Cerebral Cortex Advance Access published November 21, 2012

Cerebral Cortex
doi:10.1093/cercor/bhs361

Encoding of Rules by Neurons in the Human Dorsolateral Prefrontal Cortex

Matthew K. Mian[1], Sameer A. Sheth[1], Shaun R. Patel[1,2], Konstantinos Spiliopoulos[1], Emad N. Eskandar[1] and Ziv M. Williams[1]

[1]Nayef al-Rodhan Laboratories, Department of Neurosurgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA and [2]Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA, USA

Address correspondence to Ziv Williams, Massachusetts General Hospital–Harvard Medical School Center for Nervous System Repair, Harvard Medical School, 55 Fruit Street, WAC 745K, Boston, MA 02114, USA. Email: zwilliams1@partners.org

We use rules to extend learned behavior beyond specific instances to general scenarios. The prefrontal cortex (PFC) is thought to play an important role in representing rules, as evidenced by subjects who have difficulty in following rules after PFC damage and by animal studies demonstrating rule sensitivity of individual PFC neurons. How rules are instantiated at the single-neuronal level in the human brain, however, remains unclear. Here, we recorded from individual neurons in the human dorsolateral prefrontal cortex (DLPFC) as subjects performed a task in which they evaluated pairs of images using either of 2 abstract rules. We find that DLPFC neurons selectively encoded these rules while carrying little information about the subjects' responses or the sensory cues used to guide their decisions.

Keywords: abstract rules, deep brain stimulation, human neurophysiology, prefrontal cortex

Introduction

Many aspects of our lives are guided by rules. Rules can be concrete, where a particular stimulus prompts an automatic response (e.g. a red traffic light cues us to brake). We can also learn and follow rules that are abstract, wherein a guiding principle is not bound to a specific context and may be generalized to both familiar and novel instances (e.g. understanding that a knife rather than a spoon can enable us to cut a steak). Our ability to learn and then apply abstract rules endows us with a broad, flexible behavioral repertoire, supporting complex pursuits such as tool use (Penn and Povinelli 2007).

The prefrontal cortex (PFC), the putative neural substrate of working memory (Fuster and Alexander 1971; Miller et al. 1996; Rao et al. 1997), is thought to represent rules, in part because damage to this region impairs one's ability to follow rules (Comalli et al. 1962; Milner 1963; Luria 1966; Gershberg and Shimamura 1995; Wise et al. 1996)—even when those rules can be articulated (Shallice and Burgess 1991). Consistently, neurons in the nonhuman primate dorsolateral prefrontal cortex (DLPFC), have been demonstrated to represent both concrete (Watanabe 1990; Asaad et al. 2000; Murray et al. 2000) and abstract rules (White and Wise 1999; Wallis et al. 2001). Moreover, lesion studies in primates have solidified the importance of the PFC to applying rules in the service of goals (Passingham 1993; Parker and Gaffan 1998; Buckley et al. 2009). On the other hand, primate studies involving DLPFC lesions sometimes fail to reveal deficits in rule-guided behavior (Petrides 1982; Mansouri et al. 2007; Buckley et al. 2009), and other work has emphasized the contributions of other PFC regions (Petrides 1982; Gaffan and Harrison 1989; Murray et al. 2000; Bussey et al. 2001), minimizing the role of the DLPFC.

How abstract rules are encoded at the neuronal level in humans is still poorly understood, in part because of the lack of availability of single-neuronal recordings. In addition, it is challenging to establish whether abstract rules are similarly represented in humans who often process sensory cues under complex analogical contexts that may not be easily or faithfully modeled in animals.

To investigate the role of the human DLPFC in representing abstract rules, we recorded from individual DLPFC neurons in human neurosurgical subjects (Williams et al. 2004; Sheth et al. 2012) engaged in a task that required them to switch flexibly between 2 rules on a trial-by-trial basis. We hypothesized that the activity of neurons in the human DLPFC would distinguish between these abstract rules.

In each task trial (Fig. 1*a*), subjects viewed 2 sequential images of everyday items and then used a joystick to indicate whether the images "matched." Subjects were instructed to select a green target for matching pairs of images and a red target for nonmatching images.

Images consisted of items that could either deliver an action on another item ("tools"; e.g. a hammer) or serve as the target of such action ("objects"; e.g. a nail). The subjects were instructed to evaluate image pairs using either of 2 rules: 1) Can one item act appropriately on the other or 2) are the items similar in identity? However, subjects were not told which of these criteria to apply to the individual pairings; rather, they inferred the more applicable of the 2 rules on a trial-by-trial basis. That is, based on the image pair alone, subjects determined which rule to use and then whether the pair matched under the chosen rule (Fig. 1*b*,*c*).

Matching trials consisted of either appropriate-use (e.g. hammer–nail) or similar-identity (e.g. 2 different hammers or 2 different nails) image pairs. Nonmatching trials, in contrast, presented either inappropriate-use (e.g. hammer–nut) or dissimilar-identity (e.g. hammer–wrench or nail–nut) pairs. Target selection was confirmed by the disappearance of the target that was not chosen, but subjects received no feedback on the correctness of their selections. They were also given no instruction on whether to consider items as tools or objects. Item pairings were randomly interleaved and presented with equal frequency.

Materials and Methods

Study Subjects

Subjects enrolled after providing written informed consent, and all aspects of this study were approved by the Massachusetts General Hospital Institutional Review Board. We recruited 8 study subjects (Tables 1 and 2) undergoing deep brain stimulator placement. Five of the 8 subjects performed the standard behavioral task (see "Behavioral Task," below), and 3 performed a control task that also included "object–tool" trials. The mean age was 60.3 ± 4.6 (mean ± standard

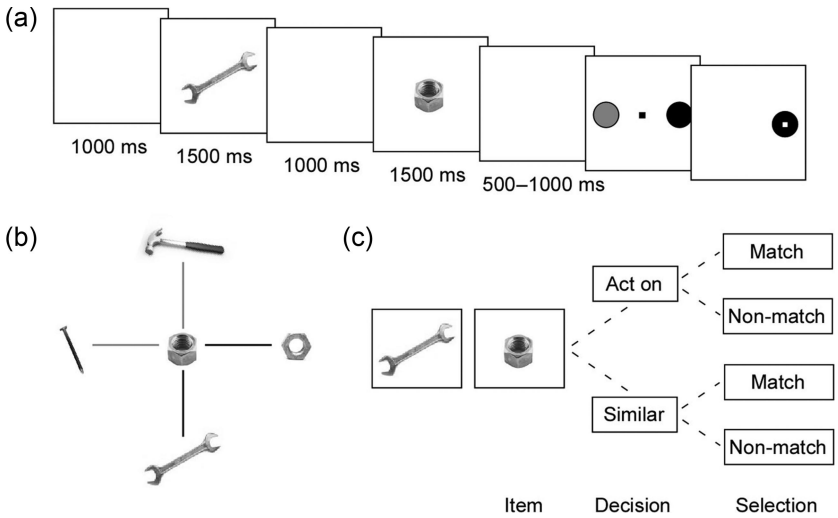

Figure 1. Behavioral task. (*a*) Subjects viewed 2 sequentially presented images (1500 msec each) an intervening fixed delay (1000 msec). After a variable second delay (500–1000 msec), they were shown a red and green target randomly positioned on either side of the screen. They then used a joystick to move a cursor to the green target to indicate that the images matched or to the red target to indicate that they did not. Images could match either as a compatible TO pair (e.g. wrench–nut, as depicted here) or as 2 items similar in identity (e.g. wrench–wrench or nut–nut). (*b*) Sample permutation of possible item combinations with a nut. Black bars represent matches, and gray bars nonmatches. (*c*) Evaluation paradigm for a sample item pair. Subjects first decided whether to invoke the act on or similar rule and then judged the pair to be either matching or nonmatching under the selected rule.

