1	Subspace partitioning in human prefrontal cortex resolves
2	cognitive interference
3	Jan Weber ^{1,2} , Gabriela Iwama ^{1,2} , Anne-Kristin Solbakk ^{3,4,5,6} , Alejandro O. Blenkmann ^{3,4} ,
4 5	Pal G. Larsson ⁵ , Jugoslav Ivanovic ⁵ , Robert T. Knight ^{7,8} , Tor Endestad ^{3,4} , Randolph Helfrich ^{1,*}
6 7	¹ Hertie Institute for Clinical Brain Research, Center for Neurology, University Medical Center Tübingen, Tübingen, Germany.
8 9 10	 ² International Max Planck Research School for the Mechanisms of Mental Function and Dysfunction, University of Tübingen, Tübingen, Germany. ³ Department of Revebulary, University of Oslo, Oslo, Nonvoy.
10 11 12 13	 ⁴ RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway. ⁵ Department of Neurosurgery, Oslo University Hospital, Oslo, Norway. ⁶ Department of Neuropsychology, Helgeland Hospital, Mosjøen, Norway.
14 15 16 17	 ⁷ Helen Wills Neuroscience Institute, UC Berkeley, Berkeley, CA, USA. ⁸ Department of Psychology, UC Berkeley, Berkeley, CA, USA.
18	*Corresponding author. Email: randolph.helfrich@gmail.com
20 21 22 23	Correspondence Randolph F. Helfrich, Hertie Institute for Clinical Brain Research, Center for Neurology, University Medical Center Tübingen, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany, phone: +49 7071 29 89189, fax: +49 7071 29 4488, eMail: randolph.helfrich@gmail.com
24	
25	
26	
27	
28	

1 Abstract

2	Human prefrontal cortex (PFC) constitutes the structural basis underlying flexible cognitive control,
3	where mixed-selective neural populations encode multiple task-features to guide subsequent
4	behavior. The mechanisms by which the brain simultaneously encodes multiple task-relevant
5	variables while minimizing interference from task-irrelevant features remain unknown. Leveraging
6	intracranial recordings from the human PFC, we first demonstrate that competition between co-
7	existing representations of past and present task variables incurs a behavioral switch cost. Our
8	results reveal that this interference between past and present states in the PFC is resolved through
9	coding partitioning into distinct low-dimensional neural states; thereby strongly attenuating
10	behavioral switch costs. In sum, these findings uncover a fundamental coding mechanism that
11	constitutes a central building block of flexible cognitive control.
12	
13	
14	
15	
16	
17	
18	
19	
20	
20	
21	
22	
23	

1 Introduction

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.

8 Multiple lines of research in humans and animal models provide converging evidence that 9 the prefrontal cortex (PFC) constitutes the key structure that enables cognitive flexibility to guide adaptive goal-directed behavior¹⁻³. To support flexible operations, PFC neurons are not feature-10 11 specific, but instead exhibit mixed selectivity and context-dependent coding of task features (i.e. 12 sensory stimuli, abstract rules or actions)⁴⁻⁷. Mixed-selective neurons contribute to multiple 13 cognitive operations by participating in different transient coalitions of cell assemblies. 14 Consequently, information encoded by mixed-selective neurons can only be understood at the level 15 of population activity. This notion is now referred to as the population doctrine, which posits that 16 neural populations reflect the fundamental unit of computation in the brain^{8,9}. The population 17 doctrine further postulates that rapid switches between distinct cognitive operations can efficiently 18 be implemented by adjusting the population geometry, i.e. the re-configuration of cell assemblies. 19 A key advantage of this model is that neural representations that evolve in parallel can be integrated into a unified, conjunctive representation¹⁰⁻¹². In support of this hypothesis, recent findings 20 21 demonstrated that conjunctive coding schemes enable the flexible context-dependent remapping between sensory inputs and behavioral outputs¹²; predicting behavioral performance on a trial-by-22 23 trial basis¹⁰.

24 However, parallel encoding of different variables implies that not all encoded features are 25 behaviorally-relevant, hence, raising the question if and how encoding of task-irrelevant (latent) 26 factors impacts subsequent behavior. Previously, it had been observed that past choices reflect typical latent factors, which impact current task-relevant representations¹³⁻¹⁶. At the behavioral level 27 previous choices modulate subsequent behavior (i.e. serial response bias^{13,16-18}). Integration of 28 29 knowledge about past choices is oftentimes desirable to correct past mistakes¹⁹, but might cause 30 interference when consecutive choices are independent, thus, giving rise to behavioral switch 31 costs²⁰⁻²². In this scenario, a conjunctive neural code that integrates information about the past is detrimental for behavioral performance¹⁵. To date, little is known about how the human brain 32 33 integrates multiple task-relevant representations, while minimizing interference from task-34 irrelevant, latent factors, in order to guide subsequent behavior. Theoretical work has proposed 35 that neural populations can reduce interference by orthogonalizing competing representations^{23,24}. 36 This change in population geometry enables downstream regions to flexibly readout information 37 about competing states from the same neural population¹². Evidence in support of this notion stems

from a recent study in mouse auditory cortex that demonstrated that past and present sensory representations reside in orthogonal subspaces²⁵. 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 higher-order association cortex and whether orthogonalization constitutes a key mechanism to reduce interference between competing neural representations²⁶.

7 Here, we addressed these outstanding questions by directly recording intracranial 8 electroencephalography (iEEG) from the human prefrontal and motor cortex while participants 9 performed a modified stop-signal task that enabled disentangling how competing neural 10 representations between past states and current goals influence human decision-making. We 11 leveraged the power of analyzing high-frequency activity (HFA) as a direct approximation of local neural population activity²⁷⁻²⁹. We specifically tested if neural population activity in PFC and motor 12 13 cortex simultaneously encode information about the past and present. The key question was 14 whether overlapping neural representations account for behavioral switch costs between 15 movement inhibition and execution. We hypothesized that efficient distributed computing at the 16 level of population activity constituted a core mechanism to avoid interference between distinct 17 latent factors. Collectively, we tested whether efficient cortical partitioning of distinct neural 18 representations enables flexible cognitive control.

