

Monkey Prefrontal Neurons Reflect Logical Operations for Cognitive Control in a Variant of the AX Continuous Performance Task (AX-CPT)

Rachael K. Blackman,^{1,2,4} David A. Crowe,^{2,3} Adele L. DeNicola,^{1,2} Sofia Sakellaridi,^{2,6} Angus W. MacDonald, III,^{5,6} and Matthew V. Chafee^{1,2,6}

¹Department of Neuroscience, ⁴Medical Scientist Training Program (MD/PhD), ⁵Department of Psychology, and ⁶Center for Cognitive Sciences, University of Minnesota, Minneapolis, Minnesota 55455, ²Brain Sciences Center, Veterans Administration Medical Center, Minneapolis, Minnesota 55417, and ³Department of Biology, Augsburg College, Minneapolis, Minnesota 55454

Cognitive control is the ability to modify the behavioral response to a stimulus based on internal representations of goals or rules. We sought to characterize neural mechanisms in prefrontal cortex associated with cognitive control in a context that would maximize the potential for future translational relevance to human neuropsychiatric disease. To that end, we trained monkeys to perform a dot-pattern variant of the AX continuous performance task that is used to measure cognitive control impairment in patients with schizophrenia (MacDonald, 2008; Jones et al., 2010). Here we describe how information processing for cognitive control in this task is related to neural activity patterns in prefrontal cortex of monkeys, to advance our understanding of how behavioral flexibility is implemented by prefrontal neurons in general, and to model neural signals in the healthy brain that may be disrupted to produce cognitive control deficits in schizophrenia. We found that the neural representation of stimuli in prefrontal cortex is strongly biased toward stimuli that inhibit prepotent or automatic responses. We also found that population signals encoding different stimuli were modulated to overlap in time specifically in the case that information from multiple stimuli had to be integrated to select a conditional response. Finally, population signals relating to the motor response were biased toward less frequent and therefore less automatic actions. These data relate neuronal activity patterns in prefrontal cortex to logical information processing operations required for cognitive control, and they characterize neural events that may be disrupted in schizophrenia.

Key words: context processing; macaque; neural activity; prefrontal; primate; schizophrenia

Significance Statement

Functional imaging studies have demonstrated that cognitive control deficits in schizophrenia are associated with reduced activation of the dorsolateral prefrontal cortex (MacDonald et al., 2005). However, these data do not reveal how the disease has disrupted the function of prefrontal neurons to produce the observed deficits in cognitive control. Relating cognitive control to neurophysiological signals at a cellular level in prefrontal cortex is a necessary first step toward understanding how disruption of these signals could lead to cognitive control failure in neuropsychiatric disease. To that end, we translated a task that measures cognitive control deficits in patients with schizophrenia to monkeys and describe here how neural signals in prefrontal cortex relate to performance.

Introduction

Cognitive control is the ability to use rules or goals to modify the response to a given stimulus, making behavior flexible and rap-

idly adaptive to changing contexts (Goldman-Rakic, 1987; Miller and Cohen, 2001; Schiffer et al., 2015). As such, cognitive control incorporates two essential capacities: (1) the ability to represent

Received Sept. 24, 2015; revised Jan. 27, 2016; accepted Feb. 15, 2016.

Author contributions: R.K.B. and M.V.C. designed research; R.K.B. and S.S. performed research; R.K.B., D.A.C., A.L.D., and M.V.C. analyzed data; R.K.B., D.A.C., A.L.D., S.S., A.W.M., and M.V.C. wrote the paper.

This work was supported by the National Institutes of Health Grants R01MH077779 and R01MH107491, the Department of Veterans Affairs, the American Brain Sciences Chair, the Wilfred Wetzel Graduate Fellowship, and Medical Scientist Training Program at the University of Minnesota Grant T32 GM008244. We thank C. Dean Evans for excellent technical assistance in surgeries and neural recording as well as animal care; Dale Boeff for computer

programming and engineering support; and Dr. Theoden Netoff for consulting on the demixed principal components analysis. The views and opinions expressed in this paper are those of the authors and not those of the United States Federal Government.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Matthew V. Chafee, Brain Sciences Center (11B), Veterans Administration Medical Center, 1 Veterans Drive, Minneapolis, MN 55417. E-mail: chafee001@umn.edu.

DOI:10.1523/JNEUROSCI.3578-15.2016

Copyright © 2016 the authors 0270-6474/16/364067-13\$15.00/0

different internal states (such as goals or rules); and (2) the ability to use internal state information to modify the mapping between stimulus and motor representations to implement behavioral flexibility. Neural recording studies in nonhuman primates trained to perform rule-based tasks have provided insight into the neural mechanisms of both aspects of cognitive control. For example, prefrontal neurons encode internal state information, such as task set (Asaad et al., 2000) and cognitive rules (Wallis et al., 2001; Buschman et al., 2012). In addition, prefrontal circuits rapidly alter how stimulus features are mapped either to actions (Mante et al., 2013) or abstract categories (Roy et al., 2010; Goodwin et al., 2012; Crowe et al., 2013) when the rule governing cognitive processing and behavior changes. These data elucidate both the neural representation of internal state in prefrontal cortex and the utilization of state information to rapidly change the mapping between stimuli and responses.

Patients with schizophrenia exhibit deficits in cognitive control (Barch et al., 2003; MacDonald et al., 2005; Richard et al., 2013), and fMRI studies have associated cognitive control deficits in patients with reduced activation of the dorsolateral prefrontal cortex (Perlstein et al., 2003; MacDonald et al., 2005). However, it is not known how the function of prefrontal cortex is disrupted at the cellular level to produce cognitive control deficits found in schizophrenia and related neuropsychiatric diseases. This is an important question to answer. Characterizing how neural activity patterns relate to cognitive function in the healthy state (the focus of the present study) is a necessary first step toward understanding how degradation of these neuronal activity patterns can lead to cognitive control failure in disease.

To characterize information processing by prefrontal neurons related to cognitive control, we translated a behavioral task that measures cognitive control deficits in patients with schizophrenia to monkeys with minimal modification and recorded neural activity in prefrontal cortex during task performance. We used the dot-pattern expectancy (DPX) task (Jones et al., 2010) for this purpose, a variant of the AX continuous performance task (AX-CPT). In the AX-CPT, subjects view a cue and then a probe stimulus. Under some trial conditions, the correct response to the probe depends on the identity of the preceding cue, requiring executive control. It is specifically on these trials that patients with schizophrenia and their first-degree relatives exhibit an elevated error rate relative to control subjects (Delawalla et al., 2008; MacDonald, 2008; Carter et al., 2012; Richard et al., 2013). We previously showed that administering ketamine (an NMDA receptor antagonist) to monkeys performing the DPX task causes them to produce the same, trial-type specific error pattern (Blackman et al., 2013) seen in patients with schizophrenia performing the same task (MacDonald, 2008; Jones et al., 2010). The context dependence of the error indicated a defect in a specific computation required for cognitive control rather than a generalized behavioral impairment. Thus, this prior work identified a strong and detailed behavioral parallel between patients with schizophrenia and a nonhuman primate translational model we seek to further develop.

As a next step, the current study characterizes how information processing operations required for cognitive control in the DPX task relate to patterns of activity in prefrontal neurons. We find the following: (1) population signals are strongly biased toward the representation of stimuli that inhibit prepotent, reflexive responses; (2) the timing and combination of signals encoding different stimuli reflect the rules by which the stimuli are combined to mediate logical 'if-then' computations for cognitive control; and (3) response-related signals are strongly bi-

ased to the representation of responses that are relatively rare and therefore less reflexive or automatic. These findings provide new insight into the neural mechanisms of cognitive control in prefrontal cortex, while at the same time characterizing the nature of information processing operations at a cellular level whose disruption could produce cognitive control deficits in schizophrenia.

Materials and Methods

Subjects

Two male rhesus macaques (8.0–10.4 kg) were used in this study. These monkeys were exposed to ketamine in a prior study (Blackman et al., 2013). All animal care and experimental procedures conformed to National Institutes of Health guidelines and were in compliance with protocols approved by the Animal Care and Use Committee at the University of Minnesota and Minneapolis Veterans Administration Medical Center.

Experimental conditions

Monkeys were seated in a primate chair in a sound-attenuated, dimly lit room during task performance. Dot patterns (2.7°–4.4°) were back-projected onto a video screen using an LCD projector (Dell 5100MP DLP). Monkeys were positioned 77.5 cm in front of the video screen. During task performance, the monkeys' eye position was tracked using an infrared eye tracking system (ISCAN), enabling the control of gaze and the retinal location of visual stimuli across conditions. To make their response each trial, monkeys moved a joystick to the right or left with their right hand.

Task

Monkeys were trained to perform the DPX task (a validated AX-CPT variant) similar to tasks used to measure deficits in executive control in patients (Jones et al., 2010) (see Fig. 1). Each trial a sequence of two dot patterns was presented (a cue followed by a probe; see Fig. 1C) separated by a delay. At the start of each trial, monkeys directed their gaze at the fixation cross presented at the center of the display for a period of 0.5 s. They were required to maintain gaze fixated within 3.3° of the central fixation target throughout the trial (if eye position moved outside this window, the trial aborted). After the initial fixation period, the cue (orange dots) was presented for 1 s, followed by a 1 s delay period, after which the probe (white dots) was presented for 0.5 s (see Fig. 1A). The intertrial interval (from the offset of the probe to the onset of the following cue, including the 0.5 s fixation period) was 1.86 s. Total trial length, from the onset of the cue of one trial to the onset of the cue in the next, was 4.36 s. Monkeys could respond anytime during the 0.5 s probe period or in the first 1.0 s of the intertrial interval afterward. A valid cue was designated A, and a valid probe was designated X. Invalid cues were collectively classified as B, and invalid probes were collectively classified as Y (see Fig. 1C). If the valid cue and probe were shown (AX cue-probe sequence), the trial was defined as a target trial and the monkeys responded with a leftward motion of the joystick (target response). If any other cue-probe sequence (AY, BX, BY) was presented, the trial was defined as a nontarget trial and the monkeys responded with a rightward motion of the joystick (nontarget response). No minimum response time limit was imposed; so, for example, monkeys could anticipate the onset of the probe on B-cue trials and respond "nontarget" immediately at probe onset (with minimal visual processing of the probe). However, median reaction time (RT) was >400 ms, even on B-cue trials (see Fig. 2A,B), suggesting that monkeys processed the probe rather than preemptively responding at probe onset. The duration of the probe stimulus was fixed at 0.5 s on all trials (regardless of the time of the motor response). This kept the timing of stimuli constant across trials and is most directly analogous to the way the task has been implemented to study cognitive control in patients with schizophrenia (Jones et al., 2010). If monkeys moved the joystick in the correct direction within the response window, they immediately received a ~0.1 ml drop of sweetened water. If monkeys moved the joystick in the incorrect direction within the response window, or did not respond, no reward was given (but otherwise the timing of stimuli was the same as on correct trials). If monkeys made

a joystick response before probe onset, the trial was aborted without reward.

