LETTER

doi:10.1038/nature22073

Thalamic amplification of cortical connectivity sustains attentional control

L. Ian Schmitt[1], Ralf D. Wimmer[1], Miho Nakajima[1], Michael Happ[1], Sima Mofakham[1] & Michael M. Halassa[1,2]

Although interactions between the thalamus and cortex are critical for cognitive function[1–3], the exact contribution of the thalamus to these interactions remains unclear. Recent studies have shown diverse connectivity patterns across the thalamus[4,5], but whether this diversity translates to thalamic functions beyond relaying information to or between cortical regions[6] is unknown. Here we show, by investigating the representation of two rules used to guide attention in the mouse prefrontal cortex (PFC), that the mediodorsal thalamus sustains these representations without relaying categorical information. Specifically, mediodorsal input amplifies local PFC connectivity, enabling rule-specific neural sequences to emerge and thereby maintain rule representations. Consistent with this notion, broadly enhancing PFC excitability diminishes rule specificity and behavioural performance, whereas enhancing mediodorsal excitability improves both. Overall, our results define a previously unknown principle in neuroscience; thalamic control of functional cortical connectivity. This function, which is dissociable from categorical information relay, indicates that the thalamus has a much broader role in cognition than previously thought.

We recorded PFC ensembles in a two-alternative forced choice (2AFC) task, in which a freely behaving mouse selected between conflicting visual and auditory stimuli based on whether one of two rules was presented, and where rule presentation was varied on a trial-by-trial basis (Fig. 1a). Broadband white noise was informative of trial availability, prompting trial initiation by snout-protrusion into a centrally located port. Subsequently, either low-pass (10 kHz) or high-pass (11 kHz) noise was presented for 100 ms to signal the task rules (rule 1 (low pass), attend to vision; rule 2 (high pass), attend to audition), and after a brief delay where the head position was stably maintained, the animal selected between spatially conflicting visual and auditory stimuli to receive a reward. Animals achieved a balanced and robust performance across modalities (Extended Data Fig. 1a) and no stereotypical behaviour, which indicates a modality or location preference, was observed during the delay. These findings suggest that animals were capable of holding the task rule 'in mind' and using it to map onto sensory targets.

Given that prelimbic cortex[7] activity during the delay (referred to as prefrontal cortex (PFC)) is necessary for task performance[8], we directly interrogated its neural substrates (Extended Data Fig. 1b–d). We found that certain PFC neurons signalled the task rule by an increase in spiking for a brief moment during the delay (Fig. 1b). This temporal sparseness is analogous to that observed in the primate dorsolateral PFC during task rule encoding and/or maintenance[9]. The majority of such neurons signalled only one rule, but a minority (17%) signalled both rules at different temporal offsets (Extended Data Fig. 1e, f). Most tuned neurons exhibited regular spiking waveforms (regular spiking, 82% of total tuned neurons; 19% of all recorded regular-spiking neurons) consistent with these neurons being pyramidal neurons. Only a minority were fast spiking (fast-spiking neurons, 18% of total tuned neurons; 11% of all fast-spiking neurons; proportion of tuned fast-spiking out of total tuned neurons versus proportion of tuned regular-spiking neurons out of total tuned neurons, $P = 0.001$, binomial test; Extended Data Fig. 1g and Supplementary Note 1). Tuned peaks tiled the entire delay (Fig. 1c), and tuning was independent of delay length (Extended Data Fig. 1h).

Linear decoding[10] showed that tuned neurons encoded task rule but not movement (Fig. 1d and Supplementary Discussion 1). Rule information was exclusively represented by neurons we identified as tuned, because spike trains derived from the remaining neurons (untuned) did not contain rule information (or movement, Fig. 1d (bottom) and Extended Data Fig. 1i). As such, distinct PFC populations represent the two task rules used for sensory selection. The conclusion that these population codes reflect sensory selection rules, rather than rules directing the selection or avoidance of one sensory target, is supported by multiple findings. First, performance on trials with single-target modality presentation was identical to performance with sensory conflict (Extended Data Fig. 2a). Second, in sessions with sufficient errors (more than 20 trials), PFC neurons that were appropriately tuned to one rule displayed identical activity during error trials of the opposite trial type, indicating that inappropriate rule encoding was the major source of errors in this task (Extended Data Fig. 2b, c). To fully test this idea, we developed a four-alternative forced choice (4AFC) task (Extended Data Fig. 2d) in which animals separately indicated their choice for target modality type as well as its spatial location, thereby distinguishing errors that are related to rule encoding (executive) from those related to target stimulus perception (sensory) (Extended Data Fig. 2e). The largest source of 4AFC errors was executive (Extended Data Fig. 2f), consistent with our interpretation that similar misattribution of task rules explains most of the errors in the 2AFC task (Supplementary Discussion 2).

When multiple cells encoding the same rule were simultaneously recorded (Extended Data Fig. 3a), their consistent temporal order indicated that a neural sequence maintained the relevant categorical representation over time[11–15]. Consistent with this, we found that many PFC neuronal pairs exhibited robust short-latency cross-correlations, indicating their co-modulation at synaptic timescales[16,17]. This co-modulation was related to similarity in both categorical and temporal tuning (Extended Data Fig. 3b–f). Analysis at a higher temporal resolution, which is required for inferring putative monosynaptic connections[18,19] (example in Fig. 1e), showed significantly higher probability and strength among same-rule encoding pairs (Fig. 1f, g) and was only observed among pairs with overlapping temporal fields (Fig. 1h). As such, our data are consistent with a synaptic chain mechanism for categorical rule representation in the PFC. This inferred relationship between connectivity and coding is reminiscent of the one directly demonstrated in the mouse primary visual cortex, where neurons with similar orientation tuning are more likely to be synaptically connected[20].

To causally test this synaptic chain model, we used local optogenetic activation of inhibitory cortical neurons[21] to produce temporally

[1]NYU Neuroscience Institute, Department of Neuroscience and Physiology, NYU Langone Medical Center, New York, New York 10016, USA. [2]Center for Neural Science, New York University, New York, New York 10016, USA.