Table 1
Subject characteristics

Subject	Number of neurons	Age (years)	Gender	Diagnosis	Recording side
1	9	56	M	Thalamic pain syndrome	R
2	8	65	M	Essential tremor	L
3	19	65	F	Essential tremor	L
4	1	48	F	MS-related tremor	L
5	5	44	M	SCA type 3	R
6	1	86	F	Essential tremor	L
7	2	55	F	Parkinson's disease	L
8	3	63	M	Parkinson's disease	L

Note: Subjects 1–5 performed the main task, and subjects 6–8 performed the control task that included object–tool trials.
SCA: spinocerebellar ataxia; MS: multiple sclerosis; R: right; L: left.

Table 2
Subject recording locations

Subject	Left–right (mm)	Anterior–posterior (mm)	Superior–inferior (mm)
1	23.0	30.0	59.0
2	−22.0	22.0	63.0
3	−25.0	31.0	54.0
4	−25.0	34.0	52.0
5	36.0	21.0	58.0
6	−30.0	13.0	63.0
7	−38.0	26.0	51.0
8	−37.0	12.0	61.0

Note: Coordinates given in MNI 152 space. Subjects 1–5 performed the main task, and subjects 6–8 performed the control task that included OT trials.

error of the mean [SEM]) years. Single-unit recordings were made from the left hemisphere in 6 subjects and from the right hemisphere in 2 (Fig. 2).

Microelectrode recordings are routinely performed prior to the implantation of deep brain stimulating electrodes to aid with localization (Amirnovin et al. 2006; Gross et al. 2006). Consideration for surgery was unrelated to the research protocol and was conducted by a multidisciplinary team of neurologists, neurosurgeons, a neuropsychologist, and a nurse practitioner. After a patient had been evaluated, consented, and scheduled for surgery, an independent member of the research team approached the patient about study inclusion. Subjects enrolled voluntarily, providing informed consent under guidelines approved by the Massachusetts General Hospital Institutional Review Board, and were free to withdraw from the study at any time, including during surgery, without consequence to the operative approach or clinical care.

Behavioral Task

Subjects viewed a computer monitor mounted at the eye level, and a joystick was positioned near the hand contralateral to the hemisphere of recordings. The task was presented using a customized software package written in MATLAB (MathWorks, Inc., Natick, MA, USA) that provided millisecond temporal precision of all task events (Asaad and Eskandar 2008a, 2008b).

Each trial (Fig. 1*a*) began with the appearance of a white circle in the center of the screen (1000 msec). Next, an image of an item identifiable as either a tool or an object was presented (1500 msec). The subjects viewed a white circle during an intervening delay (1000 msec), and then a second image of either a tool or an object was presented (1500 msec). After a variable delay (500–1000 msec), 2 circular targets (one red and one green) appeared on either side of the white circle. The target locations (left vs. right) were shuffled randomly, so subjects could not anticipate either the upcoming target location or the appropriate movement. After appearance of the targets, the subject had up to 5 s to use the joystick to move a cursor from the central circle to one of the targets.

To familiarize the subjects with task requirements, they were allowed 5–10 min of practice (with an entirely different set of images) prior to the recordings. At the onset of the practice session, subjects were given the following instructions: 1) You will be shown sequential pairs of items. 2) You will need to decide whether each pair of items "matches." 3) Two items match if one can be used appropriately on the other or if the 2 items are similar in identity. 4) To indicate a match, move the joystick to the green target; to indicate a nonmatch, move to the red target.

Therefore, as in natural human behavior, the subjects were not given explicit cues to instruct them which rule to use on each trial (i.e. a tone at the beginning of the trial), but rather inferred which rule to use based on the presented images.

The image pairs used in the main task included random combinations of tools and objects (i.e. "tool–tool," "object–object," and "tool–object" combinations; object–tool [OT] combinations were added in a

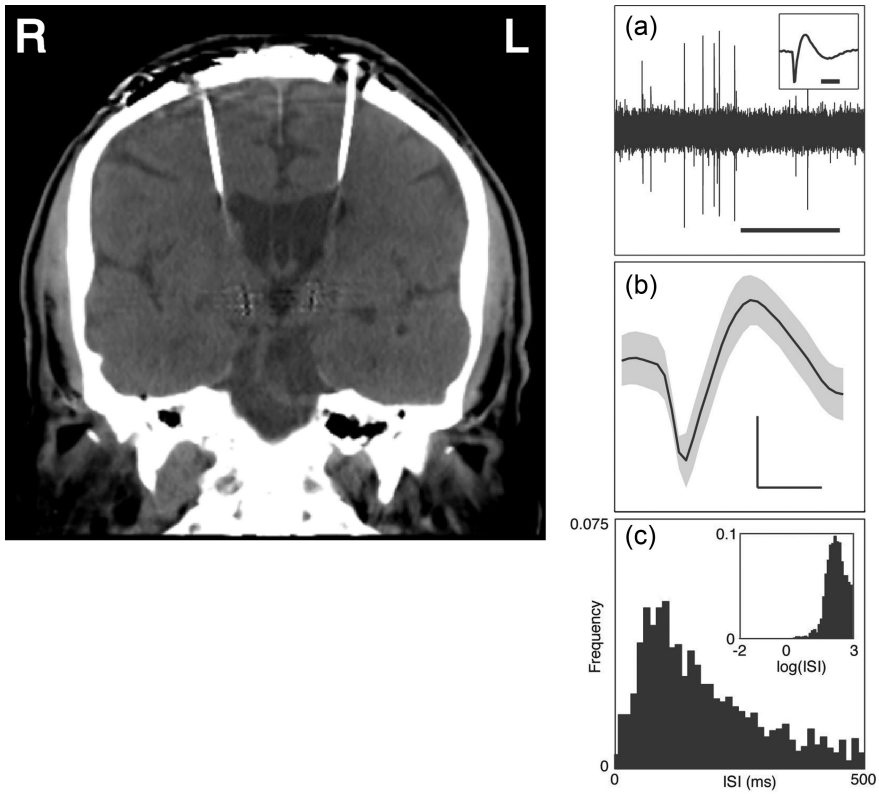

Figure 2. Neuronal recordings. On the left, CT scan (coronal slice) demonstrating the final position of bilateral deep brain stimulating electrodes in a study subject (subject 2 in Tables 1 and 2; note: $y = 22.0$ mm for the slice shown). Single-neuronal recordings in the DLPFC were possible since the superior aspect of each electrode trajectory traverses Brodmann area 9 of the DLPFC. On the right, single-unit isolation. (*a*) Example of a raw voltage tracing, with the inset demonstrating one of the action potentials. The horizontal bars represent 1 s and 1 ms in the large tracing and inset, respectively. (*b*) Voltage tracing demonstrating the mean waveform ± SD of an example neuron. Horizontal and vertical bars correspond to 1 ms and 0.1 mV, respectively. (*c*) Interspike interval (ISI) of all isolated spikes for the example neuron in (*b*). The inset depicts the same ISI distribution on a logarithmic time scale.

separate control task described below). Images were drawn from a common pool such that an image presented first during one trial could also be presented second in another. Equal proportions of matching and nonmatching pairs were given. Furthermore, the images in all pairings were always of different shape, size, and orientation.