We administered the DPX task in both prepotent and balanced trial sets. In prepotent trial sets, 69% of the trials in the set presented the AX cue-probe sequence, establishing a motor habit to respond “target.” The remaining 31% of trials were nontarget trials, made up of 12.5% AY, 12.5% BX, and 6% BY trials (see Fig. 1D, middle). We adopted this proportion of trial types to maintain consistency with the task as it has been used to measure cognitive control in patients with schizophrenia (Jones et al., 2010). We recorded the activity of 42 neural ensembles during DPX task performance (see below). Prepotent sets included either 301 trials (in 28 neural ensembles) or 400 trials (11 ensembles). In the remaining 3 ensembles, we recorded only balanced trial data. In balanced sets, we presented the four trial types (AX, AY, BX, BY) in equal proportion (see Fig. 1D, right) to decorrelate cue and probe identity within trials, making it easier to identify neurons that coded cue identity (A vs B) and probe identity (X vs Y) as independent task variables. Balanced sets included either 80 trials (37 neural ensembles) or 300 trials (4 neural ensembles). In the remaining 1 ensemble, we recorded only prepotent data. In the 38 neural ensembles recorded on both balanced and prepotent trial sets, the balanced trial set was presented first followed by the prepotent set in 28 ensembles. In the remaining 10 ensembles, prepotent trial set data were collected first followed by balanced trial set data. Balanced and prepotent trial sets were run as separate experiments on the same neural ensembles with a short delay (~30 s) in between. Other than the delay, monkeys were not explicitly cued as to the switch between trial sets or whether balanced or prepotent trial sets were underway.

Neural recordings

Once the monkeys achieved high ($\geq 80\%$ accuracy on all trial types) and stable performance on the final version of the task, they underwent an aseptic surgical procedure in which a recording chamber (13 mm inner diameter) was implanted over the prefrontal and posterior parietal cortex in the left cerebral hemisphere (prefrontal neural recordings are described here). MRI-based localization was used to place the chambers. The prefrontal chamber was centered over Brodmann’s area 46 in the principal sulcus in both monkeys (see Fig. 1B). In addition to the recording chambers, titanium screws and posts were implanted on the surface of the skull during the surgical procedure. The posts were fastened to a halo and allowed for the ability to secure head position. During the surgical procedure, the monkeys were maintained under gas anesthesia (isoflurane, 1%–2%) and were administered analgesia for several days postoperatively (Buprenex; 0.05 mg/kg twice a day, i.m.). Following postoperative recovery, neurophysiological recordings commenced.

As monkeys performed the DPX task, neural activity was recorded from the PFC (areas 9 and 46) using a computer-driven 16 microelectrode Eckhorn motor drive (Thomas Recording, GmbH). Fine glass-coated platinum iridium fiber microelectrodes (70 μm outer diameter; 1–2 M Ω impedance) were independently lowered through the intact dura mater into the cerebral cortex, and their depths were adjusted to isolate the electrical activity of single neurons. The electrical signals were amplified (gain of 2500) and bandpass filtered (cutoff frequencies of 2 and 5 kHz). Waveforms for the individual neurons’ action potentials were isolated on-line using waveform discriminators (Alpha Omega Engineering). It was possible to routinely isolate the action potentials of ~20 neurons per recording ensemble. The timing of spike occurrence for the isolated action potentials and behavioral events were stored to disk with 40 μs resolution (DAP 5200a Data Acquisition Processor; Microstar Laboratories). The monkeys received an intramuscular hindlimb injection of sterile saline (0.41–0.89 ml) on some days (Monkey 1: 25%; Monkey 2: 70%) to provide control data for a subsequent study.

Data analyses

Data were collapsed across monkeys and restricted to correct trials only. An α level of $p < 0.05$ was adopted to define significance.

Behavioral analysis. To determine whether there was an effect of trial type on RT, we conducted a Kruskal–Wallis test (as the RT data were not normally distributed) followed by a *post hoc* multicomparison test for both balanced and prepotent trial sets. A technical issue in the recording

of RTs by the behavioral control program led to a surplus of RTs at 500–502 ms time point after probe onset, although trials with this RT comprised $< 6\%$ of trials, and their removal (or inclusion) did not influence the results of analyses of the RT data. The analysis of RT indicated that RTs were longer on AY trials compared with the other trial types (see Fig. 2A, B), suggesting a behavioral cost associated with cognitive control when presentation of the A-cue established a target response expectation that then had to be reversed after presentation of the Y-probe. To determine whether this behavioral cost varied as a function of trial proportion, we compared RT on AY trials between balanced and prepotent sets.

Neuron classification. To identify neural activity that differed as a function of the cue (A vs B), probe (X vs Y), and response (target vs nontarget), we conducted ANCOVA on firing rates in the balanced datasets. Neuron inclusion criteria included a minimum baseline rate requirement of 1 Hz. To identify neurons encoding the cue during the cue period, we performed a one-way ANCOVA on firing rates using cue identity as the single factor. We entered firing rates during the fixation period as a covariate to control for trial-to-trial fluctuations in baseline firing rate. Neurons with activity varying as a function of cue identity during the cue period were designated “early cue” neurons. “Delay cue” neurons were identified using the same analysis applied to firing rates in the delay period.

Following presentation of the probe, neural activity could potentially reflect the identity of the probe or the identity of the preceding cue. To analyze the potential influence of these factors simultaneously, we performed an ANCOVA on firing rates during the probe period using cue identity (A or B), probe identity (X or Y), and their interaction as factors in this analysis. The covariate was firing rate during the fixation period. Neurons with neural activity during the probe period that varied as a function of the identity of the preceding cue were designated “late cue” neurons. To identify response neurons with activity related to response direction (left or right), we applied ANCOVA to firing rates measured within a response window starting 200 ms before and extending to 200 ms after the initiation of the response. Response direction was the single factor in this analysis, and firing rate during the fixation period was the covariate.

Neurons were always classified on the basis of the results of the ANCOVA (above) applied to neural data recorded on balanced trial sets. This allowed us to differentiate neurons that encoded the cue and probe stimuli under conditions that the two were statistically independent. We illustrate the counts of cue, probe, and response neurons as well as the average activity patterns of these populations based on the classification produced by ANCOVA applied to neural activity on balanced trial sets in Figures 2C and 3. In Figures 4–8, we analyze the activity of the same groups of neurons on prepotent trial sets, using the neural classifications derived from the balanced data.

To determine whether populations of neurons were biased in terms of the numbers of neurons preferring each of the two alternative cues, probes, or responses, we applied a χ^2 test using the Yates correction for continuity.

Spike density functions (SDFs). SDFs were constructed from single trials of neural activity by convolving the vector of spike times with a Gaussian kernel (using the “ksdensity” function in MATLAB with a kernel width of 40 ms). In subsets of statistically defined neural populations (see Neuron classification, above), we then separated single-trial SDFs into groups according to the identity of the cue and probe shown (producing four trial groups in which the cue-probe sequence was either AX, AY, BX, or BY). We next averaged the SDFs in each trial group across neurons within each population to produce a population average SDF illustrating the activity pattern as a function of cue-probe sequence.

Many neurons responded to multiple task variables in different task epochs, with the result that these neurons were allocated to multiple ANCOVA-defined functional groups, several of which consequently exhibited similar patterns of activity (see Fig. 3). Neuronal groups that exhibited similar activity profiles were subsequently combined (see Fig. 5).

We found that many cue-selective neurons preferred the B-cue during the cue period (“early cue” response) and switched their preference to the A-cue during the probe period (“late cue” response). Late cue activity was

often augmented when the A-cue was followed by the Y-probe. We termed neurons that were significantly influenced by cue identity in both the cue and probe period but changed their preference from B-cues to A-cues between these periods “switch” neurons. Switch neurons appeared to be involved in aspects of cognitive control required by the DPX task because their activity was maximal when cognitive control demand was greatest and the prepotent target response had to be inhibited (e.g., after B-cues in the cue period, and Y-probes following A-cues in the probe period). To examine the influence of varying cognitive control demand on the activity of switch neurons, we computed the difference in firing rate on A-cue and B-cue trials to measure cue preference and determined whether cue preference was significantly stronger in the prepotent relative to the balanced trial sets (see Fig. 9). To evaluate the significance of the difference in cue preference between balanced and prepotent trial sets, we conducted a permutation test in which we randomly shuffled trials of spike data between the balanced and prepotent sets and recomputed the cue preference (100 iterations). The change in cue preference between balanced and prepotent trial sets in the original data was considered significant if it exceeded 95% of the difference measures obtained in the shuffled data.

Cue preference index. To further quantify the preference of each neuron for the A and B cues, we computed a cue preference index, $(A - B)/(A + B)$, where A is the average firing rate of the cell during the epoch in response to A-cues and B is the average firing rate in response to B-cues. We calculated the difference in the cue preference index between the cue and probe periods ($\text{Index}_{\text{probe}} - \text{Index}_{\text{cue}}$). We then determined whether the difference in cue preference between the cue and probe periods in individual neurons was significant by performing a permutation test in each neuron in which we shuffled firing rates by epoch between the cue and probe periods, and recalculated the difference in the cue preference index in the cue and probe periods (1000 repetitions). The initial change in cue preference in each neuron was considered significant if the index was either greater than the 97.5 percentile or less than the 2.5 percentile of the shuffled distribution ($p < 0.05$, two-tailed).

Decoding analysis. To quantify the strength and timing with which population activity represented different task variables, we decoded the cue (A vs B), probe (X vs Y), and response (target vs nontarget) from patterns of neural population activity using pattern classification analyses (Klecka, 1980; Goodwin et al., 2012). The decoding analysis was applied to single-trial patterns of activity observed in populations of neurons. To decode each variable, neurons were included in the population if their activity varied significantly as a function of the decoded variable. Significant neurons were aggregated across ensembles, and trials were matched according to the repetition number of each task condition. We restricted the populations used in the decoding to neurons that were recorded for a minimum of at least 75 trials for balanced sets and 230 trials for prepotent sets. We then truncated the number of trials in neurons recorded for more trials to these minimums. This held the number of trials constant across neurons in spite of small variation in the numbers of trials performed by animals between experiments. That, in turn, made it possible to represent the firing rate data over the population as a rectangular matrix, in which the number of columns (neurons) did not vary over trials (rows), which was necessary for the decoding.

We quantified the accuracy of decoding every 50 ms throughout the trial by decoding task variables using population activity within a sliding 150 ms window consisting of firing rates measured in three consecutive 50 ms time bins. We used leave-one-out cross validation in the decoding analyses to ensure the neural data used to decode each behavioral variable did not contribute directly to the definition of the classification functions applied to perform the decoding. We used firing rates at each time step to define the activity patterns used to perform the classification; therefore, the patterns of activity coding each variable were free to change over the trial in the decoding analysis. We used the “classify” function in the MATLAB Statistical toolbox (The MathWorks) using empirical prior probabilities to perform the classification.