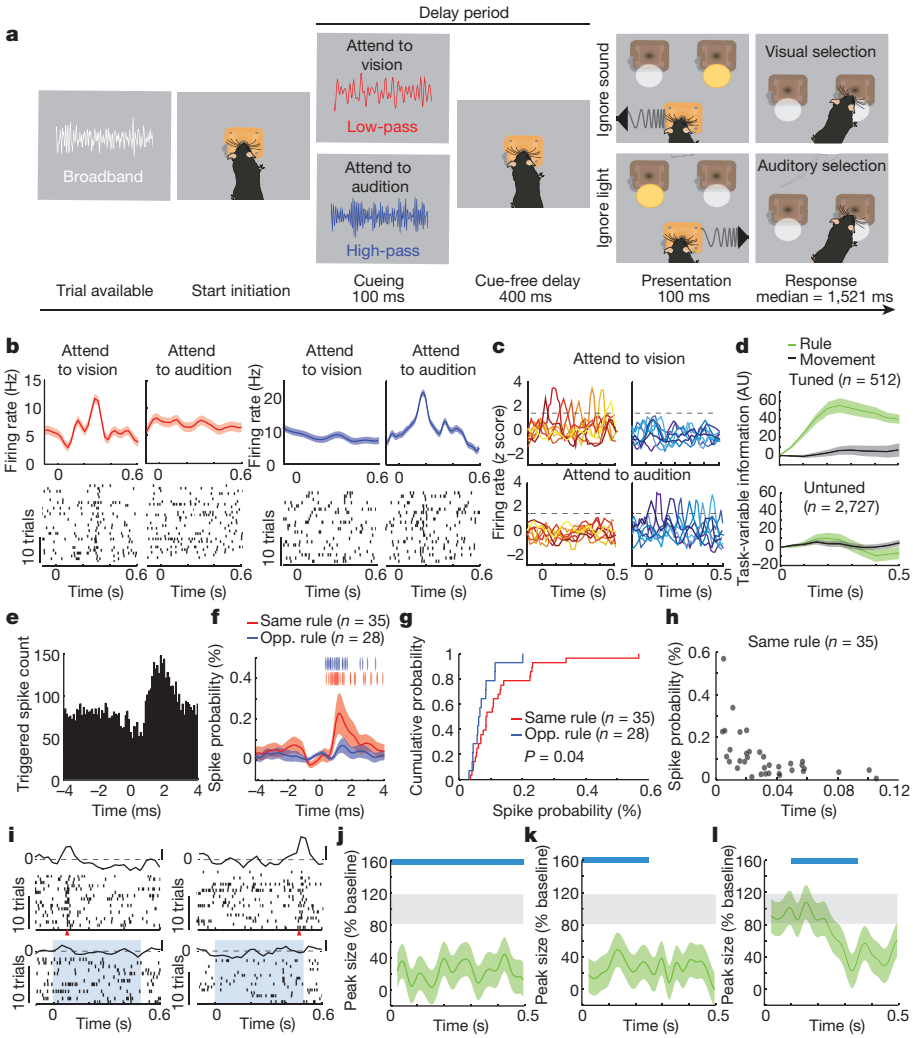


Figure 1 | Task-specific sequential PFC activity maintains rule representation. a, Schematic of task design. **b**, Example peri-stimulus time histogram (PSTH) and rasters for neurons tuned to either attend to vision (red) or attend to audition (blue) rules. **c**, Examples of tuning peaks across multiple sessions. **d**, Task-variable information, indicates that tuned neurons ($n = 512$ neurons from four mice) reflect rule information (top, green), but not movement (top, grey), whereas unturned neurons do not reflect rule information ($n = 2,727$, bottom). AU, arbitrary units. **e**, Example spike-time cross-correlation between two neurons (50-μs bins), indicating a putative monosynaptic connection. **f**, Putative monosynaptic connections in same rule tuned pairs showed a significantly larger average peak. Vertical ticks indicate peak times. Opp. rule, opposite rule. **g**, Cumulative plot showing cross-correlation values for each pair. Kolmogorov–Smirnov test. **h**, Same rule tuned pairs with putative monosynaptic connections had overlapping tuning peaks. **i**, Raster and PSTH examples showing diminished tuning during optogenetic activation of inhibitory neurons (blue shading indicates laser on). **j**, Quantification of laser effects on peak sizes ($n = 94$ neurons, three mice; example in Extended Data Fig. 3j). **k**, **l**, Temporally limited optogenetic manipulations indicate that later tuning depends on earlier activity. Blue line, laser on; green line, mean; green shading, 95% confidence interval (CI); grey shading, 95% CI for the baseline.

precise chain disruption (Extended data Fig. 3g). For each mouse, we used the minimum light intensity that, when delivered specifically during the delay, was sufficient to render it incapable of appropriate sensory selection, as has been shown previously[8]. Under those conditions, we found evidence for driving fast-spiking neurons, but overall, regular-spiking neurons were generally only slightly inhibited (Extended Data Fig. 3h, i). Nonetheless, laser delivery over the entire delay resulted in diminished rule tuning (Fig. 1i, j and Extended Data Figs 3j, 4a). Temporally limited manipulations revealed that early PFC manipulation diminished late task rule representation, even when the rule presentation period itself was spared (Fig. 1k, l and Extended Data Fig. 4b–d).

While synaptic PFC chains are probably necessary for sustaining rule representation, they are not sufficient, as presenting the two rule-associated cues outside of the task did not generate PFC tuning (Extended Data Fig. 4e, f). This indicated that additional factors are required for PFC populations to represent task rules. Given previous work that has shown a notable role for the mediodorsal thalamus in executive function[2,22], and its heavy reciprocal connectivity with the PFC[23], we investigated whether its interaction with the PFC was a factor.

Bilateral optogenetic suppression of the mediodorsal thalamus during the delay rendered mice incapable of appropriate sensory selection in the 2AFC task (Fig. 2a). Similar suppression in the 4AFC task resulted in identical error patterns to those resulting from PFC suppression (executive errors; Fig. 2b and Extended Data Fig. 5a, b), and ones that were distinct from those resulting from visual thalamic suppression (lateral geniculate nucleus (LGN); Fig. 2b and Extended Data Fig. 5c). Consistent with this behavioural dissociation, LGN suppression did not affect PFC rule-tuning (Extended Data Fig. 6a), whereas suppression of the mediodorsal thalamus diminished rule maintainence(Fig. 2c–g and Extended Data Fig. 6b–e) while largely sparing rule initiation (first 100 ms, during rule presentation; Fig. 2d, e). Suppression of the mediodorsal thalamus limited to the latter half of the delay was less effective at eliminating population coding and behaviour compared to an equivalent period of local PFC suppression (Extended Data Fig. 6e). These differences indicate that mediodorsal activity may not be required for initial rule encoding, and that the mediodorsal thalamus may be recruited by the PFC to sustain rule representation in a manner that outlasts mediodorsal neuronal spiking. Consistent with this, optogenetic PFC suppression in 100-ms bins across the delay resulted in identical behavioural effects throughout, whereas corresponding mediodorsal suppression resulted in a weaker effect at the earliest and latest bins (Fig. 2h). Notably, mediodorsal dependency was linked to delay length (Fig. 2i).