The different trial types were given with equal frequency and pseudorandomly interleaved so subjects could not anticipate the type or compatibility of forthcoming pairs. Subjects viewed combinations of a total of 8 tool types and 8 object types. Pairings were presented an average of 2.0 ± 0.5 times in a given task session, and subjects performed 1.8 ± 0.7 sessions for which different sets of cells were recorded.

Electrophysiologic Recordings

Subjects underwent standard burr hole placement and dural opening. A set of tungsten microelectrodes (500–1500 kΩ; FHC, Inc., Bowdoin, ME, USA) was attached to a motorized microdrive (Alpha Omega Engineering, Nazareth, Israel) as described previously (Williams et al. 2004). Up to 5 microelectrodes were placed in a ben-gun configuration. These were advanced into the cortex under direct visualization, starting approximately 2 mm above the cortical surface. Once the electrodes were in contact with the cortex, the burr hole was covered with a layer of fibrin sealant (Tisseel; Baxter, Inc., Deerfield, IL, USA) to limit cortical pulsation.

Electrodes were advanced in increments of 10–100 μm. Once putative neurons were noted on one or more channels, the electrodes were left in place and monitored for signal stability. Neurons were not screened for task responsiveness. Recordings were amplified, bandpass filtered between 300 Hz and 6 kHz, and sampled at 20 kHz using an Alpha Omega (Alpha Omega Engineering, Nazareth, Israel) recording system. Activity was captured using a Spike2 software package (Cambridge Electronic Design, Cambridge, UK) and saved for offline sorting (Offline Sorter, Plexon, Inc., Dallas, TX, USA). Once recordings were complete, the surgery proceeded as planned. No long-acting sedatives were given prior to or during recordings, as per standard operative protocol.

A histogram of peak heights was constructed from the raw voltage tracings, and a minimum threshold of 3 standard deviations (SDs) was applied to filter out the majority of the low-voltage background activity. Classification of action potentials was performed using template matching and principal component analysis. Putative neurons were required to separate clearly from any channel noise, to demonstrate waveform morphology consistent with that of a cortical neuron, and to have at least 99% of spikes separated by a minimum refractory interspike interval of 1 msec. This process is outlined in Figure 2.

We recorded an average of 0.4 well-isolated neurons per electrode. When an individual electrode recorded more than one neuron, a high degree of isolation was required in order to include each as a single unit ($P < 0.01$, multivariate analysis of variance across the first 2 principal components). Candidate neurons were required to fire at a rate of at least 1.0 spikes/s and be stably active for at least 30 task trials. No multiunit activity was used.

Confirming Recording Locations

Stereotactic localization for surgical planning was performed using a magnetic resonance imaging scan. We confirmed recording locations with a high-resolution postoperative computed tomography (CT) scan (slice thickness 0.5 mm), referenced with the sites of the burr hole and the permanent deep brain stimulating electrode. CT data were

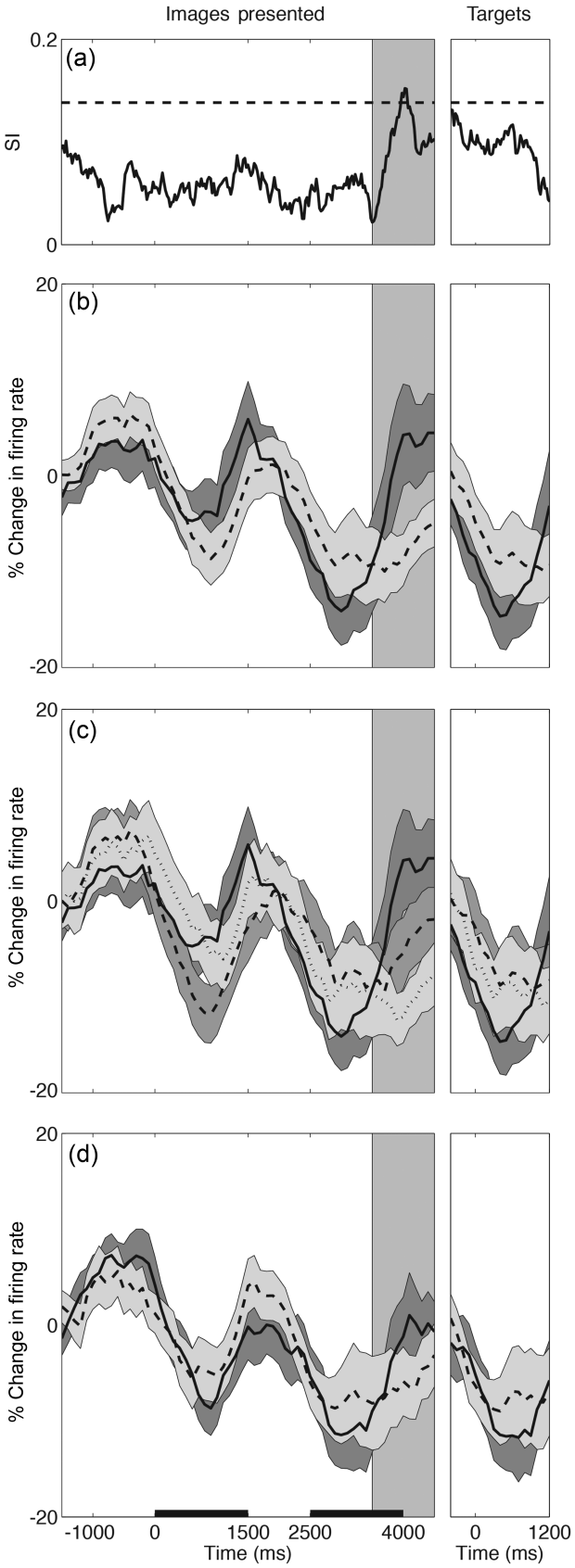

Figure 3. Time courses of DLPFC population responses ($n = 42$). (*a*) SI for the population activity across all conditions with the dotted horizontal line indicating the upper 95% bound by bootstrap analysis. The normalized population firing rate ± SEM (shaded region) is shown for (*b*) act on (i.e. TO; solid line) versus similar criterion (i.e. TT/OO; dashed line) trials; (*c*) TO (solid) versus TT (dashed) versus OO (dotted) trials; and (*d*) trials in which items were matched (solid) versus nonmatched (dashed). Activity is aligned to the first image presentation (left), and red–green target presentation (right). The horizontal scale bars in *(d)* indicate the first and second image presentation periods. The shaded vertical region indicates the second image delay period, the center of which corresponds to the maximal SI (peak in *a*).

linearly registered to Montreal Neurological Institute (MNI) 152 standard space (1 mm resolution) using the cost function of mutual information and 12 degrees of freedom. The automated linear registration tool is based around a multistart, multiresolution global optimization method (Jenkinson and Smith 2001; Jenkinson et al. 2002). Recording locations were visualized on the standard MNI brain, and stereotactic coordinates were noted for each patient. Recording sites in all subjects (Table 2) were located within the caudal portion of Brodmann area 9 of the DLPFC.