Demixed principal component analysis. We applied a demixed principal component analysis developed by Machens and colleagues (Brendel et al., 2011) and recently used by Wallis and colleagues (Lara and Wallis, 2014) to characterize temporal profiles of neuronal activity in prefrontal

cortex without a priori assumptions about the timing of activity changes. The analysis identifies a set of principal components (weighted linear combinations of neurons) that specify axes through the rate space of the population that are oriented so as to capture the greatest amount of variance in population activity patterns with the added constraint that the axes capture the maximal variance attributable to task variables (Brendel et al., 2011). Once the components were derived, population activity patterns were measured in a sequence of 50 ms bins, and then were projected onto each component (by multiplying the rate of each neuron by its coefficient, or weight, in the component, and summing the products). This produced a time series indicating variation of population activity along the component of interest throughout the trial. Data were demeaned, and the demixed principal component analysis was conducted using the methods outlined by Brendel et al. (2011). For each component, the relative contribution of time, trial type, and their interaction was examined to investigate the amount of variance explained by these factors relative to the total variance explained of that component.

Results

Monkeys performed the DPX variant of AX-CPT in which letters were substituted by dot patterns (Fig. 1). The stimulus and response contingencies of the AX-CPT and the DPX task are otherwise identical. Use of dot-patterns allows comparison of behavior between monkeys and humans without introducing prior visual experience with letters as a variable confounding comparison of behavior across species. The DPX task requires cognitive control by virtue of the feature that the response to a probe stimulus is contingent on the identity of a preceding cue stored in working memory. This requires that monkeys flexibly select different responses to the same probe stimulus contingent upon an internal representation of context, in this case provided by the cue stimulus.

In the DPX task, monkeys viewed a cue stimulus (designated A or B) followed by a probe stimulus (designated X or Y) in each trial (Fig. 1A,C). Trials with the AX sequence were designated target trials, requiring a leftward joystick response, and all other cue-probe sequences were nontarget, requiring a rightward joystick response. Based on the above contingencies, B-cue trials were always nontarget (independent of the following probe), and Y-probe trials were always nontarget (independent of the preceding cue). However, on A-cue trials, the response could be either target or nontarget, depending on the subsequent probe. Similarly, on X-probe trials, the response could be either target or nontarget, depending on the preceding cue. On these trials, visual information about the cue stimulus stored in working memory and the probe stimulus had to be combined to determine the correct motor response. The result of this logical operation (e.g., IF A-cue AND X-probe, THEN target response) identified the correct conditional motor response to the probe stimulus.

DPX trials were administered in either balanced or prepotent sets. In the balanced trial sets (80 or 300 trials), the four trial types defined by the possible combinations of cue and probe stimuli (AX, AY, BX, and BY) were administered in equal proportion (Fig. 1D, right). On prepotent trial sets (301 or 400 trials), the majority of trials (69%) were of the AX target sequence (Fig. 1D, middle), establishing a sensory-motor habit to respond target whenever the X-probe appeared. The remaining trials were divided between AY (12.5%), BX (12.5%), and BY (6%) cue-probe sequences (Fig. 1D, middle). This is the same proportion of trial types that has been used to characterize cognitive control deficits in patients with schizophrenia (Jones et al., 2010). Prepotent trial sets are considered to increase cognitive control demand on BX trials because the B-cue stored in working memory must override

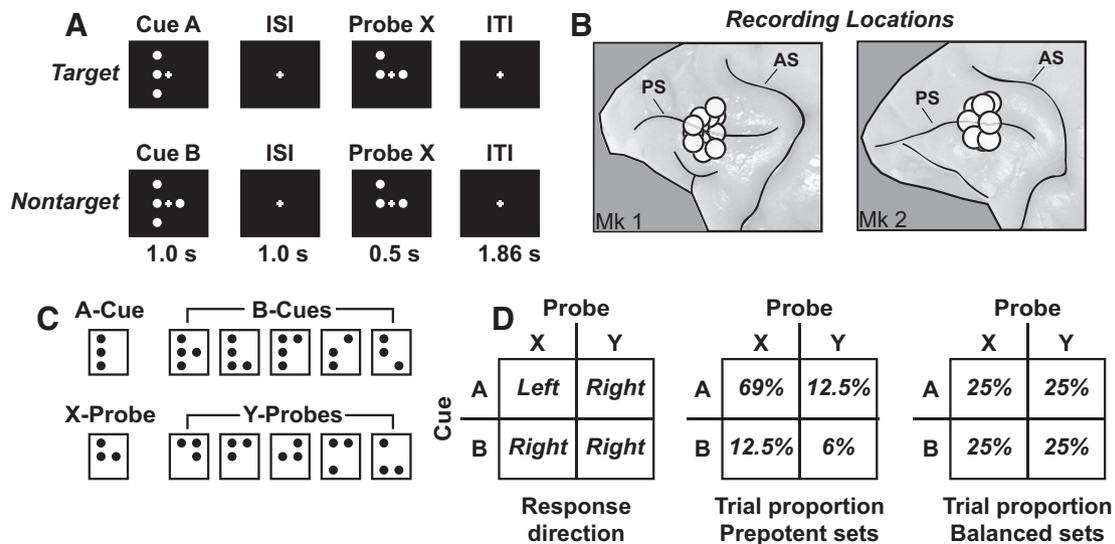


Figure 1. The DPX task. **A**, Cue-probe sequence on AX (top row) and BX (bottom row) trials. The AX sequence was the target sequence and required a leftward joystick movement. All other cue-probe sequences were nontarget and required a rightward joystick movement. ISI, Interstimulus interval; ITI, intertrial interval. **B**, Recording locations. Circles encompass the cortical area sampled by the 16 electrode recording matrix placed at different positions within the prefrontal recording chambers. PS, Principal sulcus; AS, arcuate sulcus. **C**, Dot patterns comprising the set of cue and probe stimuli in the DPX task. One cue dot pattern was designated A; several cue dot patterns were designated B (top row). Likewise, one probe dot pattern was designated X, and several probe dot patterns were designated Y (bottom row). **D**, Left, Response direction as a function of the combination of cue and probe stimuli. **D**, Middle, Proportion of trial types defined by cue-probe combinations in prepotent trial sets. **D**, Right, Proportion of trial types defined by cue-probe combinations in balanced trial sets.

Table 1. Accuracy of behavioral performance on the DPX task^a

Monkey	Trial type	Balanced	Prepotent
1	AX correct	90.1 (29.9)	94.4 (23.0)
	AY correct	100 (0.0)	99 (9.2)
	BX correct	91.5 (27.9)	92.4 (26.6)
	BY correct	100 (0.0)	100 (0.0)
2	AX correct	93.8 (24.2)	98.1 (13.8)
	AY correct	97.9 (14.4)	98.3 (12.8)
	BX correct	92.4 (26.5)	96.0 (19.6)
	BY correct	99.6 (6.0)	99.8 (4.3)

^aData are mean (SD) percentage correct as a function of cue-probe sequence (AX, AY, BX, BY) during balanced and prepotent trial sets for each monkey.

the habitual target response to the subsequent X-probe to command a nontarget response.

Behavioral performance

Both monkeys performed with a high level of accuracy (>95% correct overall) during the balanced trial sets (Monkey 1: 95.4%; Monkey 2: 95.9%) as well as the prepotent trial sets (Monkey 1: 95.1%; Monkey 2: 97.9%) of the DPX task. Performance was >90% correct on all trial types (Table 1). There was a tendency for performance to be more accurate on Y-probe trials than X-probe trials (Table 1). This may have reflected the benefit of a sensorimotor habit to respond “nontarget” whenever the Y-probe appeared, and a cost associated with the decision to respond “target” or “nontarget” (depending on the preceding cue) when the X-probe appeared. RTs varied across trial types (Kruskal–Wallis test, Balanced: $\chi^2_{(3)} = 1311.8$, $p < 0.001$; Prepotent: $\chi^2_{(3)} = 2546.3$, $p < 0.001$) such that monkeys responded the slowest on AY trials ($p < 0.05$) relative to the other trial types (AX, BX, BY), a pattern that was seen both in the balanced and prepotent trial sets (Fig. 2*A, B*). AY trials require overriding the tendency to respond “target” following the A-cue when the Y-probe appears, as on the majority of trials the X-probe followed the A-cue, and a target response could therefore be anticipated.

The prolonged RT on AY trials appeared to reflect the greater degree of cognitive control required.

Neural database

We recorded the activity of 819 neurons in prefrontal cortex (Fig. 1*B*) in 42 ensembles consisting of between 6 and 33 individually isolated neurons (mean = 19.5) recorded simultaneously. We classified neurons by applying ANCOVA to their firing rates during balanced trial sets. This allowed us to clearly differentiate neural signals that encoded the cue and the probe under conditions that the two stimuli were statistically independent. We characterize the activity of neurons on the balanced trial sets in Figures 2*C* and 3, and on prepotent trial sets in Figures 4–8. Because neurons could encode the cue any time in the trial after its presentation, we identified “early cue,” “delay cue,” and “late cue” neurons as those in which firing rate varied as a function of cue identity during the cue, delay, and probe periods of the trial, respectively. Of the 819 recorded neurons, 509 exhibited activity that varied significantly as a function of the cue, probe, and/or response (62% of recorded neurons). Cue neurons (early, delay, and late combined) comprised by far the largest group (Fig. 2*C*; “Cue,” 432 of 509 task related neurons, or 85%). The numbers of neurons classified as early cue, delay cue, and/or late cue are indicated in the corresponding panels of Figure 3*A*. Smaller groups of neurons exhibited activity that varied as a function of the probe and response (Fig. 2*C*, “probe,” “response”). Most often, modulations in firing rate relating to the probe or the response rode on top of separable modulations in firing rate relating to the cue in single neurons, pointing to the predominance of cue representation in the neural data (Fig. 2*C*; most neurons in the probe and response circles fall in the zones of intersection with the cue circle).

Neural signals encoding the cue

Population SDFs (Fig. 3) illustrate the average activity of statistically defined populations of neurons identified as early cue (Fig.