To understand how mediodorsal neurons sustained PFC representations, we recorded their spiking during the task (Fig. 2j). Certain mediodorsal neurons displayed temporally limited enhanced spiking during the task delay, but were non-selective to rule as the vast majority showed identical activity for both the attend to vision and attend to audition trials (Fig. 2k, l, Extended Data Fig. 7a and Supplementary Note 2). As such, it was not surprising that this mediodorsal population was uninformative to the task rule (Fig. 2m and Extended Data Fig. 7b–e). Notably, peaks were only encountered in the lateral mediodorsal thalamus (Extended Data Fig. 7f), consistent with their reciprocal connectivity pattern with the PFC[24] (Extended Data Fig. 7g, h). In fact, 58% of neurons recorded in the lateral mediodorsal

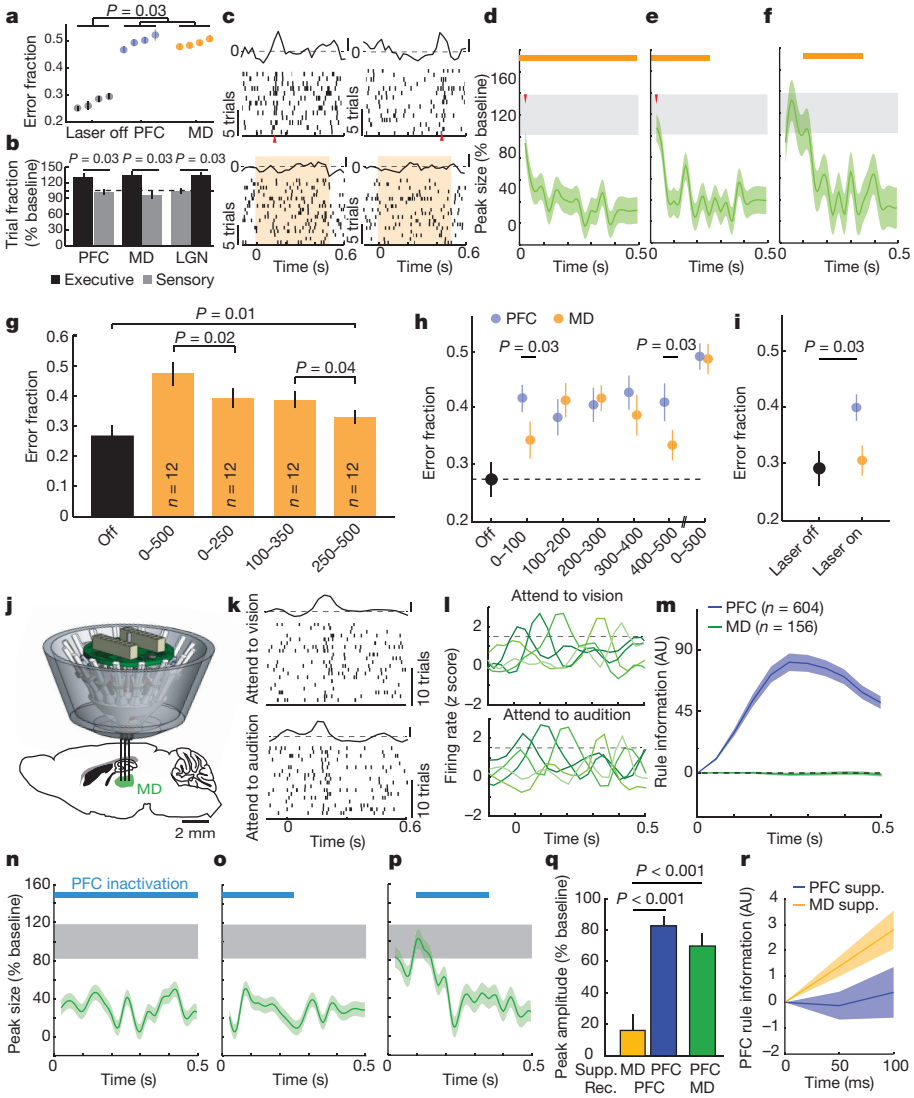


Figure 2 | Categorically-free mediodorsal activity is required for PFC rule representation and task performance. **a**, Delay-limited PFC or mediodorsal (MD) inhibition diminishes task performance. (four sessions, $n = 4$ mice each). **b**, Delay-limited PFC or mediodorsal inhibition in the 4AFC task selectively increases executive errors, whereas LGN suppression selectively increases sensory ones ($n = 4$ mice per group). **c**, Raster and PSTH examples of PFC rule tuning with mediodorsal suppression (shading denotes laser). **d**, Population quantification of data shown in **c** as in Fig. 1j ($n = 58$ neurons). **e**, **f**, Temporally limited mediodorsal suppression effects on population tuning (**e**, $n = 43$ neurons; **f**, $n = 46$ neurons; three mice with four sessions per condition). **g**, Comparison of behavioural performance with full and temporally limited mediodorsal suppression ($n = 3$ mice with four sessions each). **h**, Effect of a short (100 ms) suppression of the PFC or mediodorsal thalamus across the delay on performance. **i**, Effect of mediodorsal suppression in short delay (20 ms) trials. **j**, Schematic for mediodorsal recordings. **k**, Raster and PSTH example of a mediodorsal neuron showing similar peaks in both trial types. **l**, Example PSTHs of five mediodorsal neurons showing consistent lack of rule specificity. **m**, Linear decoding fails to reveal rule information in mediodorsal neurons with peaks (PFC, $n = 604$ neurons, six mice; mediodorsal thalamus, $n = 156$ neurons, three mice). **n**–**p**, Mediodorsal peak elimination by PFC suppression (full, $n = 47$ neurons; early, $n = 34$ neurons; middle, $n = 36$ neurons; two mice with 4–5 sessions each). **q**, Effects of mediodorsal (yellow) and PFC (blue) suppression on the first 50 ms of PFC tuning. Mediodorsal suppression produces a small effect on early PFC peaks ($n = 101$ neurons, three mice) relative to local PFC suppression ($n = 146$ neurons, three mice), whereas PFC suppression strongly reduced early mediodorsal tuning (green, $n = 81$ neurons, two mice). Supp., suppression; rec., recording. **r**, Early rule information is preserved with mediodorsal, but not PFC, suppression (decoding as in Fig. 1) (mediodorsal thalamus, $n = 101$ neurons, three mice; PFC, $n = 146$ neurons, three mice). Shading indicates 95% CI. Wilcoxon rank-sum test was used for all comparisons. Data are presented as mean ± s.e.m.

thalamus showed rule non-selective peaks dependent on PFC activity (Fig. 2n–p). However, in contrast to the impact of mediodorsal inactivation on PFC tuning, the very first mediodorsal peaks were eliminated by PFC suppression (Fig. 2q, r). Overall, these data indicate that mediodorsal engagement in the delay is not to relay categorical rule information, but rather to sustain existing cortical representations. Consistent with this, neither overall spike rates nor tuning peaks were changed in error trials (Extended Data Fig. 7i, j), confirming the conclusions that most errors in this task are due to rule misattribution and that this information is generated cortically. To our knowledge, this is the first direct mechanistic demonstration of a non-relay function of the thalamus.