Selectivity Index

The window following the second image presentation represents the time during which we expected information about evaluative rules to be represented. To choose a temporal window relevant to neuronal activity in this task without introducing a time-of-interest bias, we calculated a selectivity index (SI) (Moody et al. 1998) using the neuronal population responses of all 6 primary task conditions. The SI was binned in 1000 msec time windows centered on times t_i and advanced in 100 msec increments across all n trial conditions $X_1 \ldots X_n$, whereby:

$$\mathrm{SI}(t_i) = \frac{[n - \sum_{j=1\ldots n} X_j(t_i)/\max(X_1(t_i) \ldots X_n(t_i))]}{(n-1)}.$$

Therefore, an SI of 0 represents no discrimination between trial conditions and an SI of 1 represents complete discrimination of one condition. We designated the center of our epoch of interest as the time point of the maximum SI, and this choice was independent of which of the tested conditions contributed to the SI. Again, this method for determining SI is insensitive to which conditions are discriminated by the neuronal activity; it is not biased toward any particular task condition.

Significance of SI was determined by bootstrap analysis. The average neuronal response to each condition was determined, shuffling was performed across all time points, and a new shuffled SI was computed. This was repeated 5000 times. Confidence bounds were then determined from this set of shuffled SIs. Importantly, neuronal responses were normalized (divided by each neuron's average response during a 1000-msec window prior to the trial) so that variability in the firing rate between neurons did not bear on this analysis.

Although we focused our analysis around the peak of the SI, this window corresponds to a period during which PFC neurons would be posited to encode differences between task conditions (i.e. the delay period prior to the movement cue). From this perspective (as well as the visually apparent nature of our main finding; Fig. 3*b*), our results were not specifically affected by the method chosen to select the period of interest.

Neuronal Analysis and Statistical Methods

Peristimulus histograms and rasters were constructed for all isolated neurons. To permit comparisons between cells with different firing rates, neuronal activity was normalized (divided) by the mean baseline firing rate during a 1000 msec window prior to the appearance of the central point at the onset of the trial (during which time subjects viewed a blank screen). Activity during the small proportion of incorrect trials (5.2%) was excluded. Differences in neuronal response between task conditions were assessed for both the single units and the population using Mann–Whitney and *t*-tests, with $P < 0.05$ considered statistically significant.

Sequential feature selection with stepwise regression was used to examine whether variations in neuronal activity were predicted by particular task or subject features. This statistical approach creates a

multilinear model in which the explanatory powers of different features are sequentially tested (Richard Draper and Smith 1998). Feature terms are added and removed from the model in a stepwise manner using regression *P*-values of 0.05 and 0.10, respectively. Task-related features tested in our model included pairing type (tool–object, TO; tool–tool, TT; and object–object, OO), category of the first and second images (tool and object), and matching status of the image pair (match and nonmatch). Subject-related features included disease process (essential tremor [ET] and non-ET) and hemisphere of recordings (right and left). In addition, a standard multilinear regression was performed using the statistical computing package in MATLAB (MathWorks, Inc., Natick, MA, USA).

Results

Behavior and Recordings

Five study subjects (Tables 1 and 2) performed the main behavioral task, and 3 additional subjects performed a control task, described below. The subjects completed 95.4 ± 2.3% (mean ± SEM) of trials correctly; incorrect trials were excluded from our analysis. In total, we recorded 48 well-isolated single neurons from Brodmann area 9 of the DLPFC (Fig. 2), with a mean baseline firing rate of 3.0 ± 0.4 (mean ± SEM) spikes per second (importantly, these recordings were made acutely in the operating room rather than from chronic multielectrode arrays).

To first determine the time period during which neural activity carried the greatest degree of discrimination across tested pairings, we calculated a population SI. The period centered on the end of the second image presentation, termed here the "second image delay," had the highest selectivity ($P = 0.018$, bootstrap test; Fig. 3*a*, shaded area).

DLPFC Neurons Encode Abstract Rules

During the second image delay, DLPFC neurons responded differentially to the rules that were used to evaluate the item pairings. To examine trials in which the second presented item was the same but the rule varied, we compared activity on TO and OO trials.

We found that, across the entire pool of neurons recorded for the main task ($n = 42$), the population-averaged firing rate differed significantly based on the abstract rule (i.e. "act on" vs. "similar" criterion; $P = 0.010$, Mann–Whitney test; Figs 3*c* and 4*a*). Rule selectivity developed 300 msec prior to the end of the second image presentation and persisted until 175 msec before the colored targets appeared (bootstrap test at $P < 0.05$). Consistently, 24% of individual cells demonstrated a difference in firing during this interval based on the rule that was used (Mann–Whitney test at $P < 0.05$; examples in Fig. 5), with 17% and 7% firing preferentially for the act on and similar criteria, respectively.

When considering all trials in which the rule differed without regard to the category of the second image (i.e. TO vs. the composite of all TT and OO trials), we similarly find a significant difference in population-averaged activity (Figs 3*b* and 4*a*; $P = 0.04$; Mann–Whitney test). Further, there is no difference in aggregate activity between TT and OO trials (Figs 3*b* and 4*a*; $P = 0.16$; Mann–Whitney test)—the 2 trial types in which the same rule (the similar criterion) is invoked.

At the individual neuron level, many cells discriminated between TO and each TT/OO and TT trials. Twelve of the 42 recorded cells (29%) demonstrated differences in firing during the second image delay period between TO and TT trial types at $P < 0.05$, with 10 (24%) increasing their firing and 2 (5%) decreasing. Including all TT and OO trials, we similarly find that 13 of the 42 cells (31%) distinguish rule types (TO vs. TT/OO), with 11 (26%) increasing their firing for TO versus TT/OO trials and 2 (5%) decreasing.

Neuronal responses were heterogeneous (Fig. 4*b*), with 19 of the 42 neurons (45%) demonstrating at least a 20% difference in firing for the TO versus OO manipulation (18 of the 42 neurons, or 43%, when considering TO vs. TT/OO). Thus, nearly half of the recorded neurons discriminated between the 2 rules with a 20% or more difference in firing in the window of interest, with an approximate 3:1 ratio for the act on versus the similar criterion.

DLPFC Neurons Do Not Encode Item Categories or Response Selection

Neuronal activity did not reflect the identities of individual items themselves as tools or objects. During the second image delay, there was no difference in population activity between trials in which the second image was an object versus a tool ($P = 0.83$, Mann–Whitney test).

In other words, the presentation of a tool versus an object in itself did not modulate neuronal responses. In accordance, only 1 of the 42 cells (2.4%) discriminated between tools and objects during the second image delay (Mann–Whitney at $P < 0.05$). We similarly found no difference in population response between tools versus objects during first image delay (Mann–Whitney test, $P = 0.83$), with 1 (2.4%) demonstrating differential activity during each of the first image presentation and the first delay (Mann–Whitney test at $P < 0.05$).