19

20 Results

21 We recorded intracranial EEG (iEEG) from 19 pharmaco-resistant patients with epilepsy (33.73 22 years ± 12.52, mean ± SD; 7 females) who performed a predictive motor task (Fig. 1a). Participants 23 were instructed to closely track a vertically moving target and respond as soon as the moving target 24 reached a predefined spatial location (go-trials). In a subset of trials, the moving target stopped 25 prematurely and participants were instructed to withhold their response (stop-trials). A contextual 26 cue indicated the likelihood of a premature stop (referred to as predictive context; Methods). 27 Previously, we demonstrated that the human PFC integrates current contextual information to guide 28 goal-directed behavior³⁰. Additionally, human behavior is also strongly modulated by a variety of 29 latent factors, such as the tendency to systematically repeat choices, known as history-dependent serial biases^{13,17,31}. We here investigated the neural mechanisms underlying such history-30 31 dependent serial biases, quantified as the behavioral switch costs between distinct action demands 32 initiation or inhibition of a goal-directed movement. Thus, task-switching was defined based on 33 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). 34

35

36 Shifting between competing task demands has a behavioral switch cost

37 To assess whether task-switching modulated behavioral performance, we contrasted reaction time,

1 accuracy and d-prime (d') between switch- and no-switch-trials. Participants were significantly 2 slower (P = 0.0006; Cohen's d = 1.29; +19.02 ± 14.73ms; mean ± SD; Wilcoxon rank sum test; **Fig.** 3 **1b**) and made significantly more mistakes after *switch*- as compared to *no-switch-trials* (*P* = 0.0016; 4 Cohen's d = -0.9; +8.99 ± 9.94%; mean ± SD; Fig. 1b). Furthermore, task-switching across 5 successive trials reduced participants' sensitivity to correctly decide between two action 6 alternatives (d'; P = 0.0012; Cohen's d = -1.04; Fig. 1b). We further confirmed that our behavioral 7 analyses were not confounded by regressor collinearity (Methods; Supplementary Fig. 1). We 8 observed a main effect of predictive context and task-history on reaction time (task-history: 95% CI 9 $= [0.016 \ 0.026], P < 0.0001; predictive context: 95% CI = [0.0005 \ 0.0006], P < 0.0001) and accuracy$ (task-history: 95% CI = [-0.99 -0.37], P < 0.0001; predictive context: 95% CI = [-0.03 -0.02], P < 10 11 0.0001). Thus, predictive context and task-history independently modulated behavior. 12 13



14 15

Experimental design, behavioral performance and schematic hypotheses. a. Top: participants performed 16 a predictive motor task where they had to track a vertically moving target and respond as soon as the target 17 reached a pre-defined lower limit (go-trials: hit-lower limit, HLL: black horizontal line). At the start of every trial. 18 participants received a contextual cue indicating the likelihood (0, 25 and 75% likelihood; green, orange, red

19 circle, respectively) that the moving target would stop prematurely requiring participants to withhold their 20

response (stop-trials). Feedback was provided at the end of each trial. Bottom: task-switching was defined as

1 the trial-type (stop/go-trial) congruency between two successive trials (Methods). b, Behavioral performance 2 as a function of task-switching. Left: reaction time was significantly increased after switch trials (go-trial 3 preceded by a stop-trial) as compared to no-switch trials (go-trial preceded by a go-trial). Accuracy (middle) 4 5 and *d-prime (right)* significantly decreased after switch as compared to no-switch trials. Grey lines display individual participants, density-plots display the data distribution and boxplots show the median (horizontal 6 7 line), the first/third quartile (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 8 active top-down control processes in the prefrontal cortex to reconfigure the cognitive system. In the second 9 scenario, switch costs could reflect persistent inhibition (passive inertia) of motor areas after withholding a 10 response.

11

12 Overlapping neural populations encode past and present states in human PFC

13 To dissect the neural mechanisms underlying behavioral switch costs, we simultaneously recorded 14 neural activity from electrodes located in the human PFC (26 ± 18.74 electrodes per participant; 15 mean \pm SD) and motor cortex (14.83 \pm 15.23 electrodes). Based on theoretical models³²⁻³⁴, two 16 possible scenarios were considered: (1) Switch costs could reflect the time needed to engage 17 active top-down control processes in PFC (Fig. 1c) or alternatively, (2) switch costs might result 18 from prolonged inhibition of motor cortex between switching from movement-inhibition to 19 movement-execution (Fig. 1c). In order to differentiate these models, we quantified the univariate 20 neural information (unsigned, bias-corrected percent explained variance; Methods) about task-21 history and predictive context in PFC and motor cortex. We orthogonalized the different factors to 22 disentangle their unique behavioral relevance. We observed significant context- and history-23 dependent neural information in PFC (**Fig. 2a-d**; *task-history*: $P_{cluster} < 0.0001$, Cohen's d = 1.81; 24 predictive context: $P_{cluster} < 0.0001$, Cohen's d = 2.07; cluster permutation test) and motor cortex 25 (task-history: $P_{cluster} < 0.0001$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$, Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$; Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$; Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$; Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$; Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$; Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$; Cohen's d = 1.46; predictive context: $P_{cluster} = 0.005$; Predictive context: $P_{cluster} = 0.005$ 26 2.41).

Having established a robust coding of *predictive context* and *task-history* in the prefrontal-motor network, we then investigated when neural information coding predicted individual switch costs in behavior using a sliding window correlation analysis (*Methods*). We hypothesized anti-correlated 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.

33 In line with our hypothesis, strong neural coding of context predicted lower switch costs, whereas 34 strong neural coding of task-history indexed increased behavioral costs (Fig. 2e; Supplementary 35 Fig. S2). Importantly, a significant neuro-behavioral correlation was only observed in PFC 36 (predictive context: P = 0.039, r = 0.59; task-history: P = 0.013, r = -0.57; cluster permutation test), 37 but not in motor cortex (predictive context: no cluster; task-history: P = 0.434). Furthermore, this 38 association differed significantly between predictive context and task-history for PFC (P = 0.028; motor cortex: P = 0.854; FDR-corrected³⁵; black horizontal line in Fig. 2e; Methods). 39 40 Next, we quantified the link between information coding of predictive context and task-history (Fig.

41 **2f**; *Methods*), which revealed a robust negative correlation between context- and history-dependent

- 1 neural information in PFC (P = 0.004, r = -0.64; cluster permutation test), but not in motor cortex (P
- 2 = 0.227). Note, however, that the correlation between PFC and motor cortex did not significantly
- 3 differ (*P* = 0.183; *Methods*).
- 4 To further illustrate the group-level correlation between information coding and behavioral switch
- 5 costs, we performed a median split analysis (**Fig. 2g**). Participants with low switch costs revealed
- 6 stronger neural coding for predictive context, but weaker neural coding for task-history. The
- 7 opposite relationship was observed for participants with high switch costs (predictive context: Pcluster
- 8 = 0.026, Cohen's d = -1.6; *task-history*: $P_{cluster} = 0.084$, Cohen's d = 1.49; cluster permutation).
- 9 Collectively, these findings indicate that limited neural resources impose a competition between
- 10 feature-coding of past and present states and that over-representation of past states comes at
- 11 substantial behavioral costs.
- 12