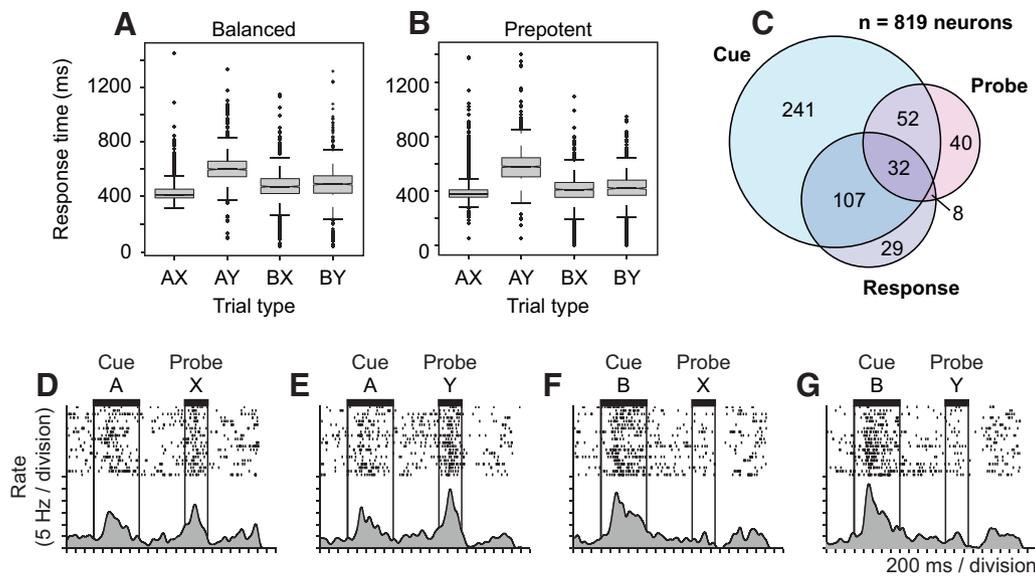


Figure 2. Behavioral performance, neuron counts, and example “switch” neuron. **A, B**, Box-whisker plots illustrate the distribution of RTs on the DPX task separated by trial type (defined by cue-probe sequence) on balanced (**A**) and prepotent (**B**) trial sets. **C**, Venn diagram illustrates the number of neurons in which firing rate varied significantly ($p < 0.05$, ANCOVA) as a function of the cue (light blue circle), probe (red), and response (light gray) in the DPX task of a total sample of 819 prefrontal neurons. **D–G**, Activity of a single prefrontal neuron on the DPX task on (**D**) AX, (**E**) AY, (**F**) BX, and (**G**) BY trials.

3A–C), delay cue (Fig. 3D–F), late cue (Fig. 3G–I), probe (Fig. 3J–L), and response (Fig. 3M–O) neurons by ANCOVA (see Materials and Methods). Population activity is represented separately on subsets of trials in which the cue-probe sequence was AX, AY, BX, or BY. Neural activity in several populations exhibited unique temporal features as a function of the cue-probe sequence.

The DPX task requires, under some trial conditions, that the brain integrate information about the cue and the probe to select the correct response. This implies that signals encoding the previous cue have to exist during the probe period in order for information about the two stimuli to be integrated. One way to achieve this would be the activation of classical “delay neurons” (Goldman-Rakic, 1987; Miller et al., 1996; Chafee and Goldman-Rakic, 1998), hypothetically selective either for the A-cue or B-cue, which would sustain an elevated firing rate after their preferred cue disappeared until the probe period when the cue information they encoded could be combined with probe information to compute the correct response. A population of neurons exhibiting this form of sustained cue-selective activity was indeed observed (Fig. 3A, E). However, a much larger population of neurons encoding the cue stimulus exhibited a strikingly different activity pattern. This population of “switch” neurons (Fig. 3D, G, M) was most strongly activated during the cue period by presentation of B-cues, after which activity decayed toward baseline in the delay period, indicating a B-cue preference early in the trial. The same neurons were most strongly activated during the probe period on trials that the A-cue had been presented, with activity ramping up in the delay period and peaking in the probe period, indicating an A-cue preference late in the trial. Thus, the same neural population switched its stimulus preference from B-cues in the cue period to A-cues between the cue and the probe period. Because “switch” neurons constituted a large fraction of cue-selective neurons, populations of prefrontal neurons were strongly biased overall in terms of the numbers of neurons that preferred the A-cue versus the B-cue during the cue and probe periods. The large majority of cue-selective neurons activated

during the cue period (“early cue” neurons) preferred the B-cue (Fig. 3C; $\chi^2_{(1)} = 84.78$, $p < 0.001$). This bias disappeared during the delay period (Fig. 3F) and then reversed during the probe period, such that the majority of cue-selective neurons activated during the probe period (“late cue” neurons) preferred the A-cue (Fig. 3I; $\chi^2_{(1)} = 7.28$, $p < 0.01$). The switch in cue preference between task epochs was evident not only at the population level (Fig. 3) but was seen also in single prefrontal neurons (Fig. 2D–G) and was a prominent feature of neural activity in prefrontal cortex during the DPX task.

Figure 4 plots the cue preference index computed for single neurons using cue period (x -axis) and probe period (y -axis) activity. Colored symbols indicate neurons in which the cue preference index significantly changed from the cue to the probe periods ($p < 0.05$ two-sided, bootstrap test; see Materials and Methods). The proportion of neurons that increased their cue preference index between the cue and probe periods, indicating a shift in cue preference toward A-cues (Fig. 4, red; 54 of 106 neurons) significantly exceeded the proportion of neurons that decreased their cue preference index, indicating the opposite shift in preference toward B-cues (Fig. 4, blue; 9 of 106 neurons; z test of proportions; $z = -6.8$, $p < 0.0001$).

Aggregating ANCOVA-defined neuronal types into groups with common activity patterns and alignment of activity to the time of the motor response

The ANCOVA-defined classification of neurons led to several groups with highly similar patterns of population activity (e.g., Fig. 3B, D, G, K, M). This was a consequence of the fact that the firing rate of single neurons often exhibited a significant relationship to different task factors during different epochs in the trial. As a result, the same individual neurons were often allocated to multiple ANCOVA-defined neuronal classes, which ended up exhibiting similar patterns of activity. To reduce this redundancy, we combined ANCOVA-defined classes that exhibited a similar pattern of activity into a single group. The assignment of the 10 ANOVA-defined neuronal classes to 6 groups sharing a common

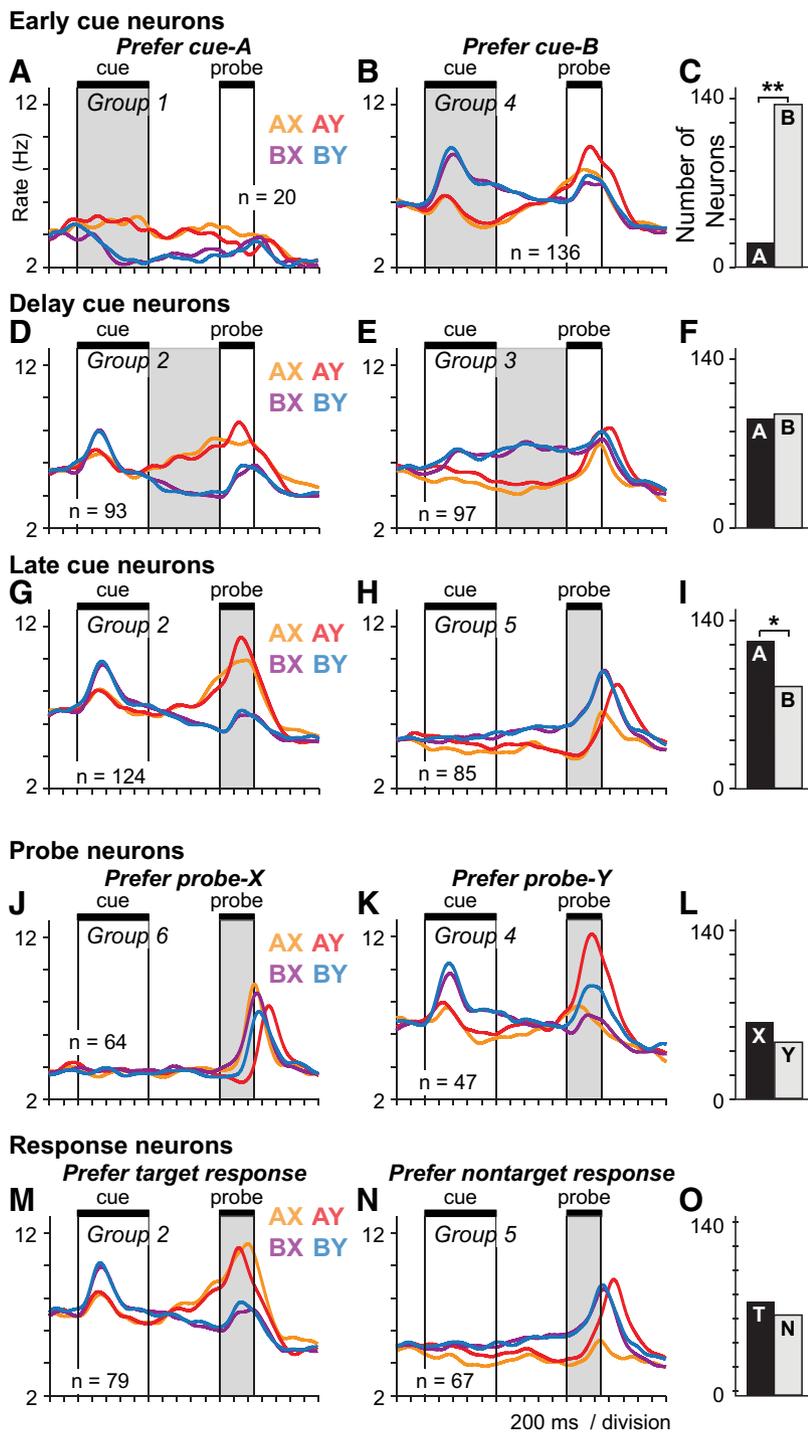


Figure 3. Prefrontal population activity patterns encoding the cue, probe, and response. ANCOVA was used to identify populations of neurons with activity varying significantly ($p < 0.05$) as a function of the cue (A vs B), the probe (X vs Y) or the response (target vs nontarget), during different periods of the DPX trial. Gray represents the trial period used to classify neurons by ANCOVA. Trials are restricted to the balanced trial sets to illustrate the neural data entered into the ANCOVA. Neurons are limited to those recorded on both balanced and prepotent sets in this figure; counts differ slightly from those in Fig. 2C. SDF ($\sigma = 40$ ms) of different colors illustrate the average activity time course of each population broken out into four trial groups based on the cue-probe sequence: orange represents AX; red represents AY; purple represents BX; blue represents BY. Neuronal populations defined by ANCOVA that had similar patterns of activity were collapsed into a single group as indicated by the group numbers in each panel. **A, B**, “Early cue” neurons as defined by ANCOVA preferring the A-cue (**A**) or B-cue (**B**) during the cue period (shaded gray). **C**, Counts of early cue neurons preferring A-cues (black bar) and B-cues (gray bar). **D, E**, “Delay cue” neurons as defined by ANCOVA preferring the A-cue (**D**) or B-cue (**E**) during the delay period (shaded gray). **F**, Counts of delay cue neurons preferring A-cues (black bar) and B-cues (gray bar). **G, H**, “Late cue” neurons as defined by ANCOVA preferring the A-cue (**G**) or B-cue (**H**) during the probe period (shaded gray). **I**, Counts of late cue neurons preferring A-cues (black bar) and B-cues (gray bar). **J, K**, “Probe” neurons as defined by ANCOVA preferring the X-probe (**J**) or Y-probe (**K**) during the probe period (shaded gray).