DLPFC activity did not encode response selections (i.e. match vs. nonmatch). Population activity did not discriminate between matched versus nonmatched pairs for any trial context: TO ($P = 0.66$, Mann–Whitney test), TT ($P = 0.54$, Mann–Whitney test), or OO ($P = 0.63$, Mann–Whitney test). There was also no difference in activity when comparing match versus nonmatch trials across all contexts ($P = 0.49$, Mann–Whitney test; Figs 3*d* and 4*a*). Similarly, only a small proportion (2 of 42 or 4.7%) of individual cells reflected matching status during the second image delay (Mann–Whitney at $P < 0.05$).

Additionally, during the second image delay, neurons were insensitive to the direction of forthcoming movement ($P = 0.27$, Mann–Whitney test), and in relation to response selection during target presentation ($P = 0.25$, Mann–Whitney test). Thus, while DLPFC neurons displayed selective responses based on the rule invoked in the trial, they carried little or no information about the outcome of rule evaluation.

Predictive Activity

We performed sequential feature selection analysis to model the extent to which variations in neuronal activity predicted individual features of the trials. Tested features included the abstract rule (act on vs. similar), categorical identities of the first and second items (tool vs. object), matching status (match vs. nonmatch), and direction of the upcoming motor response (left vs. right). Of these, neuronal responses across the population predicted only the rule ($\beta = 0.15$; $P = 0.0021$). We also performed a multilinear regression, which confirmed that rule type alone was correlated with normalized neural activity (coefficient = 0.19; 95% confidence interval = 0.08–0.30).

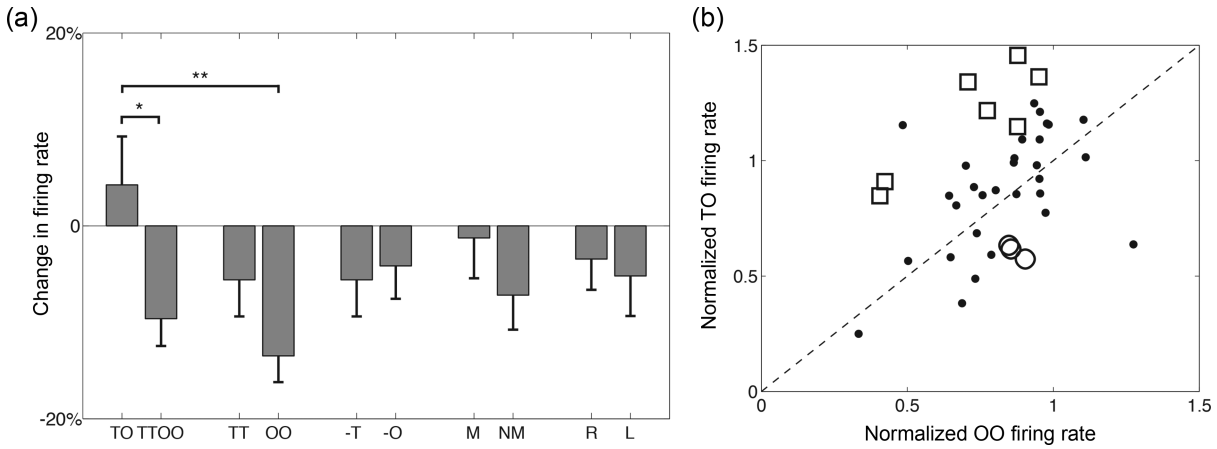

Figure 4. (*a*) Differences in population response ($n = 42$) across trial types during the second image delay. Normalized population activity ± SEM. TO, tool–object; OO, object–object; -T, second image is tool; -O, second image is object; M, match; NM, nonmatch; R, right; and L, left. Note: $^{*}P = 0.010$, and $^{**}P = 0.040$, Mann–Whitney test. (*b*) Distribution of individual neuronal responses. Normalized activity during the second image delay for TO trials is plotted against activity for OO trials; each point corresponds to a single neuron. Neurons with significant increases in the firing rate for TO over OO trials are highlighted with open squares (□), while those firing preferentially for OO over TO trials are represented with open circles (○).

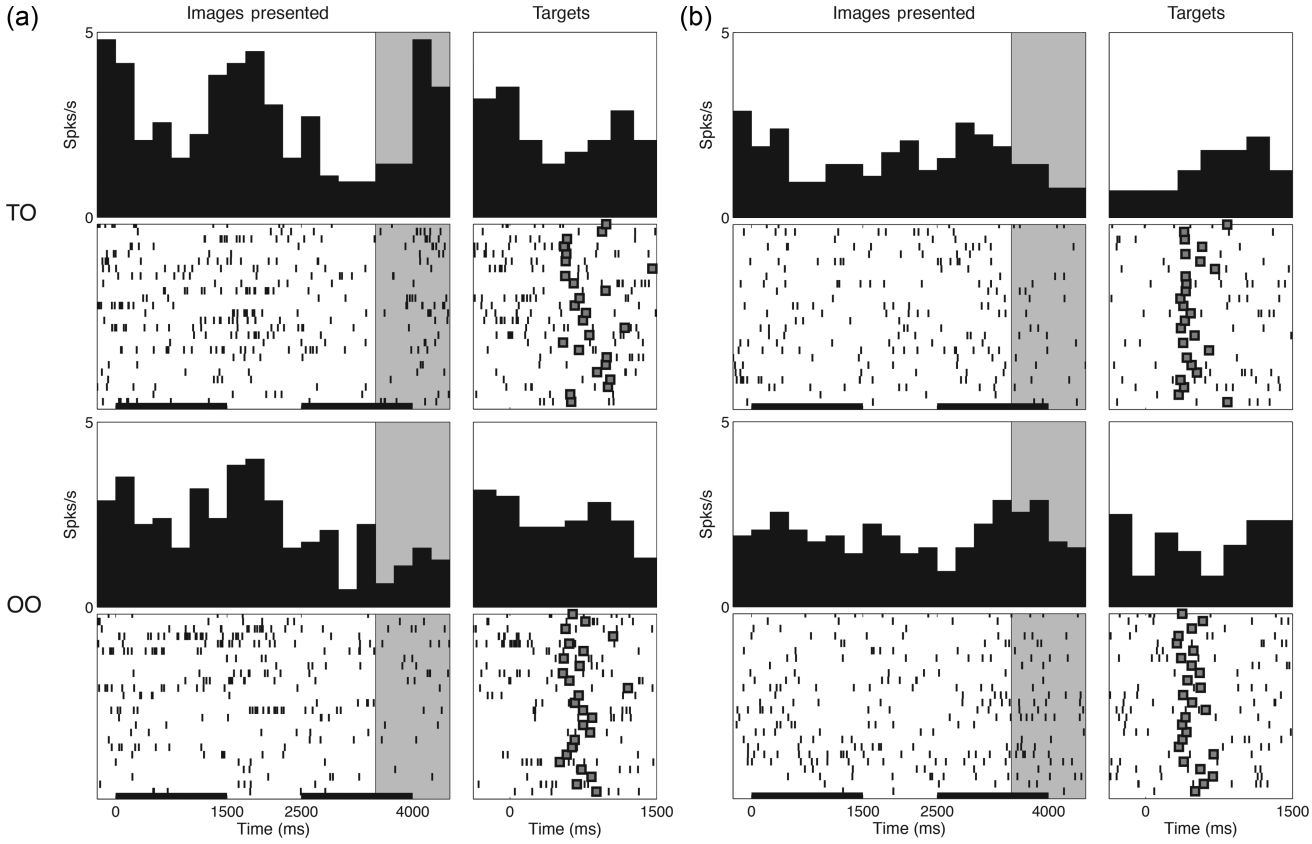

Figure 5. Peristimulus time histograms and raster plots for 2 DLPFC neurons that represented differences in abstract rules. Neural activity for TO (upper panels) and OO trials (lower panels) is shown for (*a*) a neuron firing preferentially for TO trials as well as (*b*) a neuron firing preferentially for OO trials. Within each panel, activity is aligned to the presentation of the task images (left) and appearance of the red–green targets (right). The black horizontal scale bars indicate the image presentation period (1500 ms each), and the filled boxes indicate the times of joystick movement (target selection). The shaded regions indicate the second image delay period (see text), during which these neuron demonstrated differential firing on TO versus OO trials. Note: (*a*) $P = 0.0020$, and (*b*) $P = 0.023$; Mann–Whitney test.

