Neuron
Article

Neural Correlates of Task Switching in Prefrontal Cortex and Primary Auditory Cortex in a Novel Stimulus Selection Task for Rodents

Chris C. Rodgers[1,3,*] and Michael R. DeWeese[1,2,3]
[1]Helen Wills Neuroscience Institute, University of California, Berkeley, 132 Barker Hall, Berkeley, CA 94720, USA
[2]Department of Physics, University of California, Berkeley, 132 Barker Hall, Berkeley, CA 94720, USA
[3]Redwood Center for Theoretical Neuroscience, University of California, Berkeley, 132 Barker Hall, Berkeley, CA 94720, USA
*Correspondence: xrodgers@gmail.com
http://dx.doi.org/10.1016/j.neuron.2014.04.031

SUMMARY

Animals can selectively respond to a target sound despite simultaneous distractors, just as humans can respond to one voice at a crowded cocktail party. To investigate the underlying neural mechanisms, we recorded single-unit activity in primary auditory cortex (A1) and medial prefrontal cortex (mPFC) of rats selectively responding to a target sound from a mixture. We found that prestimulus activity in mPFC encoded the selection rule—which sound from the mixture the rat should select. Moreover, electrically disrupting mPFC significantly impaired performance. Surprisingly, prestimulus activity in A1 also encoded selection rule, a cognitive variable typically considered the domain of prefrontal regions. Prestimulus changes correlated with stimulus-evoked changes, but stimulus tuning was not strongly affected. We suggest a model in which anticipatory activation of a specific network of neurons underlies the selection of a sound from a mixture, giving rise to robust and widespread rule encoding in both brain regions.

INTRODUCTION

Humans can select and respond to one person's voice while many others are speaking at the same time. We do this effortlessly, yet no known algorithm can solve this "cocktail party problem" in realistic settings, perhaps because we do not fully understand the relevant computations performed in the brain (Cherry, 1953; Sayers and Cherry, 1957; Ding and Simon, 2012; McDermott, 2009). Other social animals such as birds and rodents demonstrate a similar ability (Bee and Micheyl, 2008); for instance, mother mice respond to distinct pup calls when several are calling at once (Geissler and Ehret, 2002). Humans use selective attention, the cognitive process of selecting and responding to a single target stimulus among simultaneous distractors (Desimone and Duncan, 1995), to solve the cocktail party problem (Ahveninen et al., 2011). Experiments in visual selective attention have revealed that the prefrontal cortex (PFC) sends top-down "bias signals" to sensory cortex (Miller and Cohen, 2001; Moore et al., 2003) to select the target stimulus, enhancing its neural representation while suppressing the representation of distractors. Similar mechanisms may be at work in the auditory cortex: electrocorticographic (Mesgarani and Chang, 2012; Zion Golumbic et al., 2013) and magnetoencephalographic (Ding and Simon, 2012) recordings show that brain activity is dominated by the attended voice. Ultimately, recordings from single units (individual neurons) will be needed to understand the circuit. In addition, many models of visual selection are not obviously applicable to the auditory modality—for instance, the idea that visual attention co-opts the neural mechanisms for shifting gaze (Moore et al., 2003). Establishing an animal model of auditory selective attention could shed more light on whether the known mechanisms of visual selection are universal or specific to one modality.

Nonhuman primates are the traditional model organism for studying complex cognition (Gold and Shadlen, 2007) but rodents are also capable of sophisticated decision-making (Raposo et al., 2012; Brunton et al., 2013; Zariwala et al., 2013). The behavioral flexibility of rodents is thought to be mediated by the PFC (Karlsson et al., 2012; Kvitsiani et al., 2013), even though this region is not necessary for simple sensory discriminations (Pai et al., 2011). The medial PFC (mPFC) in particular is critical for task switching (Birrell and Brown, 2000; Floresco et al., 2008; Durstewitz et al., 2010; Ragozzino et al., 1999), such as switching the navigational strategy used to solve a maze (Rich and Shapiro, 2009). Rodent mPFC thus appears to maintain a representation of the current task rule, analogous to the rule-encoding neurons observed in primate PFC (Wallis et al., 2001; Asaad et al., 2000; Johnston et al., 2007), although large parts of the monkey PFC appear to be functionally and anatomically unique to primates (Wise, 2008).

Frontal areas have been shown to play a role in directing flexible auditory processing in the primary auditory cortex (A1). For example, training ferrets to detect tones at a specific target frequency (Fritz et al., 2003) produces rapid tuning changes in A1 (i.e., the neurons responded more to the target frequency) and modulates functional connectivity between A1 and frontal areas (Fritz et al., 2010). Moreover, when these ferrets switch between different auditory tasks, the observed tuning changes match the demands of each task (Fritz et al., 2005). These experiments

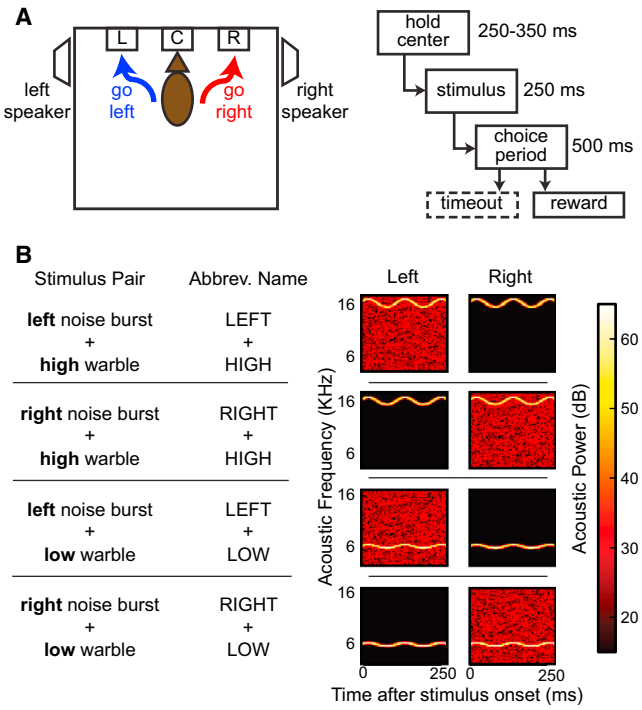

Figure 1. Behavioral Paradigm

(A) Left: a schematic of the behavioral arena with left (L), center (C), and right (R) ports (or nose-pokes), and left and right speakers. Right: timeline of each trial. The rat initiates a trial by nose-poking the center port as shown. After a hold period, an auditory stimulus is played. Following this, the rat may choose to go to the left port (blue arrow), go to the right port (red arrow), or do neither of those (a "nogo" response).

(B) Task stimuli (left, description; right, spectrogram of the auditory waveform). On each trial, the rat hears one of four possible auditory stimulus pairs: LEFT+HIGH, RIGHT+HIGH, LEFT+LOW, or RIGHT+LOW. Each is a simultaneous combination of a broadband noise burst played from either the left or right speaker, and a low-pitched or high-pitched warble played with equal intensity from both speakers.

(C) Task rules. The session consists of alternating localization and pitch discrimination blocks. In localization blocks, the rat must go left for sounds containing LEFT and it must nogo for sounds containing RIGHT; the low- or high-pitched warble is an irrelevant distractor. In pitch discrimination blocks, the rat must go right if the stimulus pair contains LOW and it must nogo if the stimulus pair contains HIGH; the noise burst is an irrelevant distractor.

revealed A1 to be surprisingly dynamic; however, subjects were not required to select a target sound in the presence of simultaneous distractors, a critical aspect of the cocktail party problem (McDermott, 2009; Ding and Simon, 2012).

In this study, we have taken advantage of the relative ease and speed with which rodents can be trained on demanding tasks (Carandini and Churchland, 2013) to develop a novel behavioral task in which rats hear two simultaneous sounds, but select and respond only to one. This requires cognitive flexibility because, depending on which sound the experimenter instructs the subject to select, the same pair of sounds can elicit a different behavioral response ("same stimulus; different response"). The subjects must alternate which sound they select multiple times within each session. We are aware of no purely auditory single-unit studies in any animal with these properties. The analogous ability in vision—to respond to a behaviorally relevant stimulus in the presence of competing distractors—has been referred to as stimulus "selection" (Knudsen, 2007; Reynolds and Chelazzi, 2004; Pestilli et al., 2011); following this, we refer to our task as auditory stimulus selection.

Similar visual and cross-modal tasks have been termed set shifting (Stoet and Snyder, 2004), task switching (Sasaki and Uka, 2009), and selective attention (Moran and Desimone, 1985; Hocherman et al., 1976; Otazu et al., 2009). Other studies have investigated "response selection": how decisions are translated into appropriate motor actions, following stimulus selection or even in the absence of an explicit stimulus (Young and Shapiro, 2011; Turken and Swick, 1999). We also note a similarity between our task and the Wisconsin Card Sorting Task for diagnosing disorders of executive function (Monchi et al., 2001).

We recorded from individual mPFC and A1 neurons in rats performing our task. We found that the prestimulus, anticipatory activity of our recorded neurons in mPFC encoded the selection rule—which sound the subject should select. Surprisingly, we also found this prestimulus effect in a sizable fraction of the neurons we recorded in A1. Disruption of mPFC through electrical microstimulation significantly impaired task performance. Finally, although changes in prestimulus baseline correlated with changes in stimulus-evoked activity in both brain regions, this did not appear to alter tuning properties in a way that would be obviously beneficial for responding to the selected sound.

RESULTS

A Novel Behavioral Task for Rodents: Auditory Stimulus Selection

We developed an auditory stimulus selection task for rats, in which the subject was trained to respond to either of two simultaneously presented sounds. The rat initiated each trial (Figure 1A) by holding its nose in the center port of a three-port behavior box—the "hold period." This triggered speakers on the left and right to play one of the following four equally likely stimulus pairs: LEFT+HIGH, RIGHT+HIGH, LEFT+LOW, or RIGHT+LOW (Figure 1B). Each stimulus pair was a simultaneous combination of (1) a broadband noise burst, played from either the LEFT or RIGHT speaker; and (2) either a HIGH- or LOW-pitched warble (frequency-modulated tone), played from both speakers simultaneously. After the onset of stimulus presentation, the rat could then choose to "go left" (poke its nose in the left port), "go right" (poke its nose in the right port), or "nogo" (not poke either side). Correct pokes into the side ports were rewarded with water; incorrect pokes were penalized with a 2–6 s timeout (see Supplemental Experimental Procedures available online).

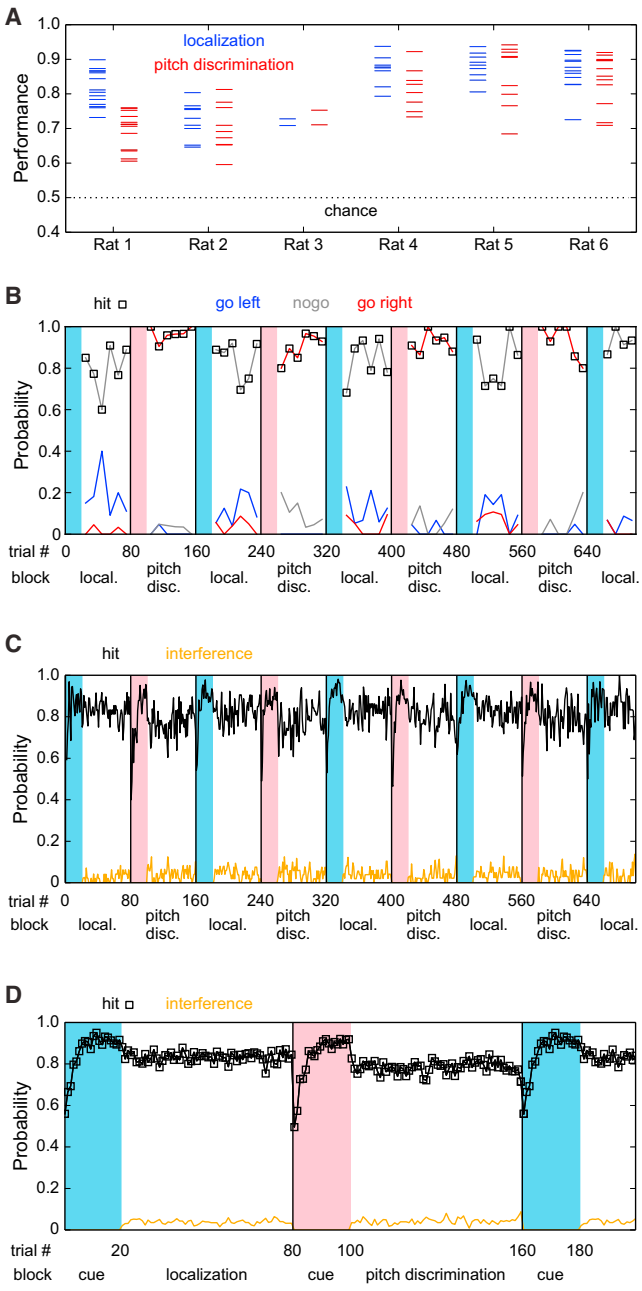

Figure 2. Trained Rats Select and Respond to the Target Sound, Not the Distractor

(A) Behavior performance during recording sessions. Each hash mark is the performance during localization (blue) or pitch discrimination (red) in a single recording session. Performance is well above chance (black dotted line; see Supplemental Experimental Procedures).

(B) Distribution of behavioral responses to an example stimulus pair (RIGHT+LOW) over the course of an average session. We averaged all sessions from a single rat (rat 5) and binned the trials into groups of ten. The x-axis shows both trial number and block type. The correct response to this stimulus pair is to go right during pitch discrimination and to nogo during localization. Each trace shows the probability that the rat will go right (red), nogo (gray), or go left (blue); black open squares mark the correct response for that block. The rat responds correctly most of the time, even though the required action changes abruptly at the block boundaries. This stimulus pair does not occur during cue trials, which begin each block and are shaded in cyan and pink throughout this figure.

(C) Similar to (B), but averaged over all sessions, rats, and stimuli. Most trials are correct (black trace). Interference trials (orange; see text) are rare.

(D) Performance briefly dips during cue trials at the beginning of a block but recovers within a few trials. All localization blocks from (C) are averaged together, as are all pitch discrimination blocks. To emphasize block transitions, the x axis repeats itself after trial 160; the cyan shaded areas are identical because the block structure is cyclical.

On each trial, one of the sounds in the stimulus pair (the "target") indicated the correct response; the other sound (the "distractor") was uninformative. The behavioral session alternated between "localization" blocks of trials, during which the noise burst was the target, and "pitch discrimination" blocks, during which the warble was the target (Figure 1C). Each block consisted of 80 trials, the first 20 trials of which were reserved to indicate the block change. During these 20 "cue trials," the rat heard only target sounds without any distractor.

The entire training process required approximately 10 weeks. Trained rats performed many trials per session (median, 698). We verified that the rats were performing significantly above 50% in both blocks, which meant that their behavioral response was driven by the target sound, rather than by the distractor or by a combination of target and distractor (Figures S1B, S1C, and Supplemental Experimental Procedures). Our best rats' typical performance during recording sessions was approximately 85% in both blocks (Figure 2). After each block change, rats rapidly and correctly switched to selecting the new target sound. Performance was typically better on go trials than on nogo trials (Figure S1A).

Anticipatory Neuronal Activity in mPFC and A1 Encodes the Selection Rule

We next asked whether neuronal activity changed according to which target the rats selected. We implanted tetrodes into the brain, targeting A1 and/or the prelimbic region of mPFC, and recorded single-unit action potentials (spikes) during behavior. By analogy with the rule-encoding neurons in the primate PFC, we hypothesized that the firing rates of single mPFC neurons would differ significantly between localization and pitch discrimination blocks. We first confined our analysis to the hold period, the interval before stimulus onset when the rat is holding its nose in the center port and presumably preparing to select the target sound from the imminent stimulus pair.

We found that the hold period activity of a majority of mPFC neurons robustly encoded the selection rule on "correct trials," those trials on which the rat gave the correct response. An example unit (Figure 3A) fired significantly more in the hold period during localization trials than it did during pitch discrimination trials. A different but simultaneously recorded single unit in mPFC (Figure 3B) fired significantly more during pitch discrimination than during localization. In both cases, the effect persisted across the entire session of over 1,300 trials, alternating with each block just as the behavior did. Across our recorded population of mPFC neurons, 63% (76/121) of the neurons individually and significantly encoded the selection rule during the hold period (Figure 3C). Of these, 36 neurons preferred (i.e., fired

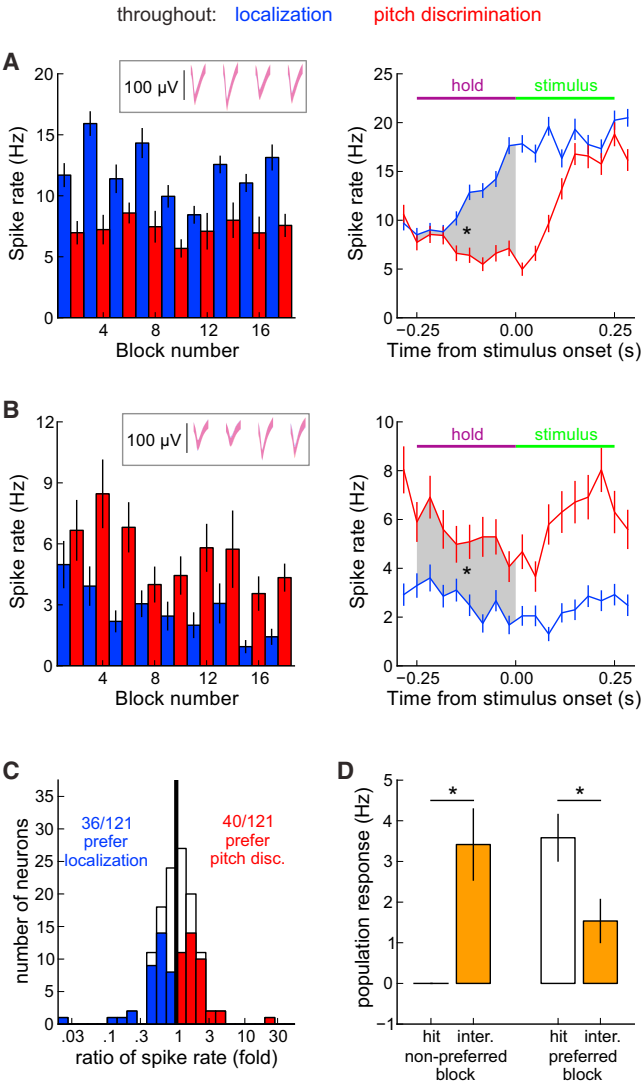

Figure 3. Prestimulus Activity in mPFC Encodes the Selection Rule

(A) Left: An example mPFC single unit that fires more during the hold period for localization (blue bars throughout this figure) than for pitch discrimination (red bars). For all figures, error bars represent SEM unless otherwise noted. Inset: Extracellular waveforms (mean ± SD) with duration of 0.8 ms on each channel of the tetrode. The waveforms are colored red and blue based on the block in which they were recorded, but are almost entirely overlapping (purple). Right: peristimulus time histogram (PSTH) of the same unit, averaged over all correct trials from each block. During the hold period (gray), the firing rate is significantly ($p < 0.001$, unpaired Mann-Whitney U-test) higher on localization (mean 12.1 Hz, n = 483) than on pitch discrimination (mean 7.2 Hz, n = 295) trials.

(B) Another example mPFC single unit, this one preferring pitch discrimination. The hold period firing rate is significantly ($p < 0.001$) higher on pitch discrimination (mean 5.4 Hz) than on localization (mean 2.7 Hz) trials. Trial counts are the same as the simultaneously recorded unit in (A). This neuron's firing rate is persistently elevated at all points plotted.

(C) Stacked histogram of the ratio of hold period firing rate (pitch discrimination over localization) for all mPFC neurons. Red and blue bars represent significantly modulated neurons. We used an unpaired Mann-Whitney U-test for all neurons and controlled for multiple comparisons using the Benjamini-Hochberg false discovery rate.

(D) Rule encoding is diminished on interference trials. We averaged together the firing rates of each rule-encoding neuron during either correct (white) or interference (orange) trials in their preferred and nonpreferred blocks. Firing rates are normalized by subtracting the firing rate on correct trials in the nonpreferred block. The population response on interference trials is significantly increased during the nonpreferred block and decreased during the preferred block (n = 57 neurons, paired Mann-Whitney test).

more during) localization and 40 preferred pitch discrimination; neither preference was significantly more common (binomial test, $p > 0.05$).

Surprisingly, we also found a similar effect in A1 (Figure 4). Although encoding of selection rule was our hypothesized result in mPFC, this was unexpected in A1, especially given the absence of auditory stimulation during the hold period. Across our recorded population, 36% (36/99) of A1 neurons encoded selection rule. As with mPFC, neither population was significantly larger (13 preferring localization, 23 preferring pitch discrimination; binomial test, $p > 0.05$). Because A1 encodes sounds sparsely (DeWeese et al., 2003; Hromádka et al., 2008; Carlson et al., 2012), we were not surprised to observe that only some (49/99) of our recorded neurons in A1 significantly responded to our task stimuli (Figures S5A–S5E). Rule encoding was not significantly more widespread in either stimulus-responsive (14/49) or nonresponsive (22/50) neurons ($p > 0.05$, Fisher's exact test). This finding is reminiscent of evidence from human imaging that attention affects strongly stimulus-driven regions of auditory cortex less than it affects other, more poorly tuned regions (Petkov et al., 2004).

These effects were strong: among rule-encoding neurons, the median increase in firing rate during the preferred block was 74.7% in mPFC and 99.7% in A1. These results are unlikely to be due to firing rate drift over the course of the session or spike sorting errors arising from small differences in spike waveform shape between blocks (see Supplemental Experimental Procedures). In sum, these results demonstrate widespread and robust encoding of selection rule in the prestimulus activity of both mPFC and A1 neurons.

Error Trial Analysis

In the previous section, we analyzed only correct trials. We next considered "interference" trials, during which the rat appeared to select the wrong sound. On such trials, the rat heard a "go" distractor (i.e., a sound that should have elicited a go response, had it been presented during the other block) and incorrectly went to the choice port associated with that distractor instead of doing what the target sound indicated. If anticipatory encoding of selection rule is important for successful stimulus selection, then this encoding should have been weaker or even reversed when the rat selected the wrong sound.

Indeed, in mPFC, the encoding of selection rule was significantly weakened on interference trials, as compared with correct trials (Figure 3D). In A1, we observed a more extreme effect (Figure 4D): the rule encoding was actually reversed on interference trials (i.e., firing rates were greater during the nonpreferred block on such trials). These observations are consistent with the idea that anticipatory activity predicted which sound the rat would select, even for trials on which the rat appeared to respond to the distractor by going to the wrong choice port.