To gain mechanistic insight into how the mediodorsal thalamus sustains categorical cortical representations, we performed multi-site recordings within the mediodorsal thalamus–PFC loop. We found that mediodorsal neurons increased spiking upon task engagement, but that only inhibitory cortical (fast-spiking) neurons and not excitatory (regular-spiking) neurons showed a similar increase (Fig. 3a). This was also true for changes seen in the task delay; mediodorsal and cortical fast-spiking neurons showed additional spiking enhancement, whereas

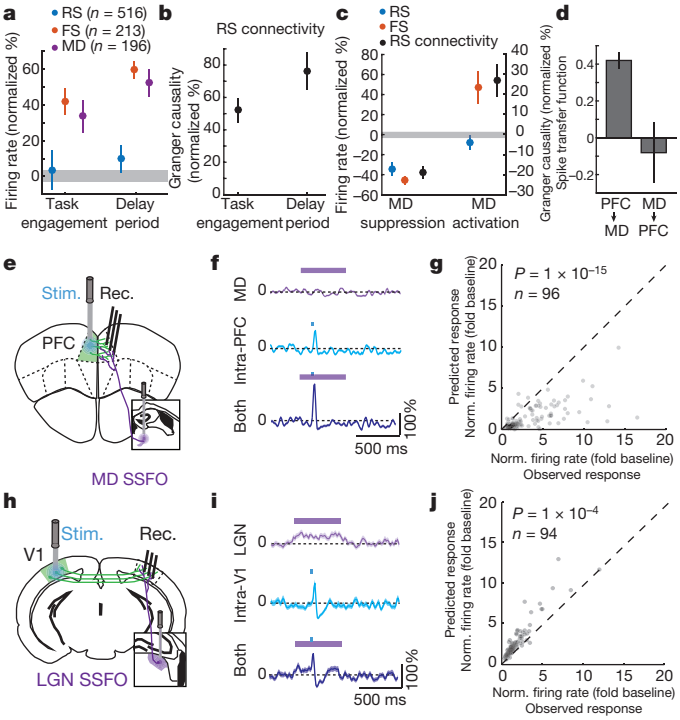


Figure 3 | Mediodorsal input amplifies local PFC connectivity. **a**, Mediodorsal and cortical fast-spiking (FS), but not regular-spiking (RS), neurons show increased firing rates upon task engagement and further increase during the delay (normalized to values outside of the task; PFC, six mice, regular-spiking neurons, 516 cells, fast-spiking neurons, 213 cells; mediodorsal thalamus, three mice, 196 neurons). Grey shading indicates 95% CI of null distribution. **b**, Regular-spiking neuronal network connectivity, assessed by Granger causality of spike trains (normalized to values outside of the task, $n = 6$ mice, 43 sessions; median, 13 neurons per session). **c**, Suppressing the mediodorsal thamalus reduces firing rates of cortical fast-spiking and regular-spiking neurons during task delay (regular-spiking neurons, $n = 245$; fast-spiking neurons, $n = 114$; $n = 2$ mice), as well as regular-spiking neuronal connectivity ($n = 2$ mice, 19 sessions; median 13 neurons per session). Enhancing mediodorsal excitability increases the firing rates of cortical fast-spiking neurons, connectivity among cortical regular-spiking neurons, but not neuronal firing rates of regular-spiking neurons (regular-spiking neurons, $n = 303$, fast-spiking neurons, $n = 131$; $n = 2$ mice; regular-spiking connectivity ($n = 17$ sessions; median, 18 neurons per session)). **d**, Spike transfer function (Methods) of the PFC to mediodorsal thalamus is significantly higher compared to transfer of the mediodorsal thalamus to PFC ($n = 17$ sessions, median, 11 PFC and 10 mediodorsal neurons per session for PFC to mediodorsal thalamus, median, 18 PFC and 15 mediodorsal neurons per sessions for mediodorsal thalamus to PFC, 2 mice per condition). Error bars are 95% CI estimated across sessions. **e**, Experimental setup for testing the effect of mediodorsal activation on local intra-PFC connectivity. **f**, Example regular-spiking neuron responses (normalized PSTH, mean ± s.e.m.) to mediodorsal activation alone, intra-PFC activation alone or the combination. **g**, Comparison of the observed combined response with the arithmetic sum of its individual components shows supra-linearity ($P < 10^{-15}$, signed-rank test). **h**, **i**, As in **e**, **f**, but for SSFO-mediated activation of LGN and recordings from V1. **j**, Combined stimulation results in a sub-linearity ($P < 10^{-4}$, signed-rank test).

regular-spiking neuronal spike rates remained relatively unaltered (Fig. 3a). Next, we investigated what could maintain spike rates of regular-spiking neurons constant across these conditions, but increase spiking of fast-spiking neurons. Because PFC neurons generate functional sequences following rule presentation, we suspected that their local connectivity might be enhanced by mediodorsal inputs, balancing increased inhibition. Consistent with this hypothesis, Granger causality of cortical regular-spiking neuronal network spike trains, a proxy for functional connectivity[25,26], increased upon task engagement and further increased during the task delay (Fig. 3b). These findings suggested that the firing rates of regular-spiking neurons became more dependent on local regular-spiking neuronal connections, and by extension mediodorsal inputs, as the animal engaged in the task. Consistent with this, optogenetic suppression of mediodorsal activity resulted in reduced spiking of regular-spiking neurons during, but not outside, the task (Fig. 3c and Extended Data Fig. 8a-c). Broadly enhancing mediodorsal excitability through stabilized step function opsin (SSFO) activation[27] selectively increased spiking of fast-spiking neurons and increased functional connectivity of regular-spiking neurons, without significantly changing spike rates of these putative excitatory neurons (Fig. 3c and Extended Data Fig. 8d–f). This modulatory mediodorsal-to-PFC input is very different from the driving PFC to mediodorsal thalamus input we observe after SSFO-mediated activation of the PFC (Fig. 3d and Extended Data Fig. 8g–i), and is consistent with the theoretical prediction that strong (driver–driver) thalamocortical loops are not implemented in the brain[28]. To further test whether mediodorsal inputs enhance functional cortical connectivity, we examined the effect of mediodorsal activation on intra-cortical evoked responses (Fig. 3e). In this context, a non-driving mediodorsal input amplified an intracortical evoked response (Fig. 3f). The supralinear nature of this amplification is evident by comparing the observed functional connection to the one predicted based on the arithmetic sum of its individual components (thalamic and cortical; Fig. 3g). Notably, these results were different from the ones observed in the thalamocortical loop containing the LGN and primary visual cortex (V1) (Fig. 3h); LGN inputs drove robust spiking in ipsilateral V1 and combining this driving input with contralateral V1 stimulation led to sublinear responses (Fig. 3i, j and Extended Data Fig. 9a).