Task Difficulty and Switch Costs

Various metrics of task performance, including error rates and response times (RTs), can be used to assess whether task conditions differ in difficulty. Such metrics have been shown to vary in tasks that are optimized to tax cognitive resources (Bush and Shin 2006) and can modulate activity in the DLPFC (MacDonald et al. 2000). In the present study, we found no evidence of variations in task performance to

support differences in difficulty associated with the different rules.

The proportion of correct trials (main task and OT control, described below) was 94.8 ± 1.7% (mean ± SEM). RT, defined as the interval from target presentation to joystick movement, was 815 ± 261 msec (mean ± SD) across all trials. Subjects completed task trial types with similar proportions of errors (χ^2 test, $P = 0.40$), demonstrating no difference in performance based on which rule was applied. Further, there was no difference in RTs between rule types ($P = 0.31$, t-test). Nor was there any significant correlation between reaction time and neuronal activity across all trials ($P > 0.1$), or for any of the 3 primary trial types: TO, TT, and OO (each $P > 0.1$). These observations together suggest that there was no difference in task difficulty to account for the observed neuronal selectivity.

Prior studies using other tasks have found evidence of behavioral "switch costs," which sometimes correlate with changes in neuronal firing—particularly when trial types are of varying difficulty (Sheth et al. 2012). When we screened for similar features in this task, however, we did not find strong effects in either the behavioral or neuronal data. Specifically, we found no significant modulation of current trial reaction time by the type of the preceding trial. Reaction times did not vary for either similar ($P = 0.59$, t-test) or act on ($P = 0.17$, t-test) trials based on whether the preceding trial had been similar versus act on. Further, reaction times for trials in which rules had "switched" with respect to the preceding trial were no different from when they did not ($P = 0.57$, t-test). In the same vein, current trial reaction time was insensitive to whether the previous trial had been matching versus nonmatching.

From the neuronal perspective, we did not find that the type of trial transition was a significant determinant of neuronal activity. The population response did not modulate based on whether the current trial involved a switch in the active rule ($P = 0.74$, Mann–Whitney test). Five individual neurons (12%) did demonstrate a difference in firing, with 3 of the 5 increasing their activity for switch versus nonswitch trials. For similar trials, 3 neurons (7%) modulated their firing based on the rule invoked in the previous trial (2 increasing and 1 decreasing their activity with a previous similar trial). There were no neurons that modulated their activity for act on trials based on the rule invoked in the preceding trial.

Motor-Related Activity

The present findings were not attributable to potential differences in unconstrained movement by the subjects during the image or delay periods. While Brodmann area 9 of the DLPFC is not considered a motor area per se, many areas across the brain, including the DLPFC (Kubota and Funahashi 1982), can demonstrate differential responses based on the direction of an intended movement. As noted above, subjects could not anticipate the target locations or upcoming movement direction until after the targets were presented.

Nevertheless, we examined neuronal responses during the movement period itself (defined as a 1000 msec window centered on target selection). We found that only 6% of cells demonstrated tuning to the direction of joystick movement (right vs. left; t-test at $P < 0.05$,). In addition, there was no difference in population activity during the movement period between trials in which the right versus left target was selected (Mann–Whitney test, $P = 0.41$). The difference in population activity between TO and OO pairings was also no longer present at the time of movement (Mann–Whitney test, $P = 0.56$; Fig. 3*c*). Finally, when comparing neuronal activity during target selection, we found no neuronal modulation for the selection of the red versus green target ($P = 0.49$, Mann–Whitney test). There was also no difference in green versus red selections for either the TO ($P = 0.66$, Mann–Whitney test) or OO ($P = 0.63$, Mann–Whitney test) trial contexts.

Sequence of Presentation

We next considered whether neuronal modulation was associated with item novelty. Half of the OO trials (i.e. matching trials) presented 2 items that were similar in identity yet different in appearance (e.g. 2 nails of distinct shapes, sizes, and orientation), whereas TO trials always depicted 2 different items. To determine whether item dissimilarity elicited a novelty or "priming" effect (Grill-Spector et al. 2006), we compared activity from all TO pairs to activity from the subset of nonmatching OO pairs. We find that there remained highly significant modulation based on the abstract rule even when the identities of paired items were distinct ($P = 0.0072$, Mann–Whitney test).

Selective neuronal responses were also not dependent on the particular sequence in which items were displayed. The relationship between a tool and an object is not commutative; that is, a tool may be used to act on an object, but an object may not be used to act on a tool. Therefore, in the main task, tools preceded objects in TO trials, rather than vice versa. To examine whether the particular sequence in which tools and objects were presented contributed to the observed neuronal selectivity, we used a separate control task in 3 additional subjects (subjects 6, 7, and 8 in Tables 1 and 2).

In this task, the tested conditions included matching and nonmatching TO, OT, and OO pairings. Subjects were instructed that 2 items formed a match if 1) one could be act appropriately on the other or 2) if they were similar in identity, regardless of the order in which they were presented. The behavioral requirements of this control task were otherwise identical to those of the main task.

The population of 6 neurons recorded from in this control did not distinguish between TO and OT trials during the second image delay ($P > 0.1$, sequential feature selection). In comparison, neuronal responses for the same set of neurons were significantly predictive of the abstract rule invoked, regardless of which item was presented first ($P = 0.007$, sequential feature selection). These findings suggest that neurons in the DLPFC represent rules irrespective of the sequence in which the constituent items are presented.

Disease Process and Side of Recordings

Subject clinical disorder was not associated with differences in neural activity. Subjects participating in the study presented with a number of conditions (Table 1), of which the most common (3 of the 8 subjects; 27 of the 42 neurons) was ET. In general, tremor symptoms were unlikely to have systematically altered our results as the present task was not based on the subjects' ability to manipulate or handle physical items, but rather engaged their ability to evaluate pairings. Consistently, disease process was not found to be an independent predictor

of the neuronal responses ($P = 0.30$, ET vs. non-ET by sequential feature selection). Further, the population-averaged rule discrimination described above (i.e. TO vs. OO) remains robust (but not strictly significant) when these patients are excluded ($P = 0.06$; Mann–Whitney test).

Finally, it is possible that there may be a sidedness with regard to rule representation in the DLPFC. In total, 28 neurons were recorded from the left hemisphere, and 14 from the right. In aggregate, the left-sided neuron population distinguished between TO and OO trials during the second image delay, while the right-sided pool did not (though this difference may also be a function of statistical power).