overall activity pattern is shown by the group labels indicated in each panel of Fig. 3. Group 1 neurons (Fig. 5A,B) significantly preferred the A-cue during the cue period by ANCOVA (Figs. 3A). (Numbers of neurons differ in Figures 3 and 5 because each neuron belonged to only one group in Fig. 5 but could belong to multiple classes in Fig. 3.) Group 1 (Fig. 5A,B) was a small group with a weak signal. Neurons in Groups 2 and 4 were “switch” neurons preferring the B-cue early and the A-cue late in the trial. Group 2 neurons (Fig. 5C,D) significantly preferred the A-cue during the delay period (Fig. 3D) or the A-cue during the probe period (Fig. 3G) by ANCOVA. Group 4 neurons (Fig. 5G,H) significantly preferred the B-cue during the cue period (Fig. 3B) or the Y-probe during the probe period (Fig. 3K) by ANCOVA. Together, neurons in Groups 2 and 4 constituted the bulk of task-related neurons in prefrontal cortex. In both Groups 2 and 4, the strongest activity occurred on AY trials (Fig. 5C,G, red SDF). Activity on BX and BY trials was weaker, and SDFs tended to overlap on these trials (Fig. 5C,G, purple and blue SDF). Groups 2 and 4 were differentiated on the basis of activity on AX trials (Fig. 5C,G, orange SDF) compared with activity on BX and BY trials (Fig. 5C,G, blue and purple) considered together. In Group 2, peak activity on AX trials exceeded peak activity on BX and BY trials (Fig. 5C,D, orange SDF above purple and blue SDF). In Group 4, peak activity on AX trials occurred at approximately the same level as peak activity on BX and BY trials (Fig. 5G,H, peaks of orange, purple, and blue SDF are of approximately equal height). Group 3 neurons (Fig. 5E,F) significantly preferred the B-cue during the delay period (Fig. 3E) by ANCOVA. The activity of Group 3 neurons took the form of classical delay period activity associated with working memory as the activity reflecting the preferred cue was tonically maintained in the delay period. (Fig. 5E,F). Neurons in Group 5 (Fig. 5I,J) significantly preferred the B-cue during the probe period (Fig. 3H) or the non-

L, Counts of probe neurons preferring X-probes (black bar) and Y-probes (gray bar). **M, N**, “Response” neurons as defined by ANCOVA preferring the target response (**M**) or nontarget response (**N**) during the response period. **O**, Counts of response neurons preferring target responses (black bar) and nontarget responses (gray bar). Asterisks above bar charts indicate a significant bias in the number of neurons preferring A- versus B-cues, X- versus Y-probes, and target versus nontarget responses: * $p < 0.05$; ** $p < 0.01$.

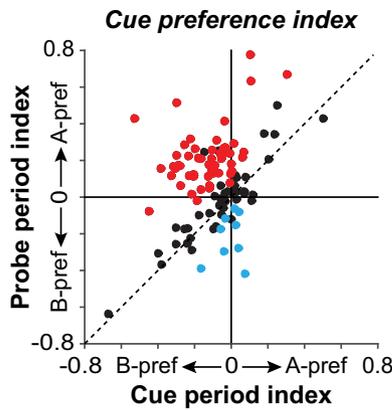


Figure 4. Quantification of switch in cue preference from the cue to the probe period in single neurons. Each point plots the cue preference index of a single neuron using firing rates during the cue period (Cue period index; x-axis) and probe period (Probe period index; y-axis). Negative values indicate B-cue preference; positive values indicate A-cue preference (index ranges from -1 to 1). Data are from prepotent trial sets. Neurons were included if their activity related significantly to the cue during the cue (and/or delay) period, and also the probe period. Points in the upper left quadrant indicate neurons that switch their cue preference from B-cues during the cue period (negative x-coordinate) to A-cues during the probe period (positive y-coordinate). Red circles represent neurons in which the cue preference index in the probe period was shifted significantly toward A-cues compared with the cue period ($p < 0.05$). Blue circles represent the converse (neurons in which the cue preference index in the probe period was shifted significantly toward B-cues compared with the cue period). Black circles represent neurons with no significant shift in their cue-preference index between the cue and probe periods.

target response during the response period (Fig. 3N) by ANCOVA. Finally, neurons in Group 6 (Fig. 5K,L) significantly preferred the X-probe during the probe period (Fig. 3J).

Neural signals encoding the probe

Prefrontal neurons encoded the identity of the probe stimulus in a manner that depended on the preceding cue. On A-cue trials, when the cue and probe had to be considered together to select the correct response, prefrontal population activity differentiated Y-probes (Fig. 5C,E,G, red SDF) from X-probes (Fig. 5C,E,G, orange SDF). In contrast, on B-cue trials, when the required response was nontarget on every trial and the probe was irrelevant to response selection, population activity essentially overlapped on Y-probe (Fig. 5C,E,G, blue SDF) and X-probe (Fig. 5C,E,G, purple SDF) trials. These data indicate that prefrontal neurons encoded the identity of the probe in a context-dependent manner, selectively representing the probe stimulus under conditions (specifically following A-cues) when probe information was useful to response selection. This suggests that stimulus processing reflected the role that stimuli played in the logical computations required for cognitive control. In addition, the neural groups encoding the probe stimulus typically preferred the Y-probe over the X-probe (Fig. 5C–H, red over orange SDF). This provides additional evidence that processing of probe stimuli reflected their significance for cognitive control. The appearance of the Y-probe following the A-cue required overriding the expectation that the subsequent probe would be the X-probe requiring the target response (as most trials were AX). The fact that RT was significantly longer on AY trials when this expectation had to be overridden relative to the other trial types (Fig. 2), is consistent with the view that AY trials imposed greater cognitive control demand. Stronger activation on AY trials in prefrontal cortex (Fig. 5C–H, red) may therefore reflect the

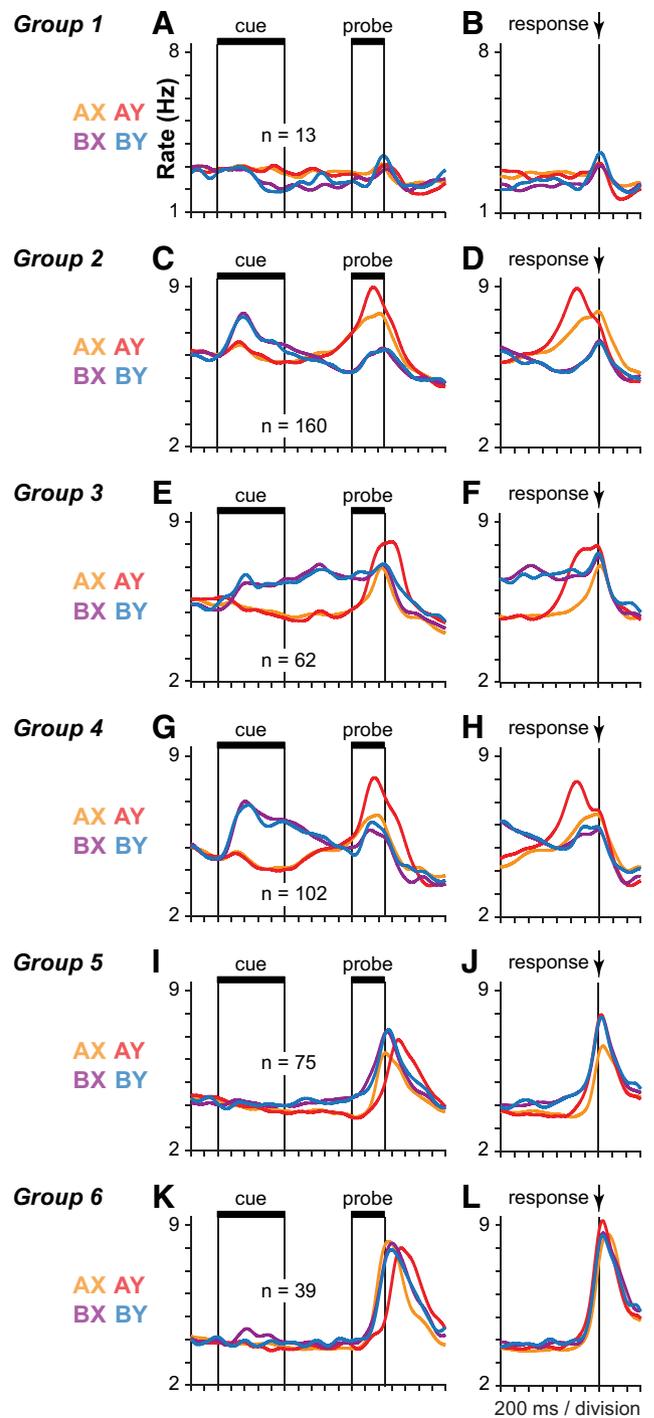


Figure 5. Neuronal groups defined by common activity patterns on prepotent trial sets. A–L, SDFs ($\sigma = 40$ ms) illustrate the average population activity of six groups of neurons (one group per row). Left column, Activity of each group aligned to cue onset (A, C, E, G, I, K). Right column, Activity of the same neurons aligned to the time of the motor response (B, D, F, H, J, L). Neuronal groups were formed by aggregating ANOVA defined neuronal classes that exhibited similar profiles of activity in the initial classification (as labeled in Fig. 3).

greater level of cognitive control required on AY trials. To quantify the cue dependence of probe encoding, we compared the accuracy of decoding the probe stimulus from patterns of population activity on A-cue and B-cue trials (Fig. 6). Probe decoding was more accurate on A-cue trials (Fig. 6A) compared with B-cue trials (Fig. 6B).

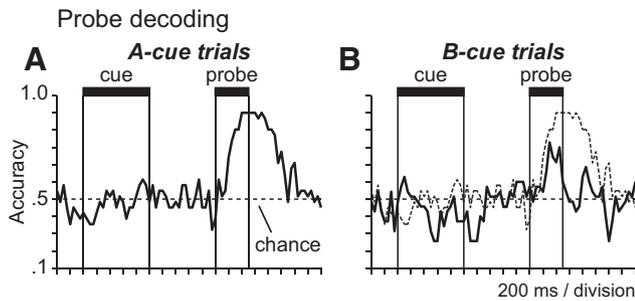


Figure 6. Cue-contingent probe encoding. *A, B*, Time courses plot the percentage of trials on which the probe stimulus (X vs Y) was correctly decoded from population activity patterns in probe-selective neurons at different times in the trial (50 ms time bins) when the probe was preceded by the A-cue (*A*) and when the probe was preceded by the B-cue (*B*; dotted time course shows A-cue trial decoding in *A* above for comparison).

