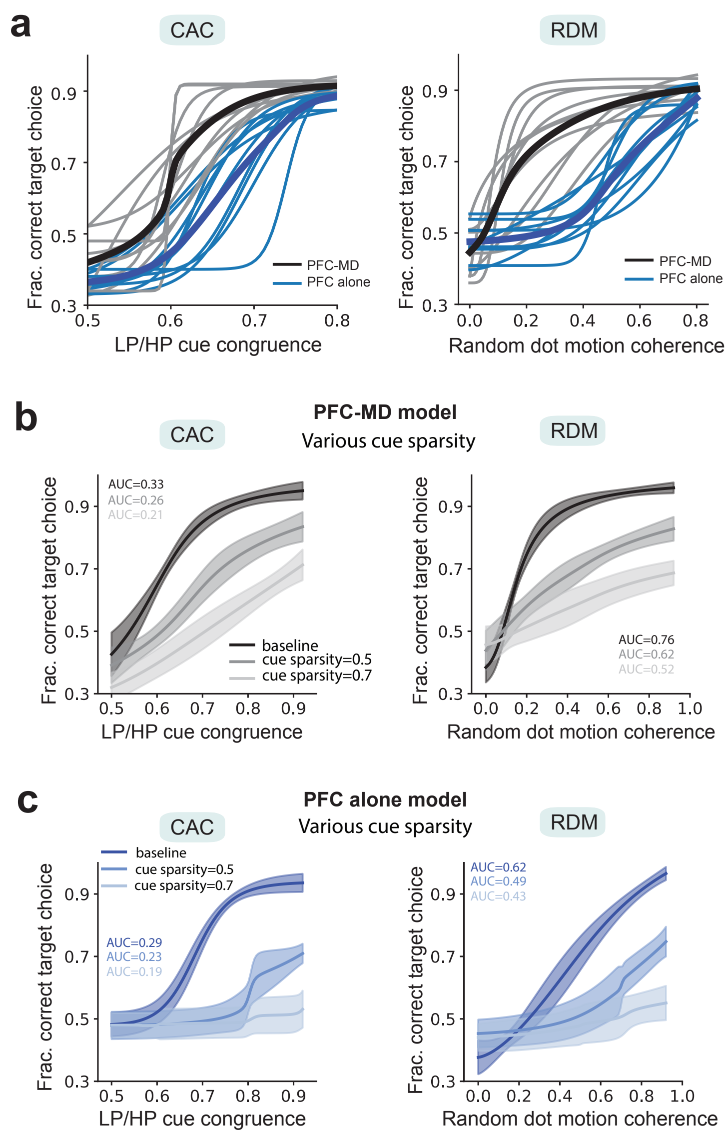
Supplementary Information

Mediodorsal thalamus regulates task uncertainty to enable cognitive flexibility

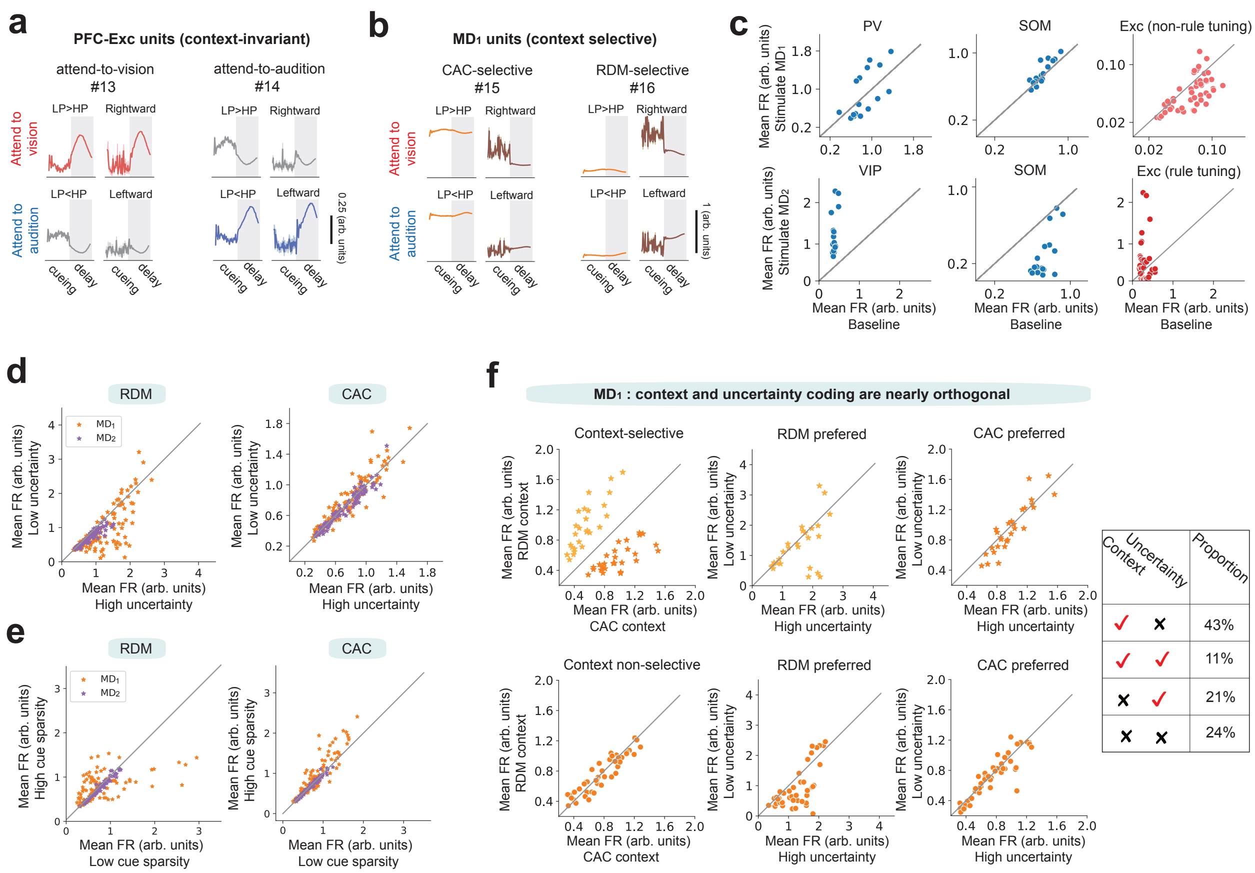
Xiaohan Zhang[1], Arghya Mukherjee[2], Michael M. Halassa[2], Zhe Sage Chen[1,3,4,5,*]
[1]Department of Psychiatry, New York University Grossman School of Medicine, New York, NY, USA
[2]Department of Neuroscience, Tufts University School of Medicine, Boston, MA, USA
[3]Department of Neuroscience and Physiology, New York University Grossman School of Medicine, New York, NY, USA
[4]Neuroscience Institute, New York University Grossman School of Medicine, New York, NY, USA
[5]Department of Biomedical Engineering, New York University Tandon School of Engineering, Brooklyn, NY, USA.