13 14

15 Fig. 2. Behavioral dissociation of neural dynamics encoding the past and present. a, Left: Time course 16 of context-dependent information averaged across all context-encoding electrodes in PFC. Right: Time course 17 of history-dependent neural information averaged across all history-encoding electrodes in PFC. Lines and 18 shaded regions show the mean and SEM. Grey traces indicate the time course of neural information across 19 non-encoding electrodes. The lower horizontal black line shows the temporal extent of significant neural 20 information. b, Same as (a), but for motor cortex. c, Context-encoding electrodes overlaid on a standardized 21 brain in MNI coordinates for PFC (red) and motor cortex (blue). Overall, 165 electrodes in PFC and 89 22 electrodes in motor cortex carried significant context-dependent information. d. History-encoding electrodes, 23 same conventions as in (c). Task-history was significantly encoded in 96 electrodes in PFC and 48 electrodes 24 in motor cortex. e, Left: Temporally resolved correlation between neural information and individual switch costs

(accuracy; see Supplementary Fig. 2 for reaction time) for PFC (left) and motor cortex (right). Lines and 1 2 shaded regions show the mean and 95% CIs of bootstrapped correlation coefficients (Methods). The lower 3 colored horizontal lines show the temporal extent of significant correlation for context- (purple) and history-4 dependent (green) neural information. The black line shows the temporal extent of significant differences in 5 neuro-behavioral correlation between predictive context and -history. f, Top: Temporally resolved correlation 6 7 between context- and history-dependent neural information for PFC (red) and motor cortex (blue). The red horizontal line shows the temporal extent of significant correlation between context- and history-dependent 8 information in PFC. Bottom: correlation between context- and history-dependent information based on the 9 significant temporal cluster shown in the top panel of (f). Filled dots represent individual participants; colorcoded by their individual switch costs. g, Median split analysis based on individual switch costs (shown for 10 11 accuracy) for history- (left) and context-dependent information (right) in PFC. The lower horizontal black line 12 highlights the temporal extent of significant differences between individuals with a low (yellow) versus high 13 (red) switch cost. Lines and shaded regions show the mean and SEM. The small insets depict the averaged 14 neural information across significant clusters for individuals with high versus low switch costs.

15

16 **Competitive coding of past and present states in human PFC predicts behavior**

17 Having established that neural information coding about past and present states is anti-correlated and exerts dissociable effects on behavior, we tested whether a competitive coding scheme 18 19 between past and present states in the human PFC could account for individual switch costs. Based 20 on the highly distributed nature of feature-coding in human PFC (cf. Fig. 2c/d), a multivariate data 21 analysis approach was used to test this prediction (Fig. 3a; Methods). We employed a variant of 22 targeted-dimensionality reduction (TDR³⁶). In brief, linear regression was used to determine how 23 predictive context and task-history modulate neural activity at every electrode. Subsequently, low-24 dimensional subspaces that capture context- and history-dependent variance in neural activity, 25 were identified using principal component analysis (PCA). Consistent with previous findings 26 demonstrating low-dimensional neural representations of task features³⁶⁻³⁸, we observed that 27 neural coding of predictive context and task-history was restricted to a low-dimensional subspace 28 (Fig. 3b). The activity subspaces in PFC spanned by the first five PCs captured $97.35 \pm 0.76\%$ 29 (mean \pm SD) of the variance for predictive context and 97.33 \pm 0.83% of the variance for task-30 history without significant differences in dimensionality between the subspaces (Fig. 3b; Wilcoxon 31 rank sum test, P = 0.952; comparable for different numbers of components). 32 Activity within these subspaces reflects time-varying neural population dynamics predictive of past and present states. Based on the principle of communication subspaces³⁹, we tested whether the 33 34 magnitude of partitioned information between the feature-subspaces predicts individual switch 35 costs. 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. 3b** right; *Methods*). In line

38 with our hypothesis, a stronger coding subspace partition (less overlap between coding subspaces)

- 39 predicted reduced switch costs (**Fig. 3c**; P = 0.015, Spearman r = -0.64, N = 14; **Supplementary**
- 40 Fig. 3).





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 (see 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 **Supplementary Fig. 3** for group-level correlation)

13 14

10

11

12

15

16 Flexible change in population geometry reduces switch costs

17 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 18 19 information? A viable mechanism would be a flexible, feature-dependent set of weights between 20 upstream and downstream neural populations supporting robust read-out of feature-specific 21 information. To test this prediction, we quantified the subspace alignment (Fig. 3a; Methods) 22 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 23 24 population code for past and present states (e.g. similar local populations contributing to the 25 respective coding). Consequently, downstream populations would not be able to untangle 26 information about past and present states, ultimately leading to information interference between 27 them. In line with this proposed mechanism, we observed that a stronger alignment between 28 subspaces representing past and present states comes with a stronger individual switch cost (P = 29 0.002, Spearman r = 0.78, N = 14; Fig. 4a-c). Taken together, this set of observations suggests

- 1 that conjunctive coding of the past and present leads to mutual interference between the past and
- 2 present in the human PFC, ultimately deteriorating behavioral performance; whereas, efficient
- 3 coding in distinct population subspaces benefits behavior and enables rapid switching and might
- 4 therefore, constitute a central mechanism underlying cognitive flexibility.
- 5
- 6



7 8 Fig. 4. Low-dimensional coding partition predicts individual switch costs. a, Example of subspace 9 alignment in PFC for a participant with a low behavioral switch cost. Top Left: Channel weights obtained for 10 the dominant mode (PC1) across context (left) and history axes (right). Top Right: Channel weights obtained 11 for the dominant mode across context (top) and history axes (bottom) overlaid on an individual brain. Bottom 12 Left: Strong negative correlation between channel weights projecting onto the dominant mode across context 13 and history axes. Bottom Right: Temporal evolution of low-dimensional context- and history-coding 14 subspaces, shown for the dominant mode. Note the anticorrelated traces. b, Example of subspace alignment 15 in PFC for a participant with low behavioral switch costs. Same conventions as in (a), c. Group-level correlation 16 reveals a positive relationship between subspace alignment strength and individual switch cost. Strong 17 subspace alignment is associated with a high switch cost whereas weak subspace alignment is associated 18 with a low switch cost. Outlined circles in grey indicate the example subjects in (a) and (b). The inset highlights 19 the strength of subspace alignment based on a median split for participants with a low (yellow) and high (red) 20 switch cost. 21

22

23 Discussion

24 Humans rapidly switch between actions to meet shifting task demands or internal goals⁴⁰. Flexibly 25 shifting between competing task demands incurs a switch cost reflected in slower and more error-26 prone behavioral responses^{21,41}. Despite robust behavioral evidence of switch costs^{22,41-43}, their 27 underlying neural mechanisms and neurocomputational principles to overcome interference and 28 diminish switch costs remain elusive. Prior studies have implicated various regions in switching, including prefrontal^{1,32,42,44,45} and sensorimotor regions⁴⁶⁻⁴⁹. Yet, the majority of evidence was 29 30 obtained from single unit recordings or local field potentials in rodents⁴⁵ or non-human 31 primates^{1,32,50-52}. Thus, it remains unclear if similar principles apply to the human brain. Here, we 32 bridged this gap using intracranial recordings from the human prefrontal and motor cortex. Our 33 findings demonstrate that (I) neural coding of past and present states coexists in the human 34 prefrontal and motor cortex (Fig. 2a-d), and that (II) coding competition between past and present

states manifests in switch costs (Fig. 2e-g). Crucially, (III) the human PFC solves the interference problem by efficiently partitioning information about the past and present into distinct lowdimensional neural states (Fig. 3c; Fig. 4a-b). Finally, (IV) inter-individual variability of lowdimensional coding partition predicts inter-individual variability in switch costs (Fig. 4c). In sum, these findings reveal a fundamental coding principle that might constitute a central building block of human cognitive flexibility.