Neural signals encoding response direction

Realigning the activity of these 6 groups of neurons to the time of the motor response (Fig. 5*B, D, F, H, J, L*) revealed essentially two sequential peaks in activity that occurred toward the end of the trial near the time of the motor response. The earlier peak in activity occurred prominently in Groups 2, 3, and 4, and in each case this activity was greatest on AY trials (Fig. 5*D, F, H*, red SDF). The later peak in activity occurred prominently in Groups 5 and 6, and this peak was better aligned to response onset (Fig. 5*J, L*) than to probe onset (Fig. 5*I, K*). As noted above, Group 6 neurons were classified by the ANCOVA as probe neurons that preferred X-probes (Fig. 3*J*). However, realignment of the activity to the response suggests that this was an artifact of the faster RT on X-probe trials relative to Y-probe trials (particularly on AY trials; Fig. 2*A, B*), so that more response-related activation fell within the probe window on X-probe trials than on Y-probe trials, artificially inflating firing rates within this window as a result. Probe-selective activity in Group 6 neurons essentially disappeared when the activity was realigned to the time of the motor response (Fig. 5*L*). Group 5 and 6 neurons exhibited modulations in firing rate that were most clearly time-locked to the time of the motor response. Activity in Group 5 neurons was greater for all nontarget responses (Fig. 5*J*, red, purple, blue; AY, BX, and BY trials) compared with target responses (Fig. 5*J*, orange; AX trials). Aligned to response time, Group 6 neurons exhibited a strong nonselective phasic response that was time-locked to movement onset but was not selective for the cue, probe, or direction of the response (Fig. 5*L*). These neurons may provide a type of “go” signal that releases the habitual target response unless overridden by the activity of the other neural classes, many of which were most active when this response needed to be suppressed. Notably, there were no neural groups found to be more active for the habitual target response when activity was realigned to response onset.

Decoding the cue, probe, and response from mixed-selectivity neurons

The activity of most prefrontal neurons was modulated in relation to multiple DPX task factors simultaneously (e.g., cue, probe, and/or response), consistent with the mixed selectivity that has been reported in prefrontal neurons in other task contexts (Rigotti et al., 2013). This raises the question whether the mixture of signals found in single prefrontal neurons supports an accurate representation of the cue, probe, and response in the DPX task at the population level. To quantify the degree to which the identity of the cue and probe stimuli as well as the response

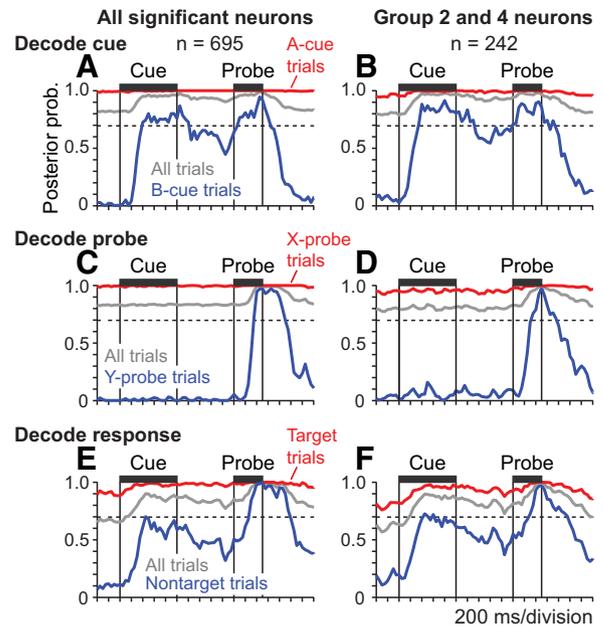


Figure 7. Cue, probe, and response decoding on prepotent trial sets. Lines plot the posterior probability associated with the correct cue (*A, B*), probe (*C, D*), and response direction (*E, F*) of each trial derived by applying pattern classification to neural firing rates observed in a window of 3 consecutive 50 ms time bins advanced through the trial in 50 ms steps. The smaller numbers of neurons included in the population decoding relative to the ANCOVA classification in Fig. 3 reflect the exclusion of some neurons with very low spike rates. Left column, Posterior probability time courses obtained when including all neurons with activity significantly related to any task variable (cue, probe, or response) in the decoding analysis (*A, C, E*). Right column, Posterior probability time courses obtained when including only Group 2 and 4 neurons in the decoding analysis (*B, D, F*). Dashed horizontal lines indicate the prior probability of A-cues, X-probes, and target responses based on their proportion relative to all trials in the prepotent trial sets. *A, B*, The posterior probability time course associated with the correct cue (A or B) derived from decoding applied to the activity of all significant neurons (*A*) or restricted to neurons in Groups 2 and 4 (*B*). Decoding results are plotted separately on A-cue trials (red), B-cue trials (blue), or all trials combined (gray). *C, D*, The posterior probability time course associated with the correct probe (X or Y) derived from decoding applied to the activity of all significant neurons (*C*) or restricted to neurons in Groups 2 and 4 (*D*). Decoding results are plotted separately on X-probe trials (red), Y-probe trials (blue), or all trials combined (gray). *E, F*, The posterior probability time course associated with the correct response direction (target or nontarget) derived from decoding applied to the activity of all significant neurons (*E*) or restricted to neurons in Groups 2 and 4 (*F*). Decoding results are plotted separately on target trials (red), nontarget trials (blue), or all trials combined (gray).

direction were represented by the activity of the neural populations with mixed selectivity, we decoded each task variable from population activity patterns measured in a sequence of 50 ms time bins. We used a decoder which took into account the prior probability associated with values of each variable based on their frequency of occurrence in the trial set (e.g., the number of A- vs B-cues), as well as additional information encoded by neural activity patterns, when computing the posterior probability associated with each potential outcome. In the prepotent dataset, given the high prior probability that the cue was A, the probe was X, and the response was therefore “target” (independent of information provided by neural activity), the posterior probability associated with these values for the cue, probe, and response was very high in the intertrial interval before the presentation of stimuli (Fig. 7*A, C, E*, red decoding time courses at or near ceiling). This reflects the classifier “guessing” that the cue was A, the probe was X, and the response was target due to the dominance of the prior probabilities for these outcomes in the absence of informative

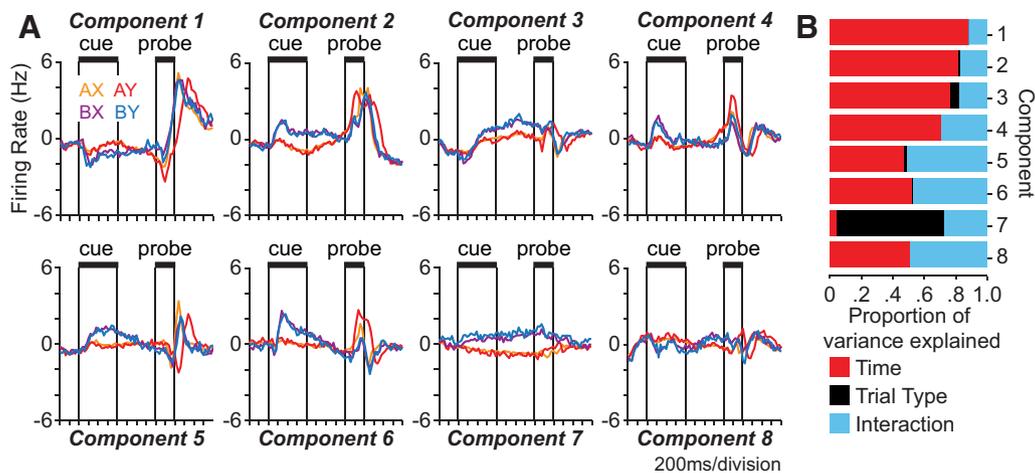


Figure 8. Demixed principal components of population activity and variance explained for each component on prepotent trial sets. **A**, Lines plot the first eight demixed principal components derived from the activity of all recorded neurons on prepotent trial sets. Color represents cue-probe sequence (trial type). **B**, Bars represent the proportion of variance in the first eight demixed principal components explained by time (red), trial type (black; cue-probe sequence), and their interaction (blue).

neural activity to the contrary. Conversely, the posterior probability associated with the B-cue, Y-probe, and nontarget response was near 0 during the intertrial interval based on the lower prior probability of these values before the onset of informative neural activity. However, on B-cue trials, the posterior probability of the B-cue climbed to ~ 0.8 after cue onset and remained elevated for the remainder of the trial, whether we included all neurons (Fig. 7A, blue) or just Group 2 and 4 neurons (Fig. 7B, blue) in the decoding analysis. On Y-probe trials, the posterior probability of Y-probes remained near 0 throughout the cue and delay periods, and then shortly after Y-probes were presented climbed to ~ 0.9 whether using all neurons (Fig. 7C, blue) or just Groups 2 and 4 (Fig. 7D, blue). Finally, on nontarget trials, the posterior probability of the nontarget response started near 0 during the intertrial interval and then climbed to an intermediate level after presentation of the cue (Fig. 7E, F, blue, cue period). This reflects that, on average, the mixture of A-cues and B-cues in the trial set provided partial information about the subsequent response direction (the response direction was ambiguous following A-cues, but following B-cues the nontarget response direction was fixed). The posterior probability of the nontarget response then climbed to nearly 1.0 during the probe period when remaining ambiguity about response direction on A-cue trials was eliminated by presentation of the probe stimulus and response direction was determined with certainty (Fig. 7E, F, blue). The posterior probability time course of nontarget responses exhibited the above features whether using all neurons (Fig. 7E, blue) or just Groups 2 and 4 (Fig. 7F, blue). The observation that highly accurate, cue, probe, and response direction decoding was obtained using only Group 2 and 4 neurons indicates that all three variables were encoded by the activity of neurons which exhibited “switch” activity. Moreover, despite the fact that the activity of these neurons was dynamically modulated in time, preferentially encoding B-cues early and A-cues late in the trial at both the single neuron and population levels, switch activity carried enough information about the identity of the cue at each time point to support a consistently elevated posterior probability and therefore a sustained neural representation of the cue that was maintained from the cue to the probe periods. Based on the activity of switch neurons, B-cue decoding, for example (Fig. 7B, blue), remained above baseline levels from the cue to the probe periods (although it dropped somewhat in the intervening delay).

Temporal patterning of PFC neural signals

Prefrontal neurons exhibited complex temporal patterns of modulation in their activity during the DPX trial, including the increasing and decreasing profiles associated with the encoding of B-cues and A-cues by switch neurons (Fig. 5C, G). To identify the set of temporal activity profiles present and quantify their relative influence on population activity, we applied a demixed principal component analysis (Brendel et al., 2011) to population activity. This analysis identified a set of orthogonal principal components (weighted linear combinations of neurons) that were oriented through the rate space to capture the maximum variance in population activity attributable to trial type (combination of cue and probe). We then projected the time-varying activity of the population onto each principal component to visualize the basis set of temporal activity patterns in relation to trial events (Fig. 8). Four of the first 8 principal components revealed a switch from early B-cue to late A-cue or AY trial type preference, which was the defining characteristic of switch neurons (Fig. 8A, components 2, 4, 5, 6). Some components resembled traditional working memory activity by demonstrating a consistent cue preference sustained throughout the delay period (Fig. 8A, components 3 and 7). The proportions of variance in each component explained by time, trial type, and their interaction are illustrated in Figure 8B. Most of the variance in the components was attributable to time (Fig. 8B, red), and the interaction between time and trial type (Fig. 8B, blue), with the exception of component 7, which was dominated by trial type as an isolated factor (Fig. 8B, black).