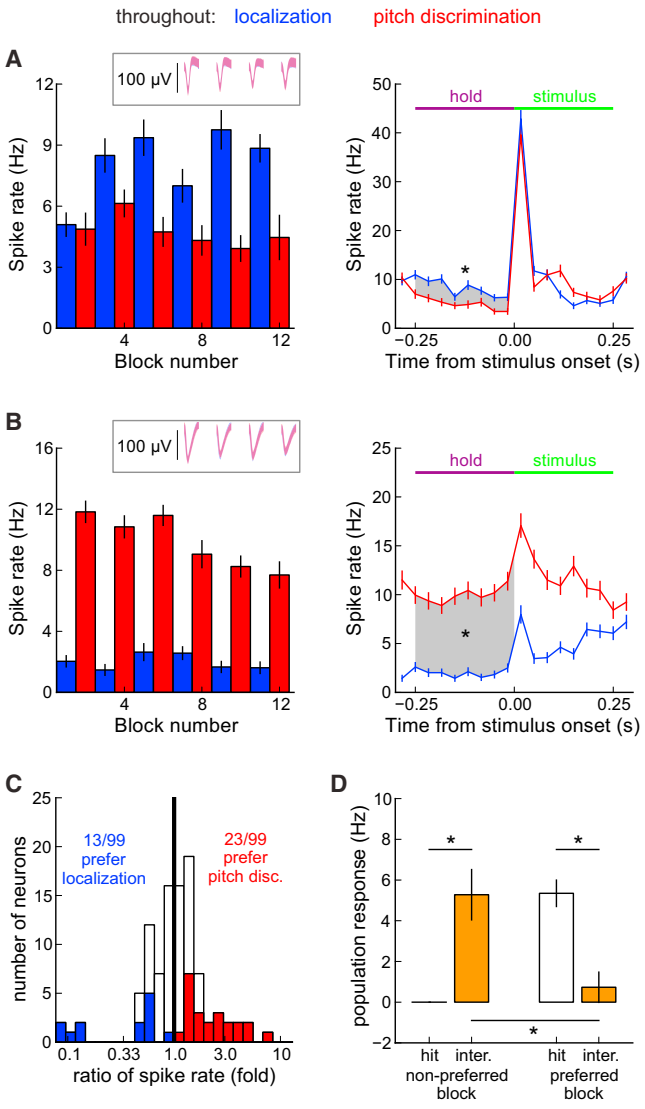

Figure 4. Prestimulus Activity in A1 also Encodes the Selection Rule

(A) An example A1 neuron that responds significantly more ($p < 0.001$) during localization (8.0 Hz, n = 312 trials) than during pitch discrimination (4.8 Hz, n = 253). Throughout, conventions and statistical procedures are as in Figure 3. Note the peak following stimulus onset, which was used to analyze the evoked response (Figure 6).

(B) Simultaneously recorded A1 neuron that significantly ($p < 0.001$) prefers pitch discrimination (10.1 Hz, n = 312 trials) over localization (2.0 Hz, n = 253).

(C) Stacked histogram of the ratio of hold period firing rate (pitch discrimination over localization) for all A1 neurons.

(D) Rule encoding during the hold period is inverted on interference trials. The population response on interference trials (orange bars) is significantly greater during the nonpreferred block than during the preferred block ($p < 0.01$, n = 16 neurons, paired Mann-Whitney U-test). In contrast, on correct trials (white bars) the firing rate is higher during the preferred block than the nonpreferred block, by definition.

Potential Role of Posture

The mPFC regulates cognitive state, but it also encodes body position and plays a role in motor planning (Erlich et al., 2011; Euston and McNaughton, 2006). We analyzed video of the rats and found evidence of preparatory head positioning that differed between blocks (see Supplemental Experimental Procedures), presumably a behavioral strategy that the rat used to prepare for the differing motor actions required in each block (i.e., go left in localization and go right in pitch discrimination).

This raised the question of whether neurons were encoding this postural difference, rather than selection rule. We found that, in the vast majority of rule-encoding neurons in both brain regions, the selection rule explained more of the variance in firing rates than did head angle (Figures S2I–S2T and S3I–S3Q). In addition, the rule encoding was largely maintained on a subset of "posture-equalized" trials, selected so that the mean head angle was the same in each block (Figures S2M–S2O and S3M–S3O). Finally, as we discuss below, the long duration of the neural effects we observed further argues against the possibility that changes in posture were the underlying cause.

Within-Trial Timescale of the Encoding of the Selection Rule

We next asked how soon before the stimulus the rule encoding emerged, and for how long afterward it persisted. For each rule-encoding neuron, we compared across blocks the smoothed firing rates in every 50 ms bin up to 3 s before and after stimulus onset. We thereby determined the largest interval around the hold period during which the neural activity significantly encoded the selection rule. Across the data set, the median intertrial interval was 4.0 s (interquartile range, 2.7–5.3 s) and so this range (± 3 s) overlapped with the previous and/or next trial in many cases.

The temporal dynamics of the encoding varied widely across neurons in both regions (Figures 5A and 5B). For some neurons, rule encoding was strictly confined to the hold period. Other neurons showed significant encoding during all time bins tested: their firing rates were persistently elevated during the preferred block. We found neurons spanning this range of timescales in both brain areas. In A1, the median rule-encoding unit first developed a significant block preference 0.55 s prestimulus; in PFC, the median was 0.625 s prestimulus. Thus, the majority of rule-encoding neurons developed this property well before the rat initiated a trial by center-poking.

To examine the population dynamics of rule encoding, we averaged the normalized activity of all rule-encoding neurons during their preferred block. On average, the population activity ramped up gradually before stimulus onset, over a timescale of several seconds, and then fell relatively quickly afterward (Figure 5C; Figures S4B–S4E). The population activity in mPFC was significantly elevated earlier than in A1, consistent with its hypothesized role as the origin of top-down bias signals to sensory cortex (Miller and Cohen, 2001). However, we note that the wide range of timescales within both regions, and the fact that only a small fraction of our data set consists of simultaneous recordings from A1 and PFC, complicates a direct comparison between brain regions.

Finally, we asked whether the rule encoding reflected an increased firing rate in one block (as compared with the

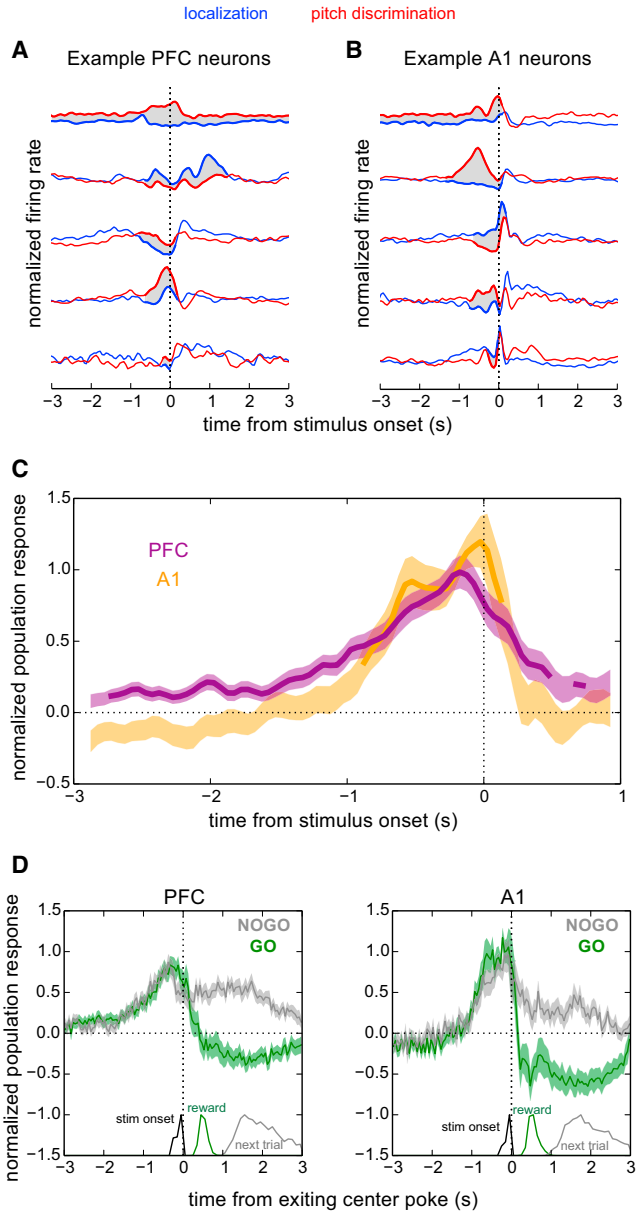

Figure 5. Within-Trial Timescale of the Encoding of Selection Rule

(A) PSTHs from example rule-encoding mPFC neurons in each block (blue, localization; red, pitch discrimination). Note that the timescale is much longer than that in previous figures. Firing rates are smoothed with a 50 ms Gaussian kernel and normalized to equal variance across neurons. Gray shading represents the maximum time interval, containing the hold period, during which the traces significantly diverge. Although these neurons were identified based on a difference in firing rate during the hold period, the traces often diverge for much longer. We observed a wide variety of timescales and dynamics. The first neuron effectively fires persistently more in one block. The third and fourth neurons demonstrate that the firing rate can either decrease during the nonpreferred block or, more commonly, increase during the preferred block.

(B) Example neurons from A1, following the conventions of (A). Again, the neurons exhibit a wide variety of dynamics, from essentially persistent block-specific activation for over 3 s preceding the stimulus (first neuron), to very brief activation well under 1 s (last neuron).

(C) Population time course: the traces represent the mean response, ± SEM, during the preferred block in mPFC (purple, n = 76) and A1 (orange, n = 36). The firing rates of all rule-encoding neurons were normalized (mean, 0; variance, 1) and then averaged together. The bold mean trace represents time points during which the population response significantly ($p < 0.05$, one-sample t test) exceeds zero, the mean firing rate. In both brain regions, the firing rate in the preferred block gradually increases, peaking around the time of stimulus onset, and then decreases more quickly back to baseline. The PFC population increases its response earlier (first significantly activated 2.7 s before stimulus onset) than the A1 population (first significantly activated 0.88 s before stimulus onset), consistent with the hypothesized role of PFC as the source of top-down modulation.

(D) Population time course plotted separately for go and nogo trials, from rule-encoding mPFC (left) and A1 (right) neurons during their preferred block. The peri-event time histograms (PETHs) are locked to the poststimulus exit from the center-port. As in (C), PETHs are mean ± SEM, and were normalized to unit variance and 0 mean before averaging across neurons. On nogo trials (gray), the firing rate remained elevated above baseline for at least several seconds, during which time the rat typically had already initiated the next trial. On go trials, the firing rate was suppressed below baseline as the rat moved to the choice port and consumed a reward, which always required at least several seconds. Latency distributions of trial events are plotted along the lower edge: stimulus onset (black), reward delivery (green, go trials only), and center-poke beginning the next trial (gray, nogo trials only). The next trial after a go trial would not be visible on this timescale due to the time required to consume the reward.

spontaneous rate while the rats were not performing the task), a decreased firing rate in the other block (versus spontaneous), or something else (for instance, a low spontaneous rate, an elevated rate for one block, and an even higher rate for the other block, which might reflect an encoding of task difficulty). Individual neurons exhibited a diversity of effects and we observed single units showing each of these possibilities (Figures 5A and 5B). However, across the population of rule-encoding neurons, the firing rate was significantly higher than spontaneous during the preferred block and significantly lower than spontaneous during the nonpreferred block (Figures S2F and S3F). These data argue against a model in which neurons encode task difficulty and instead suggest that each block actively engages two different populations of neurons, increasing the firing rate in one population and suppressing it in the other.

Encoding of Behavioral Choice

We found a prominent difference between the firing rates during go and nogo trials (Figure 5D). During a typical rule-encoding unit's preferred block, its firing rate remained elevated on nogo trials for several seconds after the center-poke, during which time the rat was often beginning the next trial. In contrast, on go trials, the typical unit's firing rate rapidly fell as the rat left the center port and remained suppressed for several seconds, during which time the rat was typically moving to the reward port and consuming reward.

One interpretation of this result is that rule encoding is particularly important for producing the nogo response, consistent with previous reports of enhanced encoding of nogo stimuli (David et al., 2012; Fritz et al., 2003). Another interpretation is that rule encoding persisted on nogo trials simply because the animal was already preparing to begin the next trial, whereas on go trials the rat was moving to the reward port to consume water and thus no longer needed to represent the stimulus selection rule.

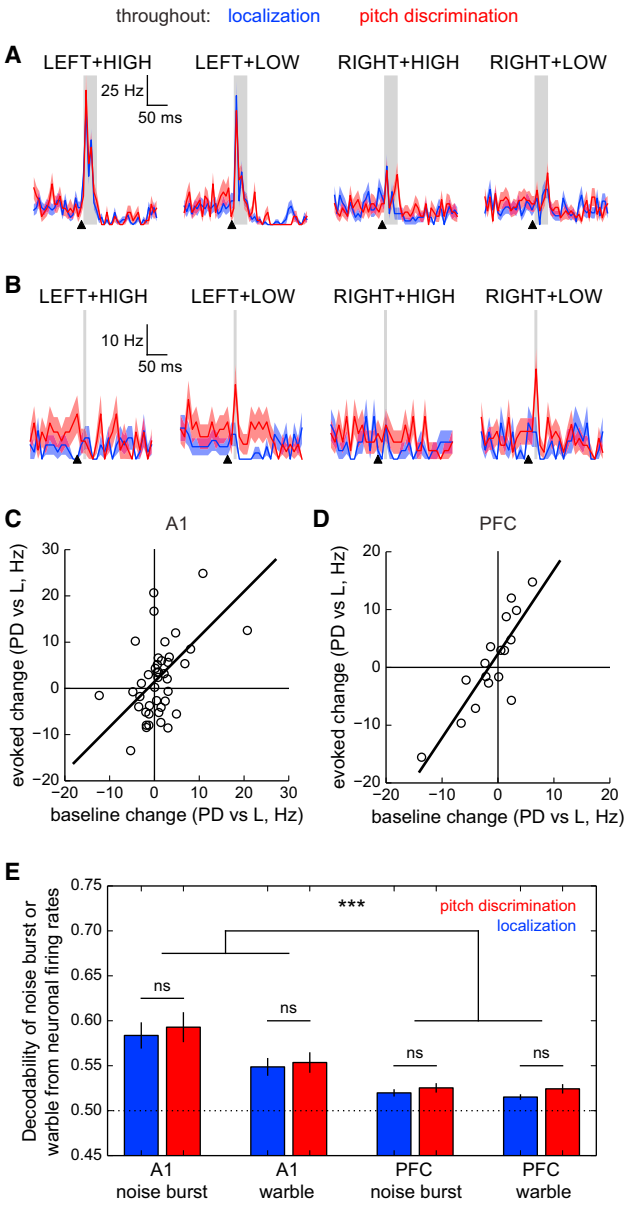

Figure 6. Limited Evidence for Modulation of Stimulus Tuning

(A) An example A1 neuron exhibiting a preference for some acoustic stimuli (LEFT+HIGH, LEFT+LOW) over others (RIGHT+HIGH, RIGHT+LOW), but no change in this tuning with block (localization, blue; pitch discrimination, red). Black triangles represent stimulus onset; shaded areas represent the response window for this neuron.

(B) An example auditory-responsive mPFC neuron. Evoked responses were weaker for mPFC neurons than for A1 neurons (Figures S5A–S5E).

(C) For A1 neurons (n = 43), increased anticipatory firing during one block significantly correlated ($p < 0.001$) with increased evoked responses during the same block. Each circle shows the change in evoked response (y-axis) versus the change in hold period firing rate (x-axis) for each neuron, quantified as mean pitch discrimination firing rate minus localization firing rate. The slope of the trend line is close to 1, suggesting that most of the evoked modulation across neurons is due to anticipatory modulation.

(D) Following the conventions of (C), but for auditory-responsive mPFC neurons (n = 17). Again, the change in anticipatory activity across neurons correlated closely with the change in evoked response ($p < 0.001$).

(E) No evidence for tuning changes that increase the decodability of the target sound. The identity of the noise burst (LEFT or RIGHT) or the warble (LOW or HIGH) could be decoded from the trial-by-trial responses of simultaneously recorded ensembles of auditory-responsive cells in either A1 or PFC. It could be decoded significantly more accurately ($p < 0.001$ for the main effect of brain region) using A1 responses (n = 21 ensembles of 49 neurons total) than using mPFC responses (n = 17 ensembles of 31 neurons total). However, for both sounds and both brain regions, the decoding was not significantly more accurate during either pitch discrimination (red) or localization (blue) trials ($p > 0.05$ for each pair of bars using a paired t test). The chance decoding level, attainable by a neuron with no information about the stimulus, was 0.5. Error bars represent SEM over ensembles. We used a three-way ANOVA on brain region, target sound, and block.

Changes in Baseline Activity Correlate with Changes in Evoked Activity

Given that the prestimulus activity encoded the selection rule, we next assessed whether the stimulus-driven activity in A1 differed between blocks. We first defined the evoked response window of each neuron as the period after stimulus onset during which the firing rate was significantly elevated above the prestimulus rate (Figures S5A–S5E). The evoked response on each trial was then defined as the number of spikes emitted during this window. We analyzed the mPFC neurons in the same way and found a population of neurons showing auditory responses to our task stimuli that were low-latency and tightly locked to stimulus onset, similar to A1 (Figures 6A and 6B). Auditory-responsive neurons in mPFC were significantly rarer (mPFC, 31/121; A1, 49/99; $p < 0.001$, Fisher's exact test) than in A1, and their responses were weaker (Figures S5C and S5D). Their median response latency was also significantly longer, though only slightly (19.75 ms versus 16.75 ms in A1; Figure S5E).

Our results show that prestimulus activity is modulated by selection rule, often persistently, consistent with a model in which rule-encoding neurons receive a higher level of tonic excitatory input during one block, for example. We expected that this task-specific modulation of baseline activity might correspond to an increased response during the stimulus-evoked response as well. In both A1 and mPFC, this was indeed the case: across the population, an increase in prestimulus firing rate during one block correlated with a comparable increase in evoked firing rate during the same block (Figures 6C and 6D; example cell: Figure 4B). However, after accounting for changes in prestimulus activity, we found very little evidence of any block-specific change in evoked firing rate (see Supplemental Experimental Procedures), which suggests that evoked activity is not strongly altered beyond an additive effect of baseline.

We also asked whether selection rule modulated stimulus tuning, for instance, to enhance the representation of the target sound. Such an effect might have been obscured in our analyses thus far, which averaged over stimuli in order to detect a change in overall response strength. To ascertain directly whether the target sound was better represented in the neural activity, we used an ideal linear decoder analysis (see Supplemental Experimental Procedures; Figure 6E) to quantify how well the stimulus-evoked activity in both brain regions encoded the identity of the noise burst or warble. As expected, the identity of each sound may be decoded more accurately from the activity of A1 neurons than from mPFC neurons, likely due to their stronger responses

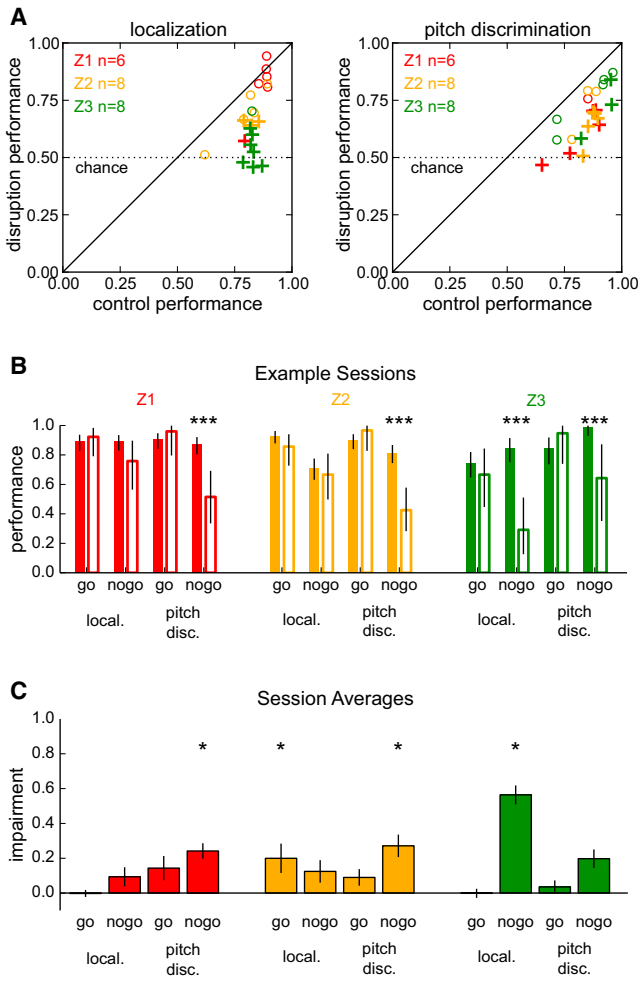

Figure 7. Disruption of mPFC Robustly Impairs Performance

(A) Electrical disruption of mPFC significantly impaired task performance (i.e., fraction of correct trials) during localization trials (left) and pitch discrimination trials (right) in most sessions. Each point represents the performance within a single session on disruption (y-axis) versus control (x-axis) trials. + represents sessions during which the performance was significantly impaired ($p < 0.05$, Fisher's exact test). Throughout this figure, colors represent different rats (red, Z1; yellow, Z2; and green, Z3).

(B) Example session from each rat. Performance is shown for each trial type (go and nogo in each block) on control (solid bars) and disruption (open bars) trials. Error bars represent 95% confidence intervals using Pearson-Klopper binomial fit. Asterisks indicate trial types for which electrical disruption significantly impairs performance (Fisher's exact test, $p < 0.001$ for all significant comparisons shown). The effect is robust within each example session, but varies between rats.

(C) Pattern of impairment (i.e., the difference in performance between control and disruption trials) across sessions for each rat. Error bars represent SEM across sessions. Asterisks show trial types that were significantly impaired ($p < 0.05$, binomial test on number of impaired sessions) for each rat. All rats were significantly impaired on nogo trials during one block or the other. One rat (Z2) also showed a significant impairment on localization go trials.

See also Figure S6.

and tighter stimulus selectivity. However, for both brain regions and for both noise bursts and warbles, we cannot decode the sound any more accurately from the responses on localization trials than on pitch discrimination trials. Moreover, we did not observe any correlation between each neuron's change in anticipatory activity and its tuning for the stimuli (Figures S5F and S5G) or any indication that some stimulus pairs (e.g., those requiring different responses in each block) elicited a greater response or a stronger modulation of that response (Figures S5H and S5I).