7 Neural coding of past and present behavioral states in human PFC and motor cortex

8 Prior studies have demonstrated that task-relevant features are encoded in a distributed network of brain regions, including frontoparietal^{36,53-55} and sensorimotor regions⁵⁶⁻⁵⁸. Recently, it had been 9 shown that the conjunctive neural representation of goal-relevant features is critical for action 10 11 selection in humans¹⁰. However, despite mounting evidence emphasizing the strong influence that 12 past states can exert over goal-relevant present states (serial biases), little is known about the 13 intricate interplay between past and present state representations in the human brain. Using direct 14 brain recordings in humans, we demonstrate that neural information about the past and present 15 co-exist in space and time in the human prefrontal and motor cortex. Moreover, univariate 16 information coding revealed similar temporal profiles distributed across both regions. These results go substantially beyond recent evidence obtained in animal models²⁶ and show that neural 17 18 information about the past and present is not unique to the PFC, but is equally represented in the 19 human motor cortex. In sum, these results demonstrate a widely distributed coding of past and 20 present states across the prefrontal-motor hierarchy.

21 Competition among past and present states in the human PFC manifest in switch costs

22 When the past does not predict the future, an optimal agent should in principle discard irrelevant 23 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 24 and choice history during perceptual decision-making, highlighting the key roles of prefrontal^{59,60}, 25 parietal^{18,61,62} or motor^{18,63} cortex in shaping subsequent behavior. Most previous studies focused 26 27 on neural signatures of past states without accounting for co-emerging neural signatures of present 28 states. This precluded strong inference about how co-existing neural dynamics of past and present 29 states jointly shape behavior. Here, we overcome this limitation by combining human iEEG 30 recordings with information theoretical approaches to quantify the neural information about past and present states directly from HFA as a proxy of multi-unit activity^{27,28,64}. We demonstrate a 31 32 behaviorally relevant dissociation between neural coding of past and present states. Specifically, 33 the strength of neural coding of past and present information predicts individual switch costs. Strong 34 neural coding of the present and weak neural coding of the past predicts low switch costs whereas 35 the opposite pattern increases switch costs. These results suggest that limited neural resources 36 give rise to a biased competition between co-existing representations of past and present states:

1 An over-representation of the past interferes with goal-directed behavior in the present and is

2 detrimental for behavior.

3 Efficient low-dimensional coding partition of past and present mitigates switch costs

4 Is there a neural mechanism that reduces mutual interference between competing representations? 5 Theoretical work has proposed that neural populations might resolve interference by orthogonalizing competing representations^{23,24}. In this way, the same neural population encodes a 6 7 competing set of stimuli, but keeps their representations separable in neural state space. This 8 enables downstream populations to optimally decode information about a particular state. Evidence 9 supporting this notion has been obtained in two recent animal model studies^{25,26}. These studies demonstrated that both sensory (auditory cortex²⁵) and association regions (medial PFC²⁶) can 10 11 maintain competing internal or external inputs by generating orthogonal representations. Yet, 12 whether these representations have immediate behavioral relevance remained unaddressed. Here, 13 our results reveal that the human PFC efficiently resolves mutual interference by partitioning 14 information about past and present states into distinct low-dimensional subspaces. Importantly, the 15 magnitude of overlap between past and present states predicts inter-individual switch costs. These 16 results provide evidence that the segregation of competing representations into distinct population 17 subspaces delineates an efficient coding mechanism that has immediate behavioral relevance 18 constituting an integral component of flexible cognitive control.

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 non-overlapping, low-dimensional coding subspaces.

1 Materials and Methods

2 **Participants.** We obtained intracranial recordings from a total of 19 pharmaco-resistant epilepsy 3 patients (33.73 years ± 12.52, mean ± SD; 7 females) who underwent presurgical monitoring and 4 were implanted with intracranial depth electrodes (DIXI Medical, France). Data from one patient 5 were excluded from neural analyses because a low-pass filter was applied at 50 Hz during data 6 export from the clinical system, thus, precluding analyses focusing on high-frequency activity. All 7 patients were recruited from the Department of Neurosurgery, Oslo University Hospital. Electrode 8 implantation site was solely determined by clinical considerations and all patients provided written 9 informed consent to participate in the study. All procedures were approved by the Regional 10 Committees for Medical and Health Research Ethics, Region North Norway (#2015/175) and the 11 Data Protection Officer at Oslo University Hospital as well as the University Medical Center 12 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 CT scans and pre-surgical
 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 recently⁶⁵. In brief, the pre-implant MRI and the post-implant CT were transformed into Talairach space. Then we segmented the MRI using Freesurfer 5.3.0⁶⁶ and coregistered the T1 to the CT. 3D electrode coordinates were determined using the Fieldtrip toolbox⁶⁷ on the CT scan. Then we warped the aligned electrodes onto a template brain in MNI space for group-level analyses.

25 **Experimental Procedure.** Participants performed a predictive motor task where they had to 26 continuously track a vertically moving target and respond as soon as the target hits or withhold their 27 response if the target stops prior a predefined spatial position using their dominant hand³⁰. Each 28 trial started with a baseline period of 500ms followed by a cue (presented for 800ms centered) that 29 informed participants about the likelihood that the moving target would stop prior to the lower limit 30 (hit lower limit; HLL; Fig. 1a). Thus, the predictive cue could be directly translated into the probability 31 that participants had to release the space button (BR) or withhold their response. Participants were 32 instructed to either release the button as soon as the target hits (*go-trials*) or withhold their response 33 if the target stops prior to the HLL (stop-trials). We parametrically modulated the likelihood of 34 stopping. A green circle indicated a 0% likelihood, an orange circle indicated a 25% likelihood and 35 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 upwards with constant velocity and reach the HLL after 560 – 580ms. The upper boundary was reached after 740 – 760ms, thus, leaving a response window of 160ms between the HLL and the upper boundary. If participants released the button within this 160ms 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 1000ms.