When stratifying the population of recorded neurons by hemisphere, we observed neural modulation to rule in the left-sided neuron pool ($P = 0.03$, Mann–Whitney test; $n = 34$ neurons from 6 subjects), but not the right ($P = 0.45$, Mann–Whitney test; $n = 14$ neurons from 2 subjects). This was confirmed with sequential feature selection analysis, which demonstrated side of recordings as a significant factor ($P < 0.0001$). Finally, while we did identify rule-discriminating individual neurons in both hemispheres, more of these neurons were on the left (11 vs. 2). Taken together, these findings suggest that the left-sided DLPFC (evaluated between subjects) may be more critical to rule representation than its right hemisphere counterpart.

Discussion

In this study, we recruited subjects undergoing deep brain stimulator placement to perform an intraoperative task in which they viewed pairs of items and then used either of 2 abstract rules to indicate whether the items matched. From these subjects, we recorded spiking activity of individual neurons in Brodmann area 9 of the DLPFC. The recorded neurons—both individually and as a population—discriminated between the 2 abstract rules used to evaluate the item pairings. Rule sensitivity was prominent in a window spanning the presentation of the second item and the subsequent premotor response delay period. DLPFC neurons represented these rules selectively and did not carry significant information about the item categories, the match selection, or the upcoming motor response.

Given the proposed complexity by which we evaluate visual stimuli and select appropriate responses, it is not surprising that many parts of the brain have been implicated in the associated neural processing. Areas within the temporal lobe, for example, have been linked to the encoding of paired associations and to the categorization of individual objects (Sakai and Miyashita 1991; Kreiman et al. 2000; Sigala and Logothetis 2002; Hung et al. 2005). Additionally, areas such as the premotor cortex, parietal lobe, and the posterior cingulate have been associated with the appropriate selection and manipulation of items based on their use (Snyder et al. 1997; Turken and Swick 1999; Tanji and Hoshi 2000; Jacobs et al. 2010).

The DLPFC is among the most caudal of PFC areas, receiving convergent input from the ventral and dorsal visual streams (Rao et al. 1997; Fuster 2008). It receives dense afferent connections from the temporal lobe and inferior frontal cortex (Pandya and Yeterian 1990; Petrides and Pandya 1999) and projects widely to structures such as the premotor, anterior cingulate, and orbitofrontal areas (Bates and Goldman-Rakic 1993; Lu et al. 1994). This situates the DLPFC in a unique position to encode rules, enabling it to modulate downstream areas responsible for modifying and executing behavior while at the same time granting it access to preprocessed, bottom-up signals that convey information about changes in the external environment. Consistently, studies in nonhuman primates using match-to-sample tasks have revealed that neurons in the DLPFC (and other PFC subregions) can encode differences in rules (White and Wise 1999; Wallis et al. 2001).

There are several human neuroimaging studies that have investigated rule instantiation in the DLPFC with differing results. Prior studies have suggested that the DLPFC plays a role in rule acquisition (Crescentini et al. 2011), in maintaining the representation of a freely selected rule (Bengtsson et al. 2009), and also in rule-encoding—but only when the forthcoming responses cannot be anticipated (as in our task; Nee and Brown 2012). Other reports, in comparison, have emphasized the rule sensitivity of the ventral PFC and other cortical regions, rather than the DLPFC (Bunge et al. 2003; Sakai and Passingham 2003).

The task in our study is distinct from previous paradigms in that rule selection was not guided by either an external pretrial cue or subsequent feedback; rather, the appropriate rule was inferred. Here, we demonstrate that human DLPFC neurons can indeed represent differences in the abstract rules chosen across trials. DLPFC neurons regarded these rules as the most salient task feature, carrying little information about the identities of the items that constituted the task pairs or the rule-based response. Such rule selectivity is compatible with prior findings, demonstrating that PFC neurons are capable of representing behaviorally relevant information at the exclusion of other task features (Rainer et al. 1998; Asaad et al. 2000).

In this study, we sought to mimic scenarios faced in natural human behavior. Accordingly, our task was designed without explicit external cues to instruct rule selection (such as a tone or visual image at the start of the trial; Wallis et al. 2001; Bunge et al. 2003). Rather, the task relied on the presented pairings alone to guide selection of the appropriate evaluative criterion (e.g. hammer–nail → act on rule). While DLPFC neurons were highly selective to the employed rules, we found no differential response to the categorical identities of the items used to make these pairings. We also found no differences in behavioral reaction times or error rates to suggest that the subjects were responding differently to the rule types being used.

The timing of DLPFC discriminatory activity observed here further argues that subjects were not prospectively encoding or anticipating the rule to be invoked prior the second task image. If this had been the case, we would have expected rule differentiation during the first image delay (or perhaps even the first image). Further, the latency of rule discrimination during the second image period was relatively long when compared with similar animal studies (more than 1 s vs. 1–2 hundred milliseconds; Miller and Desimone 1994; Rainer et al. 1998; Asaad et al. 2000; Wallis et al. 2001).

In summary, we observed differential DLPFC activity based on the rule used to evaluate pairings of everyday items. By representing such abstract rules, DLPFC neurons may function to narrow the pool of candidate responses to those that are most behaviorally relevant. Such an approach would be more

efficient than encoding all possible sensory-motor selections concurrently. By biasing downstream targets in this way, the DLPFC may provide a critical processing step, facilitating decisions under dynamic circumstances.

Funding

This work was supported by National Institutes of Health (grant number 5R01-HD059852), PECASE, and the Whitehall Foundation (to Z.M.W.); by National Institutes of Health (grant number R25-NS065743 (to S.A.S.); grant numbers 1R01-EY017658, 1R01-NS63249, 5P5MH086400), the Esther A. & Joseph Klingenstein Fund (to E.N.E.), and the Sackler Scholar Programme in Psychobiology (to S.R.P.).

Notes

The authors thank Patrick Schweder for assistance in electrode localization and Michael Campos, Mouhsin Shafi, and Keren Haroush for their insightful discussion. *Conflict of Interest*: None declared.

References

Amirnovin R, Williams ZM, Cosgrove GR, Eskandar EN. 2006. Experience with microelectrode guided subthalamic nucleus deep brain stimulation. Neurosurgery. 58(1 Suppl.):ONS96–102; discussion ONS96–102.

Asaad WF, Eskandar EN. 2008b. Achieving behavioral control with millisecond resolution in a high-level programming environment. J Neurosci Methods. 173:235–240.

Asaad WF, Eskandar EN. 2008a. A flexible software tool for temporally-precise behavioral control in Matlab. J Neurosci Methods. 174:245–258.

Asaad WF, Rainer G, Miller EK. 2000. Task-specific neural activity in the primate prefrontal cortex. J Neurophysiol. 84:451–459.

Bates JF, Goldman-Rakic PS. 1993. Prefrontal connections of medial motor areas in the rhesus monkey. J Comp Neurol. 336:211–228.

Bengtsson SL, Haynes J-D, Sakai K, Buckley MJ, Passingham RE. 2009. The representation of abstract task rules in the human prefrontal cortex. Cereb Cortex. 19:1929–1936.

Buckley MJ, Mansouri FA, Hoda H, Mahboubi M, Browning PGF, Kwok SC, Phillips A, Tanaka K. 2009. Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. Science. 325:52–58.