Thus, whereas neurons showing an increased prestimulus firing rate in one block generally showed a corresponding increase in the stimulus-evoked rate during the same block, these changes in evoked rate did not enhance the representation of the target sound, at least in any way that our decoding analysis could detect. However, as we discuss further below, it may be that in this task the brain does not need to maximize the information available about the target sound via tuning changes. (After all, even the small ensembles of neurons we recorded provided useful information about the identity of both sounds; given access to ensembles on the scale of auditory cortex, the brain should be able to decode the stimuli with virtual certainty.) Rather, the challenge of stimulus selection in a task such as ours may be to flexibly re-route the relevant stimulus information to the appropriate motor neurons at every block change.

Disruption of mPFC Significantly Impairs Task Performance

mPFC has been shown to be required for many task switching paradigms, which prompted us to ask whether it is required for our task. To answer this question, we developed an electrical disruption technique, inspired by transcranial magnetic stimulation (TMS) in humans (Dayan et al., 2013). We first implanted mPFC of three trained animals with extracellular stimulating electrodes. On 20% of trials ("disruption" trials), we injected a 10 Hz train of current pulses (see Experimental Procedures) throughout both the hold period and the auditory stimulus. Such electrical stimulation drives an extremely rapid activation of nearby neurons (Histed et al., 2009), followed by a slower suppression of firing rates (Butovas and Schwarz, 2003; Logothetis et al., 2010) for a few hundred milliseconds. Thus, the primary effect of this approach is neither to silence nor activate the brain region, but rather to disrupt the normal firing rates and patterns. Moreover, because we did not observe any spatial clustering of neurons preferring one task or the other, it is unlikely that such microstimulation would preferentially activate neurons of either preference, even if the stimulation protocol were purely excitatory.

Across all three animals, electrical disruption tended to impair performance (Figure 7) in both localization and pitch discrimination. This impairment was significant across sessions ($p < 0.05$, binomial test on the number of impaired sessions) during pitch discrimination for all (three of three) rats and during localization for most (two of three) rats. Electrical disruption largely, although not exclusively, affected performance on nogo trials. All rats were impaired on pitch discrimination nogo trials in almost all sessions. Some rats were additionally impaired on go or localization nogo trials. These effects were generally quite strong within individual sessions even though they varied between rats (Figures 7B and 7C). Taken together, these data suggest that in the absence of normal mPFC

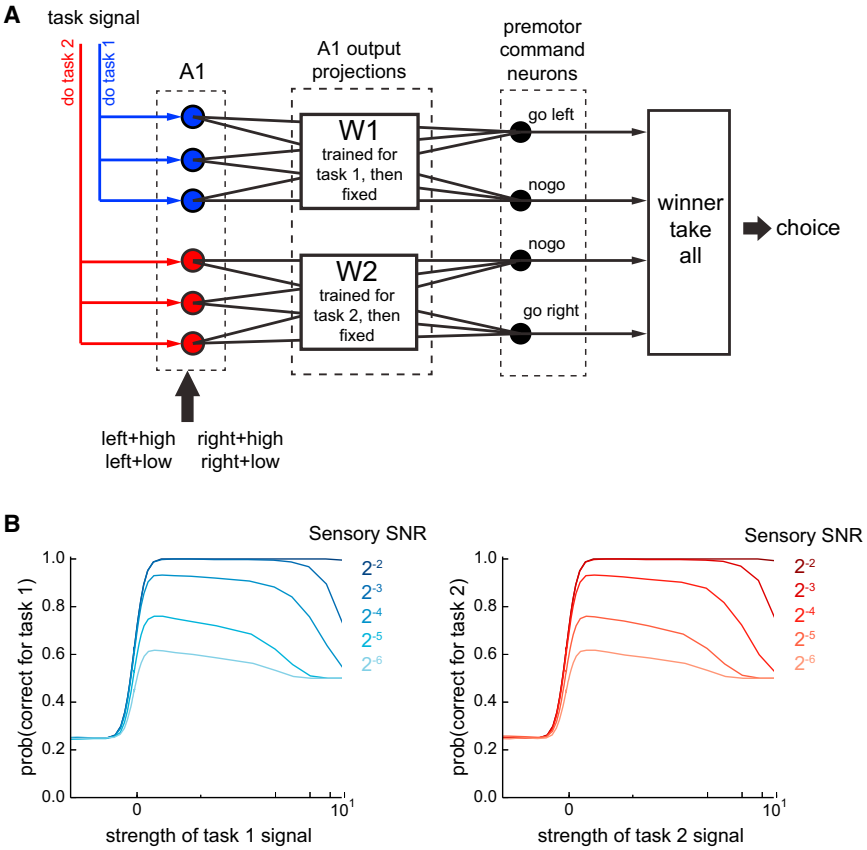

Figure 8. A Simulated Network Model Using Anticipatory Modulation Can Perform Stimulus Selection

(A) Network connectivity. Simulated A1 consists of N neurons, each with random tuning for the four task stimuli and subject to additive Gaussian noise. Red and blue populations are activated in one or the other block by an excitatory "task signal" projection. Each population connects to a set of premotor command neurons encoding the possible responses in that block. The projection weights W1 and W2 are optimized independently during an initial supervised training phase and constrained to be excitatory. The most active command neuron determines the network's choice.

(B) Performance of the model for N = 320 neurons on task 1 (left) and task 2 (right). We tested a range of values for the sensory signal-to-noise ratio (SNR), defined as the ratio of the tuning for sensory stimuli to the strength of the additive Gaussian noise in each A1 neuron. At the highest SNR of 0.25 (darkest trace), the model produces 100% correct responses for virtually any positive task switch signal. (Negative task signals correspond to activating the incorrect population for the current task.) At low SNRs, as the task signal increases in strength, the sensory input is eventually drowned out and the model's performance falls to chance (50%).

See also Figure S7.

activity, each rat resorts to its default strategy (typically "always go") in one or both blocks. Normal activity in mPFC is therefore important for good performance in our paradigm, but the strong impairment on nogo trials in particular made it difficult to ascertain whether the primary effect was on stimulus selection, as opposed to impulse control or some other aspect of the task (Figure S6).

A Simulated Network Model Demonstrates How Modulation of Anticipatory Activity Could Solve the Stimulus Selection Problem

Our data suggest a simple model of how the brain might perform stimulus selection: the PFC sends tonic excitation, perhaps indirectly, to a specific population of neurons for each task (e.g., populations for localization and pitch discrimination), which increases both their anticipatory and stimulus-evoked activity but does not affect their tuning. These populations connect to downstream motor regions that produce the appropriate response for that block; however, only the population with the increased firing rates can control behavior. We produced a quantitative simulation of this model to show that it can indeed solve the problem of stimulus selection. The simulation: (1) requires only random stimulus tuning in A1, (2) does not require tuning changes or synaptic reweighting after the initial training phase, and (3) uses only excitatory connections, consistent with the observation that most long-range projections in the brain are excitatory (Logothetis et al., 2010).

The model (Figure 8; see Supplemental Experimental Procedures for details) consists of a population of N neurons in A1, randomly tuned for each of our four stimulus pairs and subject to Gaussian noise. Half of the neurons are arbitrarily assigned to each task. Each population projects to two command neurons encoding the two possible behavioral responses during that block (e.g., go left and nogo during localization); this projection is trained to activate the correct command neuron for each stimulus pair. The actual behavioral choice is determined by which command neuron is the most active ("winner-take-all").

After the training phase, the synaptic weights are fixed and the model is tested on its ability to produce the correct response in each block. To simulate the rule encoding we observed in our data set, a "task signal" is added to the activations of the neurons in the appropriate population for the current block. Because all feed-forward weights are positive, adding this task signal translates into an excitatory boost to the premotor neurons receiving input from that population. Thus, even without any synaptic reweighting, the model tends to choose the response appropriate for the current block and stimulus. With 320 neurons, the network performs above 80% correct even with a signal-to-noise ratio (SNR) as low as 0.0625 (i.e., very weak sensory responses in each neuron relative to its internal noise). Increasing the network size can lower this SNR limit even further (Figure S7). Thus, our model demonstrates that a network can perform stimulus selection by task-specific activation of neurons, even without task-specific adaptation of their tuning.

DISCUSSION

Auditory Stimulus Selection: Task Switching between Conflicting Auditory Discriminations

When human listeners hear two simultaneous voices, they can selectively attend and respond to either one. This is a complex ability, and the rodent task we have developed models part of it—selecting and responding to one of two simultaneous sounds. Our subjects can voluntarily switch which sound they select, and do so at each block change within a single recording session.

Although previous studies of task switching in rodents did not require stimulus selection, they did require subjects to switch the navigational strategy they use to solve a maze (Rich and Shapiro, 2009) or between a sensory discrimination and a fixed response (e.g., "follow the light" versus "always go left"; Floresco et al., 2008; Durstewitz et al., 2010). It has been challenging to extend these results to task switching between distinct sensory discriminations, perhaps because this requires ignoring a previously trained stimulus. Even in cross-modal tasks, in which the targets and distractors come from entirely different modalities, strong cueing mechanisms (violating our "same stimulus; different response" condition) have been used to induce the switch: introducing novel stimuli (Birrell and Brown, 2000), deleting distractors (Otazu et al., 2009), or changing the behavioral arena completely (Haddon and Killcross, 2007). Finally, most previous studies required rats to shift no more than once per session, sometimes just once per lifetime, whereas our study requires multiple switches per session.

Despite its clinical and computational relevance (Ding and Simon, 2012), the auditory cocktail party problem remains less studied than comparable visual tasks. One multi-unit study (Lakatos et al., 2013) required primates to select a target stream of tones; however, the subjects were unable to ignore any distractor stream within one octave of the target. Human voices typically overlap extensively in acoustic frequency (McDermott, 2009), which contributes to the difficulty of the cocktail party problem, and we thus designed our stimuli to overlap in frequency. In sum, we believe our task represents an important first step toward understanding the cocktail party problem in rats, paving the way toward future studies with the modern tools available in rodent models (e.g., viral vectors for manipulating genetically identified cell types).

Anticipatory Activity in both mPFC and A1 Encodes the Selection Rule

We found that rodent mPFC robustly encodes the selection rule, analogous to the rule-encoding role of the primate prefrontal cortex (Asaad et al., 2000; Wallis et al., 2001; Johnston et al., 2007). Rule encoding develops in the mPFC population over 2.5 s before the stimulus onset, as the rat is planning to initiate a trial or even finishing the previous trial. Thus, we find that the prefrontal cortex densely and persistently codes for cognitive variables (cf. Rigotti et al., 2013), in contrast to the sparse coding of stimuli typical of sensory cortex (Hromádka et al., 2008; Olshausen and Field, 1996). This dense and widespread coding of selection rule in our data are perhaps surprising because only one bit of information needs to be encoded—pitch discrimination or localization—and this information is only necessary while making a decision on each trial. This persistent activity may represent a memory trace of the selection rule (Funahashi et al., 1989) or it may represent a shift to a completely different network state (Karlsson et al., 2012) depending on which stimulus the rat plans to select.

We also observed rule encoding in A1, a surprising result because this has traditionally been considered the domain of prefrontal areas. However, attention is known to induce anticipatory activity in sensory areas (Luck et al., 1997; Reynolds et al., 2000; Chen and Seidemann, 2012; Kastner et al., 1999; Thut et al., 2006). More generally, single neuron activity in primary sensory cortex can predict a motor response (Niwa et al., 2012) or an expected reward (Shuler and Bear, 2006), and anticipation of a visual stimulus can trigger a hemodynamic response in V1, although without a corresponding change in spiking (Sirotin and Das, 2009). Therefore, perhaps it is not surprising that primary sensory cortex could also encode selection rule. In this way, both the stimulus and the information about how that stimulus should be interpreted are encoded in the same neurons, providing a potential locus for the behavioral decision to be made.

Finally, we observed a surprising amount of similarity between A1 and mPFC, both of which showed robust encoding of the selection rule and of behavioral choice (Figure 5D). In monkeys, attention effects become more prominent higher in the visual hierarchy (Reynolds and Heeger, 2009). In contrast, our results show that rat A1 already encodes a nonsensory variable. This could be a difference between rats and monkeys, or between auditory and visual cortex, or both.

Comparison with Studies of Selective Attention and Task-Relevant Plasticity

In this study, we found limited evidence for any modulation of sensory-evoked responses in A1 beyond an additive effect of increased baseline. In particular, the neurons did not appear to change their tuning to encode the target stimulus with greater fidelity. This is consistent with some, but not all, previous studies of auditory task switching. For instance, switching between temporal and spatial auditory discriminations does not significantly change spatial tuning in A1 at the population level (Lee and Middlebrooks, 2011), although switching between passive and engaged states robustly modulates neuronal sensitivity (Otazu et al., 2009; Lee and Middlebrooks, 2011).

However, a series of pioneering experiments did demonstrate task-relevant tuning changes in A1 of ferrets trained to detect a target frequency (Fritz et al., 2003, 2010). One important methodological difference is that their study, unlike ours, made use of a large battery of probe stimuli and was therefore better optimized to detect receptive field changes. This plasticity was nuanced: it could induce both facilitation and, intriguingly, suppression at the task-relevant frequency; whether facilitation or suppression was more prevalent depended on whether positive or negative reinforcement was used (David et al., 2012). Further studies of complex auditory behaviors will be necessary to better understand the factors that determine whether a given behavioral paradigm produces tuning changes in auditory cortex.

Visual selective attention has been shown to enhance target representations and suppress distractors in V4 and other visual areas (Cohen and Maunsell, 2011; David et al., 2008; Mitchell et al., 2007; Reynolds and Heeger, 2009). However, selective attention consists of two component processes with separate behavioral measures: stimulus selection and perceptual enhancement (cf. Knudsen, 2007; Reynolds and Chelazzi, 2004; Pestilli et al., 2011). Perceptual enhancement is typically measured using faint stimuli to probe psychophysical thresholds (Cohen and Maunsell, 2009). In contrast, studies such as ours that use easily detectable stimuli far above threshold (Hocherman et al., 1976; Stoet and Snyder, 2004) often report limited evidence for enhanced representations of target stimuli in sensory cortex (Sasaki and Uka, 2009; Mante et al., 2013; Pestilli et al., 2011). In such tasks, the dominant challenge is not detecting the stimuli, but rather selecting the relevant target, which may rely on changes of baseline (Pestilli et al., 2011) perhaps due to anticipatory modulation (Chen and Seidemann, 2012). Similarly, the cocktail party problem is often difficult because all voices are of competing intensity, not because the target voice is barely audible. Thus, the nature of the task may determine whether stimulus representations are modulated or remain relatively fixed.

Some models of visual selection (Gilbert and Shallice, 2002; Mante et al., 2013) propose that stimulus selection occurs in frontal areas, not sensory cortex. Our data are similar in the sense that we do not observe tuning changes in sensory cortex (Mante et al., 2013) but different in that we do not observe strong representations of the stimuli in PFC, similar to a recent observation in primate PFC (Lara and Wallis, 2014). Our results are more consistent with a distributed processing model in which contextual information from PFC modulates activity in A1 to increase the fidelity with which the appropriate motor action can be read out (Fritz et al., 2010; David et al., 2012; Blake et al., 2002).

The Potential Roles of Motor Planning and Posture

We considered the potential roles of both posture—the angle of the rat's head relative to the behavior box in particular—and motor planning in driving the observed task-specific modulation of neuronal activity. Because each block is associated with a different choice port, it is plausible that the rats adopted a different default motor plan for the two blocks: go left for one task and go right for the other. Moreover, we observed a difference in head angle between blocks, presumably a behavioral strategy that rats used to prepare for the differing motor actions required. We note a similarity with some blocked visual spatial attention tasks, in which 80% of the trials require a saccade in the same direction (Cohen and Maunsell, 2009). In all such tasks, it can be difficult to disambiguate response selection and stimulus selection (Erlich et al., 2011; Sato and Schall, 2003; Steinmetz and Moore, 2012).

We found that some rule-encoding neurons, especially in mPFC, also encoded head angle to some extent (Figures S2K and S3K). This is consistent with previous mPFC data (Euston and McNaughton, 2006) and the idea that single prefrontal neurons simultaneously encode disparate sensorimotor and cognitive signals (Rigotti et al., 2013). However, we found that the firing rate of most neurons was better explained by block than by head angle and that rule encoding persisted even when we controlled for head angle by trial selection (Figures S2I–S2T, S3I–S3Q). These results suggest that cognitive context (i.e., task rule) drives both the observed neuronal activity and the adaptive posture, rather than posture directly driving the neuronal activity.

Even if the rule-encoding activity we observed does not simply reflect postural differences, it is possible that it represents an internal motor plan (which could be covertly present even in the absence of a measurable behavioral parameter like head angle). It is difficult to disambiguate motor planning from rule encoding because the task itself requires different sensorimotor mappings in each block. However, the time course of the rule encoding was quite protracted in many neurons, in some cases even persistent throughout the block (Figure 5), during which time the rat was engaged in various motor actions such as moving to or from the center port and harvesting reward (see example neurons in Figures S2P and S3P). It seems unlikely that neurons would continue to represent the specific action of moving from the center port to the choice port on such a long timescale. In addition, in our paradigm any default motor plan is subject to cancellation on nogo trials (sometimes called "countermanding," cf. Schall et al., 2000; Eagle and Robbins, 2003; Eagle et al., 2008), because the animal does not know during the anticipatory period whether it will be signaled to perform a go response or not. Finally, we did not observe any correlation between the anticipatory firing rate and the reaction or movement time (Figures S2G, S2H, S3G, and S3H). To summarize, our task requires remapping sensory stimuli to motor responses, and it is reasonable to expect rule encoding to incorporate both the sensory and motor planning aspects of this remapping. Anticipatory modulation may encode both the selection rule and, therefore, the motor plan required to implement that rule.

Stimulus Selection via Activation of Latent Circuits for Each Target

In light of our results, we propose a model for stimulus selection based on activation of separate, task-specific circuits. In this model, there are two neuronal populations in both A1 and mPFC—one activated during the localization block, the other during the pitch discrimination block. These populations show increased prestimulus and stimulus-evoked activity during their preferred block but do not change their tuning for specific stimuli. We hypothesize that the difference between these populations is their downstream connectivity: each may project to separate targets in a downstream effector region such as the striatum (Znamenskiy and Zador, 2013), forming distinct circuits for each task. In this model, only one circuit is activated at a time, via feed-forward excitation perhaps originating in mPFC, and only this circuit has sufficient baseline activity to drive behavior. In some ways, this model is more parsimonious than the traditional tuning change model, which requires that prefrontal (or other) brain regions be able to modulate the tuning of many A1 neurons as quickly as the subject shifts the focus of attention. Although attention does produce tuning changes (David et al., 2008; Fritz et al., 2003)

over minutes (which is typically the fastest that they can be estimated), it is unclear how known synaptic plasticity mechanisms could mediate task-specific tuning changes on a subsecond timescale.

Our model makes several testable predictions. First, there should exist "premotor" neurons (possibly in the striatum) receiving input from A1 that also show a block-dependent anticipatory modulation. In addition, specific activation of one of the subpopulations in mPFC, A1, or striatum should bias behavior toward the block preferred by that subpopulation. Such a manipulation would require targeting specific neurons based on their anticipatory firing rate, a challenging experiment that might nonetheless be feasible using activity-dependent promoters to drive light-gated ion channels, for example.

In conclusion, these results establish the rat as a model organism for auditory stimulus selection, paving the way for future investigations of the cocktail party problem with emerging optical and genetic tools amenable to rodents. We found widespread and robust rule encoding in mPFC and A1, although we observed little change in the stimulus tuning of evoked responses. We propose a simple model to explain these results: task-specific activation of latent circuits, rather than task-specific adaptation of a single circuit.

EXPERIMENTAL PROCEDURES

All procedures were approved by the Animal Care and Use Committee at the University of California, Berkeley. We used male Long-Evans rats (Harlan), housed in pairs. Training began when their body mass reached 150–225 g, at approximately 45–60 days old. Rats were given restricted access to water in the day before the training session so that they would be motivated to obtain a water reward. After each session, they were given ad lib access to water for 1 hour. We monitored body weight and other markers to ensure they remained healthy. We used standard behavioral shaping and surgical implantation techniques (see Supplemental Experimental Procedures).

Electrical Disruption Protocol

We began with a very low current, ~10 uA per electrode, which was typically too low to produce any behavioral effect. We wanted to use a minimal perturbation to ensure that the effects were as localized as possible in both time and space, and so we used pilot sessions to increase the amount of current until performance on the task became moderately impaired. During the testing sessions that we report in the main text, the mean currents used were 37 uA, 41 uA, and 23 uA per electrode for Z1, Z2, and Z3 respectively. See Supplemental Experimental Procedures for further details and comparisons with other studies.

Data Analysis

We preprocessed the data using the open-source OpenElectrophy software suite (Garcia and Fourcaud-Trocmé, 2009) built on the Neo object model (Garcia et al., 2014). We used KlustaKwik (Kadir et al., 2013) and Klusters (Hazan et al., 2006), while blind to the experimental variables, to identify single units.

We analyzed the data with Python within the IPython environment (Pérez and Granger, 2007) and the modules numpy, scipy, matplotlib, scikits-learn, statsmodels, and pandas. We also conducted some statistical analyses in R using the rpy2 module. Except where otherwise noted in the text, we observed consistent results across all subjects and therefore pooled the data (Figures S2D, S2E, S3D, and S3E). All of the data and code necessary to recapitulate the analyses presented here are available online at https://github.com/cxrodgers/Rodgers2014 and at the data-sharing website CRCNS.org through link http://dx.doi.org/10.6080/K0W66HPJ.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and seven figures and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2014.04.031.

ACKNOWLEDGMENTS

The authors thank Mohammad Dastjerdi and Christian Fernandez for developing the initial rodent behaviors in the lab; Alfonso Apicella for helping set up the lab; Anita Devineni, Tony Zador, Uri Livneh, Erin Rich, Nicholas Steinmetz, and Aaron Koralek for comments on the manuscript; Sarah Kochik and Vuong Vu for managing behavioral training; Trevor Dolinajec, Lily Lin, Danielle Jobe, and Yassi Sabahi for electrode fabrication; Yuliy Tsank, Sharon Lin, Ambika Rustagi, Amish Shah, Carol Pham, Francine Foo, Helene Moorman, Jason Murphy, Jason Zhang, Josh Chong, Kelly Clancy, Tanling Hsu, Pooja Upadhyaya, and Tommy Li for help with rat behavior; and Vivek Ayer for technical assistance.

Accepted: April 3, 2014
Published: June 4, 2014

REFERENCES

Ahveninen, J., Hämäläinen, M., Jääskeläinen, I.P., Ahlfors, S.P., Huang, S., Lin, F.-H., Raij, T., Sams, M., Vasios, C.E., and Belliveau, J.W. (2011). Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. Proc. Natl. Acad. Sci. USA *108*, 4182–4187.

Asaad, W.F., Rainer, G., and Miller, E.K. (2000). Task-specific neural activity in the primate prefrontal cortex. J. Neurophysiol. *84*, 451–459.