8 Behavioral Analysis. The effect of *predictive context* on behavioral performance has been 9 extensively reported in a previous study³⁰. Here, we defined task-switching based on the 10 congruency between the trial type (stop/go-trial) at trial N and at trial N-1 (congruent = go-trial 11 followed by go-trial; incongruent = stop-trial followed by go-trial). Note that, given the experimental 12 setup, the reverse step (go-trial followed by stop-trial) was not feasible since our analyses required 13 time-locking relative to the motor response, which was withheld in case of stop-trials. To account 14 for collinearity between various factors (Supplementary Fig. 1) on behavioral performance, we 15 employed a generalized linear mixed effect model including current predictive context (likelihood of 16 stop), past predictive context (likelihood of stop in past trial), task-history (stop/go-trial), past choice 17 (button release/withhold response) and past feedback (correct/incorrect) as fixed-effect predictors, 18 participants as random-effects and response time/accuracy as response variables.

iEEG Preprocessing. Intracranial EEG data were demeaned, linearly de-trended, 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 (IEDs) and other artifacts were removed. We segmented the cleaned data into 10 seconds long, partially overlapping trials to prevent edge artifacts due to subsequent filtering. Unless stated otherwise, trials were event locked to the participants' response.

26 Extraction of High-Frequency Activity (HFA). The extraction of the high-frequency activity time 27 series was conducted in a three-step process. First, we bandpass-filtered the raw data epochs (10 28 seconds) between 70-150 Hz into eight, non-overlapping 10 Hz wide bins. We then applied the 29 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^{68,69}. This involved 30 31 randomly resampling baseline values (from -0.2 to -0.01s relative to cue onset) 1000 times with 32 replacement and normalizing single high-frequency traces by subtracting the mean and dividing by 33 the standard deviation of the bootstrap distribution. The eight individual high-frequency traces were 34 then averaged to yield a single time series of high-frequency band activity.

1 Regions of Interest. We classified electrodes into discrete prefrontal or (pre-)motor electrodes

2 based on anatomical and functional characteristics using the human Brainnetome Atlas⁷⁰ (see

3 **Supplementary Table 1** for details about electrode classification).

4 **Univariate Information Dynamics.** We quantified the information encoded in a neural population 5 about two main factors of interest, predictive context and task-history, using a well-established information theoretical approach^{55,71-73}. We employed a six-way analysis of variance (ANOVA) to 6 7 quantify the percentage of HFA variance explained by the following task factors: predictive context 8 (likelihood of stop), choice, past predictive context (likelihood of stop in last trial), task-history (stop 9 or go trial), past choice, past feedback. Importantly, due to collinearity between the task factors, an 10 unbalanced ANOVA⁵⁵, that implicitly orthogonalized the different factors, was employed. Thus, 11 variance explained by task-history or predictive context could not be explained by any other residual 12 regressor. The amount of percent explained variance was quantified using the debiased effect size 13 ω^2 , which is defined as

14
$$\omega^2 = \frac{SS_{between-groups} - (df \times MSE)}{SS_{total} + MSE}$$

15 where *SS*_{total} reflects the total sum of squares across *n* trials,

16
$$SS_{total} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

17 *SS*_{between-aroups} the sum of squares between *G* groups (e.g. factor levels),

18
$$SS_{between-groups} = \sum_{group}^{G} n_{group} (\bar{x}_{group} - \bar{x})^2$$

19 *MSE* the mean square error,

20
$$MSE = \sum_{i=1}^{n} (x_i - \bar{x}_{group})^2$$

and *df* the degrees of freedom specified as df = G - 1. We estimated ω^2 using a sliding window 21 22 of 50ms that was shifted in steps of 2ms to obtain a time course of neural information. This 23 approach is insensitive with respect to the time of task-related activation and to the direction of 24 encoding (i.e., HFA increases or decreases). Electrodes that revealed a significant main effect of 25 task-history and/or predictive context for at least 10% of the trial length were considered as 26 information-encoding electrodes^{68,71,74,75}. Finally, to minimize inter-individual variance and 27 maximize the sensitivity to identify a temporally consistent pattern that accounts for most of the variance across participants, we used principal component analysis (PCA)^{72,75}. PCA was applied 28

1 to the *F* value time series concatenated across participants (channel x time matrix^{30,72,76}). In order 2 to define PCs that explain a significant proportion of variance in the data, we employed non-3 parametric permutation testing (1000 iterations) to determine the proportion of variance explained 4 by chance. Electrodes that exhibited a strong weight (75th percentile) on any of the high variance-5 explaining PCs were used for further analyses as outlined recently⁷². 6 The orthogonalized computation of neural information allowed us to further quantify the similarities 7 of information dynamics linked to predictive context and task-history in a time-resolved manner 8 using spearman correlation. To further link the strength of regressor information to the individual 9 switch costs, we first computed the individual switch cost by subtracting reaction time/accuracy on 10 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. Significance of correlation across time was assessed using cluster-based permutation testing to correct for multiple comparisons (1000 iterations; randomly shuffling participant indices without replacement). To compare two distributions of correlation coefficients, we used Fisher's z-transformation to convert Pearson's *r* to the normally distributed variable *z*, based on which the *p-value* was derived.

17

18 Multivariate Population Dynamics. We characterized low-dimensional coding dynamics with 19 respect to predictive context and task-history using a variant of targeted-dimensionality reduction 20 (TDR³⁶). A multi-variable linear regression was employed to determine how predictive context and 21 task-history contribute to the response of every electrode in a temporally resolved manner. Here, 22 we included all electrodes with no prior constraint on information strength to estimate latent coding 23 dynamics across the entire sampled population. Subsequently, principal component analysis (PCA) 24 was employed to identify low-dimensional subspaces capturing variance due to predictive context 25 and task-history. The subsequent population analysis focused on the time window prior to the 26 participants' response (300ms to 0ms; cf. Fig 2e/f) to maximize the temporal sensitivity of 27 population coding analysis.

<u>Neural coding trajectories.</u> 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 = 50ms; shift size = 10ms) as

32

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

where *P* represents a point in the context- or history-trajectory embedded a n-dimensional space where *n* reflects the number of PCs required to cumulatively explain at least 95% of the total variance. Importantly, we first projected the *history-subspace* into the *context-subspace*^{77,78}. We then determined the magnitude of maximal divergence between the two coding trajectories across

1 time to further infer the behavioral relevance of coding separation between *predictive context* and

2 task-history on a group level. We excluded participants with less than 10 electrodes per ROI from

3 this analysis to ensure robust PCA estimates.