Influence of cognitive control load on neural activity

We presented both balanced and prepotent trial sets of the DPX task. In the balanced trial sets, AX, AY, BX, and BY cue-probe combinations were presented with equal frequency. In the prepotent trial sets, 69% of the trials presented the AX cue-probe sequence so that the large majority of trials required the target response, making this the prepotent response. We investigated whether the greater cognitive control required to inhibit the prepotent target response on prepotent relative to balanced trial sets modulated neural signals in prefrontal cortex. Cognitive control in the DPX task is most directly related to the necessity to inhibit the prepotent “target” response to the probe. Monkeys had to inhibit the prepotent target response under two conditions: whenever the B-cue was pre-

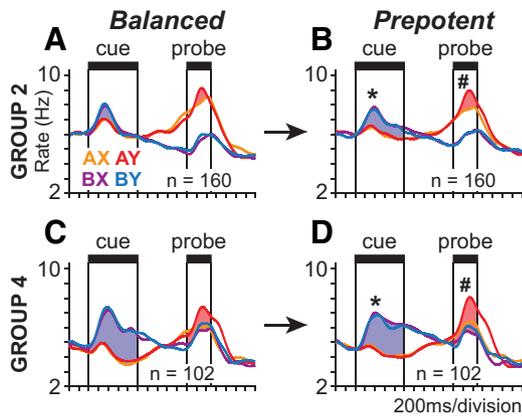


Figure 9. Relation of population activity to cognitive control demand. **A–D**, SDFs ($\sigma = 40$ ms) illustrate the activity of Group 2 (**A, B**) and Group 4 (**C, D**) neurons during balanced (**A, C**) and prepotent (**B, D**) trial sets (activity aligned to cue onset). Blue regions represent B-cue preference during the cue period (increase in activity on B-cue relative to A-cue trials). Red shaded regions represent the AY bias during the probe period measured as the increase in activity on AY trials relative either to AX trials (Group 2; **A, B**) or the three other trial types (Group 4; **C, D**). B-cue bias (*) and AY bias (#) were both significantly greater during prepotent (**B, D**) relative to balanced (**A, C**) trial sets (permutation test, $p < 0.05$).

sented and when the Y-probe was presented after the A-cue. The activity of Group 2 and 4 neurons (Fig. 5) was strongest under these conditions, suggesting that their activity plays a role in the neural circuit mediating cognitive control in the DPX task. If Group 2 and 4 neurons are part of a neural circuit mechanism for cognitive control, one might predict that the activity of these neurons would be enhanced during the prepotent relative to the balanced trial sets, reflecting the increase in cognitive control required. To test this, we computed the increase in activity evoked by B-cues relative to A-cues during the cue period (Fig. 9, blue shaded regions), as well as the increment in activity elicited by the presentation of Y-probes following A-cues during the probe period (Fig. 9, red shaded regions), and found that these differences in activity were significantly larger ($p < 0.05$) on prepotent (Fig. 9B,D) than balanced (Fig. 9A,C) trial sets in both Group 2 and 4 neurons. (Significance was determined by a permutation test shuffling trials between the prepotent and balanced sets and recalculating the relevant differences in population firing rates). These data indicate that the activity of switch neurons is enhanced under conditions of greater cognitive control. Although there is a correspondence therefore between the activation of switch neurons and conditions of greatest cognitive control demand, additional experiments would be needed to establish a causal link (by showing that the activity of switch neurons was uniquely modulated by changes in cognitive control demand relative to other neuronal types in prefrontal cortex, for example, or that selectively modulating the activity of switch neurons modulated cognitive control at the behavioral level).

Discussion

Cognitive control is the ability to rapidly switch between alternative responses to the same stimulus by adopting a new goal, rule, or strategy. It has been proposed that cognitive control is mediated by the activation of prefrontal neurons coding contextual information (such as a goal or rule) and that prefrontal output alters how sensory inputs are mapped to motor outputs in target cortical areas (Miller and Cohen, 2001). Some evidence in support of this framework has been

reported. Prefrontal neurons encode abstract rules (Wallis et al., 2001; Buschman et al., 2012; Eiselt and Nieder, 2013; Ott et al., 2014) and modulate their activity in relation to a change in cognitive strategy (Seo et al., 2014). Rule signals can modify which feature dimensions of a stimulus are integrated to make motor decisions in prefrontal local circuits (Mante et al., 2013), and prefrontal cortex transmits signals related to cognitive control to its cortical targets (Crowe et al., 2013).

The neural mechanisms of cognitive control are of substantial importance to the study of human neuropsychiatric disease because cognitive control is selectively disrupted in patients with schizophrenia (Barch et al., 2003; MacDonald et al., 2005; Jones et al., 2010; Richard et al., 2013), their first-degree relatives (Snitz et al., 2006), as well as patients with other neuropsychiatric diagnoses (Brambilla et al., 2007; Msetfi et al., 2009). However, the relationship between the neural correlates of cognitive control as studied in monkeys and the types of cognitive control that are disrupted in patients with schizophrenia has remained obscure because often, cognitive control is studied using very different behavioral tasks involving different stimuli, responses, and cognitive operations in monkeys and patients. To close this gap, we selected a task that measures cognitive control impairment in patients with schizophrenia, and translated it to monkeys directly with minimal modification. The task we selected for translation is a variant of the AX-CPT, termed the DPX task, which is identical to the AX-CPT, except that dot patterns replace letters as stimuli. The AX-CPT and DPX tasks both measure a specific cognitive control impairment in schizophrenia (Barch et al., 2003; MacDonald, 2008; Jones et al., 2010; Carter et al., 2012), and performance deficits in patients are associated with focally reduced activation of the dorsolateral prefrontal cortex (MacDonald and Carter, 2003; Perlstein et al., 2003; MacDonald et al., 2005; Yoon et al., 2008). It is unknown, however, how schizophrenia distorts information processing in prefrontal cortex at a cellular level. By translating the same task used to measure cognitive control deficits in patients with schizophrenia to monkeys directly and characterizing neural activity in prefrontal cortex at a cellular level during task performance, we hoped to model human cognitive control and its neural basis in the healthy brain. This serves as a necessary first step toward describing how a distortion of these neuronal activity patterns could lead to cognitive control failure in human neuropsychiatric disease. We report is that the neural representation of cue and probe stimuli is strongly influenced by their relation to logical operations required for cognitive control. This computation can be described as evaluating the logical relation: IF cue-A AND probe-X, THEN target response, else nontarget response. We describe how features of this logical operation were reflected in prefrontal neural activity below.

Impact of cognitive control on cue encoding

“Switch” neurons were the predominant type of neuron that encoded the identity of the cue stimulus in prefrontal cortex. Both the temporal dynamics and stimulus specificity of activity in switch neurons suggested they encoded the logical significance of cue stimuli for the computations that mediate cognitive control in the DPX task. During the cue period (“early cue” activity), switch neurons were preferentially activated by the presentation of the B-cue (Figs. 3B–D,G,K,M, 5C,G). The B-cue signaled the necessity to override the habitual “target” response to the upcoming probe stimulus in the DPX task, regardless of the identity of the subsequent probe.

Activation of switch neurons in response to presentation of the B-cue may reflect the fact that enough information had been provided by the B-cue during the cue period to definitively ascertain the “nontarget” status of the trial. The strong overrepresentation of B-cues during the cue period further suggests that prefrontal circuits are tuned to detect conditions under which habitual responses will not be successful, and cognitive control is needed to select and execute responses that run counter to habit. Patients with schizophrenia performing the AX-CPT make most errors on B-cue trials (Delawalla et al., 2008; MacDonald, 2008; Jones et al., 2010; Carter et al., 2012; Richard et al., 2013), and the dorsolateral prefrontal cortex is less active in patients than controls on B-cue trials (MacDonald et al., 2005). Cognitive control failures in schizophrenia could therefore partly reflect a failure to activate switch neurons when the B-cue is presented.

During the probe period in contrast, switch neurons were most strongly active on A-cue trials. Presentation of the A-cue imposed different logical constraints on behavior. On A-cue trials, the brain had to integrate information about the cue and probe to determine the response, and the A-cue by itself did not provide sufficient information to definitively ascertain the “nontarget” status of the trial. On A-cue trials, activity in switch neurons remained low throughout the cue period and started to increase during the delay period in advance of the presentation of the probe stimulus (Fig. 5C,G), in effect anticipating the time when additional information provided by the probe would definitively determine the “target” or “nontarget” status of the trial. In this way, the temporal dynamics of activity in switch neurons was tailored to the differing time points on B-cue and A-cue trials at which information provided by incoming stimuli became available to definitively ascertain the “nontarget” status of the trial. This corresponded to detecting the necessity to engage cognitive control versus just releasing the habitual “target” response. Consistent with the hypothesis that the physiological dynamics of switch neurons bear a functional relationship to cognitive control, we found that switch neuron activity was significantly elevated under conditions of increased cognitive control demand (Fig. 9).

Impact of cognitive control on probe encoding

Encoding of probe stimuli was also modulated according to the logical constraints of the task. Differential (X vs Y) probe signals were strong when the cue was ambiguous and probe information had to be integrated to make the correct contingent choice (Fig. 6A). Otherwise, probe signals were weak when the cue by itself determined response direction and the probe was irrelevant (Fig. 6B). Following an A-cue, monkeys could anticipate a “target” response because most A-cues were followed by X-probes (the target sequence). Presentation of the Y-probe after an A-cue would therefore require inhibiting this prepotent response, increasing the demand on cognitive control. Evidence of the behavioral cost associated with increased cognitive control was provided by the significantly elevated RT on AY trials relative to the other trial types (Fig. 2A,B). The observation that population activity was generally stronger on AY trials (Fig. 5C,E,G, red) than AX trials (Fig. 5C,E,G, orange) suggests that this activity was modulated in relation to the degree of cognitive control required. It is noteworthy that, although RT was longest and population neural activity strongest on AY trials, monkeys were slightly more accurate on Y-probe than X-probe trials overall (Table 1; performance on AY and BY trials is slightly better than perfor-

mance on AX and BY trials). The prolonged RT on AY trials may have indicated that monkeys traded speed for accuracy on AY trials when cognitive control increased. The overall better level of performance on Y-probe relative to X-probe trials may have additionally reflected the benefit of a sensorimotor habit to respond “nontarget” in response to the Y-probe in general. In general, both early cue encoding (biased toward B-cues) and probe encoding (biased toward Y-probes) reflected the preferential activation of prefrontal neurons under stimulus conditions that were associated with increased cognitive control demand defined as the necessity to prohibit a habitual response.