[*]Correspondence: zhe.chen@nyulangone.org

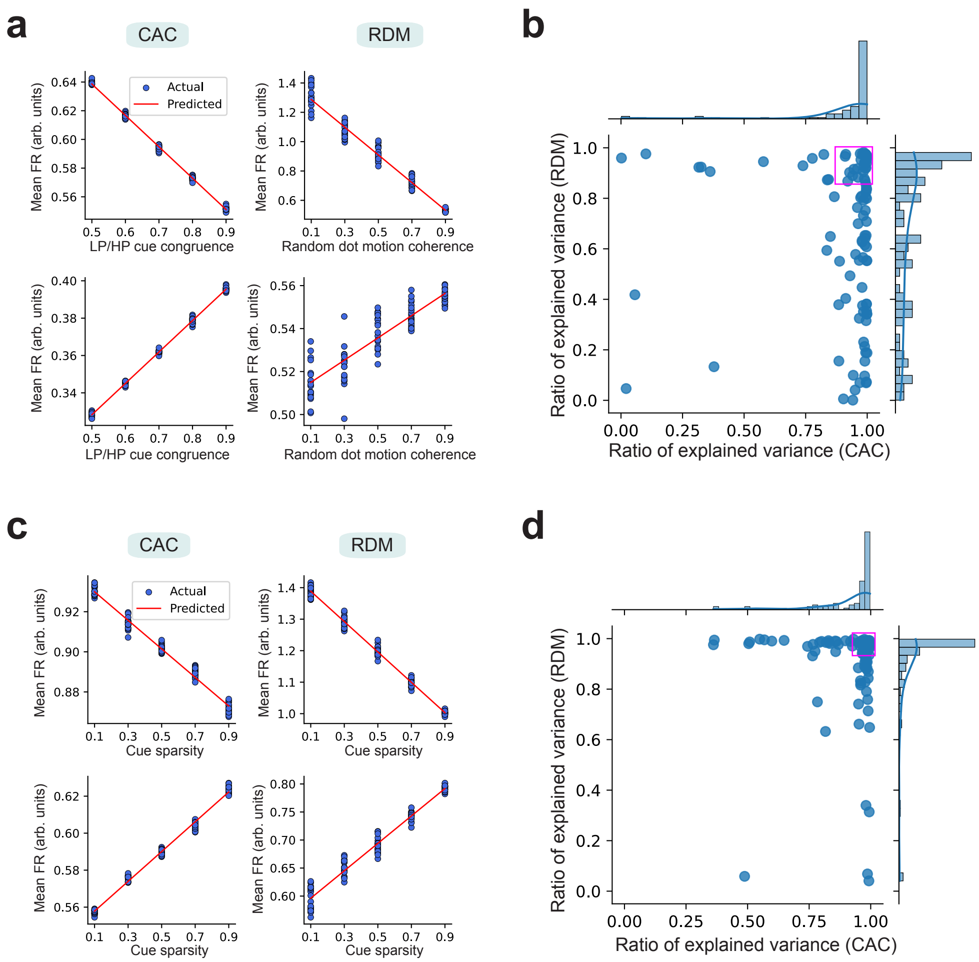


Supplementary Fig. 1. **Psychometric curve comparison between PFC-MD and PFC-alone models that were both performance-optimized.**

a, Selected psychometric curves of PFC-MD and PFC-alone models ($n = 10$ for each group) for both CAC and RDM contexts. The thick solid curves represent the respective ensemble averages. **b,** Psychometric curve of the PFC-MD model under three different cue sparsity levels in the CAC and RDM contexts. The AUC statistics decreased with an increasing sparsity level (CAC: black vs dark grey, p=0.00016; black vs light grey, p=0.00016, dark grey vs light grey, p=0.0008; RDM: black vs dark grey, p=0.00016; black vs light grey, p=0.00016, dark grey vs light grey, p=0.0008, rank-sum test). **c,** Psychometric curve of the PFC-alone model under three different cue sparsity levels in the CAC and RDM contexts. The AUC statistics decreased with an increasing sparsity level. (CAC: dark blue vs blue, p=0.00016; dark blue vs light blue, p=0.00016, blue vs light blue, p=0.02; RDM: dark blue vs blue, p=0.0003; dark blue vs light blue, p=0.00016, blue vs light blue, p=0.049, rank-sum test).