To formalize our notion of this non-relay thalamic function, we built a data-driven spiking PFC–mediodorsal thalamus neural model (Extended Data Fig. 9b–f and Methods). Progressively increasing mediodorsal excitability to mimic task engagement enhanced correlated spiking among model PFC neurons, but resulted in rule-specific neural sequences only when co-tuned 'starter' neurons were synchronized by a common input (Extended Data Fig. 9g, h). The theoretical prediction that a combinination of mediodorsal amplification of cortical connectivity with a specific input that drives initial synchrony is sufficient to generate a sequential categorical representation, was experimentally validated (Extended Data Fig. 9i, j).

Our model shows that enhancing mediodorsal excitability increases PFC rule information content by improving tuning of individual cortical neurons, and by recruiting previously untuned ones (Extended Data Fig. 9i). This enhancement comes in contrast to rule information reduction when PFC excitability itself is increased because of chain cross-talk (Extended Data Fig. 9j). Consistent with the model, we found that SSFO-mediated enhancement of mediodorsal excitability led to increased PFC rule information (Fig. 4a–f), while enhancing PFC excitability diminished it (Fig. 4b–f). Changes in the neural code were also reflected in behaviour, as enhancing PFC excitability substantially diminished task performance, whereas boosting mediodorsal excitability improved it, with individual mice consistently performing at levels not typically observed in our cohorts without manipulation (26% reduction in lapse rate, Fig. 4h, i). Such enhancement was not found when primary auditory thalamus (medial geniculate body (MGB)) activity was activated in the context of auditory discrimination (Extended Data Fig. 10), highlighting the utility of the SSFO approach as a diagnostic test for categorical information encoded within neural circuits and the idea that, in contrast to the MGB, the mediodorsal thalamus does not relay such information during attentional control. Even mediodorsal neurons that apparently signalled one rule showed emergence of a peak in the opposite trial type with local SSFO activation, confirming that this information is not used for task performance (Extended Data Fig. 9k, l).

Overall, our study describes how a thalamic circuit amplifies local cortical connectivity to sustain attentional control. This function

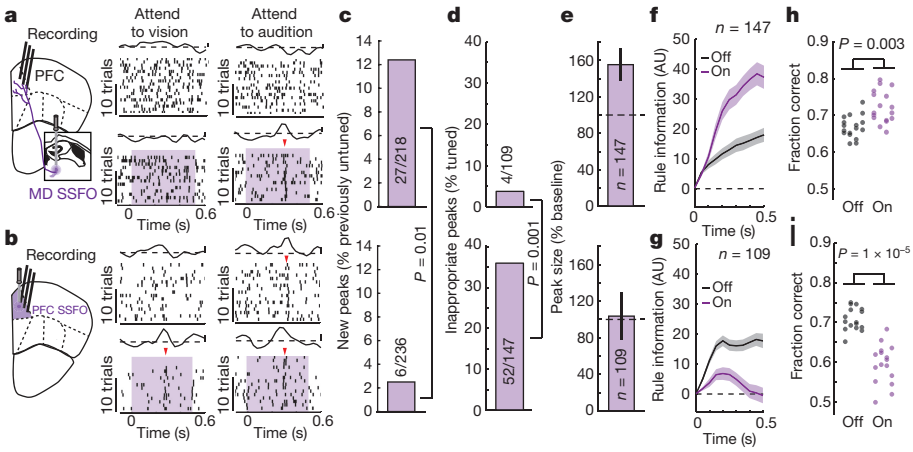


Figure 4 | Enhancing mediodorsal excitability strengthens PFC rule representation and improves performance. a, Optogenetic mediodorsal thalamus activation causes tuning of previously untuned PFC neurons in the 2AFC delay period. **b**, Optogenetic PFC activation generates inappropriate PFC tuning peaks (shading indicates laser on). **c**, **d**, Quantification of examples in **a**, **b** (binomial test). **e**, Quantification of existing peak size change. Data are mean ± 95% CI. **f**, **g**, Quantification of rule information within the PFC following mediodorsal or PFC activation ($n = 2$ mice). **h**, **i**, Opposing performance effects of the two manipulations ($n = 16$ sessions each from four mice). Shading indicates 95% confidence intervals (see Methods). Behavioural data are presented as session averages and compared using Wilcoxon rank-sum test.

may be general to cognitive processes that require extended cortical representations over time (Supplementary Discussion 4, 5). Because other thalamic circuits are thought to enhance connectivity between cortical areas[29,30], the precise engagement of the thalamus in a cognitive process may determine which cortical regions process information locally and which are engaged when processing spans multiple cortical nodes. Future studies aimed at testing this idea will undoubtedly provide a new way of thinking about cognition, where the thalamus forms a functional backbone that sustains, coordinates and switches distributed cortical computations.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 November 2016; accepted 15 March 2017.
Published online 3 May 2017.

