al., The human brainnetome atlas: A new brain atlas based on connectional architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).

61. S. Durschmid et al., Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6755–6760 (2016).

62. I. Saez et al., Encoding of multiple reward-related computations in transient and sustained high-frequency activity in human OFC. *Curr. Biol.* **28**, 2889–2899.e3 (2018).

63. S. B. M. Yoo, B. Y. Hayden, The transition from evaluation to selection involves neural subspace reorganization in core reward regions. *Neuron* **105**, 712–724.e4 (2020).

64. E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).

65. J. Weber et al., *PNAS* Subspace Partitioning. *GitHub*. https://github.com/JanWeber-neuro/PNAS-Subspace-Partitioning. Deposited 19 June 2023.