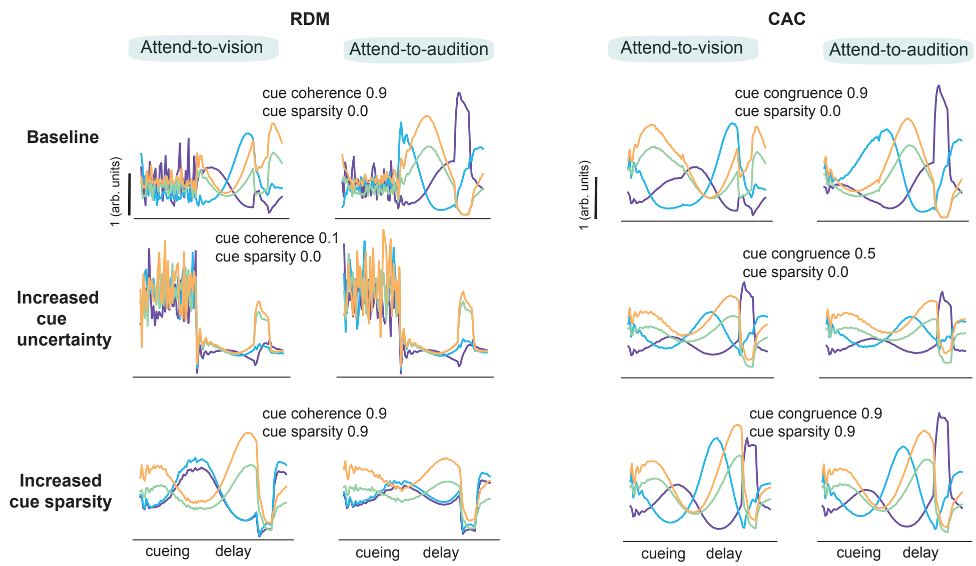


Supplementary Fig. 2. **PFC and MD single-unit modulation in the performance-optimized PFC-MD model.**
a, Two additional examples (similar to **Fig. 2a**) of context-invariant PFC excitatory units encoding two rules under two contexts. **b,** Two additional examples (similar to **Fig. 2b**) of context-selective MD units encoding context information. **c,** Stimulating MD_1 suppressed the firing rate of task-irrelevant PFC units (*upper panel*), and stimulating MD_2 amplified the firing rate of rule-tuned PFC units (*lower panel*) **d,** Examples of MD_2 units showed no firing modulation with respect to cue uncertainty. This was in contrast to MD_1 units shown in **Fig. 2f**. **e,** Examples of MD_2 units showed no firing modulation with respect to cue sparsity. This was in contrast to MD_1 units shown in **Fig. 2g**. **f,** Among MD_1 units, tuning properties with respect to the context and cue uncertainty were orthogonal.