Bee, M.A., and Micheyl, C. (2008). The cocktail party problem: what is it? How can it be solved? And why should animal behaviorists study it? J. Comp. Psychol. *122*, 235–251.

Birrell, J.M., and Brown, V.J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. J. Neurosci. *20*, 4320–4324.

Blake, D.T., Strata, F., Churchland, A.K., and Merzenich, M.M. (2002). Neural correlates of instrumental learning in primary auditory cortex. Proc. Natl. Acad. Sci. USA *99*, 10114–10119.

Brunton, B.W., Botvinick, M.M., and Brody, C.D. (2013). Rats and humans can optimally accumulate evidence for decision-making. Science *340*, 95–98.

Butovas, S., and Schwarz, C. (2003). Spatiotemporal effects of microstimulation in rat neocortex: a parametric study using multielectrode recordings. J. Neurophysiol. *90*, 3024–3039.

Carandini, M., and Churchland, A.K. (2013). Probing perceptual decisions in rodents. Nat. Neurosci. *16*, 824–831.

Carlson, N.L., Ming, V.L., and Deweese, M.R. (2012). Sparse codes for speech predict spectrotemporal receptive fields in the inferior colliculus. PLoS Comput. Biol. *8*, e1002594.

Chen, Y., and Seidemann, E. (2012). Attentional modulations related to spatial gating but not to allocation of limited resources in primate V1. Neuron *74*, 557–566.

Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. *25*, 975–979.

Cohen, M.R., and Maunsell, J.H.R. (2009). Attention improves performance primarily by reducing interneuronal correlations. Nat. Neurosci. *12*, 1594–1600.

Cohen, M.R., and Maunsell, J.H.R. (2011). Using neuronal populations to study the mechanisms underlying spatial and feature attention. Neuron *70*, 1192–1204.

David, S.V., Hayden, B.Y., Mazer, J.A., and Gallant, J.L. (2008). Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. Neuron *59*, 509–521.

David, S.V., Fritz, J.B., and Shamma, S.A. (2012). Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc. Natl. Acad. Sci. USA *109*, 2144–2149.

Dayan, E., Censor, N., Buch, E.R., Sandrini, M., and Cohen, L.G. (2013). Noninvasive brain stimulation: from physiology to network dynamics and back. Nat. Neurosci. *16*, 838–844.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. *18*, 193–222.

DeWeese, M.R., Wehr, M., and Zador, A.M. (2003). Binary spiking in auditory cortex. J. Neurosci. *23*, 7940–7949.

Ding, N., and Simon, J.Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. USA *109*, 11854–11859.

Durstewitz, D., Vittoz, N.M., Floresco, S.B., and Seamans, J.K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. Neuron *66*, 438–448.

Eagle, D.M., and Robbins, T.W. (2003). Inhibitory control in rats performing a stop-signal reaction-time task: effects of lesions of the medial striatum and d-amphetamine. Behav. Neurosci. *117*, 1302–1317.

Eagle, D.M., Baunez, C., Hutcheson, D.M., Lehmann, O., Shah, A.P., and Robbins, T.W. (2008). Stop-signal reaction-time task performance: role of prefrontal cortex and subthalamic nucleus. Cereb. Cortex *18*, 178–188.

Erlich, J.C., Bialek, M., and Brody, C.D. (2011). A cortical substrate for memory-guided orienting in the rat. Neuron *72*, 330–343.

Euston, D.R., and McNaughton, B.L. (2006). Apparent encoding of sequential context in rat medial prefrontal cortex is accounted for by behavioral variability. J. Neurosci. *26*, 13143–13155.

Floresco, S.B., Block, A.E., and Tse, M.T.L. (2008). Inactivation of the medial prefrontal cortex of the rat impairs strategy set-shifting, but not reversal learning, using a novel, automated procedure. Behav. Brain Res. *190*, 85–96.

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat. Neurosci. *6*, 1216–1223.

Fritz, J.B., Elhilali, M., and Shamma, S.A. (2005). Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. J. Neurosci. *25*, 7623–7635.

Fritz, J.B., David, S.V., Radtke-Schuller, S., Yin, P., and Shamma, S.A. (2010). Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat. Neurosci. *13*, 1011–1019.

Funahashi, S., Bruce, C.J., and Goldman-Rakic, P.S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J. Neurophysiol. *61*, 331–349.

Garcia, S., and Fourcaud-Trocmé, N. (2009). OpenElectrophy: an electrophysiological data- and analysis-sharing framework. Front. Neuroinform. *3*, 14.

Garcia, S., Guarino, D., Jaillet, F., Jennings, T., Pröpper, R., Rautenberg, P.L., Rodgers, C.C., Sobolev, A., Wachtler, T., Yger, P., and Davison, A.P. (2014). Neo: an object model for handling electrophysiology data in multiple formats. Front. Neuroinform. *8*, 10.

Geissler, D.B., and Ehret, G. (2002). Time-critical integration of formants for perception of communication calls in mice. Proc. Natl. Acad. Sci. USA *99*, 9021–9025.

Gilbert, S.J., and Shallice, T. (2002). Task switching: a PDP model. Cognit. Psychol. *44*, 297–337.

Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. Annu. Rev. Neurosci. *30*, 535–574.

Haddon, J.E., and Killcross, S. (2007). Contextual control of choice performance: behavioral, neurobiological, and neurochemical influences. Ann. N Y Acad. Sci. *1104*, 250–269.

Hazan, L., Zugaro, M., and Buzsáki, G. (2006). Klusters, NeuroScope, NDManager: a free software suite for neurophysiological data processing and visualization. J. Neurosci. Methods *155*, 207–216.

Histed, M.H., Bonin, V., and Reid, R.C. (2009). Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation. Neuron *63*, 508–522.

Hocherman, S., Benson, D.A., Goldstein, M.H., Jr., Heffner, H.E., and Hienz, R.D. (1976). Evoked unit activity in auditory cortex of monkeys performing a selective attention task. Brain Res. *117*, 51–68.

Hromádka, T., Deweese, M.R., and Zador, A.M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. PLoS Biol. *6*, e16.

Johnston, K., Levin, H.M., Koval, M.J., and Everling, S. (2007). Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. Neuron *53*, 453–462.

Kadir, S., Goodman, D., and Harris, K. (2013). High-dimensional cluster analysis with the masked EM algorithm. http://arxiv.org/abs/1309.2848.

Karlsson, M.P., Tervo, D.G.R., and Karpova, A.Y. (2012). Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. Science *338*, 135–139.

Kastner, S., Pinsk, M.A., De Weerd, P., Desimone, R., and Ungerleider, L.G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron *22*, 751–761.

Knudsen, E.I. (2007). Fundamental components of attention. Annu. Rev. Neurosci. *30*, 57–78.

Kvitsiani, D., Ranade, S., Hangya, B., Taniguchi, H., Huang, J.Z., and Kepecs, A. (2013). Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. Nature *498*, 363–366.

Lakatos, P., Musacchia, G., O'Connel, M.N., Falchier, A.Y., Javitt, D.C., and Schroeder, C.E. (2013). The spectrotemporal filter mechanism of auditory selective attention. Neuron *77*, 750–761.

Lara, A.H., and Wallis, J.D. (2014). Executive control processes underlying multi-item working memory. Nat. Neurosci. http://dx.doi.org/10.1038/nn.3702.

Lee, C.-C., and Middlebrooks, J.C. (2011). Auditory cortex spatial sensitivity sharpens during task performance. Nat. Neurosci. *14*, 108–114.

Logothetis, N.K., Augath, M., Murayama, Y., Rauch, A., Sultan, F., Goense, J., Oeltermann, A., and Merkle, H. (2010). The effects of electrical microstimulation on cortical signal propagation. Nat. Neurosci. *13*, 1283–1291.

Luck, S.J., Chelazzi, L., Hillyard, S.A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J. Neurophysiol. *77*, 24–42.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.

McDermott, J.H. (2009). The cocktail party problem. Curr. Biol. *19*, R1024–R1027.

Mesgarani, N., and Chang, E.F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. Nature *485*, 233–236.

Miller, E.K., and Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. *24*, 167–202.

Mitchell, J.F., Sundberg, K.A., and Reynolds, J.H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron *55*, 131–141.

Monchi, O., Petrides, M., Petre, V., Worsley, K., and Dagher, A. (2001). Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. J. Neurosci. *21*, 7733–7741.

Moore, T., Armstrong, K.M., and Fallah, M. (2003). Visuomotor origins of covert spatial attention. Neuron *40*, 671–683.

Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science *229*, 782–784.

Niwa, M., Johnson, J.S., O'Connor, K.N., and Sutter, M.L. (2012). Activity related to perceptual judgment and action in primary auditory cortex. J. Neurosci. *32*, 3193–3210.

Olshausen, B.A., and Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature *381*, 607–609.

Otazu, G.H., Tai, L.-H., Yang, Y., and Zador, A.M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. Nat. Neurosci. *12*, 646–654.

Pai, S., Erlich, J.C., Kopec, C., and Brody, C.D. (2011). Minimal impairment in a rat model of duration discrimination following excitotoxic lesions of primary auditory and prefrontal cortices. Front Syst Neurosci *5*, 74.

Pérez, F., and Granger, B.E. (2007). IPython: a system for interactive scientific computing. Comput. Sci. Eng. *9*, 21–29.

Pestilli, F., Carrasco, M., Heeger, D.J., and Gardner, J.L. (2011). Attentional enhancement via selection and pooling of early sensory responses in human visual cortex. Neuron *72*, 832–846.

Petkov, C.I., Kang, X., Alho, K., Bertrand, O., Yund, E.W., and Woods, D.L. (2004). Attentional modulation of human auditory cortex. Nat. Neurosci. *7*, 658–663.

Ragozzino, M.E., Detrick, S., and Kesner, R.P. (1999). Involvement of the prelimbic-infralimbic areas of the rodent prefrontal cortex in behavioral flexibility for place and response learning. J. Neurosci. *19*, 4585–4594.

Raposo, D., Sheppard, J.P., Schrater, P.R., and Churchland, A.K. (2012). Multisensory decision-making in rats and humans. J. Neurosci. *32*, 3726–3735.

Reynolds, J.H., and Chelazzi, L. (2004). Attentional modulation of visual processing. Annu. Rev. Neurosci. *27*, 611–647.

Reynolds, J.H., and Heeger, D.J. (2009). The normalization model of attention. Neuron *61*, 168–185.

Reynolds, J.H., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of V4 neurons. Neuron *26*, 703–714.

Rich, E.L., and Shapiro, M. (2009). Rat prefrontal cortical neurons selectively code strategy switches. J. Neurosci. *29*, 7208–7219.

Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature *497*, 585–590.

Sasaki, R., and Uka, T. (2009). Dynamic readout of behaviorally relevant signals from area MT during task switching. Neuron *62*, 147–157.

Sato, T.R., and Schall, J.D. (2003). Effects of stimulus-response compatibility on neural selection in frontal eye field. Neuron *38*, 637–648.

Sayers, B.M., and Cherry, E.C. (1957). Mechanism of binaural fusion in the hearing of speech. J. Acoust. Soc. Am. *29*, 973–987.

Schall, J.D., Hanes, D.P., and Taylor, T.L. (2000). Neural control of behavior: countermanding eye movements. Psychol. Res. *63*, 299–307.

Shuler, M.G., and Bear, M.F. (2006). Reward timing in the primary visual cortex. Science *311*, 1606–1609.

Sirotin, Y.B., and Das, A. (2009). Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. Nature *457*, 475–479.

Steinmetz, N.A., and Moore, T. (2012). Lumping and splitting the neural circuitry of visual attention. Neuron *73*, 410–412.

Stoet, G., and Snyder, L.H. (2004). Single neurons in posterior parietal cortex of monkeys encode cognitive set. Neuron *42*, 1003–1012.

Thut, G., Nietzel, A., Brandt, S.A., and Pascual-Leone, A. (2006). Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. J. Neurosci. *26*, 9494–9502.

Turken, A.U., and Swick, D. (1999). Response selection in the human anterior cingulate cortex. Nat. Neurosci. *2*, 920–924.

Wallis, J.D., Anderson, K.C., and Miller, E.K. (2001). Single neurons in prefrontal cortex encode abstract rules. Nature *411*, 953–956.

Wise, S.P. (2008). Forward frontal fields: phylogeny and fundamental function. Trends Neurosci. *31*, 599–608.

Young, J.J., and Shapiro, M.L. (2011). The orbitofrontal cortex and response selection. Ann. N Y Acad. Sci. *1239*, 25–32.

Zariwala, H.A., Kepecs, A., Uchida, N., Hirokawa, J., and Mainen, Z.F. (2013). The limits of deliberation in a perceptual decision task. Neuron *78*, 339–351.

Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". Neuron *77*, 980–991.

Znamenskiy, P., and Zador, A.M. (2013). Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. Nature *497*, 482–485.

Neuron, Volume *82*

Supplemental Information

Neural Correlates of Task Switching in Prefrontal Cortex and Primary Auditory Cortex in a Novel Stimulus Selection Task for Rodents

Chris C. Rodgers and Michael R. DeWeese

Supplemental Figures

Figure S1. Behavioral performance, related to Figure 2

Refer to “Behavioral training” and “Chance performance on the task” in Supplemental Experimental Procedures for details.

A) Performance of each rat in greater detail, with go and nogo trials separately considered in each block. Rats generally did better on go (first and third columns) than on nogo trials (second and fourth columns). Some rats did better on localization (first and second columns) than on pitch discrimination (third and fourth columns). Error bars show SEM across sessions.

B) Rats respond to the distractor, not the target/distractor combination. Performance (fraction of correct responses) of one rat that performed a slightly modified “catch trial” task on the last day of recordings. This task was designed to probe whether the rats learn to respond to a unified stimulus pair, or whether they learn to respond just to the target sound regardless of the identity of the distractor. On a small proportion (15%) of trials, we replaced the distractor with a neutral sound that the rat had never heard on any prior training session or behavioral testing session. For example, on catch trials during localization the rat heard the same target as before (LEFT or RIGHT) with a novel mid-range warble with no behavioral relevance. If the rats had memorized each of the four possible stimulus pairs and were unable to generalize properly to cases with the usual target sound combined with a novel distractor, they should have performed at chance on these novel stimulus pairs. The performance on catch trials (red) and standard trials (white) for each trial type is shown, with 95% confidence intervals (Pearson-Klopper binomial fit). The rats perform just as well on catch trials as on standard trials (Fisher’s Exact Test on the trial outcomes, $p > 0.05$ in all cases). This suggests that the rats are selecting the target stimulus, not memorizing a fixed set of four stimulus pairs.

C) Same as panel B, but for a different rat.

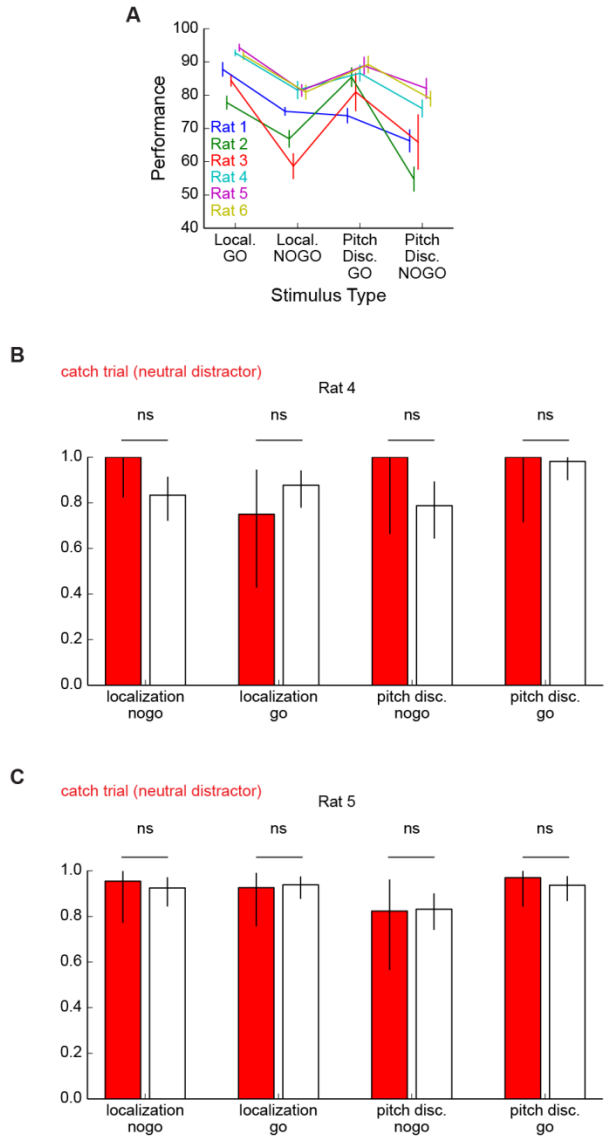
A
100
90
80
70
60
50
40
Performance
Rat 1
Rat 2
Rat 3
Rat 4
Rat 5
Rat 6
Local. GO
Local. NOGO
Pitch Disc. GO
Pitch Disc. NOGO
Stimulus Type
B
catch trial (neutral distractor)
Rat 4
ns
ns
ns
ns
1.0
0.8
0.6
0.4
0.2
0.0
localization nogo
localization go
pitch disc. nogo
pitch disc. go
C
catch trial (neutral distractor)
Rat 5
ns
ns
ns
ns
1.0
0.8
0.6
0.4
0.2
0.0
localization nogo
localization go
pitch disc. nogo
pitch disc. go

FIGURE S1

Figure S2. More information on mPFC anticipatory effect, related to Figure 3

Refer to "Analysis of waveform variation and firing rate drift" and "Power analysis" in Supplemental Experimental Procedures for details.

A) Stacked histogram of the difference, rather than ratio, of hold period firing rate between blocks for all mPFC neurons.

B) Alternative presentation of the hold period effect across mPFC neurons. The hold period firing rate in each block is shown as the x- and y-coordinate of each point (red and blue: significant block preference; gray: not significant). Note the logarithmic scaling, to avoid crowding the points with low firing rate. Error bars are 95% confidence intervals obtained by bootstrapping and were truncated at the edge of the plot. Significance was assessed with a Mann-Whitney U-test as described in the text.

C) Analysis of the hold period effect in mPFC on various types of correct and error trials. The trials are grouped by the meaning of the target sound, the rat's response, and the meaning of the distractor sound. Neurons are grouped by their preferred block. White bars represent correct trials, on which the rat's response matches the target sound. The gray bar represents "go-on-nogo" error trials when the target required a nogo response but the rat went to the choice port anyway. (The opposite error, nogo-on-go, was too rare to include in this analysis. We only analyzed neurons from sessions with at least 3 trials of each type, where type refers to each bar in this panel.) The orange bar represents "interference" trials on which the rat heard a distractor sound cuing a go response and went to the choice port associated with that distractor (WP, or wrong port). To aid in visualization, firing rates were normalized by subtracting the mean response on correct trials during the non-preferred block, and then averaged across neurons. There is no significant difference in the hold period activity between correct go, correct nogo, and incorrect go-on-nogo trials. If the hold period activity actually encoded a simple motor plan to go to the choice port regardless of the upcoming stimulus, then it should have been lower on correct nogo trials when the rat did not perform this action. In fact there is no significant difference between correct-nogo and correct-go trials. Thus, if the hold period activity represents a motor plan, it must be subject to change ("countermanding") after the stimulus. However, there is a significant difference between interference trials and all other trial types in that block. That is, when the rat gives the response that would be appropriate in the other block, the anticipatory activity is higher in the non-preferred block and lower in the preferred block — *i.e.*, the block modulation is attenuated, trending toward reversal of the effect. Significance between each pair of bars was assessed with a paired Mann-Whitney U-test across neurons, which is invariant under the subtractive normalization performed, and the p-values were Bonferroni corrected.

D) Histogram across rats of the number of neurons preferring pitch discrimination (red), localization (blue), and neither (black). In some rats (marked N/A), no mPFC neurons were recorded.

E) Proportion of rule-encoding neurons in mPFC is consistent across rats. The data from (D) are expressed here as a percentage of total neurons. No percentages are plotted for rats with fewer than 8 neurons total recorded in mPFC.

F) On average, firing rates in both blocks are bidirectionally modulated, compared with the spontaneous rate. We defined the spontaneous rate as the average rate during epochs more than 2 s from the nearest stimulus onset. Localization-preferring neurons (left) fire significantly more than spontaneous in localization and significantly less than spontaneous in pitch discrimination. A similar statement holds for pitch discrimination-preferring neurons (center). Neurons that do not prefer either block (right) — that is, neurons for which the pre-stimulus firing rate does not significantly differ between blocks — tend to be suppressed (relative to their spontaneous rates) in both blocks, though this trend was not statistically significant during when compared with pitch discrimination ($p = 0.06$). The bars show the average and SEM across neurons of the hold period firing rates minus the spontaneous firing rate for that neuron. We assessed significance with a paired Mann-Whitney test, which is invariant to this form of subtractive normalization.

G) No evidence for a correlation between the pre-stimulus firing rate of PFC neurons and the animal's reaction time, defined as the time between stimulus onset and withdrawal from center port. We plot here the

distribution of correlation coefficients obtained in both blocks, with red representing significantly correlated neurons ($p < 0.05$ after correction with the false discovery rate). Only a small minority of neurons showed a significant correlation, and the mean of the full distribution was not significantly different from zero ($p > 0.05$, one-sample t-test). Similar results were obtained when considering only rule-encoding neurons (data not shown).

H) Similar to (G) except that here we are correlating the anticipatory firing rate with the "motion time," which we defined as the time necessary for the rat to move to the choice port on successful go trials. Again, only a small minority of neurons show an individually significant correlation and the population distribution is not significantly different from zero. Similar results were obtained when considering only rule-encoding neurons (data not shown).

I) The azimuthal (left/right) angle of the rat's head during center poke, relative to the behavior box, differs by approximately 30 degrees between blocks. Red: pitch discrimination. Blue: localization. The mean head angle over all trials is defined here as zero degrees. These data are from an example session but all analyzed sessions yielded similar results. This preparatory motor activity is presumably an adaptive behavioral strategy in response to the fact that the choice port differs between blocks. See "Video analysis of preparatory head positioning" in Supplemental Experimental Procedures for further detail on panels I-L.

J) Example PFC neuron recorded during the session shown in (I). Each point shows the square root of the number of spikes fired and head angle on a single trial. (Throughout this figure, we use the square root transformation to stabilize the variance of the counts and thus to ensure homogeneity of variance independent of the mean to facilitate our regression analysis.) Across all trials, these variable are significantly correlated (black trend line). However, this correlation between firing rate and head angle is almost entirely explained by block type. Within each block, there is no such correlation — the red and blue trend lines are not statistically significantly different from horizontal.

K) Fraction of explainable variance (FEV) in the spike counts (again square root-normalized) that the least-squares linear fit attributes to block (orange) and head angle (black), individually for each rule-encoding PFC neuron during the analyzed video sessions. The magenta horizontal line indicates 50%. The bars are sorted in order of increasing head angle FEV. Most bars are mostly orange, indicating that block is the major explanatory factor in the spike count. Only three out of 16 bars are more than 50% black, corresponding to neurons whose firing rate was mostly explained by head angle.