Bunge SA, Kahn I, Wallis JD, Miller EK, Wagner AD. 2003. Neural circuits subserving the retrieval and maintenance of abstract rules. J Neurophysiol. 90:3419–3428.

Bush G, Shin LM. 2006. The Multi-Source Interference Task: an fMRI task that reliably activates the cingulo-frontal-parietal cognitive/attention network. Nat Protoc. 1:308–313.

Bussey TJ, Wise SP, Murray EA. 2001. The role of ventral and orbital prefrontal cortex in conditional visuomotor learning and strategy use in rhesus monkeys (Macaca mulatta). Behav Neurosci. 115:971–982.

Comalli PE, Wapner S, Werner H. 1962. Interference effects of Stroop color-word test in childhood, adulthood, and aging. J Genet Psychol. 100:47–53.

Crescentini C, Seyed-Allaei S, De Pisapia N, Jovicich J, Amati D, Shallice T. 2011. Mechanisms of rule acquisition and rule following in inductive reasoning. J Neurosci. 31:7763–7774.

Fuster J. 2008. The prefrontal cortex. London: Elsevier, p. 410

Fuster JM, Alexander GE. 1971. Neuron activity related to short-term memory. Science. 173:652–654.

Gaffan D, Harrison S. 1989. A comparison of the effects of fornix transection and sulcus principalis ablation upon spatial learning by monkeys. Behav Brain Res. 31:207–220.

Gershberg FB, Shimamura AP. 1995. Impaired use of organizational strategies in free recall following frontal lobe damage. Neuropsychologia. 33:1305–1333.

Grill-Spector K, Henson R, Martin A. 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci. 10:14–23.

Gross RE, Krack P, Rodriguez-Oroz MC, Rezai AR, Benabid A-L. 2006. Electrophysiological mapping for the implantation of deep brain stimulators for Parkinson's disease and tremor. Mov Disord. 21 (Suppl 14):S259–S283.

Hung CP, Kreiman G, Poggio T, DiCarlo JJ. 2005. Fast readout of object identity from macaque inferior temporal cortex. Science. 310:863–866.

Jacobs S, Danielmeier C, Frey SH. 2010. Human anterior intraparietal and ventral premotor cortices support representations of grasping with the hand or a novel tool. J Cogn Neurosci. 22:2594–2608.

Jenkinson M, Bannister P, Brady M, Smith S. 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage. 17:825–841.

Jenkinson M, Smith S. 2001. A global optimisation method for robust affine registration of brain images. Med Image Anal. 5:143–156.

Kreiman G, Koch C, Fried I. 2000. Category-specific visual responses of single neurons in the human medial temporal lobe. Nat Neurosci. 3:946–953.

Kubota K, Funahashi S. 1982. Direction-specific activities of dorsolateral prefrontal and motor cortex pyramidal tract neurons during visual tracking. J Neurophysiol. 47:362–376.

Lu MT, Preston JB, Strick PL. 1994. Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. J Comp Neurol. 341:375–392.

Luria A. 1966. Higher cortical functions in man. New York: Basic Books.

MacDonald AW, Cohen JD, Stenger VA, Carter CS. 2000. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science. 288:1835–1838.

Mansouri FA, Buckley MJ, Tanaka K. 2007. Mnemonic function of the dorsolateral prefrontal cortex in conflict-induced behavioral adjustment. Science. 318:987–990.

Miller EK, Desimone R. 1994. Parallel neuronal mechanisms for short-term memory. Science. 263:520–522.

Miller EK, Erickson CA, Desimone R. 1996. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J Neurosci. 16:5154–5167.

Milner B. 1963. Effects of different brain lesions on card sorting: the role of the frontal lobes. Arch Neurol. 9:100–110.

Moody SL, Wise SP, di Pellegrino G, Zipser D. 1998. A model that accounts for activity in primate frontal cortex during a delayed matching-to-sample task. J Neurosci. 18:399–410.

Murray EA, Bussey TJ, Wise SP. 2000. Role of prefrontal cortex in a network for arbitrary visuomotor mapping. Exp Brain Res. 133:114–129.

Nee DE, Brown JW. 2012. Rostral-caudal gradients of abstraction revealed by multi-variate pattern analysis of working memory. Neuroimage. 63:1285–1294.

Pandya DN, Yeterian EH. 1990. Prefrontal cortex in relation to other cortical areas in rhesus monkey: architecture and connections. Prog Brain Res. 85:63–94.

Parker A, Gaffan D. 1998. Memory after frontal/temporal disconnection in monkeys: conditional and non-conditional tasks, unilateral and bilateral frontal lesions. Neuropsychologia. 36:259–271.

Passingham RE. 1993. The frontal lobes and voluntary action. USA: Oxford University Press.

Penn DC, Povinelli DJ. 2007. Causal cognition in human and nonhuman animals: a comparative, critical review. Annu Rev Psychol. 58:97–118.

Petrides M. 1982. Motor conditional associative-learning after selective prefrontal lesions in the monkey. Behav Brain Res. 5:407–413.

Petrides M, Pandya DN. 1999. Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. Eur J Neurosci. 11:1011–1036.

Rainer G, Asaad WF, Miller EK. 1998. Memory fields of neurons in the primate prefrontal cortex. Proc Natl Acad Sci U S A. 95:15008–15013.

Rao SC, Rainer G, Miller EK. 1997. Integration of what and where in the primate prefrontal cortex. Science. 276:821–824.
Richard Draper N, Smith H. 1998. Applied regression analysis. Vol. 1. New York: Wiley, p. 706.
Sakai K, Miyashita Y. 1991. Neural organization for the long-term memory of paired associates. Nature. 354:152–155.
Sakai K, Passingham RE. 2003. Prefrontal interactions reflect future task operations. Nat Neurosci. 6:75–81.
Shallice T, Burgess PW. 1991. Deficits in strategy application following frontal lobe damage in man. Brain. 114(Pt 2): 727–741.
Sheth SA, Mian MK, Patel SR, Asaad WF, Williams ZM, Dougherty DD, Bush G, Eskandar EN. 2012. Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. Nature. 488:218–221.
Sigala N, Logothetis NK. 2002. Visual categorization shapes feature selectivity in the primate temporal cortex. Nature. 415:318–320.
Snyder LH, Batista AP, Andersen RA. 1997. Coding of intention in the posterior parietal cortex. Nature. 386:167–170.
Tanji J, Hoshi E. 2000. Integration of target and body-part information in the premotor cortex when planning action. Nature. 408:466–470.
Turken AU, Swick D. 1999. Response selection in the human anterior cingulate cortex. Nat Neurosci. 2:920–924.
Wallis JD, Anderson KC, Miller EK. 2001. Single neurons in prefrontal cortex encode abstract rules. Nature. 411:953–956.
Watanabe M. 1990. Prefrontal unit activity during associative learning in the monkey. Exp Brain Res. 80:296–309.
White IM, Wise SP. 1999. Rule-dependent neuronal activity in the prefrontal cortex. Exp Brain Res. 126:315–335.
Williams ZM, Bush G, Rauch SL, Cosgrove GR, Eskandar EN. 2004. Human anterior cingulate neurons and the integration of monetary reward with motor responses. Nat Neurosci. 7:1370–1375.
Wise SP, Murray EA, Gerfen CR. 1996. The frontal cortex-basal ganglia system in primates. Crit Rev Neurobiol. 10:317–356.