1. Ito, H. T., Zhang, S. J., Witter, M. P., Moser, E. I. & Moser, M. B. A prefrontal–thalamo–hippocampal circuit for goal-directed spatial navigation. *Nature* **522,** 50–55 (2015).
2. Parnaudeau, S. *et al.* Inhibition of mediodorsal thalamus disrupts thalamofrontal connectivity and cognition. *Neuron* **77,** 1151–1162 (2013).
3. Xu, W. & Südhof, T. C. A neural circuit for memory specificity and generalization. *Science* **339,** 1290–1295 (2013).
4. Kuramoto, E. *et al.* Individual mediodorsal thalamic neurons project to multiple areas of the rat prefrontal cortex: a single neuron-tracing study using virus vectors. *J. Comp. Neurol.* **525,** 166–185 (2016).
5. Rubio-Garrido, P., Pérez-de-Manzo, F., Porrero, C., Galazo, M. J. & Clascá, F. Thalamic input to distal apical dendrites in neocortical layer 1 is massive and highly convergent. *Cereb. Cortex* **19,** 2380–2395 (2009).
6. Sherman, S. M. Thalamus plays a central role in ongoing cortical functioning. *Nat. Neurosci.* **19,** 533–541 (2016).
7. Hoover, W. B. & Vertes, R. P. Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. *Brain Struct. Funct.* **212,** 149–179 (2007).
8. Wimmer, R. D. *et al.* Thalamic control of sensory selection in divided attention. *Nature* **526,** 705–709 (2015).
9. Lundqvist, M. *et al.* Gamma and beta bursts underlie working memory. *Neuron* **90,** 152–164 (2016).
10. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503,** 78–84 (2013).
11. Goldman, M. S. Memory without feedback in a neural network. *Neuron* **61,** 621–634 (2009).
12. Harvey, C. D., Coen, P. & Tank, D. W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484,** 62–68 (2012).
13. Ikegaya, Y. *et al.* Synfire chains and cortical songs: temporal modules of cortical activity. *Science* **304,** 559–564 (2004).
14. Long, M. A., Jin, D. Z. & Fee, M. S. Support for a synaptic chain model of neuronal sequence generation. *Nature* **468,** 394–399 (2010).
15. Rajan, K., Harvey, C. D. & Tank, D. W. Recurrent network models of sequence generation and memory. *Neuron* **90,** 128–142 (2016).
16. Barthó, P. *et al.* Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *J. Neurophysiol.* **92,** 600–608 (2004).
17. Csicsvari, J., Hirase, H., Czurko, A. & Buzsáki, G. Reliability and state dependence of pyramidal cell–interneuron synapses in the hippocampus: an ensemble approach in the behaving rat. *Neuron* **21,** 179–189 (1998).
18. Hatsopoulos, N. G., Ojakangas, C. L., Paninski, L. & Donoghue, J. P. Information about movement direction obtained from synchronous activity of motor cortical neurons. *Proc. Natl Acad. Sci. USA* **95,** 15706–15711 (1998).
19. Young, E. D. & Sachs, M. B. Auditory nerve inputs to cochlear nucleus neurons studied with cross-correlation. *Neuroscience* **154,** 127–138 (2008).
20. Cossell, L. *et al.* Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* **518,** 399–403 (2015).
21. Halassa, M. M. *et al.* State-dependent architecture of thalamic reticular subnetworks. *Cell* **158,** 808–821 (2014).
22. Browning, P. G., Chakraborty, S. & Mitchell, A. S. Evidence for mediodorsal thalamus and prefrontal cortex interactions during cognition in macaques. *Cereb. Cortex* **25,** 4519–4534 (2015).
23. Preuss, T. M. & Goldman-Rakic, P. S. Crossed corticothalamic and thalamocortical connections of macaque prefrontal cortex. *J. Comp. Neurol.* **257,** 269–281 (1987).
24. Alcaraz, F., Marchand, A. R., Courtand, G., Coutureau, E. & Wolff, M. Parallel inputs from the mediodorsal thalamus to the prefrontal cortex in the rat. *Eur. J. Neurosci.* **44,** 1972–1986 (2016).
25. Barnett, L. & Seth, A. K. The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *J. Neurosci. Methods* **223,** 50–68 (2014).
26. Kim, S., Putrino, D., Ghosh, S. & Brown, E. N. A Granger causality measure for point process models of ensemble neural spiking activity. *PLOS Comput. Biol.* **7,** e1001110 (2011).
27. Yizhar, O. *et al.* Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* **477,** 171–178 (2011).
28. Crick, F. & Koch, C. Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature* **391,** 245–250 (1998).
29. Saalmann, Y. B., Pinsk, M. A., Wang, L., Li, X. & Kastner, S. The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* **337,** 753–756 (2012).
30. Zhou, H., Schafer, R. J. & Desimone, R. Pulvinar–cortex interactions in vision and attention. *Neuron* **89,** 209–220 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. A. Movshon, D. J. Heeger, X.-J. Wang, M. A. Wilson, C. D. Brody and E. K. Miller for helpful discussions. L.I.S. is supported by a NARSAD Young Investigator award and R.D.W. by a fellowship from the Swiss National Science Foundation. M.N. is supported by a JSPS fellowship. M.M.H. is supported by grants from NIMH, NINDS, Brain and Behavior, Sloan and Klingenstein Foundations as well as the Human Frontiers Science Program.

Author Contributions L.I.S. designed experiments, performed behavioural studies, analysed the physiological data and contributed to writing the manuscript. R.D.W. designed the 4AFC task, performed the physiological recordings, analysed behavioural data and contributed to writing the manuscript. M.N. validated viral tools, performed tracing studies and contributed to behavioural training. M.H. assisted L.I.S. with analysis. S.M. performed the modelling. M.M.H. conceived experiments and analyses, interpreted the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.M.H. (michael.halassa@nyumc.org).

METHODS

Animals. A total of 43 male mice were used in this study and the figure contribution of each mouse is summarized in Supplementary Table 1. All mice tested were between 2 and 12 months of age. C57BL/6J mice were purchased from Taconic Biosciences and VGAT–channel rhodopsin-2 (ChR2) mice were obtained from the Jackson laboratories. VGAT-cre mice were backcrossed to C57BL/6J mice for at least six generations. All mice were kept on a 12 h–12 h light–dark cycle. All animal experiments were performed according to the guidelines of the US National Institutes of Health and the Institutional Animal Care and Use Committee at the New York University Langone Medical Center.

Behavioural training and testing. *Behavioural setup.* Behavioural training and testing took place in gridded floor-mounted, custom-built enclosures made of sheet metal covered with a thin layer of antistatic coating for electrical insulation (dimensions in cm: length, 15.2; width, 12.7; height, 24). All enclosures contained custom-designed operant ports, each of which was equipped with an IR LED/IR phototransistor pair (Digikey) for nose-poke detection. Trial initiation was achieved through an 'initiation port' mounted on the grid floor 6 cm away from the 'response ports' located at the front of the chamber. Task rule cues and auditory sweeps were presented with millisecond precision through a ceiling-mounted speaker controlled by an RX8 Multi I/O processing system (Tucker-Davis Technologies). Visual stimuli were presented by two dimmable, white-light-emitting diodes (Mouser) mounted on each side of the initiation port and controlled by an Arduino Mega microcontroller (Ivrea). For the 2AFC and 4AFC tasks, two and four response ports were mounted at the angled front wall 7.5 or 5 cm apart, respectively. Response ports were separated by 1-cm divider walls and each was capable of delivering a milk reward (10 μl evaporated milk delivered by a single syringe pump (New Era Pump Systems) when a correct response was performed. For the auditory Go/No-go task environment, response and reward ports were dissociated, with the reward port placed directly underneath the response port. In the 4AFC, the two outermost ports were assigned for 'select auditory' responses, whereas the two innermost ports were assigned for 'select visual' responses. Access to all response ports was restricted by vertical sliding gates which were controlled by a servo motor (Tower Hobbies). The TDT Rx8 sound production system (Tucker Davis Technologies) was triggered through MATLAB (MathWorks), interfacing with a custom-written software running on an Arduino Mega (Ivrea) for trial logic control.

Training. Prior to training, all mice were food restricted to and maintained at 85–90% of their *ad libitum* body weight.

2AFC. Training was largely similar to our previously described approach[8].

First, 10 μl of evaporated milk (reward) was delivered randomly to each reward port for shaping and reward habituation. Making response ports accessible signalled reward availability. Illumination of the LED at the spatially congruent side was used to establish the association with visual targets on half of the trials. On the other half, association was established with the auditory targets where an upsweep (10 to 14 kHz, 500 ms) indicated a left and a downsweep (16 to 12 kHz, 500 ms) indicated a right reward. An individual trial was terminated 20 s after reward collection, and a new trial became available 5 s later.