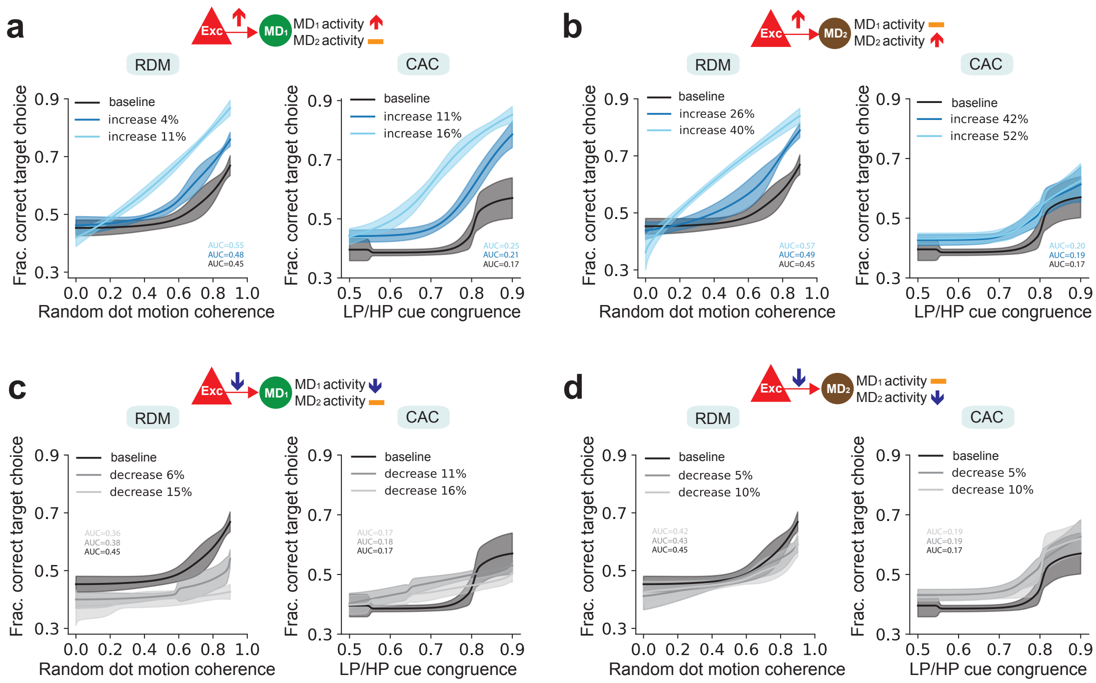


Supplementary Fig. 3. **$\mathbf{MD_1}$ units modulate their firing rates with respect to cue uncertainty and cue sparsity.**

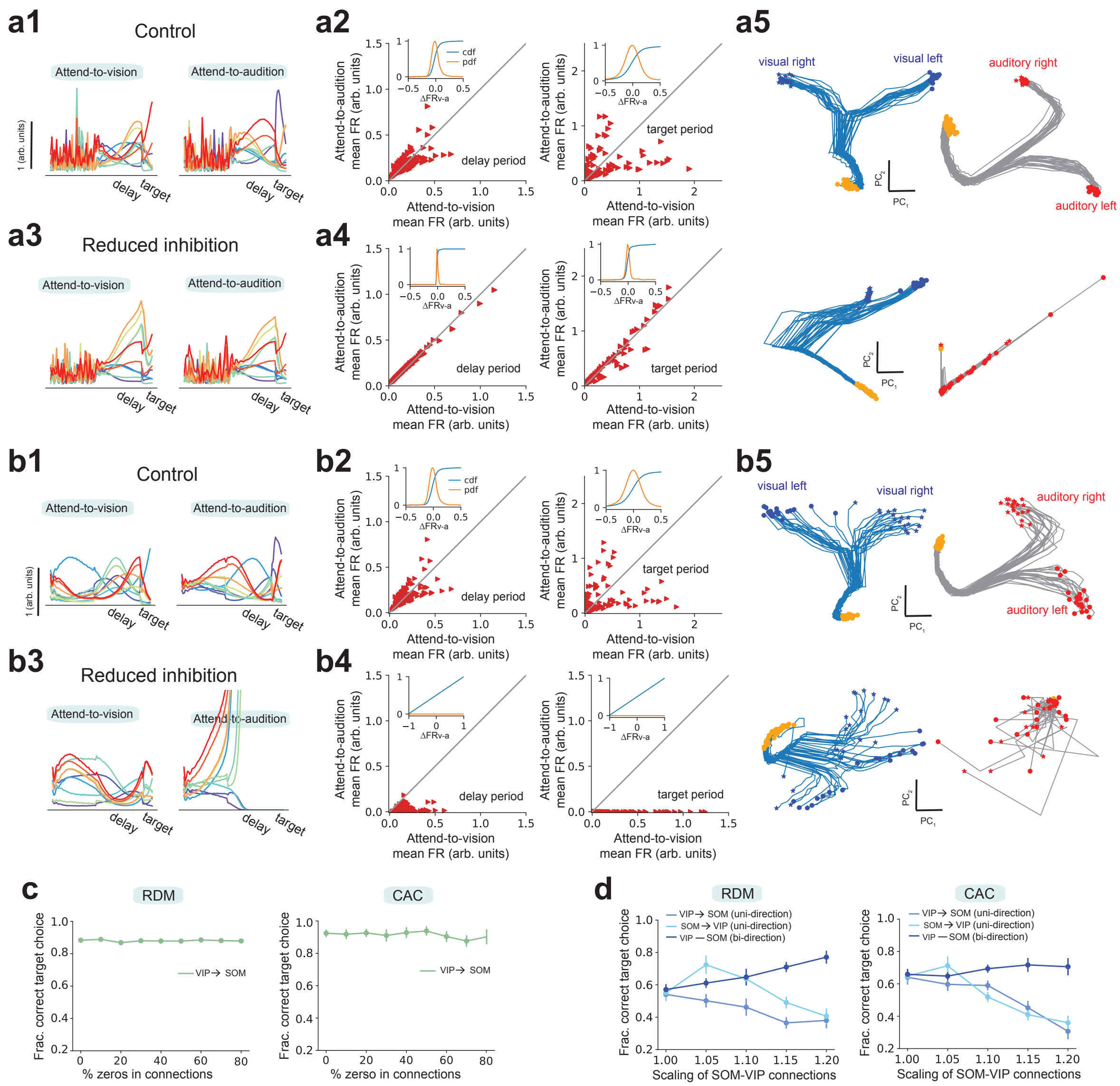
a, The line fitting results of cue uncertainty from two MD_1 units. **b,** The explained variance from both CAC and RDM contexts. **c,** The fitting result of cue sparsity from two MD_1 units. **d,** The explained variance from both CAC and RDM contexts.