L) Summary plot of the data in panel (K). The distribution of FEV across neurons is plotted for each factor (head angle and block). Individual crosses represent individual rule-encoding PFC neurons. The red line shows the median; the gray box outlines the inter-quartile range. Across the population, most of the variance (median: 94.1%) is explained by block; the small amount that remains is explained by head angle (median: 5.9%).

M) Rule-encoding in mPFC neurons is broadly similar to that of the full data set when only "posture-equalized" trials are considered. Here, we analyze only rule-encoding neurons (*i.e.*, those with a significant preference for one block) recorded during the three sessions for which we analyzed the video and obtained head angle measurements. Throughout this panel, the x-axis shows the magnitude of rule encoding. We quantified rule encoding in panels M, N, and O of this figure as the base-10 logarithm of the ratio of the firing rate in pitch discrimination over localization. Top: histogram of rule encoding over neurons, when considering only posture-equalized trials. Middle: same as above, but for the rest of the trials. Bottom: distribution of rule encoding on the posture-equalized trials after shuffling the block labels. Rule encoding is similarly distributed on both posture-equalized (top) and other trials (middle). Rule encoding is significantly ($p < 0.01$; shuffle test) stronger on posture-equalized trials (σ of 0.47) than would be obtained by random chance (mean σ of 0.10), as in the bottom row. See "Posture equalization by trial selection" in Supplemental Experimental Procedures for additional details.

N) Mean magnitude of rule-encoding in mPFC cells is not significantly altered on the posture-equalized trials. The mean and S.E.M. of the magnitude of rule encoding is shown for both localization- and pitch

discrimination-preferring cells (n = 8 for both groups), on posture-equalized trials (blue bars) or all other trials (orange). There is no significant ($p > 0.05$, paired Mann-Whitney U-test) difference in the mean magnitude of rule encoding, though there is a non-significant trend toward weaker rule encoding on posture-equalized trials.

O) Posture-equalization results for individual PFC cells (n=16). The rule encoding of each cell is plotted for posture-equalized trials (left) and all other trials (right) and connected with a line. The red line indicates a rule encoding of zero: no difference between blocks. In some cells the rule encoding is weaker (closer to zero) on posture-equalized trials. If head angle were the only determinant of the difference in firing rate, then we would expect rule encoding on posture-equalized trials to be around zero, randomly greater in localization or pitch discrimination with 50% probability. Instead, we see that in only one out of 16 mPFC cells does the preferred block reverse on the posture-equalized trials, versus the rest of the trials. If reversals were random, then the probability of observing such a small ratio of reversals would be 0.0005 (two-tailed binomial test).

P) Rule encoding persisted throughout reward consumption in some neurons. This panel shows example PSTHs during localization (blue trace; mean plus or minus SEM over trials) and pitch discrimination (red trace) of two example mPFC neurons that maintain rule-encoding throughout reward consumption. The traces are locked to exiting the center port at time zero. Along the bottom of the panel, distributions of the latencies to stimulus onset (black) and entering reward port (green) are shown. Only go trials are included. Reward consumption began upon entry into the reward port and continued for several seconds.

Q) Simulated posture-encoding neurons are correctly detected by the multiple-regression/ANOVA approach used to analyze the real data. For each rule-encoding mPFC neuron in (K) and (L), we designed a model neuron with a firing rate linearly related to the head angle, but not the block, on each trial. We added Gaussian noise to its firing rate to match the variance of the actual counts. Finally we analyzed each model neuron using the same multiple-regression/ANOVA approach that we used in (K) and (L). Here, the model neurons are correctly identified as primarily encoding head angle: the median FEV of head angle is 0.87 and it is above 0.5 for every neuron. Compare with the results in (K) and (L) in which few neurons were identified as primarily encoding head angle (median FEV of head angle: 0.059). Neurons are ordered in the same way as in (K). This demonstrates that the analysis is capable of detecting postural encoding where it exists. See “Verification of ANOVA on simulated posture-encoding neurons” in Supplemental Experimental Procedures for further details.

R) Estimated FEV (y-axis) of HeadAngle versus its true FEV (x-axis), using the empirical distributions for head angle and block identity and a simulated model neuron that linearly combines these two variables. Error bars represent the standard deviation over 150 runs. Black unity line shows perfect estimate. The estimate closely tracks the true value. See “Verification of ANOVA with correlated regressors” in Supplemental Experimental Procedures for further details on panels R-T.

S) Same as (R), except that head angle and block identity are drawn from a random normal and binomial distribution, respectively, subject to the constraint that the correlation between these two values is 0.72 (similar to the empirical correlation between head angle and block). The estimated FEV still closely tracks the true FEV.

T) Same as (R), except that the output of the model neuron is converted into a discrete, quantized representation more closely resembling spiking data.

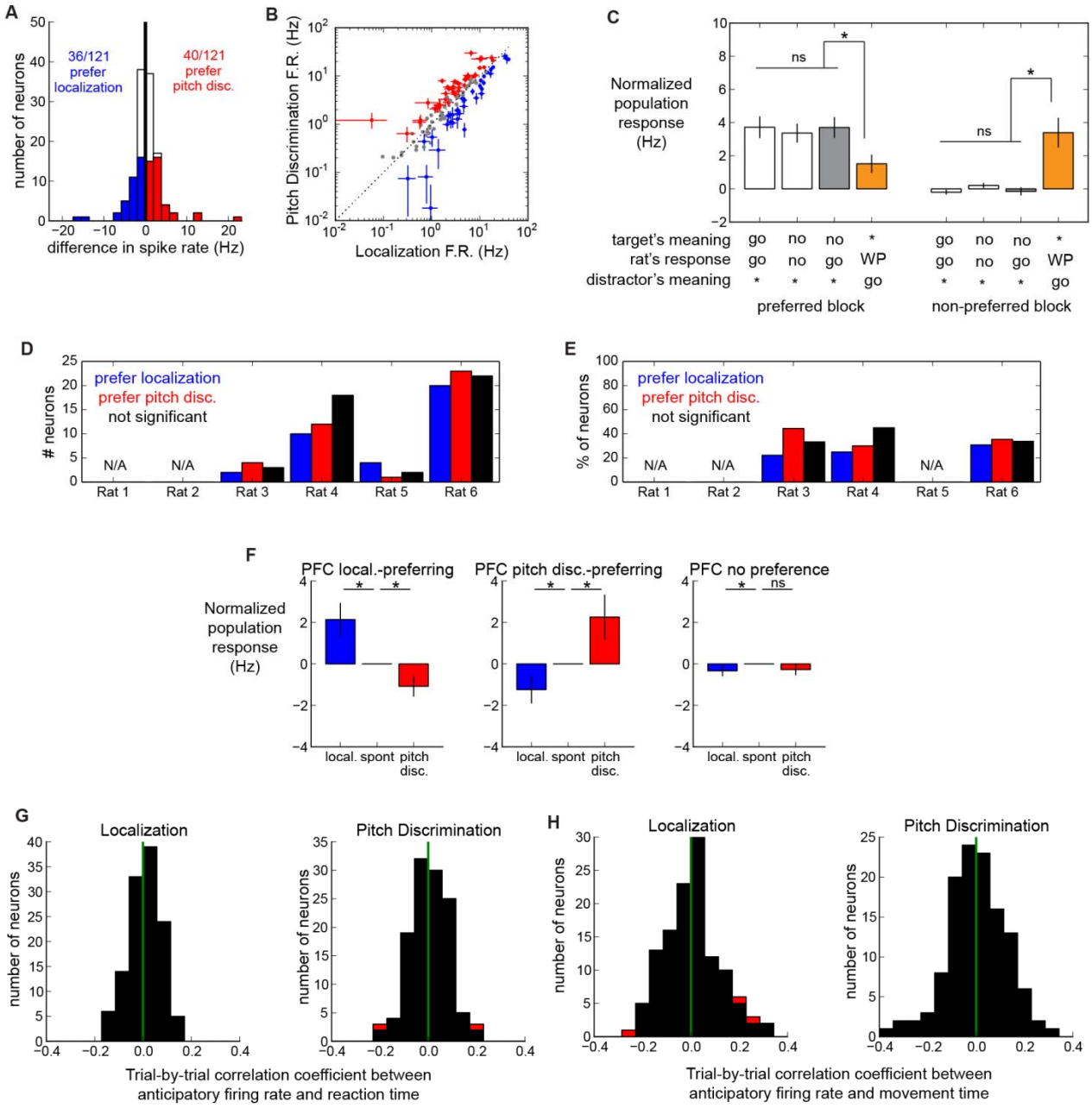

FIGURE S2 (continued on next page)

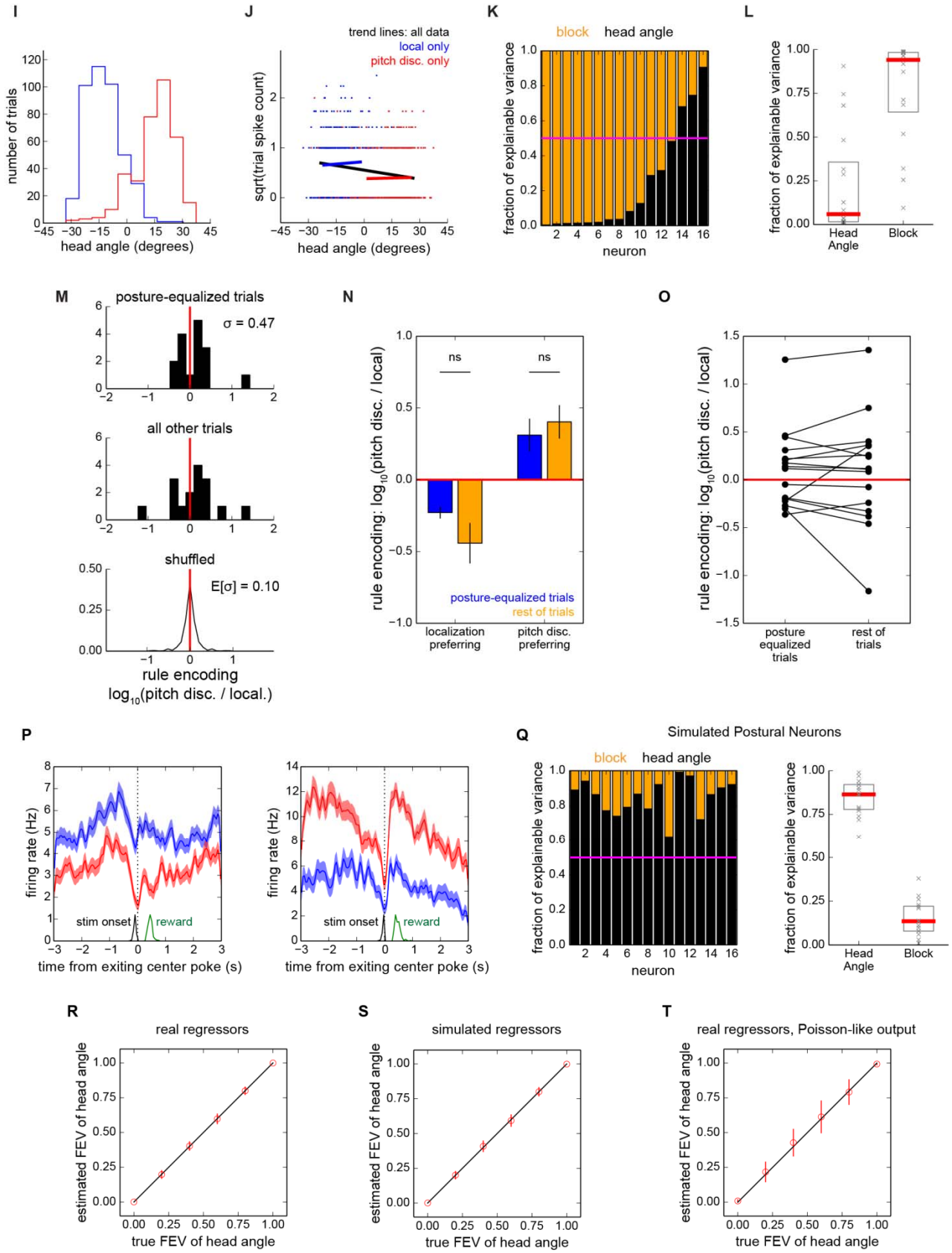

FIGURE S2 (continued from previous page)

Figure S3. More information on A1 anticipatory effect, related to Figure 4

Refer to "Analysis of waveform variation and firing rate drift" and "Power analysis" in Supplemental Experimental Procedures.

A, B) Alternative presentations of rule encoding across neurons. Same as Figure S2A and S2B, but for A1 neurons instead of mPFC neurons.

C) Analysis of rule encoding on various types of error trials. Same as Figure S2C (correct trials: white; go-on-nogo errors: gray; interference trials: orange), but for A1 neurons instead of mPFC neurons. The effects are similar to those for mPFC, but note that the effect on interference trials is stronger. For A1 neurons, the direction of hold period modulation is significantly reversed — the firing rate is higher during such trials in the non-preferred block than in the preferred block. However, there is no difference between the other trial types, regardless of whether the rat performed a go or nogo response.

D) Histogram across rats of the number of A1 neurons preferring pitch discrimination (red), localization (blue), and neither (black). In one rat, no A1 neurons were recorded (marked N/A).

E) Proportion of rule-encoding neurons in A1 is consistent across rats. The data from (D) are here expressed as a percentage of total neurons. No percentages are plotted for rats with fewer than 8 neurons total recorded in A1.

F) On average, firing rates in both blocks are bidirectionally modulated, in comparison to the spontaneous rate. We defined the spontaneous rate as the average rate during epochs more than 2 s from the nearest stimulus onset. Localization-preferring neurons (left) fire significantly more than spontaneous in localization and significantly less than spontaneous in pitch discrimination. A similar statement holds for pitch discrimination-preferring neurons (center). Neurons that do not prefer either block (right) — neurons for which the pre-stimulus firing rate does not significantly differ between blocks — showed no change in their firing rates relative to their spontaneous rates. The bars show the average and SEM across neurons of the hold period firing rates minus the spontaneous firing rate for that neuron. We assessed significance with a paired Mann-Whitney test.

G) No evidence for correlation between pre-stimulus firing rate of A1 neurons and the animal's reaction time, defined here as the time between stimulus onset and withdrawal from center port. We plot here the distribution of correlation coefficients obtained in both blocks, with red representing significantly correlated neurons ($p < 0.05$ after correction for multiple comparisons with the false discovery rate). Only a small minority of neurons showed a significant correlation, and the mean of the full distribution is not significantly different from zero ($p > 0.05$, one-sample t-test). Similar results were obtained when considering only rule-encoding neurons (data not shown).

H) Similar to (G), but here we correlate the anticipatory firing rate with the "motion time" — the time necessary for the rat to move to the choice port on successful go trials. Again, only a small minority of neurons show an individually significant correlation and the population distribution is not significantly different from zero. Similar results were obtained when considering only rule-encoding neurons (data not shown).

I) The azimuthal (left/right) angle of the rat's head during center poke, relative to the behavior box, differs by approximately 30 degrees between blocks. Red: pitch discrimination. Blue: localization. The mean head angle over all trials is defined here as zero degrees. These data are from an example session but all analyzed sessions yielded similar results. This preparatory motor activity is presumably an adaptive behavioral strategy in response to the fact that the choice port differs between blocks. See "Video analysis of preparatory head positioning" in the Supplemental Experimental Procedures for further detail on panels I-L.

J) Example A1 neuron recorded during the session shown in (I). Each point shows the square root of the number of spikes fired and head angle on a single trial. (Throughout this figure, we use the square root transformation to stabilize the variance of the counts and thus to ensure homogeneity of variance independent

of the mean to facilitate our regression analysis.) Across all trials, these variable are significantly correlated (black trend line). However, this correlation between firing rate and head angle is almost entirely explained by block type. Within each block, there is no such correlation — the red and blue trend lines are not significantly different from horizontal.

K) Fraction of explainable variance (FEV) in the spike counts (again square root-normalized) that the least-squares linear fit attributes to block (orange) and head angle (black), individually for each rule-encoding A1 neuron during the analyzed video sessions. The magenta horizontal line indicates 50%. The bars are sorted in order of increasing head angle FEV. Nearly every bar is mostly orange, indicating that block is the major explanatory factor for the spike count across the population. Only one bar is more than 50% black, corresponding to a neuron whose firing rate was mostly explained by head angle.

L) Summary plot of the data in panel (K). The distribution of FEV across neurons is plotted for each factor (head angle and block). Individual crosses represent individual rule-encoding A1 neurons. The red line shows the median; the gray box outlines the inter-quartile range. Across the population, most of the variance (median: 83.0%) is explained by block; the small amount that remains is explained by head angle (median: 17.0%).

M) Rule-encoding in A1 neurons is broadly similar to that of the full data set when only "posture-equalized" trials are considered. Here, we analyze only rule-encoding neurons (*i.e.*, those with a significant preference for one block) recorded during the three sessions for which we analyzed the video and obtained head angle measurements. Throughout this panel, the x-axis shows the magnitude of rule encoding. We quantified rule encoding in panels M, N, and O of this figure as the base-10 logarithm of the ratio of the firing rate in pitch discrimination over localization. Top: histogram of rule encoding over neurons, when considering only posture-equalized trials. Middle: same, but for the rest of the trials. Bottom: distribution of rule encoding on the posture-equalized trials after shuffling the block labels. Rule encoding is similarly distributed on posture-equalized trials (top) as it is on other trials (middle). Rule encoding is significantly ($p < 0.01$; shuffle test) stronger on posture-equalized trials (σ of 0.32) than would be obtained by random chance (mean σ of 0.09), as in the bottom row. See "Posture equalization by trial selection" in Supplemental Experimental Procedures for additional details.

N) Mean magnitude of rule-encoding in A1 cells is not significantly altered on the posture-equalized trials. The mean and S.E.M. of the magnitude of rule encoding is shown for both localization- and pitch discrimination-preferring cells (n=4 for both groups), on posture-equalized trials (blue bars) or all other trials (orange). There is no statistically significant ($p > 0.05$, paired Mann-Whitney U-test) difference in the mean magnitude of rule encoding, though there is a non-significant trend toward weaker rule encoding on posture-equalized trials.

O) Posture-equalization results for individual A1 cells (n=8). The rule encoding of each cell is plotted for posture-equalized trials (left) and all other trials (right) and connected with a line. The red line shows a rule encoding of zero: no difference between blocks. In some cells the rule encoding is weaker (closer to zero) on posture-equalized trials. If head angle were the only determinant of the difference in firing rate, then we would expect rule encoding on posture-equalized trials to be close to zero, randomly greater in localization or pitch discrimination with 50% probability. Instead, we see that in none of the 8 A1 cells does the preferred block reverse on the posture-equalized trials, as compared with the rest of the trials. If reversals were random, then the probability of observing such a small ratio of reversals would be 0.008 (two-tailed binomial test).

P) Rule encoding persisted throughout reward consumption in some neurons. This panel shows example PSTHs during localization (blue trace; mean plus or minus SEM over trials) and pitch discrimination (red trace) of two example A1 neurons that maintain rule-encoding throughout reward consumption. The traces are time-locked to exiting the center port at time zero. Along the bottom of the panel, distributions of the latencies to stimulus onset (black) and entering reward port (green) are shown. Only go trials are included. Reward consumption began upon entry into the reward port and continued for several seconds.

Q) Simulated posture-encoding neurons are correctly detected by the multiple-regression/ANOVA approach used to analyze the real data. For each rule-encoding A1 neuron in (K) and (L), we designed a model neuron with a firing rate linearly related to the head angle, but not the block, on each trial. We added Gaussian noise to its firing rate to match the variance of the actual counts. Finally we analyzed each model neuron using the same multiple-regression/ANOVA approach that we used in (K) and (L). Here, the model neurons are correctly identified as primarily encoding head angle: the median FEV of head angle is 0.88 and it is above 0.5 for every neuron. Compare with the results in (K) and (L) in which few neurons were identified as primarily encoding head angle (median FEV of head angle: 0.170). Neurons are ordered in the same way in (K). This demonstrates that the analysis is capable of detecting postural encoding, where it exists. See "Verification of ANOVA on simulated posture-encoding neurons" in Supplemental Experimental Procedures for further details.