4 Subspace alignment analysis. In a final step, we computed the alignment of low-dimensional coding 5 subspaces. We used the coefficient matrix based on the PCA approach outlined above which 6 contains the individual channel weights per principal component. We quantified alignment A 7 between low-dimensional context- and history-coding subspaces as the correlation coefficient 8 (Spearman's Rho) between vector C containing the weights of all electrodes n contributing to the 9 subspace defined by predictive context and vector H containing the weights of all electrodes n 10 contributing to the subspace defined by task-history $A = corr(C_n, H_n)$. We excluded participants with less than 10 electrodes per ROI from this analysis. 11

12

Statistics. Unless stated otherwise, we used non-parametric cluster-based permutation testing⁷⁹
 to analyze data in the time domain (Fig. 2a/b/e/f).

15 <u>Neural information</u>. Clusters for neural information time series (**Fig. 2a/b/g**) were formed by 16 thresholding a dependent (**Fig. 2a/b**) or independent (**Fig. 2g**) t-test at a critical alpha of 0.05. We 17 generated a permutation distribution by randomly shuffling encoding vs. non-encoding electrode 18 labels (**Fig. 2a/b**) or condition labels (**Fig. 2g**) and recomputing the cluster statistic. The 19 permutation p-value was obtained by comparing the cluster statistic to the random permutation 20 distribution. Clusters were considered to be significant at P < 0.05.

21 <u>Correlation analyses.</u> Clusters for correlation time series (**Fig. 2e/f**) were formed by thresholding 22 the resulting correlation p-values at 0.05. We generated a permutation distribution by randomly 23 shuffling participant labels and recomputing the cluster statistic. The permutation p-value was 24 obtained by comparing the cluster statistic to the random permutation distribution. Clusters were 25 considered to be significant at P < 0.05.

26

- 28
- 29
- 30
- 31
- 32

1 Acknowledgments

2 We thank Anais Llorens and Ingrid Funderud for their help with data collection. This work was

- 3 funded by the Baden Wuerttemberg Foundation (Postdoc Fellowship; RFH), German Research
- 4 Foundation, Emmy Noether Program (DFG HE8329/2-1; RFH), Hertie Foundation, Network for
- 5 Excellence in Clinical Neuroscience (RFH), the International Max Planck Research School for the
- 6 Mechanisms of Mental Function and Dysfunction (GI), the Research Council of Norway (grant
- 7 number 240389; AKS, TE, PGL; grant number: 314925; AOB), the Research Council of Norway
- 8 (Centre of Excellence scheme, grant number 262762; RITMO, RITPART International Partnerships
- 9 for RITMO Centre of Excellence, grant number 274996; AKS, TE, PGL) and by a NIMH Conte
- 10 Center Grant (1 PO MH109429, RTK) and the NINDS (2 R01 NS021135, RTK).
- 11

12 Author contributions

- 13 Conceptualization: JW, GI, RFH
- 14 Methodology: JW, GI
- 15 Investigation: AKS, TE, PL, JI, AOB
- 16 Visualization: JW, GI
- 17 Funding acquisition: AKS, RTK, TE, RFH
- 18 Project administration: TE, AKS, RFH
- 19 Supervision: RFH
- 20 Writing original draft: JW, RFH
- 21 Writing review & editing: GI, AKS, RTK, TE, AOB, JI, PL
- 22

23 **Competing interests**

- 24 The authors declare no competing financial interests.
- 25

26 Data availability

- 27 Source data is included as Supporting Information. Raw data are available upon request from
- 28 Anne-Kristin Solbakk (a.k.solbakk@psykologi.uio.no) or Tor Endestad (tendesta@uio.no).
- 29

30 **Code availability**

- 31 Freely available software and algorithms used for analysis are listed where applicable. All code will
- 32 be made publicly available upon publication on GitHub.
- 33
- . .
- 34
- 35
- 36 37