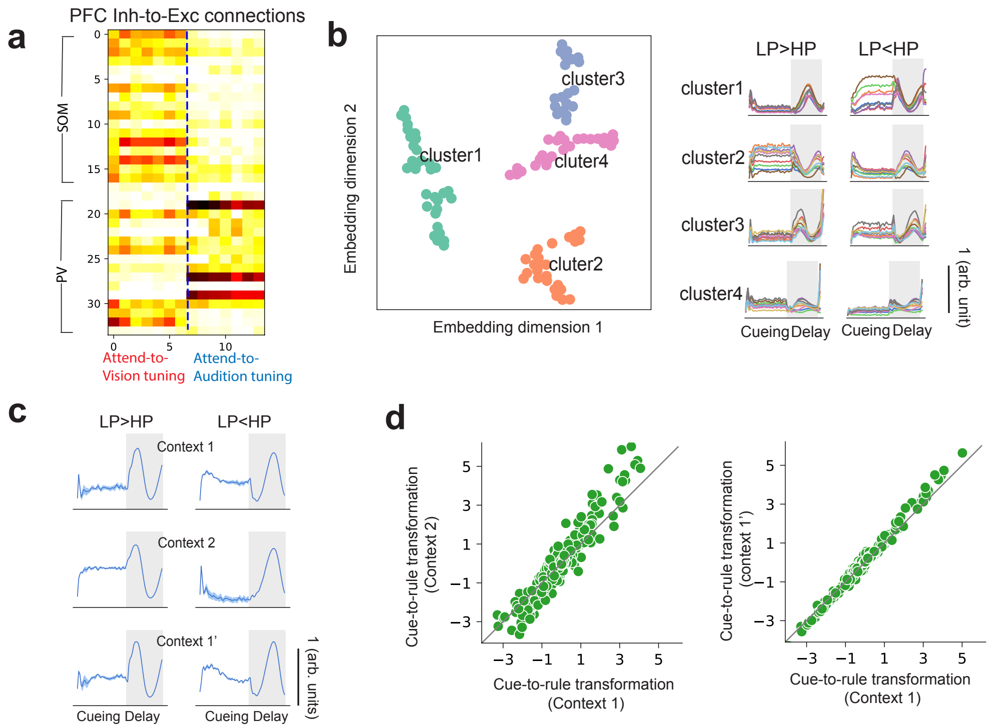
Supplementary Fig. 4. Demonstrations of PFC excitatory units that reduce rule tuning selectivity with increasing cue uncertainty or cue sparsity.