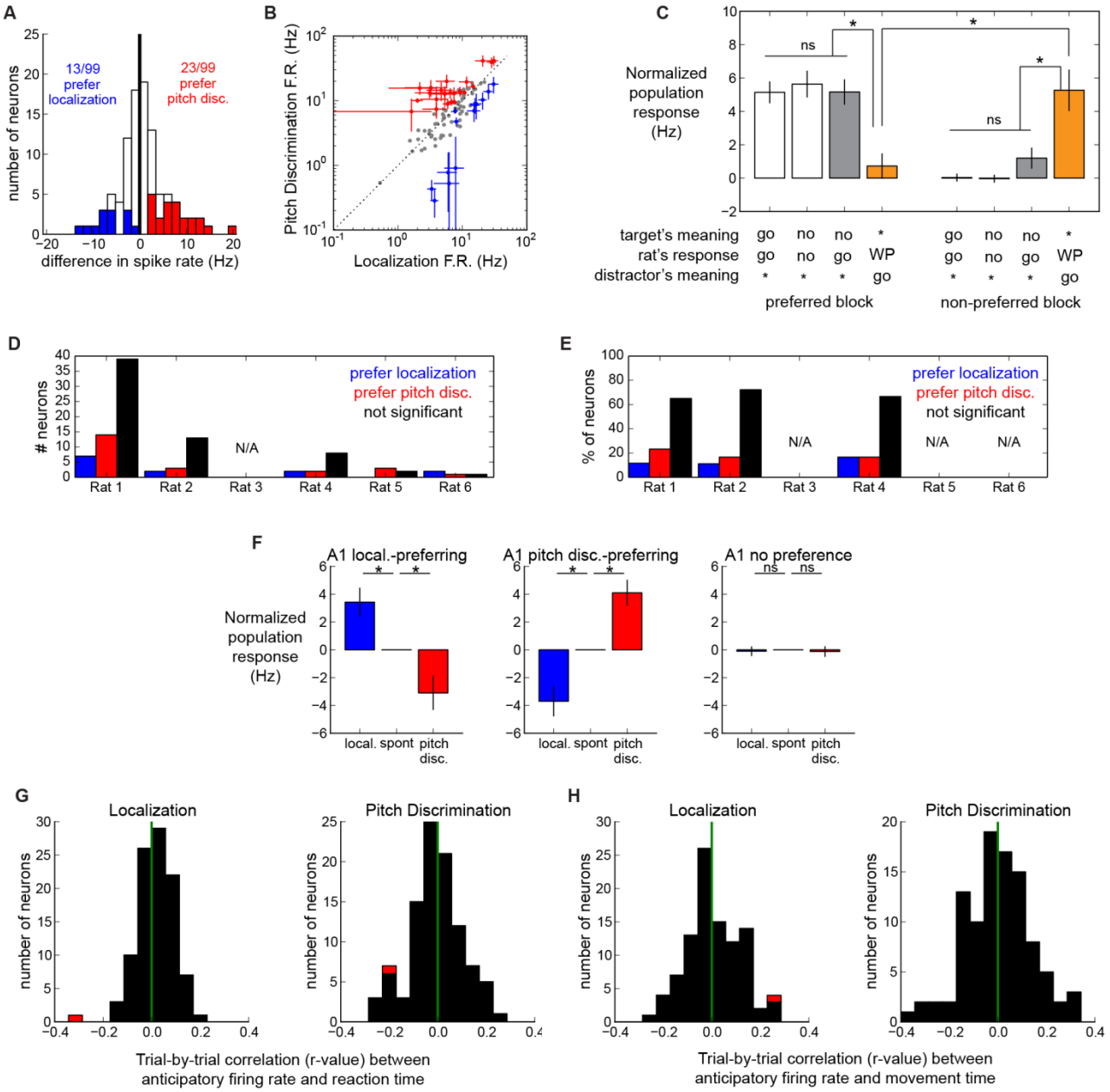

FIGURE S3 (continued on next page)

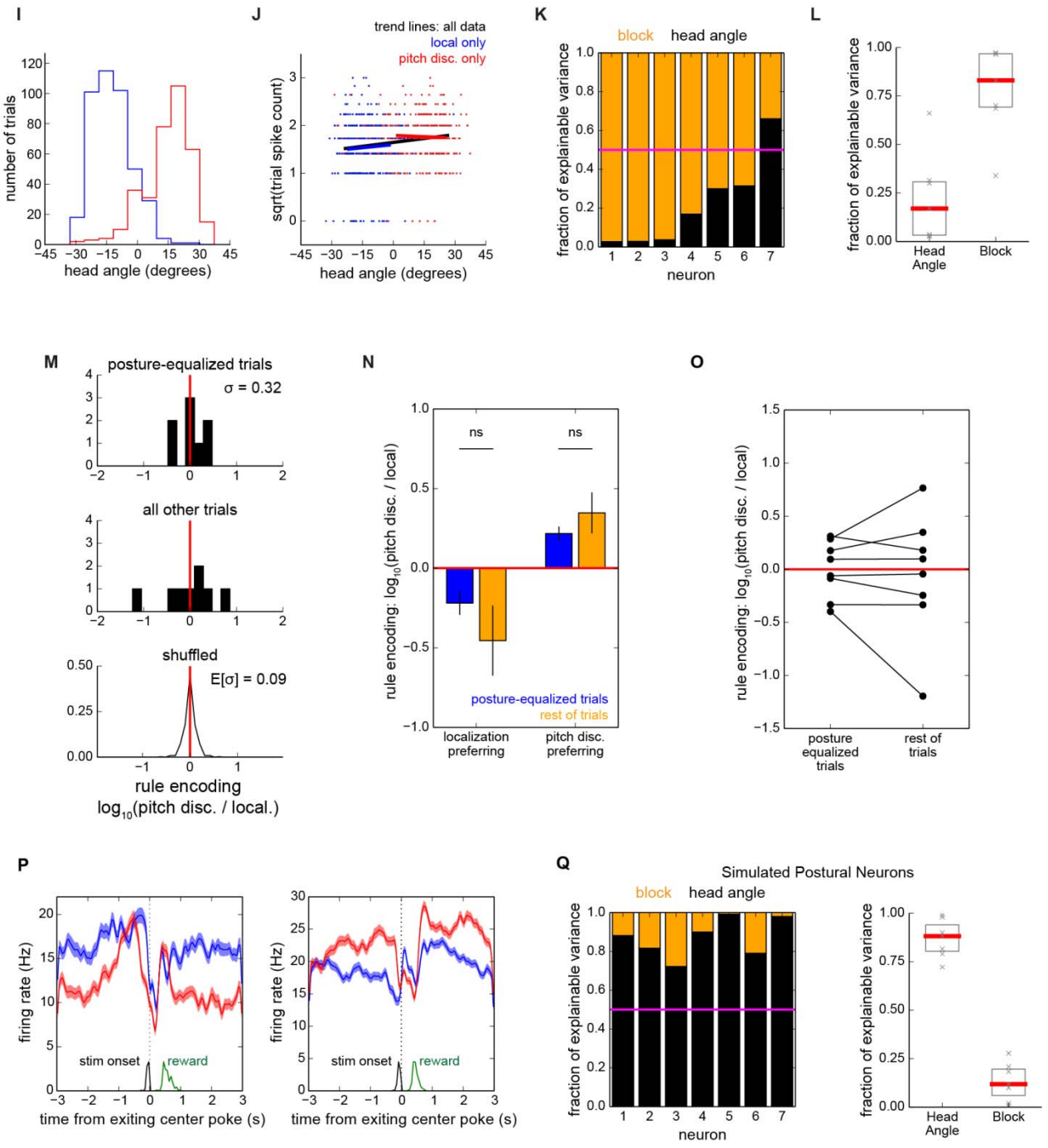

FIGURE S3 (continued from previous page)

Figure S4. More information on timecourse, related to Figure 5

Refer to “Timecourse analysis” in Supplementary Experimental Procedures for details.

A) Anticipatory modulation precedes center-poke entry. Similar to Figure 5C in the main text but with each trace time-locked to entry into the center port that initiated the trial. (In Figure 5C the data were locked to stimulus onset instead. Because these events are very close in time, this panel is very similar to Figure 5C.) This shows the population time course during the preferred block, averaged over all rule-encoding neurons in each region (purple: mPFC; orange: A1). Throughout this figure, firing rates were first normalized to zero mean and unit variance and then averaged over neurons; the thickness of the trace represents SEM over neurons. This demonstrates more clearly that the increase in anticipatory activity definitely precedes the center-poke entry.

B) Anticipatory modulation persists throughout a successful nogo trial. Population time course during the preferred block for rule-encoding neurons in mPFC (left) and A1 (right). In this panel, we include only trials *following* a successful nogo response. Along the lower edge of the figure we also plot the distribution of latencies to the end of the previous trial, defined as the withdrawal from the center port following the nogo stimulus onset. The increased activity persists between the trials, even though the rat is typically remaining motionless near the center port during this time. This demonstrates that the anticipatory effect is not due to the rat’s motion toward the center port.

C) Similar to Figure 5D in the main text, but now also showing activity in the non-preferred block (green) in addition to activity during the preferred block (purple). Also, only correct nogo trials are included to demonstrate the persistence of the effect. The population time course is locked to exit from center-port after the stimulus onset. Before exiting the center port, average activity in the non-preferred block is suppressed below baseline; average activity in the preferred block is increased above baseline. This remains true for at least several seconds after the rat exits the center port, during which time the rat typically had already initiated the next trial. To illustrate this last point, along the lower edge of the figure we also plot the distribution of latencies to the relevant trial events: stimulus onset (black) and the center-poke initiating the next trial (gray). See also Supplemental Experimental Procedures, “Rule encoding is not dependent on the motion history of the animal”.

D) Similar to panel (C), except that here we include only correct go trials. After exiting the center port, the firing rate during the preferred block (again averaged over neurons) falls below baseline and remains there as the rat moves to the choice port and drinks a reward (which always required at least several seconds). In contrast, during the non-preferred block A1 neurons are actually activated above baseline during this period, thus inverting the usual rule encoding. Along the lower edge of the figure we also plot the distribution of latencies to the relevant trial events: stimulus onset (black) and reward delivery (green). Finally, we note that these results depict the population average; individual neurons displayed a wide variety of dynamics (Figure 4).

E) Timecourse results in localization-preferring and pitch discrimination-preferring populations are similar. Similar to (C) and (D), but here we separately plot the population activity during localization (blue) and pitch discrimination (red) blocks from neurons preferring localization (left column) and pitch discrimination (right column) on correct nogo trials (top two rows) and correct go trials (bottom two rows). The dynamics of the pitch discrimination-preferring and localization-preferring neurons are quite similar (compare left and right columns), other than the fact that the red and blue traces are roughly reversed (by definition of preferred block). Along the lower edge of the figure we also plot the distributions of latencies to stimulus onset (black), reward delivery (green, go trials only), and center-poke beginning next trial (gray, nogo trials only).

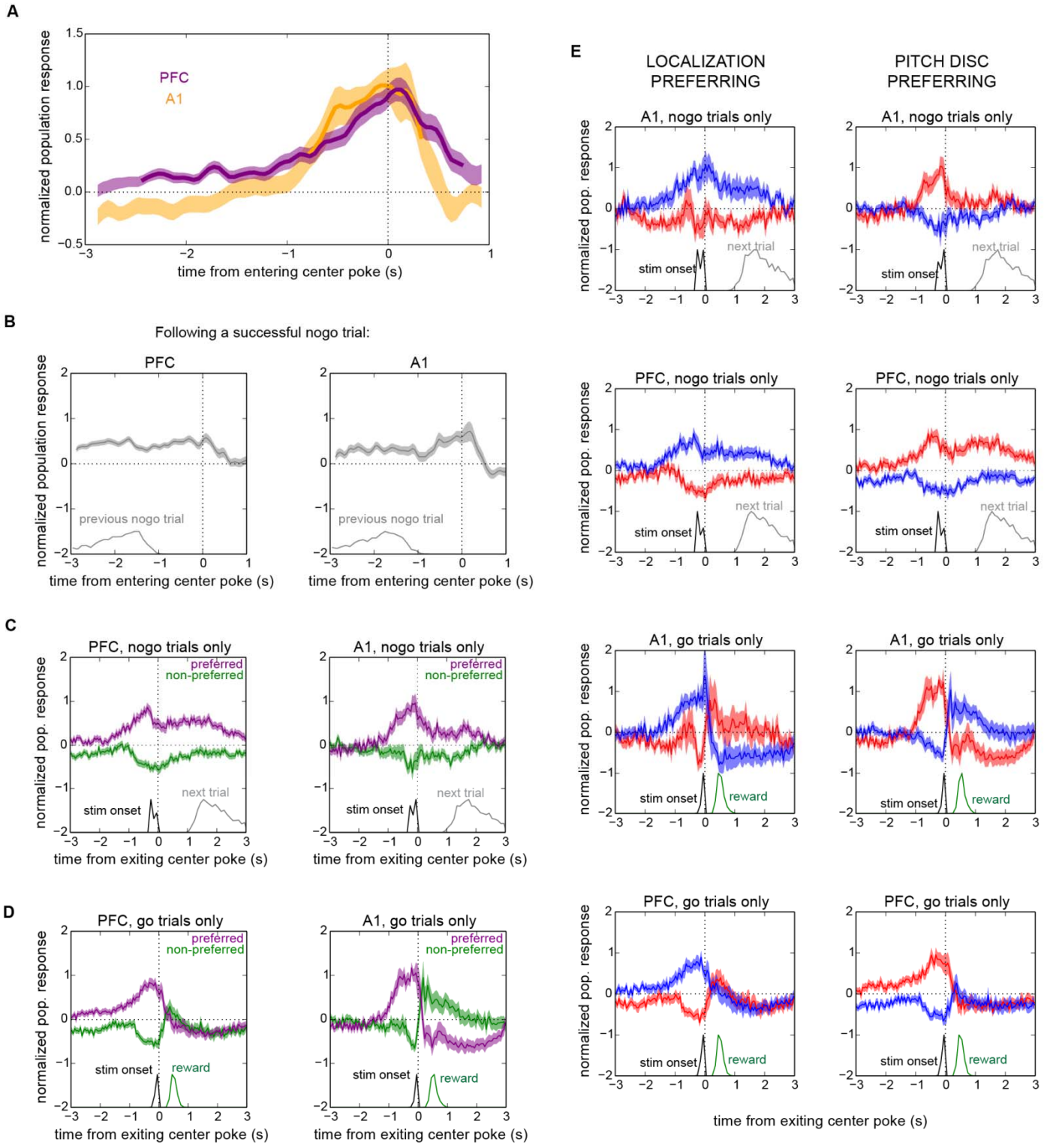
A
normalized population response
PFC
A1
time from entering center poke (s)
B
Following a successful nogo trial:
PFC
A1
previous nogo trial
C
PFC, nogo trials only
A1, nogo trials only
preferred
non-preferred
stim onset
next trial
time from exiting center poke (s)
D
PFC, go trials only
A1, go trials only
reward
E
LOCALIZATION PREFERRING
PITCH DISC PREFERRING
A1, nogo trials only
PFC, nogo trials only
A1, go trials only
PFC, go trials only
normalized pop. response

FIGURE S4

Figure S5. More information on evoked activity, related to Figure 6

Refer to “Analysis of evoked responses” in Supplemental Experimental Procedures for details.

A) PSTH of a typical auditory-responsive A1 neuron, to illustrate the notion of "evoked response strength." This neuron’s response strength is near the median of the A1 auditory-responsive population. All stimuli and trials are included in this PSTH. Note that the response to the onset of the sound is rapid, short-latency, tightly stimulus-locked, and brief. The onset window is shaded. This was defined as the continuous set of time bins post-onset over which the firing rate was significantly greater than the rate preceding the stimulus. Some mPFC units also showed similar, though typically weaker, stimulus-driven responses.

B) An example A1 neuron showing one of the strongest and most sustained responses that we recorded.

C) Distribution of onset response strengths across n=49 auditory-responsive A1 neurons (blue) and n=31 auditory-responsive PFC neurons (orange). The response strength is expressed as the average number of additional spikes (over baseline) during the onset response window for that neuron. All trials and stimuli are included, and the baseline rate is calculated over the 50 ms preceding the stimulus onset. The responses are significantly stronger in A1 neurons (median: 0.12 spikes) than in PFC neurons (median: 0.02 spikes), $p < 0.001$, unpaired Mann-Whitney U-test. Note the long tail of the distribution: a small subpopulation fires much more strongly than the median.

D) Alternative presentation of onset response strength. The data are the same as in (C), but expressed here as percentage of baseline firing rate. Because the response window is so brief, a small number of additional spikes over baseline typically represents a large (many-fold) increase in rate. Again the responses are stronger in A1 neurons (median: 209% of baseline) than in PFC neurons (median: 171% of baseline), $p < 0.001$, unpaired Mann-Whitney U-test.

E) Distribution of latencies to center of onset response window across the same populations as (C) and (D). The PFC latencies are significantly longer (median: 19.75ms) than the A1 latencies (median: 16.75ms), though only slightly. $p < 0.05$, unpaired Mann-Whitney U-test.

F) No correlation between stimulus tuning and the change in pre-stimulus rate for auditory-responsive A1 neurons. We calculated the strength of the tuning for the noise burst (LEFT versus RIGHT) or the warble (LOW versus HIGH) using the difference in firing rate to those stimuli when they were presented on cue trials. Similar to Figure 6E in the main text, we quantified tuning using the performance of an ideal decoder. This metric has a value of 0.5 for identical responses, and 1.0 for perfectly discriminable responses. We found no correlation between how well the neurons were tuned for either stimulus and the change in pre-stimulus rate across blocks. Although these results were obtained using the cue trials to measure tuning, similar results were obtained when we calculated the tuning using the responses to the stimuli containing distractors (data not shown). To obtain a robust estimate of decodability, we used only neurons for which we had at least 25 cue trials in each block and at least 10 evoked spikes total on those cue trials.

G) Same as (F), but for auditory-responsive PFC neurons. Note the difference in the scale of the x-axis vs (F), reflecting the fact that PFC neurons tend to be more poorly tuned for the stimuli than A1 neurons. Again there is no significant correlation, though there is a weak, non-significant trend in the following direction: neurons that are well-tuned for the warble tend to show increased anticipatory firing in pitch discrimination; similarly, neurons that are well-tuned for the noise burst tend to show increased anticipatory firing during localization.

H) No evidence that stimulus selection leads to a significantly greater modulation of evoked response in one stimulus pair versus the others. Across the population of auditory-responsive neurons in both brain regions, all four stimulus pairs elicit equivalently small and non-significant changes across blocks in their evoked firing rate. Each box and whiskers plot shows the distribution across neurons of the difference in evoked firing rate between pitch discrimination and localization for each stimulus pair. (We used square root as

a normalizing transform to ensure a homogeneous variance across neurons of different firing rates.) The box's top and bottom show the inter-quartile range (IQR); the tapered area of each box shows the 95% bootstrapped confidence intervals on the median. Pluses indicate outliers, defined as more than 1.25 times the IQR from the median; these points were still included in the analysis. We also repeated the analysis after subtracting the square root of the pre-stimulus rate from the square root of the evoked rate in both blocks; this is labeled "baseline-subtracted". A one-way Kruskal-Wallis test revealed no significant difference between the stimulus pairs. We also assessed whether the change to each individual stimulus pair was different from zero using the Wilcoxon signed-rank test, correcting for multiple comparisons within each group (*e.g.,* each group of four boxplots) using the false discovery rate. We found no significant difference from zero for any individual stimulus pair, indicating that there was no consistent trend toward increased evoked firing rates in either block for any stimulus pair.

I) Similar to (H), but here we took the absolute value of all of the data. We reasoned that some stimulus pairs might elicit increased firing rates during pitch discrimination in some neurons and in localization in other neurons, resulting in no net effect. Under this hypothesis, we would expect that the absolute value of the difference between blocks, across neurons, would be significantly greater for that stimulus pair versus the other stimulus pairs. However, a one-way Kruskal-Wallis test across stimulus pairs again revealed no significant difference: all stimulus pairs yield equivalent absolute differences in evoked firing rate across neurons.

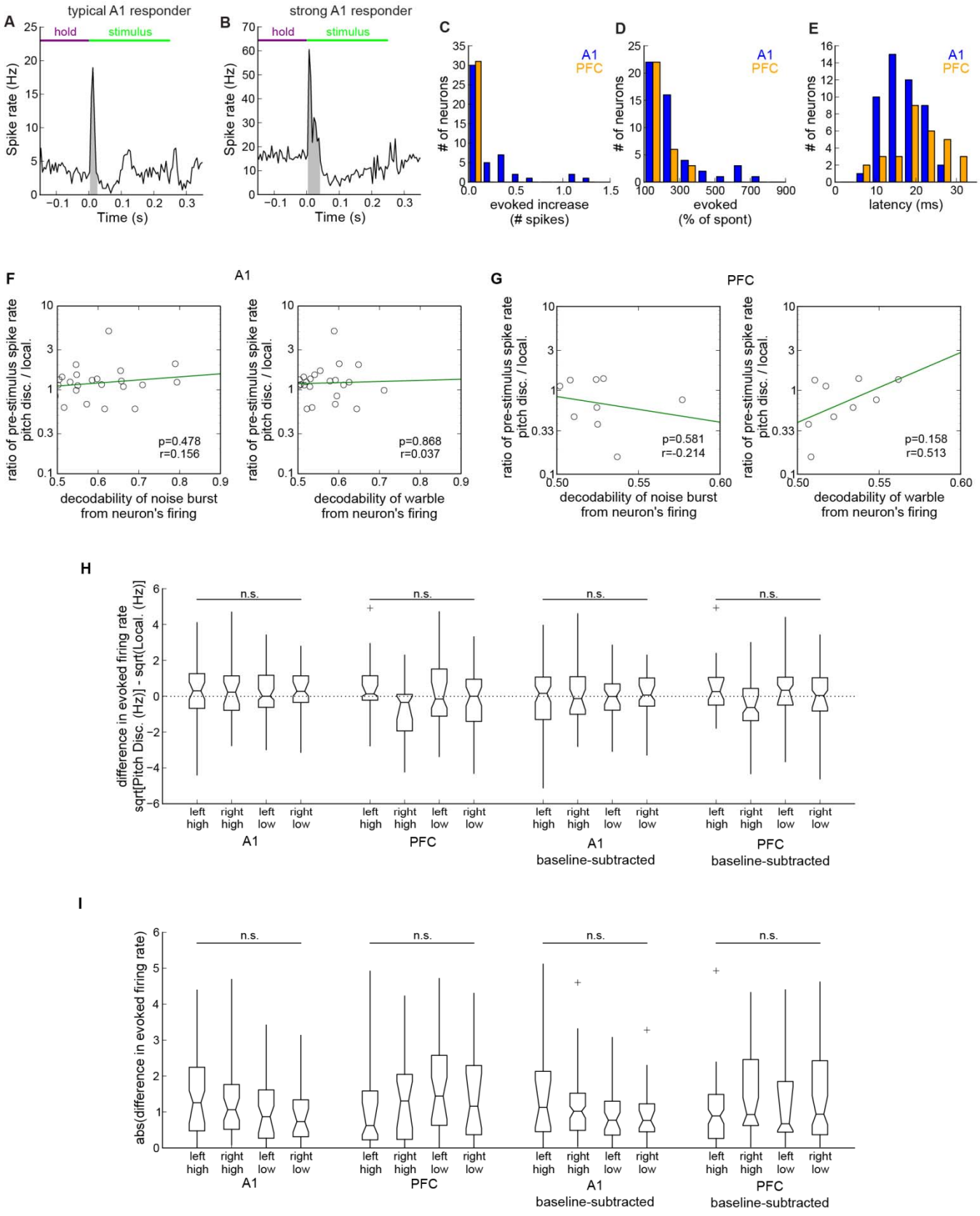
A
typical A1 responder
hold
stimulus
Spike rate (Hz)
Time (s)
B
strong A1 responder
C
# of neurons
A1
PFC
evoked increase
(# spikes)
D
evoked
(% of spont)
E
latency (ms)
F
A1
ratio of pre-stimulus spike rate
pitch disc. / local.
p=0.478
r=0.156
decodability of noise burst
from neuron's firing
p=0.868
r=0.037
decodability of warble
from neuron's firing
G
PFC
p=0.581
r=-0.214
p=0.158
r=0.513
H
difference in evoked firing rate
sqrt[Pitch Disc. (Hz)] - sqrt[Local. (Hz)]
n.s.
left high
right high
left low
right low
A1
PFC
A1
baseline-subtracted
PFC
baseline-subtracted
I
abs(difference in evoked firing rate)

FIGURE S5

Figure S6. More information on electrical disruption of mPFC, related to Figure 7

Refer to “Disruption of mPFC by electrical stimulation” in Supplemental Experimental Procedures for details.

A) Effect of electrical disruption of mPFC on performance for all individual sessions from three rats (red: Z1, yellow: Z2, green: Z3). Each line connects the performance on control trials (left) and disruption trials (right) within the same session. Plus marks represent a significant effect of disruption, which was an impairment in every case ($p < 0.05$, Fisher’s exact test). Circles represent sessions for which there was no significant effect of disruption. Trials are grouped according to go and nogo in both blocks. Asterisks indicate trial types for which the effect of disruption was significant across sessions ($p < 0.05$, sign test on the number of sessions showing impairment).