Second, mice learned to poke in order to receive reward. All other parameters remained constant. An incorrect poke had no negative consequence. By the end of this training phase, all mice collected at least 20 rewards per 30-min session.

Third, mice were trained to initiate trials. Initially, mice had to briefly (50 ms) break the infrared beam in the initiation port to trigger target stimulus presentation and render reward ports accessible. Trial rule (attend to vision or attend to audition) was indicated by 10-kHz low-pass filtered white noise (vision) or 11 kHz high-pass filtered white noise (audition) sound cues. Stimuli were presented in blocks of six trials consisting of single-modality stimulus presentation (no conflict). An incorrect response immediately rendered the response port inaccessible. Rewards were available for 15 s following correct poking, followed by a 5 s inter-trial interval (ITI). Incorrect poking was punished with a time-out, which consisted of a 30 s ITI. During an ITI, mice could not initiate new trials.

Fourth, conflict trials were introduced, in which auditory and visual targets were co-presented indicating reward at opposing locations. Four different trial types were presented in repeating blocks: (1) three auditory-only trials; (2) three visual-only trials; (3) six conflict trials with auditory target; and (4) six conflict trials with visual target. The time that mice had to break the IR barrier in the initiation port was continuously increased over the course of this training stage (1–2 weeks) until it reached 0.5 s. At the same time, duration of the target stimuli was successively shortened to a final duration of 0.1 s. Once mice performed successfully on conflict trials, single-modality trials were removed and block length was reduced to three trials.

Fifth, during the final stage of training, trial availability and task rule were dissociated. Broadband white noise indicated trial availability, which prompted a mouse to initiate a trial. Upon successful initiation, the white noise was immediately replaced by either low-pass or high-pass filtered noise for 0.1 s to indicate the rule. This was followed by a delay period (variable, but for most experiments it was 0.4 s) before target stimuli presentation. All block structure was removed and trial type was randomized. Particular steps were taken throughout the training and testing periods to ensure that mice used the rules for sensory selection (see Supplementary Discussion 2).

4AFC. The first two training steps were identical to the 2AFC training, except that auditory stimuli consisted of tone clouds (interleaved pure tones (50 ms per tone over 200 ms, 36 tones total) spanning a frequency range of 1–15 kHz) directed to the left or right ear of the mouse to indicate the side of reward delivery. In the third stage, mice were trained to recognize the difference between visual and auditory response port positions. Initially, only two reward ports were available while access to the response ports associated with the non-target modality was restricted. All other parameters were as previously described in the 2AFC. Once mice successfully oriented to both target types (about two weeks), all four response ports were made available for subsequent training. Choosing a response port of the wrong modality was punished by a brief air puff delivered directly to the response port. Mice remained on this paradigm until they reached a performance criterion of 70% accuracy on both modalities.

During the fourth training stage, sensory conflict trials were introduced using the same parameters as in the 2AFC. Trial types and locations were randomized (spatial conflict was also random). Responses were scored as correct or one of three different error types (see confusion matrix in Extended Data Fig. 2e).

Auditory Go/No-go. A total of four mice were trained. A pair of electrostatic speakers (Tucker Davis Technologies) producing the auditory stimuli were placed outside of the training apparatus and sound stimuli were conveyed by cylindrical tubes to apertures located at either side of the initiation port, allowing stereotypical delivery of stimuli across trials. Trial availability was indicated by a light positioned at the top of the box and trial initiation required a 200-ms continuous interruption of the IR beam in the initiation port to ensure that the animals head was properly positioned to hear the stimuli. Following trial initiation, a second port (the response port) was opened and a pure tone stimulus was played. A 20-kHz tone signalled a 'Go' response, whereas frequencies above or below 20 kHz signalled a 'No-go' response. The pure tone stimuli were presented for 300 ms before response time, and were pseudo-randomly varied on a trial-by-trial basis, with trials divided between the Go stimulus (approximately 40% of trials) and two No-go stimuli (16 and 24 kHz, appproximately 30% of trials per frequency). After stimulus presentation, the response port was made accessible for a 3-s period. In Go trials, correct poking within the trial period (hit) rendered the reward port accessible, and reward was subsequently delivered upon poking. For a 'miss' in which the mouse failed to poke within the 3-s period, the reward port remained inaccessible. For a 'correct rejection', which involved withholding a response when No-go stimuli were played, the reward port was made accessible at the end of the 3-s period. For a 'false alarm', which involved a poke in the response port on a No-go trial, the reward port remained inaccessible and the next trial was delayed by a 15-s time-out, as opposed to the regular 10-s inter-trial interval.

Testing. For electrophysiological recordings and experiments with optical manipulation, testing conditions were equivalent to the final stage of training.

The first cohort of PFC recordings involving 'manipulation-free mice' included three C57BL/6 wild-type mice and one VGAT-cre mouse. The VGAT-cre mouse in this cohort, which was also used for experiments involving PFC manipulations, was initially run for an equivalent number of laser-free sessions as the three wild-type mice before any manipulation. This design was used to confirm equivalence in electrophysiological findings across genotypes, and to strengthen the overall conclusions drawn by using transgenic animals. Equivalence across genotypes can be readily appreciated by comparing the four principal component analysis (PCA) plots in Extended Data Fig. 1j.

For laser sessions, laser pulses of either blue (473 nm for ChR2 activation) or yellow (560 nm for eNpHR3.0 activation) light at an intensity of 4–5 mW (measured at the tip of the optic fibres) were delivered pseudo-randomly on 50% of the trials. During most optogenetic experiments, laser stimulation occurred during the whole delay period (500 ms) of the task. For temporal-specific manipulations concurrent with electrophysiological recordings (Fig. 1k, l, 2e, f and Extended Data Figs 4, 6), laser pulses were delivered for 250 ms either during the first half, after 100 ms (following cue presentation) or the latter half of the delay period. In the high-resolution optogenetic inactivation experiment (Fig. 2h) laser pulses were 100 ms long, dividing the 500-ms delay period equally into five periods. During a session, only one condition was tested. For stabilized step function opsin (SSFO, hChR2(C128S/D156A)) experiments (Figs 3, 4 and Extended Data Fig. 8), a 50-ms pulse of blue (473 nm, 4 mW intensity) light at the beginning of the delay period was delivered to activate the opsins and a 50-ms pulse of red (603 nm, 8 mW

intensity) light to terminate activation at the end of the delay period. Similarly, for MGB manipulations (Extended Data Fig. 10), SSFO was activated by a 50-ms pulse of blue (473 nm, 4 mW intensity) light before stimulus delivery and its activity was terminated by a 50-ms pulse of red (603 nm, 8 mW intensity) at stimulus offset. An Omicron-Laserage lighthub system (Dudenhofen) was used for all optogenetic manipulations.