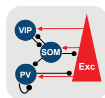
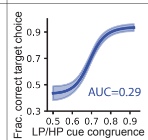
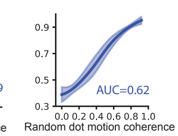
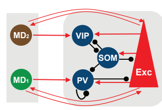
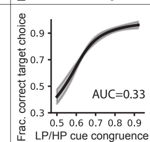
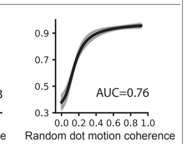
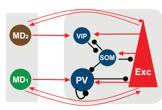
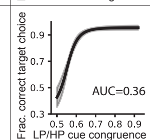
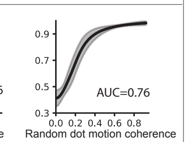
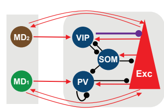
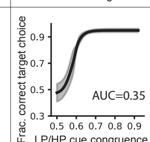
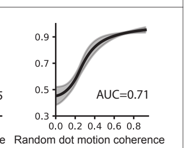
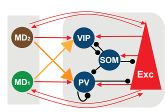
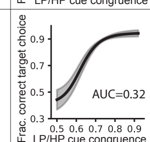
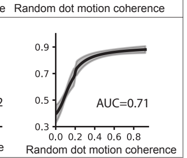
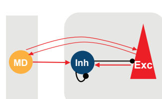
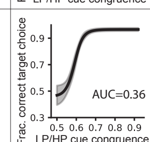
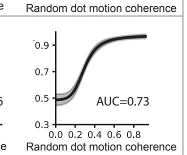
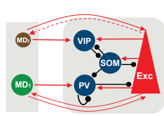
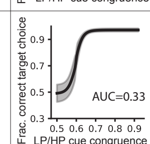
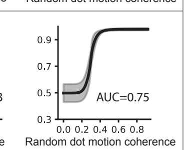
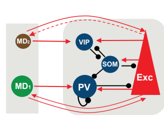
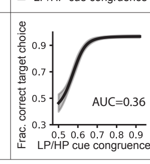
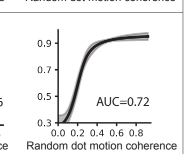
Supplementary Fig. 5. Changing directed prefrontal Exc→MD connection strengths of the PFC-MD model affects MD activity and task performance.
a, Strengthening PFC→MD_1 connection strengths (shown by a relative increase with respect to the baseline) increased the MD_1 firing rate while keeping the MD_2 firing rate unchanged, which further boosted working memory maintenance or improved task performance with elongated delay duration (RDM: dark blue vs. black, p=0.028; light blue vs. black, p=0.009; CAC: dark blue vs. black, p=0.0001; light blue vs. black, p=0.0001, rank-sum test). **b,** Strengthening PFC→MD_2 connection strengths increased the MD_2 activity while keeping the MD_1 activity unchanged, which further boosted working memory maintenance during elongated delay (RDM: dark blue vs. black, p=0.016; light blue vs. black, p=0.009; CAC: dark blue vs. black, $p=0.7\times10^{-4}$; light blue vs. black, $p=0.7\times10^{-4}$, rank-sum test). **c,** In contrast, weakening PFC→MD_1 connection strengths (shown by a relative decrease with respect to the baseline) decreased the respective MD activity, causing a degraded or inconsistent task performance (RDM: dark grey vs. black, p=0.009; light grey vs. black, p=0.009; CAC: dark grey vs. black, p=0.003; light grey vs. black, p=0.16, rank-sum test). **d,** Weakening PFC→MD_2 connection strengths decreased the respective MD activity, causing a degraded or inconsistent task performance (RDM: dark grey vs. black, p=0.25; light grey vs. black, p=0.047; CAC: dark grey vs. black, $p=0.7\times10^{-4}$; light grey vs. black, p=0.0002, rank-sum test).