1	References
2	

1 Buschman, T. J., Denovellis, E. L., Diogo, C., Bullock, D. & Miller, E. K. Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. Neuron 76, 838-846, doi:10.1016/j.neuron.2012.09.029 (2012). 2 Stokes, M. G. et al. Dynamic coding for cognitive control in prefrontal cortex. Neuron 78, 364-375, doi:10.1016/j.neuron.2013.01.039 (2013). Helfrich, R. F. & Knight, R. T. Oscillatory Dynamics of Prefrontal Cognitive Control. Trends Cogn Sci 20, 916-3 930, doi:10.1016/j.tics.2016.09.007 (2016). 4 Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. Nature 497, 585-590, doi:10.1038/nature12160 (2013). Fusi, S., Miller, E. K. & Rigotti, M. Why neurons mix: high dimensionality for higher cognition. Curr Opin 5 Neurobiol 37, 66-74, doi:10.1016/j.conb.2016.01.010 (2016). 6 Raposo, D., Kaufman, M. T. & Churchland, A. K. A category-free neural population supports evolving demands during decision-making. Nat Neurosci 17, 1784-1792, doi:10.1038/nn.3865 (2014). 7 Parthasarathy, A. et al. Mixed selectivity morphs population codes in prefrontal cortex. Nat Neurosci 20, 1770-1779, doi:10.1038/s41593-017-0003-2 (2017). Ebitz, R. B. & Hayden, B. Y. The population doctrine in cognitive neuroscience. Neuron, 8 doi:10.1016/j.neuron.2021.07.011 (2021). 9 Yuste, R. From the neuron doctrine to neural networks. Nat Rev Neurosci 16, 487-497. doi:10.1038/nrn3962 (2015). 10 Kikumoto, A. & Mayr, U. Conjunctive representations that integrate stimuli, responses, and rules are critical for action selection. Proc Natl Acad Sci U S A 117, 10603-10608, doi:10.1073/pnas.1922166117 (2020). Cellier, D., Petersen, I. T. & Hwang, K. Dynamics of Hierarchical Task Representations. J Neurosci. 11 doi:10.1523/JNEUROSCI.0233-22.2022 (2022). 12 Badre, D., Bhandari, A., Keglovits, H. & Kikumoto, A. The dimensionality of neural representations for control. Curr Opin Behav Sci 38, 20-28, doi:10.1016/j.cobeha.2020.07.002 (2021). 13 Urai, A. E., de Gee, J. W., Tsetsos, K. & Donner, T. H. Choice history biases subsequent evidence accumulation. Elife 8, doi:10.7554/eLife.46331 (2019). Sul. J. H., Kim. H., Huh. N., Lee, D. & Jung, M. W. Distinct roles of rodent orbitofrontal and medial prefrontal 14 cortex in decision making. Neuron 66, 449-460, doi:10.1016/j.neuron.2010.03.033 (2010). 15 Guise, K. G. & Shapiro, M. L. Medial Prefrontal Cortex Reduces Memory Interference by Modifying Hippocampal Encoding. Neuron 94, 183-192 e188, doi:10.1016/j.neuron.2017.03.011 (2017). 16 Pape, A. A. & Siegel, M. Motor cortex activity predicts response alternation during sensorimotor decisions. Nat Commun 7, 13098, doi:10.1038/ncomms13098 (2016). 17 Braun, A., Urai, A. E. & Donner, T. H. Adaptive History Biases Result from Confidence-Weighted Accumulation of past Choices. J Neurosci 38, 2418-2429, doi:10.1523/JNEUROSCI.2189-17.2017 (2018). 18 Urai, A. E. & Donner, T. H. Persistent activity in human parietal cortex mediates perceptual choice repetition bias. Nat Commun 13, 6015, doi:10.1038/s41467-022-33237-5 (2022). 19 Desender, K., Boldt, A., Verguts, T. & Donner, T. H. Confidence predicts speed-accuracy tradeoff for subsequent decisions. Elife 8, doi:10.7554/eLife.43499 (2019). 20 Rogers, R. D. & Monsell, S. Costs of a Predictable Switch between Simple Cognitive Tasks. J Exp Psychol Gen 124, 207-231, doi:Doi 10.1037/0096-3445.124.2.207 (1995). Monsell, S. Task switching. Trends Cogn Sci 7, 134-140, doi:10.1016/s1364-6613(03)00028-7 (2003). 21 22 Yeung, N., Nystrom, L. E., Aronson, J. A. & Cohen, J. D. Between-task competition and cognitive control in task switching. J Neurosci 26, 1429-1438, doi:10.1523/JNEUROSCI.3109-05.2006 (2006). White, O. L., Lee, D. D. & Sompolinsky, H. Short-term memory in orthogonal neural networks. Phys Rev Lett 23 92, 148102, doi:10.1103/PhysRevLett.92.148102 (2004). Bouchacourt, F. & Buschman, T. J. A Flexible Model of Working Memory. Neuron 103, 147-160 e148, 24 doi:10.1016/j.neuron.2019.04.020 (2019). Libby, A. & Buschman, T. J. Rotational dynamics reduce interference between sensory and memory 25 representations. Nat Neurosci 24, 715-726, doi:10.1038/s41593-021-00821-9 (2021). Maggi, S. & Humphries, M. D. Activity Subspaces in Medial Prefrontal Cortex Distinguish States of the World. J 26 Neurosci 42, 4131-4146, doi:10.1523/JNEUROSCI.1412-21.2022 (2022). 27 Rich, E. L. & Wallis, J. D. Spatiotemporal dynamics of information encoding revealed in orbitofrontal highgamma. Nat Commun 8, 1139, doi:10.1038/s41467-017-01253-5 (2017). 28 Leszczynski, M. et al. Dissociation of broadband high-frequency activity and neuronal firing in the neocortex. Sci Adv 6, eabb0977, doi:10.1126/sciadv.abb0977 (2020). Gallego-Carracedo, C., Perich, M. G., Chowdhury, R. H., Miller, L. E. & Gallego, J. A. Local field potentials 29 reflect cortical population dynamics in a region-specific and frequency-dependent manner. Elife 11, doi:10.7554/eLife.73155 (2022). Weber, J., Solbakk, A. K., Blenkmann, A. O., Llorens, A., Funderud, I., Leske, S., Larsson, P. G., Ivanovic, J., 30 Knight, R. T., Endestad, T., Helfrich, R. F. Population coding and oscillatory subspace synchronization integrate context into actions. bioRxiv (2022).

64 (2022).
65 31 Urai, A. E., Braun, A. & Donner, T. H. Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nat Commun* 8, 14637, doi:10.1038/ncomms14637 (2017).