A prior study recording local field potentials in monkey prefrontal cortex related neural activity to the performance of an AX continuous performance task (Dias et al., 2006). This approach provided the advantage that neural signals (local field potential) in monkeys could be directly compared with neural signals (EEG) recorded in patients with schizophrenia (Dias et al., 2006). That study reported that local field potential modulation amplitude was greater during the cue period on B-cue trials and during the probe period on AY trials (Dias et al., 2006), a pattern qualitatively similar to the activity profiles of several neuronal groups in the current data (Fig. 5C,G). The present study extends this prior work by characterizing how information processing for cognitive control relates to neural activity at the single-cell level.

Impact of cognitive control on response encoding

When activity was aligned to movement onset, neural signals reflecting the direction of the motor response favored the nontarget response direction over the target response direction (Fig. 5J). Other neurons with activity that peaked at the time of movement onset did not clearly differentiate between response directions (Fig. 5L), indicating a weak prefrontal representation of the habitual target response direction in general. Therefore, the neural representation of responses in prefrontal cortex favored responses that were not automatic or habitual.

Describing how cognitive control relates to patterns of action potentials in neurons is likely to provide important clues about the pathogenic mechanisms underlying schizophrenia. Specifically, results from genome-wide association studies have identified a functionally convergent group of risk mutations for schizophrenia that occur in the vicinity of genes involved in NMDA receptor mediated activity-dependent synaptic plasticity in the brain (Javitt and Zukin, 1991; Harrison and Weinberger, 2005; Beneyto and Meador-Woodruff, 2008; Lisman et al., 2008; Kirov et al., 2012; Weickert et al., 2013; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). The fact that activity-dependent synaptic plasticity can be exquisitely sensitive to the precise timing of spikes in pre- and postsynaptic neurons (Feldman, 2012) suggests that understanding how schizophrenia alters the patterning of cortical activity at the level of spiking neurons may be essential to understanding the pathogenesis of the disease. By characterizing how patterns of activity in individual neurons relate to cognitive control in the healthy brain, the present study takes a step toward understanding how a disturbance of neuronal activity patterns could drive cognitive control deficits in schizophrenia and potentially exacerbate synaptic deficits in the disease state. Our results elucidate the neural mechanisms performing logical computations for cognitive control at the level of spiking neurons in the prefrontal cortex

and characterize the neurophysiological processes that are likely to be disrupted in schizophrenia at a cellular level.

References

- Asaad WF, Rainer G, Miller EK (2000) Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol* 84:451–459. [Medline](#)
- Barch DM, Carter CS, MacDonald AW 3rd, Braver TS, Cohen JD (2003) Context-processing deficits in schizophrenia: diagnostic specificity, 4-week course, and relationships to clinical symptoms. *J Abnorm Psychol* 112:132–143. [CrossRef Medline](#)
- Beneyto M, Meador-Woodruff JH (2008) Lamina-specific abnormalities of NMDA receptor-associated postsynaptic protein transcripts in the prefrontal cortex in schizophrenia and bipolar disorder. *Neuropsychopharmacology* 33:2175–2186. [CrossRef Medline](#)
- Blackman RK, MacDonald AW 3rd, Chafee MV (2013) Effects of ketamine on context-processing performance in monkeys: a new animal model of cognitive deficits in schizophrenia. *Neuropsychopharmacology* 38:2090–2100. [CrossRef Medline](#)
- Brambilla P, MacDonald AW 3rd, Sassi RB, Johnson MK, Mallinger AG, Carter CS, Soares JC (2007) Context processing performance in bipolar disorder patients. *Bipolar Disord* 9:230–237. [CrossRef Medline](#)
- Brendel W, Romo R, Machens CK (2011) Demixed principal component analysis. Paper presented at twenty-fifth annual Conference on Neural Information Processing Systems, Granada, Spain, December.
- Buschman TJ, Denovellis EL, Diogo C, Bullock D, Miller EK (2012) Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron* 76:838–846. [CrossRef Medline](#)
- Carter CS, Minzenberg M, West R, MacDonald A 3rd (2012) CNTRICS imaging biomarker selections: executive control paradigms. *Schizophr Bull* 38:34–42. [CrossRef Medline](#)
- Chafee MV, Goldman-Rakic PS (1998) Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J Neurophysiol* 79:2919–2940. [Medline](#)
- Crowe DA, Goodwin SJ, Blackman RK, Sakellaridi S, Sponheim SR, MacDonald AW 3rd, Chafee MV (2013) Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nat Neurosci* 16:1484–1491. [CrossRef Medline](#)
- Delawalla Z, Csernansky JG, Barch DM (2008) Prefrontal cortex function in nonpsychotic siblings of individuals with schizophrenia. *Biol Psychiatry* 63:490–497. [CrossRef Medline](#)
- Dias EC, McGinnis T, Smiley JF, Foxe JJ, Schroeder CE, Javitt DC (2006) Changing plans: neural correlates of executive control in monkey and human frontal cortex. *Exp Brain Res* 174:279–291. [CrossRef Medline](#)
- Eiselt AK, Nieder A (2013) Representation of abstract quantitative rules applied to spatial and numerical magnitudes in primate prefrontal cortex. *J Neurosci* 33:7526–7534. [CrossRef Medline](#)
- Feldman DE (2012) The spike-timing dependence of plasticity. *Neuron* 75:556–571. [CrossRef Medline](#)
- Goldman-Rakic PS (1987) Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: *Handbook of physiology: the nervous system. Higher functions of the brain* (Mountcastle VB, Plum F, Geiger SR, eds), pp 373–417. Bethesda, MD: American Physiological Society.
- Goodwin SJ, Blackman RK, Sakellaridi S, Chafee MV (2012) Executive control over cognition: stronger and earlier rule-based modulation of spatial category signals in prefrontal cortex relative to parietal cortex. *J Neurosci* 32:3499–3515. [CrossRef Medline](#)
- Harrison PJ, Weinberger DR (2005) Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol Psychiatry* 10:40–68. [CrossRef Medline](#)
- Javitt DC, Zukin SR (1991) Recent advances in the phencyclidine model of schizophrenia. *Am J Psychiatry* 148:1301–1308. [CrossRef Medline](#)
- Jones JA, Sponheim SR, MacDonald AW 3rd (2010) The dot pattern expectancy task: reliability and replication of deficits in schizophrenia. *Psychol Assess* 22:131–141. [CrossRef Medline](#)
- Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, Moran J, Chambert K, Toncheva D, Georgieva L, Grozeva D, Fjodorova M, Wollerton R, Rees E, Nikolov I, van de Lagemaat LN, Bayés A, Fernandez E, Olason PI, Böttcher Y, et al. (2012) De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* 17:142–153. [CrossRef Medline](#)
- Klecka WR (1980) *Discriminant analysis*. Newbury Park, CA: Sage.
- Lara AH, Wallis JD (2014) Executive control processes underlying multi-item working memory. *Nat Neurosci* 17:876–883. [CrossRef Medline](#)
- Lisman JE, Coyle JT, Green RW, Javitt DC, Benes FM, Heckers S, Grace AA (2008) Circuit-based framework for understanding neurotransmitter and risk gene interactions in schizophrenia. *Trends Neurosci* 31:234–242. [CrossRef Medline](#)
- MacDonald AW 3rd (2008) Building a clinically relevant cognitive task: case study of the AX paradigm. *Schizophr Bull* 34:619–628. [CrossRef Medline](#)
- MacDonald AW 3rd, Carter CS (2003) Event-related fMRI study of context processing in dorsolateral prefrontal cortex of patients with schizophrenia. *J Abnorm Psychol* 112:689–697. [CrossRef Medline](#)
- MacDonald AW 3rd, Carter CS, Kerns JG, Ursu S, Barch DM, Holmes AJ, Stenger VA, Cohen JD (2005) Specificity of prefrontal dysfunction and context processing deficits to schizophrenia in never-medicated patients with first-episode psychosis. *Am J Psychiatry* 162:475–484. [CrossRef Medline](#)
- Mante V, Sussillo D, Shenoy KV, Newsome WT (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503:78–84. [CrossRef Medline](#)
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202. [CrossRef Medline](#)
- Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* 16:5154–5167. [Medline](#)
- Msetfi RM, Murphy RA, Kornbrot DE, Simpson J (2009) Impaired context maintenance in mild to moderately depressed students. *Q J Exp Psychol* 62:653–662. [CrossRef Medline](#)
- Ott T, Jacob SN, Nieder A (2014) Dopamine receptors differentially enhance rule coding in primate prefrontal cortex neurons. *Neuron* 84:1317–1328. [CrossRef Medline](#)
- Perlstein WM, Dixit NK, Carter CS, Noll DC, Cohen JD (2003) Prefrontal cortex dysfunction mediates deficits in working memory and prepotent responding in schizophrenia. *Biol Psychiatry* 53:25–38. [CrossRef Medline](#)
- Richard AE, Carter CS, Cohen JD, Cho RY (2013) Persistence, diagnostic specificity and genetic liability for context-processing deficits in schizophrenia. *Schizophr Res* 147:75–80. [CrossRef Medline](#)
- Rigotti M, Barak O, Warden MR, Wang XJ, Daw ND, Miller EK, Fusi S (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497:585–590. [CrossRef Medline](#)
- Roy JE, Riesenhuber M, Poggio T, Miller EK (2010) Prefrontal cortex activity during flexible categorization. *J Neurosci* 30:8519–8528. [CrossRef Medline](#)
- Schiffer AM, Waszak F, Yeung N (2015) The role of prediction and outcomes in adaptive cognitive control. *J Physiol Paris* 109:38–52. [CrossRef Medline](#)
- Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511:421–427. [CrossRef Medline](#)
- Seo H, Cai X, Donahue CH, Lee D (2014) Neural correlates of strategic reasoning during competitive games. *Science* 346:340–343. [CrossRef Medline](#)
- Snitz BE, MacDonald AW 3rd, Carter CS (2006) Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: a meta-analytic review of putative endophenotypes. *Schizophr Bull* 32:179–194. [CrossRef Medline](#)
- Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411:953–956. [CrossRef Medline](#)
- Weickert CS, Fung SJ, Catts VS, Schofield PR, Allen KM, Moore LT, Newell KA, Pellen D, Huang XF, Catts SV, Weickert TW (2013) Molecular evidence of N-methyl-D-aspartate receptor hypofunction in schizophrenia. *Mol Psychiatry* 18:1185–1192. [CrossRef Medline](#)
- Yoon JH, Minzenberg MJ, Ursu S, Ryan Walter BS, Wendelken C, Ragland JD, Carter CS (2008) Association of dorsolateral prefrontal cortex dysfunction with disrupted coordinated brain activity in schizophrenia: relationship with impaired cognition, behavioral disorganization, and global function. *Am J Psychiatry* 165:1006–1014. [CrossRef Medline](#)