Supplementary Fig. 6. **Circuit dissection of cognitive deficits.**
Comparisons of prefrontal excitatory single-unit tunings and population responses, and low-dimensional neural trajectories between control and reduced inhibition (by changing VIP→Exc connectivity) conditions, for RDM (**a1-a5**) and CAC (**b1-b5**) contexts. **a1,** Tuning curves of 10 PFC excitatory units during the control condition. **a2,** At the population level, mean firing rate (FR) comparison of PFC excitatory units between two rules during delay (left) and target (right) periods. Inset: Curves of probability density function (pdf, orange) and cumulative distribution function (cdf, blue). **a3,** With reduced inhibition, PFC excitatory units shown in (**a1**) decreased their discriminability in rule tuning during both delay and target periods. **a4,** Similar to panel (**a2**), except for the reduced inhibition condition. The cdfs between panels (**a2**) and (**a4**) were significantly different (delay period: p=1.71×10^{-10}; target period: p=1.71×10^{-10}, Kolmogorov-Smirnov test). **a5,** Two-dimensional neural trajectory representations in control and reduced inhibition conditions. **b1-b5** are similar to **a1-a5**, except for the CAC context. The cdfs between panels (**b2**) and (**b4**) were significantly different (delay period: p=4.1×10^{-54}; target period: p=4.1×10^{-54}). **c,** Weakening VIP→SOM connectivity of the performance-optimized PFC-MD model had little impact on task performance. **d,** When the cue signal was sparse, bidirectional SOM-VIP amplification of mutual inhibition strengths could amplify the sparse cue signal, but similar unidirectional manipulations were ineffective.



Supplementary Fig. 7. **The prefrontal activity in context-switching.**
a, PFC inhibitory-to-excitatory (Inh-to-Exc) connectivity was mapped into two classes of rule-tuned excitatory units (attend-to-vision vs. attend-to-audition). Only 14 representative PFC units were shown. **b,** PFC excitatory units ($n = 205$) were clustered in two-dimensional space by applying t-SNE embedding to combined Inh-to-Exc, Exc-to-Exc and MD-to-Exc synaptic weights. Four functional cell-type clusters emerged, each with a distinct rule-tuning profile. **c,** Some PFC excitatory units shifted their peak firing rates temporally during the delay period during context switching. **d,** Quantification and comparison of cue-to-rule transformation in prefrontal computation during context switching. Each dot corresponded to the gain related to each PFC unit ($n = 250$) (see Methods).

	Network connectivity	MD unit tuning				Psychometric Curves	
		Rule	Context	Cue Uncertainty	Cue Sparsity	CAC	RDM
PFC-alone (baseline)		N/A					
Fig. 1a $N_{Exc}=205$ $N_{PV}=17$ $N_{VIP}=17$ $N_{SOM}=17$ $N_{MD_1}=N_{MD_2}=100$		✗ ✗	✓ ✓	✗ ✓	✗ ✓		
$N_{Exc}=205$ $N_{PV}=21$ $N_{VIP}=15$ $N_{SOM}=15$ $N_{MD_1}=N_{MD_2}=100$		✗ ✗	✓ ✓	✗ ✓	✗ ✓		
w/o Inh specificity		✗ ✗	✓ ✓	✓ ✓	✓ ✓		
w/o MD specificity		✗ ✗	✓ ✓	✓ ✓	✓ ✓		
no cell type $N_{Exc}=205$ $N_{MD}=200$ $N_{Inh}=51$		✗	✓	✓	✓		
Fig. 6a $N_{Exc}=205$ $N_{PV}=17$ $N_{VIP}=17$ $N_{SOM}=17$ $N_{MD_1}=120$ $N_{MD_2}=80$		✗ ✗	✓ ✓	✓ ✓	✓ ✓		
$N_{Exc}=205$ $N_{PV}=21$ $N_{VIP}=15$ $N_{SOM}=15$ $N_{MD_1}=120$ $N_{MD_2}=80$		✗ ✗	✓ ✓	✗ ✓	✗ ✓		

Supplementary Table S1. Summary of tested computational model architectures, network size, and network connectivity and their impact on MD unit tuning and model psychometric curves in the cross-modal two-context 4AFC task.

	Rule and Context-Dependent Cueing			
	Rule 1: attend-to-audition Cue: A_{LP} or V_{right}		**Rule 2**: attend-to-audition Cue: A_{HP} or V_{left}	
Response \ Target	Auditory left	Auditory right	Visual left	Visual right
Choice #1 (auditory left)	correct	sensory	rule	both
Choice #2 (auditory right)	sensory	correct	both	rule
Choice #3 (visual left)	rule	both	correct	sensory
Choice #4 (visual right)	both	rule	sensory	correct

Supplementary Table S2. Categorization of task errors in the context-dependent 4AFC task.