B) Impairment caused by electrical disruption by stimulus pair, averaged across sessions for each rat. Error bars: SEM across sessions. Impairment is defined as the difference between performance on control trials and disruption trials. Colors represent individual rats, following (A). Two rats (Z1 and Z2) showed a significant decrease for RIGHT+HIGH during pitch discrimination across sessions ($p < 0.05$, binomial test on the number of sessions showing impairment). One rat (Z3) showed a deficit for both nogo sounds during localization.

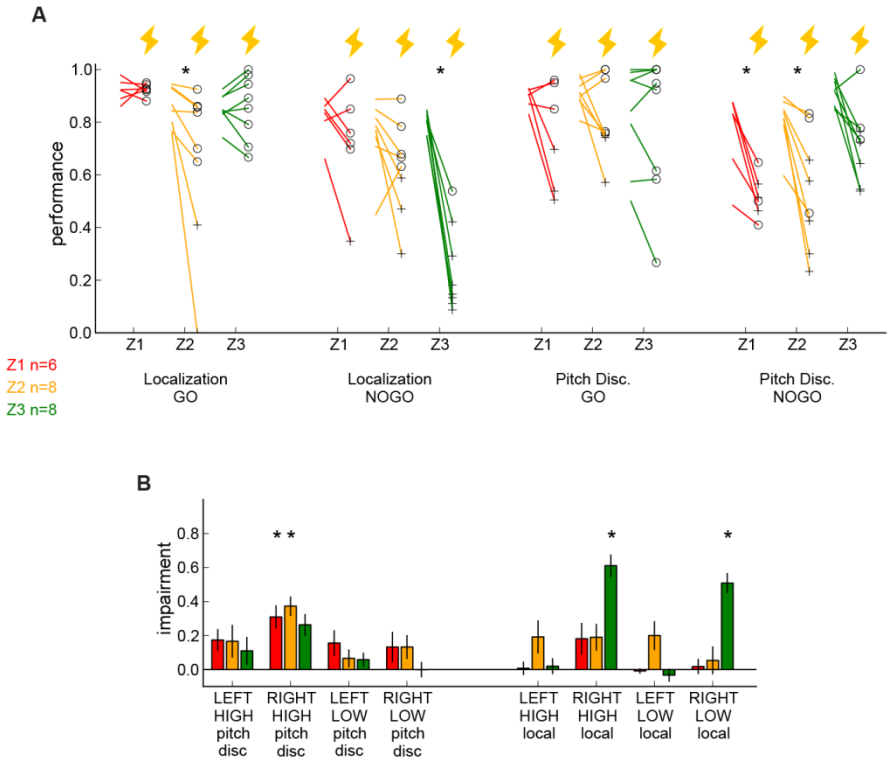

FIGURE S6

Figure S7. Behavior of the model over parameter space, related to Figure 8

Refer to “Simulated network model” in Supplemental Experimental Procedures for details.

A) The model described in Figure 8 operates well over a wide range of parameters. We quantify its performance as the strength of the task signal (x-axis in Figure 8B) necessary for the model to reach a criterion performance of 80%. As in Figure 8B, we normalize this task signal strength to the sensory noise (the variance of the Gaussian noise that was added to each A1 neuron’s activation during the simulation). This plot shows a color-coded map of the strength of the task signal necessary to reach criterion, for various values of N (x-axis) and of the sensory signal-to-noise ratio (y-axis), defined as the strength of the tuning of each neuron to the variance of the same Gaussian noise mentioned above. Darker colors are good: they indicate that a weak task signal was sufficient to reach criterion; lighter colors indicate that a stronger signal was necessary. It is desirable for the model to operate with as weak a task signal as possible, because when the task signal becomes quite strong it overwhelms the sensory input and the performance drops again. We found that larger networks invariably performed better, as expected. For a given network size (any column of the figure), the performance was relatively stable across a certain range of SNRs, but quickly degraded below a certain SNR cutoff. In the white region (small network sizes and/or very poor sensory signal-to-noise ratio), the model never reaches criterion performance.

B) Similar to panel A, but here we show the maximum magnitude of the task signal before performance falls below criterion again. Here, lighter colors (as in the upper right corner) are good, because they indicate that the network performs well over a wide range of task signal strengths; this is associated with large network sizes and/or high sensory signal-to-noise ratios.

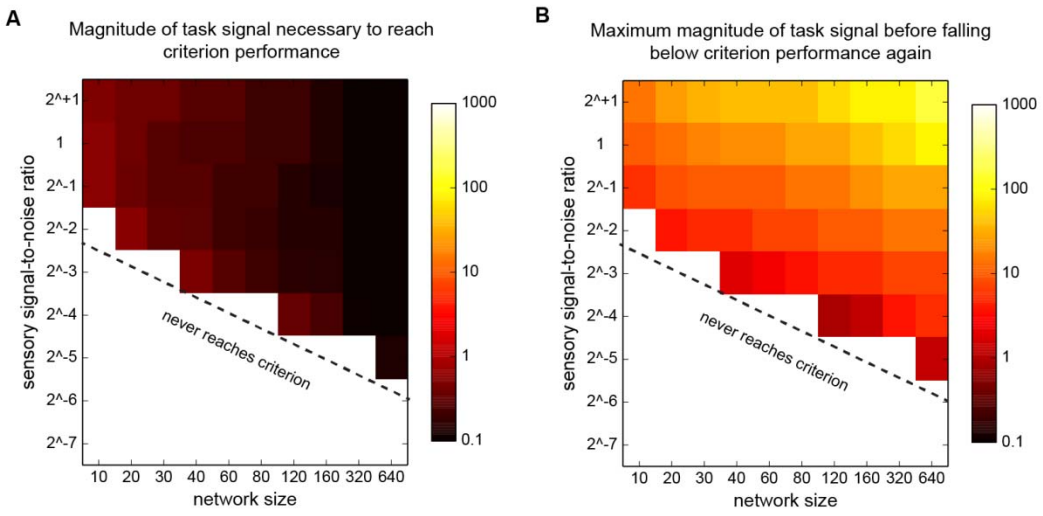

FIGURE S7

Supplemental Experimental Procedures

Behavioral training

Training procedure

We used a typical “shaping” procedure to train the rats. First they learned the localization task and pitch discrimination tasks separately and without a distractor. Next they learned to alternate between the tasks. Finally they learned to respond to the mixed stimulus containing target and distractor based on the block. Human intervention was required to determine when the rats were ready to progress to the next stage of training (generally, at least 80% hit rate). Human intervention was also required to discourage certain unwanted response strategies using the following tools: 1) increasing error timeout; 2) temporarily enforcing “all GO” or “all NOGO” trials; 3) giving water rewards out of the left or right port even in the absence of good performance in order to maintain motivation or encourage a task switch. Once the rats were sufficiently well-trained that little or no human intervention was required, they were implanted with the drive. Some rats required “retraining” after implantation using the techniques listed above; any trials thus affected were discarded from analysis. The entire training process takes about 10 weeks.

Chance performance on the task

In order for the rat to perform significantly above chance within a session, its behavior had to satisfy three criteria: 1) the rat performed significantly above 50% in each block, meaning that it must be using some information from the target sound (which is the only source of information on the correct response) to decide whether to go or nogo; 2) the rat is significantly more likely to perform the action indicated by the target than the action indicated by the distractor; 3) the rat is not using a “fixed strategy”, that is, the same mapping from stimulus pair to behavioral response in each block. (Because the target and distractor swap roles in each block, satisfying the second criterion is sufficient to satisfy the third.)

The first criterion rules out strategies like “always go left during localization”, which was common while first learning the task. We used a binomial test to compare the proportion of hits to 0.5 in both blocks and discarded any sessions that were not significantly ($p > 0.05$) above 50% in either block. The second criterion rules out certain hypothetical strategies such as always getting the congruent-nogo stimulus (RIGHT+HIGH) correct, and otherwise guessing randomly between the correct choices for that stimulus pair in each block. This fixed strategy yields 62.5% in both blocks but it uses information equally from both target and distractor; thus, it fails the second criterion. To test this, we used a paired Mann-Whitney U-test to compare whether the action on each trial was more likely to be correct for the target or correct for the distractor. In practice, none of the rats actually adopted such a hypothetical strategy: although some sessions failed the first criterion and were discarded ($p > 0.05$ for 4/55 sessions), no sessions failed the second criterion ($p < .005$ for all sessions). Therefore the first criterion (performance above 50% in both blocks) is actually the most relevant, and we mark the chance level on the plots as 50%.

Trial timings

In three rats (Rats 1-3) the hold period was drawn from a uniform distribution on 0-100 ms; after pilot results indicated pre-stimulus effects, the hold period duration was increased to 250-350 ms in the other three rats. All trials with a hold period <50 ms were discarded for the analyses in Figure 3 and Figure 4. Hold period response was counted in the minimum window that applied to all trials: 50 ms for the first 3 rats and 250 ms for the rest.

The duration of the choice period differed between sessions, but was fixed within a session (or if it was changed slightly within a session, then the trials before the change were discarded from analysis). Correct entries into the choice port on go trials were rewarded with water from the same port. Incorrect entries into the choice port on nogo trials results in a 2-6 s timeout. Correct nogo responses were not explicitly rewarded with water, although the rat avoided a timeout with this response. Poking neither port on a go trial was not explicitly punished with a timeout, other than a lost opportunity for reward.

Stimulus parameters

Each stimulus was 250 ms in duration. The warbles were frequency-modulated tones, centered at 6 KHz (LOW) and 16 KHz (HIGH) with a 10 Hz modulation frequency of amplitude 0.07 octaves, and presented with equal intensities (65 dB SPL) from both speakers. The white noise bursts were of approximately equal power at all frequencies between 5KHz and 50KHz, decaying rapidly outside of this range; the total acoustic power over this range was 55 dB SPL, delivered from only one speaker at a time (LEFT or RIGHT). We used a Lynx L22 sound card to convert digital signals to analog voltages, and Fostex FT17H tweeters to produce the sound.

In a subset of sessions, we presented task-irrelevant natural sounds during epochs when the rat was not initiating trials for the purpose of probing receptive fields. These probe sounds were terminated as soon as the rat began performing the task again. We did not observe any correlation between the neural results presented in this study and the receptive fields estimated from responses to these natural sounds.

Microdrive construction and implantation

Construction of recording electrodes

We constructed tetrodes by cutting lengths of 12.5 micron nichrome wire coated with partially annealed polyimide insulation (Kanthal Palm Coast), twisting them, and heating with a heat gun until the 4 individual strands melted together. The tetrodes were then routed through polyimide guide tubes and glued to the moveable plastic tab within a potentiometer. We pinned the individual wires using gold pins (Neuralynx) into a custom printed circuit board (custompcb.com, beta-layout.com) that we designed to connect each channel to a Mini DisplayPort connector (circuitassembly.com). To reduce the Johnson-Nyquist noise at the electrochemical interface, which scales with the square root of impedance, we gold-plated the electrode wires before implantation to 0.3 megaohm using a nanoZ (White Matter LLC). For animals with dual implants, we built separate drives to target A1 and PFC. These were connected with a custom-designed adapter to a 32-channel preamplifier/headstage (Triangle BioSystems International).

Surgical implantation

Anesthesia was induced with ketamine/xylazine and maintained with isoflurane in oxygen; depth was monitored using withdrawal reflex. Aseptic technique was maintained throughout the surgical procedure and the electrodes themselves were disinfected before implantation. Skin and fascia were folded back from the midline. Titanium screws (Small Parts) were inserted into each cranial bone. Two additional stainless steel screws, inserted into the left and right sides of the occipital plate above the cerebellum, served as separate reference and ground signals for the recordings. N-3 flux (LA-CO Industries Inc.) was used to solder wires onto these screws before implantation; subsequently, these wires were soldered to the reference and ground inputs on the microdrive. Craniotomies were performed directly dorsal to the target areas (A1: 5.25 mm posterior and 6.5 mm left from bregma; prelimbic region of mPFC: 3.0 mm anterior and 1.0 mm left from bregma). The dura was removed and the electrodes gradually inserted into each region. The craniotomy and electrodes were covered with 1.5% agarose. Methyl methacrylate (Teet's, Henry Schein) was used to affix the microdrive to the skull and screws. Rats were treated with analgesics (buprenorphine and meloxicam) both pre-operatively and for 2-3 days post-operatively. Once the rat had fully recovered, the behavioral task was resumed, now concurrent with electrophysiological recording.

See "Disruption of mPFC by electrical stimulation" for further detail on the stimulation drives.

Recording and signal processing

The electrodes were lowered by approximately 100-200 microns before most recording sessions by turning the potentiometer's screw. Before recording, we waited 30 minutes to allow the tetrodes to fully adjust. Broadband data were acquired at 30KHz and digitized and stored using a neural signal processor from Blackrock Microsystems. After the behavior, white noise bursts were presented passively to the animal to detect local field potential (LFP) and/or multi-unit auditory responses. Strong, low-latency auditory responses indicated that the electrodes were in A1 (in combination with the stereotactic coordinates used during implantation and, when possible, post-mortem histological reconstruction of electrode tracks). We only considered sessions in which we believed the electrodes to be in the correct brain regions.

We filtered the data offline to separate LFP (<200Hz) and spikes (>3 KHz). We used non-causal (temporally symmetric) filters to ensure that no phase distortion occurred. We used a detection threshold of 4.5 sigma (calculated using the more robust median absolute deviation) and a short window of 0.8 ms in order to minimize collisions between detected spikes. We extracted spike waveforms using our contributions to the open-source OpenElectrophy software suite (Garcia and Fourcaud-Trocmé, 2009) built on the Neo object model (Garcia et al., 2014), reduced the dimensionality with principle component analysis, clustered with KlustaKwik (Kadir et al., 2013), and manually reclustered as necessary with Klusters (Hazan et al., 2006) while blind to the experimental variables. Single units were identified based on the existence of a refractory period and minimal cluster overlap with other putative single units or noise.

Analysis of waveform variation and firing rate drift

We first applied standard spike-sorting procedures. Well-isolated single units should have all or most of the following characteristics: large waveforms, a well-separated cluster in feature space, a strong dip in the autocorrelation near time zero representing a refractory period, no dip in cross-correlation with other clusters (which would suggest a split unit), and stable features over the recording session. In addition, we also performed an extra analysis to check the quality of our data and to ensure that the hold period effect could not be due to sorting errors arising from small variations in spike waveform shape between blocks.

For each neuron, we identified the sub-cluster of sorted spikes that occurred during the localization hold period, and calculated the Mahalanobis distance (in the first four PCA feature dimensions) between this sub-cluster and the full cluster consisting of all spikes from the same unit. We assessed the significance of this "sub-cluster distance" by randomly permuting the labels on the subcluster and the full cluster 2000 times, and calculating the probability of observing a distance less than or equal to the true distance. We repeated this analysis for the other block (pitch discrimination). We discarded the neuron from analysis entirely if the subcluster in either block was significantly more separated from the full cluster than the permuted subclusters were ($p < 0.05$, permutation test).

In addition to the above statistical significance criterion, we also derived an effect size criterion. To derive the maximum acceptable sub-cluster distance, we simulated Gaussian sub-clusters of the same size as the actual sub-clusters and thus derived the expected distribution of mean-squared distance from the cluster center under the assumption that the sub-cluster was not different from the full cluster. We took the 95th percentile of this distribution of the maximum acceptable subcluster distance, and rejected any neuron whose sub-cluster distance exceeded this value.

Additionally, we also considered the possibility that a spurious hold period effect could arise from a slow increase or decrease in firing rate over the entire session, perhaps due to drift or motivation, even though the multiple switches between blocks within each session made such a possibility unlikely. We reasoned that, if this were true, then when taking block number into account the difference between block types should no longer be significant. We fit a linear model to the square root of the spike count in the hold period on each trial, using both block type (localization or pitch discrimination) and block number (1, 2, 3, …) as predictors. (Square root is a variance-stabilizing transform for Poisson counts.) We assessed the significance of each predictor with ANOVA. Any neuron that showed a hold period effect according to the analysis described in the text, but that failed to show a significant effect of block type or failed the overall F-test ($p > 0.05$) was discarded from the analysis. 8/231 neurons (combined across brain regions) were discarded for this reason.

For this ANOVA we used the type-III sum of squares. For all ANOVA analyses in this study, we avoided using type-I sum of squares because we found it to be much more sensitive to unequal trial counts (*e.g.*, more hits in one block than in the other).

Power analysis

We analyzed the statistical power of our method of detecting rule-encoding neurons (unpaired Mann-Whitney U-test on the spike counts across blocks) on simulated Poisson counts. We determined the total spike count in the hold period of all trials combined had to be at least 20 spikes to detect a change between blocks; therefore, neurons with fewer total spikes than this were discarded from our analysis. For the typical trial counts in our dataset and for a neuron with this minimum allowable firing rate, we would not be able to detect any increase that was less than a doubling of the firing rate (though the method becomes much more sensitive at higher firing rates). For this reason, selection rule encoding could be even more common than we have shown. Also, due to the fact that some of our A1 data was collected in animals for which the hold periods were shorter and the trial counts lower, we have less statistical power in that portion of the dataset, complicating a direct comparison between brain regions.

Timecourse analysis

For Figures 5A and 5B, we analyzed each neuron individually (as described in the text) to derive the time interval over which it significantly encoded selection rule. We included all correct trials in each block in the plotted timecourses.

For Figures 5C and 5D, as well as all of Figure S4, we analyzed the population activity averaged over neurons. We calculated this in the following way. First we calculated the PSTH of each neuron in response to each stimulus pair in each block. Second, we normalized these rates for each neuron by subtracting the mean and dividing by the standard deviation across all timepoints and stimulus pairs. Third, we grouped stimulus pairs as described in each panel (e.g., grouping all go stimuli together) and averaged within these groups. Finally, we averaged these group averages across neurons. This calculation ensures that each neuron is weighted equally in the population, and that each stimulus pair is weighted equally within its group.

Rule encoding is not dependent on the motion history of the animal

We asked whether the effects could be explained by the motion history of the animal preceding center-poke entry. We reasoned that the rat's movement history before each trial would strongly correlate with the choice on the previous trial. In particular, if the rat had just completed a successful nogo trial, it was likely to have remained relatively motionless in the center port. When we analyzed only trials following a successful nogo we found that the firing rate was still elevated during the preferred block and suppressed during the non-preferred block (Figure S4B). This demonstrates that the rule encoding cannot be purely an effect of the motion to the center port.

Analysis of evoked responses

Calculation of evoked responses

For each neuron, the spike times on each trial were smoothed with a Gaussian kernel with 1 ms standard deviation. For every 0.5 ms time bin after stimulus onset, the distribution of smoothed spike counts was compared to the combined distribution of all 0.5ms time bins in the 50 ms preceding stimulus onset with an unpaired Mann-Whitney U test. The first window of contiguous time bins that were all significantly greater than the spontaneous rates was defined as the onset response window. Windows of less than 1 ms were discarded because these neurons emitted far too few spikes to analyze statistically. A few neurons showing atypical auditory onset responses (*e.g.*, 4/108 showed a slow build rather than a short-latency peak) were discarded from the evoked-activity analyses because we were concerned that their activity might be driven by the decision rather than the stimulus.

Changes in magnitude of evoked response not explained by changes in baseline

We asked whether any individual neurons showed block-specific changes in evoked response, above and beyond what could be explained by pre-stimulus changes of baseline. To determine this, we first subtracted the block-specific baseline firing rate from the evoked response on each trial. We then used bootstrapping to draw trials equally (with replacement) from each stimulus, separately for each block. This procedure accounts for differences in the proportion of each stimulus type across blocks, arising from random chance or, since only

correct trials were included in this analysis, from better performance on some stimuli than others (e.g., better performance on go than on nogo, see also Figure S1A). Finally, we compared the distribution of bootstrapped spike counts across blocks for each neuron.

We found that a small population (6/43, or 14%) of A1 neurons increased their evoked response significantly ($p < 0.05$ from the overlap of the bootstrapped distributions in each block), above and beyond any baseline changes. However, unlike the populations contributing to the other results in the paper, these neurons were predominantly (4/6) observed in a single animal (Rat 1). This was the rat that had the most difficulty with pitch discrimination, and in these neurons the firing rate was higher during pitch discrimination. One possibility is that the greater difficulty this rat had with one block led to this block-specific increased in evoked rate; our other rats were more evenly matched in performance between blocks.

We also repeated these analyses without the use of bootstrapping using 1) an ANOVA using stimulus and block as factors, and also 2) directly comparing across blocks the number of spikes emitted in response to each stimulus individually. These analyses yielded similar results (data not shown).

Decoder analysis

We trained an ideal decoder on the evoked spike count, including baseline, on each trial. We implemented this decoder using the LogisticRegression object in scikits-learn and assessed its performance as the proportion of trials on which the identity of the noise burst or warble was correctly predicted, averaged over each stimulus. Similar results were obtained when we used a leave-one-out cross-validation scheme and when running the decoder on individual neurons instead of ensembles of neurons.

Disruption of mPFC by electrical stimulation

Stimulation hardware and parameters

In subject Z1 we implanted a pair of stimulating electrodes (FHC Inc., Model MX211FP-CR1, platinum-iridium, 240 um inter-electrode spacing, 400 kilohm impedance, rounded extra blunt taper) in each hemisphere centered near the dorsal portion of the prelimbic region, for a total of four electrodes. In subjects Z2 and Z3 we implanted an array of three electrodes (MicroProbes microelectrode array, 70/30% PtIr, 100 kilohm target impedance) in each hemisphere spanning the anterior-posterior and dorsal-ventral extent of the prelimbic region, for a total of six electrodes, each 1.4-1.5 mm from its nearest neighbor.

We estimate the exposed surface area of each MicroProbes electrode (used in Z2 and Z3) as a cone of radius 4.5 um and height 32.5um, for a total of 463 square microns. Based on impedances, we expect the exposed area of each FHC electrode (used in Z1) to be comparable; however, FHC was unable to provide a confident estimate of this parameter.

Using an isolated pulse stimulator (A-M Systems Inc., Model 2100), we injected a train of 1 ms current pulses at 10 Hz into mPFC on a subset of trials ("disruption trials") and only during the center-poke hold and stimulus presentation. Stimulation was bipolar, first a 500 us positive and then a 500 us negative pulse of equal amplitude. Two rats (Z1 and Z3) were also implanted with recording drives in auditory cortex. Stimulation was always delivered in parallel to all channels at once; a stainless steel, cranial ground screw provided the return path. Typically this ground screw was distinct from any recording ground or reference screws.