1 2 3	32	Johnston, K., Levin, H. M., Koval, M. J. & Everling, S. Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. <i>Neuron</i> 53 , 453-462, doi:10.1016/j.neuron.2006.12.023 (2007).
4 5	33	Orban de Xivry, J. J. & Lefevre, P. A switching cost for motor planning. <i>J Neurophysiol</i> 116 , 2857-2868, doi:10.1152/in.00319.2016 (2016)
6 7	34	Imburgio, M. J. & Orr, J. M. Component processes underlying voluntary task selection: Separable contributions of task-set inertia and reconfiguration. <i>Cognition</i> 212 , 104685, doi:10.1016/j.cognition.2021.104685 (2021).
8 9	35	Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. <i>J R Stat Soc B</i> 57 , 289-300, doi:DOI 10.1111/j.2517-6161.1995.tb02031.x (1995).
10 11	36	Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. <i>Nature</i> 503 , 78-84, doi:10.1038/nature12742 (2013).
12 13	37	Xie, Y. <i>et al.</i> Geometry of sequence working memory in macaque prefrontal cortex. <i>Science</i> 375 , 632-639, doi:10.1126/science.abm0204 (2022).
14 15	38	Fu, Z. <i>et al.</i> The geometry of domain-general performance monitoring in the human medial frontal cortex. <i>Science</i> 376 , eabm9922, doi:10.1126/science.abm9922 (2022).
16 17 19	39	Kohn, A. <i>et al.</i> Principles of Corticocortical Communication: Proposed Schemes and Design Considerations. <i>Trends Neurosci</i> 43 , 725-737, doi:10.1016/j.tins.2020.07.001 (2020).
18	40	Miller, E. K. & Conen, J. D. An integrative theory of prefrontal cortex function. <i>Annu Rev Neurosci</i> 24 , 167-202, doi:10.1146/annurev.neuro.24.1.167 (2001).
20 21 22	41	Allport, A., Styles, E. A. & Hslen, S. L. Shirting Intentional Set - Exploring the Dynamic Control of Tasks. Attention Perform 15 , 421-452 (1994).
22 23 24	42	Neurosci 29, 5135-5142, doi:10.1523/JNEUROSCI.2828-08.2009 (2009).
25 26	45	Frontoparietal Cortex Mediates Behavioral Flexibility. <i>J Neurosci</i> 37 , 11037-11050, doi:10.1523/JNEUROSCI.0935-17.2017 (2017).
27 28	44	Badre, D. & Wagner, A. D. Computational and neurobiological mechanisms underlying cognitive flexibility. <i>Proc</i> <i>Natl Acad Sci U S A</i> 103 , 7186-7191, doi:10.1073/pnas.0509550103 (2006).
29 30 21	45	Duan, C. A., Erlich, J. C. & Brody, C. D. Requirement of Prefrontal and Midbrain Regions for Rapid Executive Control of Behavior in the Rat. <i>Neuron</i> 86 , 1491-1503, doi:10.1016/j.neuron.2015.05.042 (2015).
31 32 33	40	parietal cortex in task switching. <i>Proc Natl Acad Sci U S A</i> 97 , 13448-13453, doi:10.1073/pnas.240460497 (2000)
34 35	47	Bunge, S. A., Kahn, I., Wallis, J. D., Miller, E. K. & Wagner, A. D. Neural circuits subserving the retrieval and maintenance of abstract rules. <i>J Neurophysiol</i> 90 , 3419-3428, doi:10.1152/in.00910.2002 (2003).
36 37	48	Smith, A. B., Taylor, E., Brammer, M. & Rubia, K. Neural correlates of switching set as measured in fast, event- related functional magnetic resonance imaging. <i>Hum Brain Mapp</i> 21 , 247-256, doi:10.1002/hbm.20007 (2004).
38 39 40	49	Korb, F. M., Jiang, J., King, J. A. & Egner, T. Hierarchically Organized Medial Frontal Cortex-Basal Ganglia Loops Selectively Control Task- and Response-Selection. <i>J Neurosci</i> 37 , 7893-7905, doi:10.1523/JNEUROSCI.3289.16.2017 (2017)
41 42	50	Stoet, G. & Snyder, L. H. Single neurons in posterior parietal cortex of monkeys encode cognitive set. <i>Neuron</i> 42 , 1003-1012, doi:10.1016/j.neuron.2004.06.003 (2004).
43 44	51	Mansouri, F. A., Buckley, M. J. & Tanaka, K. Mnemonic function of the dorsolateral prefrontal cortex in conflict- induced behavioral adjustment. <i>Science</i> 318 , 987-990, doi:10.1126/science.1146384 (2007).
45 46	52	Hussar, C. R. & Pasternak, T. Flexibility of sensory representations in prefrontal cortex depends on cell type. <i>Neuron</i> 64 , 730-743, doi:10.1016/j.neuron.2009.11.018 (2009).
47 48 40	53	Freedman, D. J. & Assad, J. A. Experience-dependent representation of visual categories in parietal cortex. <i>Nature</i> 443 , 85-88, doi:10.1038/nature05078 (2006).
49 50	54	Crowe, D. A. <i>et al.</i> Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. <i>Nat Neurosci</i> 16 , 1484-1491, doi:10.1038/nn.3509 (2013).
51 52 52	55	Siegel, M., Buschman, T. J. & Miller, E. K. Cortical information flow during flexible sensorimotor decisions. Science 348, 1352-1355, doi:10.1126/science.aab0551 (2015).
55 54	50	during perceptual decision making. Curr Biol 19 , 1581-1585, doi:10.1016/j.cub.2009.07.066 (2009).
55 56 57	58	causal effect. <i>Nature</i> 459 , 89-92, doi:10.1038/nature07821 (2009).
57 58 59	50	decision making. <i>Neuron</i> 81 , 1401-1416, doi:10.1016/j.neuron.2014.01.031 (2014).
60 61	55	and outcomes by prefrontal cortex neurons. <i>J Neurosci</i> 34 , 1970-1978, doi:10.1523/JNEUROSCI.3228-13.2014 (2014).
62 63	60	Mochol, G., Kiani, R. & Moreno-Bote, R. Prefrontal cortex represents heuristics that shape choice bias and its integration into future behavior. <i>Curr Biol</i> 31 , 1234-1244 e1236, doi:10.1016/j.cub.2021.01.068 (2021).
64 65	61	Hwang, E. J., Dahlen, J. E., Mukundan, M. & Komiyama, T. History-based action selection bias in posterior parietal cortex. <i>Nat Commun</i> 8 , 1242, doi:10.1038/s41467-017-01356-z (2017).
66 67	62	Hwang, E. J. <i>et al.</i> Corticostriatal Flow of Action Selection Bias. <i>Neuron</i> 104 , 1126-1140 e1126, doi:10.1016/j.neuron.2019.09.028 (2019).
68 69	63	de Lange, F. P., Rahnev, D. A., Donner, T. H. & Lau, H. Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. <i>J Neurosci</i> 33 , 1400-1410, doi:10.1523/JNEUROSCI.1094-12.2013 (2013).

- 64 Ray, S. & Maunsell, J. H. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol 9, e1000610, doi:10.1371/journal.pbio.1000610 (2011).
- 65 Stolk, A. et al. Integrated analysis of anatomical and electrophysiological human intracranial data. Nat Protoc 13, 1699-1723, doi:10.1038/s41596-018-0009-6 (2018).
- 66 Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179-194, doi:10.1006/nimg.1998.0395 (1999).
- 67 Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011, 156869, doi:10.1155/2011/156869 (2011).
- Helfrich, R. F. et al. Neural Mechanisms of Sustained Attention Are Rhythmic. Neuron 99, 854-865 e855, 68 doi:10.1016/j.neuron.2018.07.032 (2018).
- 69 Johnson, E. L. et al. Dynamic frontotemporal systems process space and time in working memory. PLoS Biol 16, e2004274, doi:10.1371/journal.pbio.2004274 (2018).
- 70 Fan, L. et al. The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. Cereb Cortex 26, 3508-3526, doi:10.1093/cercor/bhw157 (2016).
- 71 Voytek, B. et al. Oscillatory dynamics coordinating human frontal networks in support of goal maintenance. Nat Neurosci 18, 1318-1324, doi:10.1038/nn.4071 (2015).
- Durschmid, S. et al. Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. 72 Proc Natl Acad Sci U S A 113, 6755-6760, doi:10.1073/pnas.1525030113 (2016).
- Saez, I. et al. Encoding of Multiple Reward-Related Computations in Transient and Sustained High-Frequency 73 Activity in Human OFC. Curr Biol 28, 2889-2899 e2883, doi:10.1016/j.cub.2018.07.045 (2018).
- 74 Haller, M. et al. Persistent neuronal activity in human prefrontal cortex links perception and action. Nat Hum Behav 2, 80-91, doi:10.1038/s41562-017-0267-2 (2018).
- 75 Kam, J. W. Y. et al. Top-Down Attentional Modulation in Human Frontal Cortex: Differential Engagement during External and Internal Attention. Cereb Cortex 31, 873-883, doi:10.1093/cercor/bhaa262 (2021).
- 123456789011234567890122345678901233 1111315678901223456789012333333 76 Cohen, M. X. Analyzing Neural Time Series Data: Theory and Practice. Iss Clin Cogn Neurop, 1-578 (2014). Elsayed, G. F., Lara, A. H., Kaufman, M. T., Churchland, M. M. & Cunningham, J. P. Reorganization between 77 preparatory and movement population responses in motor cortex. Nat Commun 7, 13239,
 - doi:10.1038/ncomms13239 (2016).
- 78 Yoo, S. B. M. & Havden, B. Y. The Transition from Evaluation to Selection Involves Neural Subspace
- Reorganization in Core Reward Regions. Neuron 105, 712-724 e714, doi:10.1016/j.neuron.2019.11.013 (2020). 79 Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164, 177-190, doi:10.1016/j.jneumeth.2007.03.024 (2007).