Effect of disruption and comparison with other studies

Differences in preparation (*i.e.*, tip geometry, electrode material, stimulus delivery equipment, brain region) and uncertainty about the exact biophysical mechanisms underlying electrical microstimulation make it difficult to estimate the volume of disrupted tissue with certainty. In a previous study, simultaneous calcium imaging of 200us 25uA pulses in rodent V1 (Histed et al., 2009) primarily affected tissue within 50 microns; however, effects were visible throughout the imaging plane (>250um). However, there is no consensus on whether the important parameter for determining the stimulated volume is current (Histed et al., 2009), total charge transfer (Butovas et al., 2003), or charge density per unit area of electrode (Logothetis et al., 2010). Nonetheless, it

remains clear that microstimulation affects primarily tissue near the electrode tip and secondarily tissue at larger (millimeter-scale) distances, perhaps by activating axons.

Our goal was to disrupt mPFC, not necessarily to silence or activate it. In fact, microstimulation produces short-latency activation, followed by a long-lasting (hundreds of milliseconds) suppression of neural activity (Butovas et al., 2003). The relative magnitude of the activation and suppression depends in large part on the stimulation frequency. Stimulation in the 5Hz – 40Hz range produces synchronized pulses of neuronal activation on a background of strong suppression; stimulation at 10 Hz in particular produces a complex interaction of activation and suppression (Butovas et al., 2003). Moreover, experiments in the primate visual system (Logothetis et al., 2010) using much higher currents but a larger surface area electrode (250uA, 200us, 6000 square microns) determined 10 Hz to be an optimal frequency with which to stimulate the visual thalamus in order to maximally suppress the corresponding retinotopic location in V1 as well as its downstream projection targets. Therefore, in order to maximize the disruption induced by our stimulation, we chose a 10 Hz stimulation frequency in our experiments.

Note also that nearby neurons in mPFC are almost certainly less homogeneous in their functions and coding properties than nearby neurons in many visual sensory areas, such as MT. Thus, even if we used a stimulation protocol that strongly activated a large number of nearby neurons in mPFC, it would most likely still result in disruption of mPFC function, rather than produce the sort of coherent signal that has been used to bias decisions in primates (Salzman et al., 1992), for example.

Behavioral results

We sometimes used the same stimulation protocol during epochs in between trials when the rats were not behaviorally engaged in the task. We never noticed any overt behavioral response to stimulation under these conditions. Besides the impairment described in the main text, the only additional effect we observed during behavior was that rats appeared to have more trouble completing the center poke. The sound does not play until the rat holds for a random duration between 250ms and 350ms. Shorter ("failed") center-pokes do not initiate a trial. Consistent with the proposed role of the mPFC in behavioral inhibition (impulse control) and in estimating temporal duration, we noticed that rats exhibited more failed center-pokes during disruption (data not shown), especially at higher current levels. Typically they did not go to the choice port after a failed center-poke (that is, the stimulation did not directly elicit a choice motion); they simply repeated the center-poke until successfully initiating a trial.

For some sessions, the disruption caused a significant increase in the number of "wrong-port" responses — trials in which the rat went to the choice port associated with the other block (data not shown). This suggested a possible specific deficit in stimulus selection, or in the memory of the rules for the current block, but the effects were insufficiently consistent to draw firm conclusions. We only included sessions for which the performance on control trials was significantly above chance, using the same definitions of chance performance as we did previously for the non-stimulated animals (see: "Chance performance on the task" above).

We observed that rats appeared to be particularly impaired on the "congruent" nogo stimulus RIGHT+HIGH in one or both blocks (Figure S6B). This is interesting because this stimulus should, in theory, be the least ambiguous stimulus of all: it always requires a nogo response in either block. For this reason, there was no significant increase in the proportion of trials on which the rat gave the response that would have been appropriate in the other block, as one might have expected were stimulus selection the only cognitive ability that was affected. Future experiments will be needed to disentangle the role of the mPFC in holding the center port, interpreting the stimulus, and producing the correct motor act.

Simulated network model

In our model, simultaneously presented pairs of stimuli produce activity patterns in a population of randomly tuned A1 neurons. The A1 neurons are divided arbitrarily into two populations. The first population is used for task 1 (i.e., localization) and projects to two downstream command neurons representing the possible

responses in that block (“go left” and “nogo”). Similarly, the second population projects to two different command neurons representing the possible response in the other block (“go right” and “nogo”). This is described by the following equations:

$$S * STRF_1 * H_1 = B_1$$
$$S * STRF_2 * H_2 = B_2.$$

S: “Stimulus”. Binary matrix with shape (N_{trials}, 4) encoding the presence of two out of four possible sounds (left, right, low, or high) for each stimulus pair in N_{trials} test trials.

$STRF_1$ and $STRF_2$: “Spectrotemporal receptive field” or “receptive field” matrices with shape (4, $N_{neurons}$) encoding the fixed tuning for each of the four sounds of each neuron in each subpopulation. These are set randomly at the beginning of the simulation and subsequently held constant.

H_1 and H_2: “Projection” matrices with shape ($N_{neurons}$, 2) representing the projection from each subpopulation to the two command neurons used by that subpopulation. During training, optimal values are chosen for these matrices.

B_1 and B_2: “Output”. Binary matrices with shape (N_{trials}, 2) representing the correct command neuron activation (i.e., correct motor response) for each stimulus pair in the test set.

During training, we choose random stimulus pairs for the test trials and we fix the values of B_1 and B_2 to be the correct output activations for tasks 1 and 2, respectively. We next solve for the projection matrices H_1 and H_2 that produce the minimum squared residual, using a non-negative least squares algorithm (scipy.optimize.nnls) that produces only positive (excitatory) weights. Note that each subpopulation is essentially trained separately — the first subpopulation produces the correct responses (B_1) for task 1; the second subpopulation produces the correct response (B_2) for task 2.

During the test phase, we generate a new set of test stimulus pairs but we keep $STRF_1$ and $STRF_2$ fixed at the values chosen during training. The entire network’s choice on each test trial is defined as the command neuron with maximal activation (“winner-take-all”). The performance of the model is assessed as the percentage of the time that the network’s choice matches the correct response for that test stimulus.

Critically, during the test phase we add a “task signal” to one of the subpopulations to bias the entire network towards the task encoded by that subpopulation. Specifically, during task 1, we calculate the command neurons’ activations as follows:

$$(S * STRF_1 + T) * H_1 = B_1$$
$$(S * STRF_2) * H_2 = B_2.$$

Here, T is a matrix with dimensions (N_{trials}, $N_{neurons}$) in which every element is the same, positive value. This corresponds to adding a small positive amount on every trial to the activation of every neuron in subpopulation 1. Because H_1 is strictly positive, this translates into a positive increase in the activation of both command neurons in B_1. (During task 2, we perform the opposite manipulation by adding T to the second subpopulation. For simplicity, this is represented as a negative task signal in Figure 8 of the main text.)

For a given SNR, as the task signal increases in strength, the model’s performance eventually drops to 50%. This is because the size of the task signal begins to dominate the sensory input (the “overdriven” regime). In this regime, the model still produces responses that are appropriate to the block (*i.e.*, it does not “go to the wrong port”) but the responses are no longer related to the sensory input. In terms of the equations above, this occurs because both command neurons encoded by B_1 are increasingly activated in task 1; however, only one of these two command neurons actually encodes the correct response. When the task signal is sufficiently strong, the task-specific activation of the neurons in B_1 swamps the stimulus-specific activation of the correct

neuron, and the model chooses at random between the two neurons in B_1 (the two possible responses during that block.)

When the task signal corresponding to task 1 is activated, but the model is assessed on its performance on task 2, it performs poorly (as expected). These data are shown in Figure 8B of the main text: negative values of the task signal correspond to activation of the "wrong" network. This corresponds to doing localization instead of pitch discrimination in the real task; in such a case, only 25% of trials will be correct: those which present the RIGHT+HIGH stimulus which always means nogo.

For low network sizes and/or at very low SNRs (*e.g.*, SNR <1% at N=640), the model reached the overdriven regime before it ever produced good performance. We measured the minimum and maximum values of the task switch signal that produced good performance (>80%) over a range of network sizes and SNRs. Increasing the size of the network increases the working range of the model by 1) decreasing the minimum task signal that is necessary to reach criterion performance, and 2) increasing the maximum acceptable task signal before reaching the overdriven regime (Figure S7).

Video analysis of preparatory head positioning

Methods

We recorded video of all behavioral sessions using an infrared camera located directly above the rat. We hypothesized that the rat might use a different posture in each block since the two reward ports were in different locations. We took the video frame closest to the center of the hold period on each trial and manually scored the position of each ear. We did this by asking a human observer (author CR), who was blind to the block and outcome of each trial, to click on each ear in the frame from every trial and record the position of the clicks. Using this time consuming scoring procedure, we analyzed the three sessions during which we recorded the most rule-encoding neurons (one session each from Rats 2, 4, and 6). Using the position of each ear, and knowing that the nose was located in the center port during this interval, we were able to construct the center position and azimuthal angle of the head relative to the behavior box. We found a prominent correlation between the head angle and the block in all analyzed sessions (Figure S2I, S3I). The mean difference between blocks was 26.9, 36.6, and 24.1 degrees in each of the three analyzed sessions.

Because head angle is correlated with block, and because we analyzed rule-encoding neurons for which, by definition, the firing rate correlated with block, it stands to reason that head angle correlates with firing rate. In an example neuron (Figure S2J, S3J), the firing rate is highly significantly correlated with block (black trend line). However, for this neuron the firing rate is not correlated with head angle within each block separately (red and blue trend lines), implying that this is a mere side effect of the fact that the neuron is encoding block.

To estimate the relative contributions of head angle and block identity in driving the response of each neuron, we used multiple-regression/analysis-of-variance (ANOVA), a standard technique for estimating the contribution of each, possibly correlated, variable to an observed response (Sokal and Rohlf, 2003). This technique constructs the optimal linear model for predicting neural firing rate using both head angle and block. The amount of variance that each factor explains in the optimal linear model is taken as an estimate of the amount that it contributes to the actual response. After defining our model, we discuss further the reasoning behind this argument, assuming linearity throughout; we also present simulations we conducted to validate that the approach is accurate and effective on datasets like this one.

Our model was defined as follows:

$$\text{Response} = B_0 * \text{Block} + B_1 * \text{HeadAngle} + \text{residual}$$

Here, Block is a binary variable representing pitch discrimination or localization, HeadAngle is the measured head angle on each trial, and Response is the square root of the spike count on each trial. (Square root is a variance-stabilizing transformation for Poisson-like spike counts: it ensures the variance is homogeneous and independent of mean firing rate.)

Multiple regression yields the optimal values of B_0 and B_1, in the sense of minimizing the squared sum of the residuals. These coefficients may be seen by inspection to give the contribution of one variable while the other variable is held constant, which is exactly what we want to estimate. Analysis-of-variance (ANOVA) yields the explained variance (EV) of each variable. In this case it is simply equivalent to:

$$EV[Block] = var(B_0 * Block)$$
$$EV[HeadAngle] = var(B_1 * HeadAngle)$$

Finally, the FEV is the explained variance of each individual variable, divided by the sum of explained variances over all variables.

How does multiple-regression/ANOVA assign EV when the factors are themselves correlated? Technically, this is done by inverting the covariance matrix in the so-called normal equations:

$$B = inv(X^T X)X^T Y$$

where X is the matrix of factors, Y is the response, and B is the matrix of coefficients B_0 and B_1. That is, shared variability between the factors is removed (a process called "whitening") before calculating the coefficient of each factor. Thus the FEV is a measure of the explanatory power uniquely available to each factor in the model, after removing all shared explanatory power from the factors (Sokal and Rohlf, 2003).

An alternative calculation yields the same coefficients but provides a different intuition. To calculate the coefficient (B_i) for each variable, we first hold out that variable and find the best fit between the other variable and the response. We next fit the residuals (unexplained part of the response) to the held-out variable. The coefficient B_i thus obtained is the same B_i that is yielded by multiple regression. We then repeat for all other variables. In this way it may be seen that the coefficient of each variable (*e.g.*, HeadAngle) is determined only by how well it predicts the part of the data that cannot be linearly predicted by Block, and vice versa.

Hence, the FEV is not only a measure of the variable's contribution to the optimal linear model, but also a measure of that variable's unique explanatory power divided by the total unique explanatory power of all variables, when considering the space of all possible linear models with those variables.

Results

Across our population of rule-encoding neurons, far more of the variability was explained by block than by head position (Figure S2K, L; S3K, L). The fact that some neurons do encode head angle, not block, according to this analysis is consistent with previous results (Euston and McNaughton, 2006). It is also a proof of principle that, in at least some cases, the head angle scoring procedure is sufficiently sensitive to uncover these effects. Nonetheless, the major conclusion of this analysis was that, for most neurons in both brain regions, head angle plays only a small role in the effects we observed.

[Technical note: we also considered a 2-way model that included an interaction term, which allowed for the possibility of different head-angle encoding in each block. The general conclusion however remained the same: block explained most of the variability in most neurons. Finally, one A1 neuron was excluded from this analysis because the least-squares fit was too poor (p-value for the F statistic was greater than 0.05); however, this neuron had very little sensitivity to head angle and including it would not have changed the conclusions. We were still able to consider this neuron using the analysis in the section "Posture equalization", which does not rely on a least-squares model.]

We chose to focus on head angle because this was the most prominent and easily quantifiable postural difference between blocks and is therefore likely to be correlated with preparatory motor activity in general. No analysis can rule out the possibility that PFC is actually encoding some unknown, subtle difference in behavior between blocks. However, the effects we observe favor the hypothesis that activity of PFC neurons primarily encodes the current task. It is possible that a side effect of this difference in cognitive state is a block-specific

difference in motor planning and execution; these motor effects would thus correlate with PFC activity even though they may not be directly encoded by PFC activity. In fact, even our neurons that seem to be encoding primarily head angle may actually be encoding cognitive state: trials on which the rat most strongly plans to perform localization might also produce the strongest preparatory actions.

Verification of ANOVA for estimating the effect of head angle

Verification on simulated posture-encoding neurons

We asked whether our multiple-regression / ANOVA approach was capable of detecting neurons that encoded head angle and not block. Although we did detect a minority of neurons with this property, suggesting that our method is indeed capable of this, we further investigated this with simulation. For each rule-encoding neuron for which we had video tracking data, we designed a model neuron with a firing rate given by:

$$\text{Response} = M * \text{HeadAngle} + \text{Noise}$$

Here again, as with all of our regression analyses, we define the response as the square root of the spike count in order to stabilize the variance. This ensures homogeneity of variance independent of mean firing rate.

To ensure that the simulation best matched the real data, we defined M as the slope of the best linear fit between the response and the head angle. We then ran 1000 simulated experiments, each with the same head angles as in the real data but with the noise randomly drawn from a Gaussian distribution. We defined this noise distribution such that the simulated response of the neuron had the same mean and variance as the response of the actual neuron.

We used the same multiple-regression/ANOVA analysis on these model neurons as we did on the real neurons, including both Block and HeadAngle as factors in the model. This analysis correctly identified all model neurons as primarily driven by head angle (Figures S2Q, S3Q): the FEV of head angle was always greater than 50% in the model neurons. The median FEV of head angle was 88% in A1 and 87% in PFC, compared with 17.0% and 5.9% in these brain regions in the real data.

Note that Block still correlated with HeadAngle just as before, meaning the model could have used either factor to predict the response. However, the analysis correctly assigns high FEV to HeadAngle here, in striking contrast to the high FEVs assigned to Block in the real data. We conclude that this approach is capable of detecting neurons that primarily encode head angle.

Verification with correlated regressors

Identifying the relative contributions of two correlated variables to a third variable is an inherently difficult problem because the predictive power of the shared variability is not uniquely due to either factor. We verified by simulation that our original procedure (multiple-regression/ANOVA) could accurately determine the fraction of explainable variance (FEV) under a variety of conditions: correlated regressors, binomial factors such as block identity, and discrete output similar to spike counts.

We took the measured head angle and block identity from the video recordings of the sessions and constructed a model neuron with a response given by:

$$\text{Response} = B_0 * (\text{HeadAngle}) + B_1 * (\text{BlockIdentity}) + \text{Noise}$$
$$B_0^2 + B_1^2 = 1$$

HeadAngle and BlockIdentity were first normalized to zero mean and unit variance. We used Gaussian noise with zero mean and standard deviation of 0.5, though the results were insensitive to the exact noise level provided there was some noise present. Thus, by construction we know the true FEV of HeadAngle and BlockIdentity to be B_0^2 and B_1^2, respectively. We used our previous least-squares/ANOVA analysis to derive estimates of the FEV, and compared these estimates to the known, true value. We found that the ANOVA analysis estimated the FEV of both factors with near zero bias (Figure S2R) over the full range of values of B_0

and B_1 between 0 and 1. That is, it properly accounted for the relative contributions of head angle and block identity. The error bars on the estimates were determined by the noise level in the simulation.

We extended this analysis by replacing the empirical distributions of the factors with simulated sessions, in which HeadAngle was drawn randomly from a normal distribution and BlockIdentity was a binary variable constructed to have a certain correlation with HeadAngle (Figure S2S). (Technically this was done by first assigning BlockIdentity by thresholding HeadAngle, which produces maximal correlation, and then flipping BlockIdentity randomly until the desired correlation was reached.) In these simulated cases we also found a good match between the true and estimated values of FEV (example r=0.72 Figure S2S; similar results were found for all values of r over its entire range).

Finally, we obtained very similar results when we converted the output of the model neuron to a rectified, discrete quantity that more closely resembled spike counts (Figure S2T). Specifically, we first exponentiated the output, x, to obtain a positive quantity analogous to the expected spike count: $y = 2^{x/x0}$ for some reference value x0. We then obtained simulated spike counts by drawing random integers from a Poisson process with mean count set to y. In this case we found it necessary to use a square root variance-stabilizing transformation on the spike counts, just as we previously did with the real data. Throughout, we were able to achieve good results using either type-2 or type-3 (but not type-1) sum-of-squares in ANOVA.

We conclude that ANOVA is a robust tool for estimating the fraction of explainable variance under a wide variety of conditions: correlated regressors, binary regressors, and Poisson-like quantized output. We found that our ANOVA analysis performed well for all of these types of data, including when using simulated spike counts and real values of head angle and block identity.

Posture equalization by trial selection

We also conducted an additional analysis that directly controls for the effect of head angle by trial selection. In brief, we considered a subset of correct trials from each block ("posture-equalized trials"), selected such that the mean head angle was the same in both blocks, and compared the rule encoding on those trials with the rule encoding on the rest of the trials. If posture accounted for the entire difference between blocks, then the difference between blocks should disappear on the posture-equalized trials. For most neurons in either brain region, we found the rule encoding to be similar on posture-equalized trials, suggesting that head angle plays only a minor role overall.

We present this analysis in three parts. First, we show that the distribution of rule encoding across neurons on posture-equalized trials is significantly different from what would be expected by random chance, and is broadly similar to the result from the rest of the trials (top and middle panels of Figures S2M, S3M). Second, we show that the mean magnitude of rule encoding is not significantly altered on the posture-equalized trials (Figures S2N, S3N). Third, we show that individual rule encoding neurons rarely change their rule encoding on posture-equalized trials (Figures S2O, S3O). All results hold for both mPFC (Figure S2) and A1 (Figure S3).

We first show the distribution of rule encoding, quantified as the logarithm of the ratio of the firing rates in pitch discrimination over localization, separately for posture-equalized and all other trials (Figures S2M, S3M). These distributions appear broadly similar, suggesting that rule encoding is not strongly altered by posture equalization. If rule encoding were entirely due to postural differences, then the distribution should be tightly clustered around zero on the posture-equalized trials, but we do not observe this. To demonstrate this, we randomly shuffled the block labels on the posture-equalized trials 10000 times and calculated the rule encodings on each shuffle. The bottom panels of Figures S2M and S3M show the distribution of the (necessarily spurious) rule encodings thus obtained: it is narrowly clustered around zero, as expected, and appears much narrower than the top distribution of posture-equalized rule encoding.

The spread of the distribution of rule encoding across neurons is a measure of the "population rule encoding", which we will refer to as PRE; a broad distribution indicates the presence of many neurons that strongly prefer one or the other block. We quantified the PRE as the median absolute deviation (MAD) of the distribution of

rule encoding. The MAD is a statistical estimator of the standard deviation that uses the median and is therefore more robust to outliers (Quian-Quiroga et al., 2004). The PRE was 0.471 in PFC and 0.318 in A1 on the posture-equalized trials; the fraction of shuffles that yielded a greater PRE was <0.005 in both brain regions. Thus, the rule encoding was significantly greater on the posture-equalized trials than it was on the randomized shuffles, and we may reject the hypothesis that the observed rule encoding on posture-equalized trials is simply due to random variation. We note that this conclusion was the same when we used the standard deviation instead of the MAD (data not shown).

(Technical aside: on a small fraction of the shuffles, we obtained zero spikes in one or the other block, which yielded a rule encoding of infinity. This occurred with probability <0.005. Such points cannot be shown on the distribution in Figures S2M and S3M. One advantage of the MAD over the standard deviation is that it yielded a robust estimate of PRE across shuffles, even in this rare and extreme circumstance.)

Second, we found that the mean magnitude of rule-encoding was not significantly altered by postural equalization (Figure S2N, S3N). We compared the absolute value of rule encoding for each neuron before and after posture-equalization using a paired Mann-Whitney test and found no significant difference.

Finally, we plotted the rule encoding for each individual cell before and after posture equalization (Figure S2O, S3O) and asked how often the preferred block changed after posture equalization. (Note that “preferred block” here simply means the block with greater firing rate, while in the rest of the text it is the block with significantly greater firing rate. We did not use statistical thresholding here because the sample size is smaller after posture equalization, decreasing statistical power.) If rule encoding were entirely due to posture differences between blocks, then posture equalization should remove any difference between block; thus the preferred block after posture equalization would be random and, 50% of the time, different than the preferred block in the rest of the trials. In fact we observed that only 1/16 PFC cells and 0/8 A1 cells switch their preferred block after posture equalization. The p-value (two-tailed binomial test) of this result is 0.008 for A1 and 0.0005 in PFC. Thus, we may reject the null hypothesis that posture equalization removes any difference between the blocks.

In summary, posture equalization appeared to non-significantly weaken the rule encoding, but it did not remove it. This weakening effect is to be expected based on our result that head angle explained some of the firing rate, primarily in a small minority of our neurons (Figures S2K, S3K), and also based on previous reports that some neurons in mPFC encode head angle (Euston and McNaughton, 2006). However, we emphasize that even after posture equalization, the rule encoding was largely preserved, in terms of its population distribution (Figures S2M, S3M), its mean magnitude (Figures S2N, S3N), and its sign in individual neurons (Figures S2O, S3O).

Supplemental References

Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput. 16, 1661–1687.

Salzman, C.D., Murasugi, C.M., Britten, K.H., and Newsome, W.T. (1992). Microstimulation in visual area MT: effects on direction discrimination performance. J. Neurosci. 12, 2331–2355.

Sokal, R, and Rohlf, FJ. *Biometry*. New York: W. H. Freeman and Company, 2003. Print.