Behavioural analysis. For all experiments with optogenetic manipulations, only sessions where baseline performance was ≥65% correct were included in the analysis. For all behavioural testing, single-mouse statistics were initially used to evaluate significance and effect size followed by statistical comparisons across sessions. Performance on the auditory Go/No-go task was assessed on the basis of the number of correct responses to Go stimuli (hit rate) relative to No-go stimuli (false alarm rate) and was considered sufficient if the overall discrimination index ($d' = Z_{\text{hit}} - Z_{\text{false alarm}}$)) was greater than 2 for the baseline condition. In cases where multiple groups were compared, a Kruskal–Wallis one-way analysis of variance (ANOVA) was used to assess variance across groups, followed by post hoc testing. For pairwise comparisons a Wilcoxon rank-sum test was used. Data are presented as mean ± s.e.m. and significance levels were set to $P < 0.05$.

Virus injections. Injections were performed using a quintessential stereotactic injector (QSI, Stoelting). All viruses were obtained through UNC Chapel Hill, virus-vector core. For PFC manipulation during electrophysiological recordings, 200 nl of AAV2-hSyn-DIO-ChR2 was injected bilaterally into the PFC of VGAT-cre mice. Bilateral injections of AAV1-hSyn-eNpHR3.0-eYFP (300 nl) were used for mediodorsal thalamus and LGN manipulations. For SSFO experiments, AAV1-CamKIIa-SSFO-GFP was injected bilateral either into PFC (200 nl) or mediodorsal thalamus (400 nl). To test the effect of mediodorsal activation on functional cortical connectivity we injected the mediodorsal thalamus with AAV1-CamKIIa-SSFO-GFP (400 nl) ipsilateral and the PFC with AAV1-hSyn-ChR2-eYFP (200 nl) contralateral to the recording site. Following virus injection, animals were allowed to recover for at least two weeks for virus expression to take place before the start of behavioural testing or tissue collection.

Optic-fibre implants for behavioural experiments. Mice were deeply anaesthetized using 1% isoflurane. For each mouse, up to three pairs of optic fibres (Doric Lenses) were used in behavioural optogenetic experiments and stereotactically inserted at the following coordinates (in mm from Bregma): PFC, AP 2.6, ML ± 0.25, DV −1.25; mediodorsal thalamus, AP −1.4, ML ± 0.6, DV −1.5; LGN, AP −2.2, ML 2.15, DV 2.6. Up to three stainless-steel screws were used to anchor the implant to the skull and everything was bonded together with dental cement. Mice were allowed to recover with *ad libitum* access to food and water for one week, after which they were brought back to food regulation and behavioural training resumed. A 473-nm laser was used for ChR2 activation, whereas eNpHR3.0 activation was achieved with a laser with a wavelength of 561 nm. Laser intensities were adjusted to be 4–5 mW measured at the tip of the optic fibre, which was generally the minimum intensity required to produce behavioural effects.

Multi-electrode array construction and implantation. Custom multi-electrode array scaffolds (drive bodies) were designed using 3D CAD software (SolidWorks) and printed in Accura 55 plastic (American Precision Prototyping) as described previously[21]. Prior to implantation, each array scaffold was loaded with 12–18 independently movable microdrives carrying 12.5-μm nichrome (California Fine Wire Company) stereotrodes or tetrodes. Electrodes were pinned to custom-designed, 96-channel electrode interface boards (EIB, Sunstone Circuits) along with a common reference wire (A-M systems). For combined optogenetic manipulations and electrophysiological recordings of the PFC, optic fibres delivering the light beam lateral (45° angled tips) were embedded adjacent to the electrodes (Extended Data Fig. 3g). In the case of combined optogenetic PFC manipulations with mediodorsal recordings, the optic fibre was placed away from the electrodes at the appropriate spatial offset. For combined unilateral multi-site recordings of PFC and mediodorsal (four mice) with SSFO manipulations, two targeting arrays (0.5 × 0.5 mm for PFC and 0.5 × 0.35 mm for mediodorsal) where separated by 3.2 mm in the AP axis. For SSFO manipulations, optic fibres delivering a lateral light beam were implanted directly next to the array targeting either PFC or mediodorsal thalamus. To test the effect of mediodorsal activation on functional cortical connectivity, a single electrode array was targeted to the PFC unilaterally, whereas a 400-μm core optic fibre (Doric Lenses) was targeted to the contralateral PFC. In addition, a 200-μm core optic fibre was placed 2.8 mm behind the electrode array for activating SSFO in the ipsilateral mediodorsal thalamus. Similarly, to interrogate the same question in a sensory thalamocortical circuit, an electrode array was implanted unilaterally into V1 and an additional 400-μm core optic fibre (Doric Lenses) was targeted to the contralateral V1. In addition, a 200-μm core optic fibre was placed 0.5 mm anterior to the electrode array for activating SSFO in the ipsilateral LGN. During implantation, mice were deeply anaesthetized with 1% isofluorane and mounted on a stereotaxic frame. A craniotomy was drilled centred at AP 2 mm, ML 0.6 mm for PFC recordings (approximately 1 × 2.5 mm), at AP −3 mm, ML 2.5 mm for V1 (1.5 × 1.5 mm) or at AP −1 mm, ML 1.2 mm for mediodorsal recordings (approximately 2 × 2 mm). The dura was carefully removed and the drive implant was lowered into the craniotomy using a stereotaxic arm until stereotrode tips touched the cortical surface. Surgilube (Savage Laboratories) was applied around electrodes to guard against fixation through dental cement. Stainless-steel screws were implanted into the skull to provide electrical and mechanical stability and the entire array was secured to the skull using dental cement.

Electrophysiological recordings. Signals from stereotrodes (cortical recordings) or tetrodes (thalamic recordings) were acquired using a Neuralynx multiplexing digital recording system (Neuralynx) through a combination of 32- and 64-channel digital multiplexing headstages plugged into the 96-channel EIB of the implant. Signals from each electrode were amplified, filtered between 0.1 Hz and 9 kHz and digitized at 30 kHz. For thalamic recordings, tetrodes were lowered from the cortex into the mediodorsal thalamus over the course of 1–2 weeks where recording depths ranged from −2.8 to −3.2 mm DV. For PFC recordings, adjustments accounted for the change of depth of PFC across the AP axis. Thus, in anterior regions, unit recordings were obtained –1.2 to −1.7 mm DV, whereas for more posterior recordings electrodes were lowered −2 to −2.4 mm DV. Following acquisition, spike sorting was performed offline on the basis of the relative spike amplitude and energy within electrode pairs using the MClust toolbox (http://redishlab.neuroscience.umn.edu/mclust/MClust.html). Units were divided into fast spiking and regular spiking on the basis of the waveform characteristics as previously described[21]. In brief, the peak to trough time was measured in all spike waveforms, and showed a distinct bimodal distribution (Hartigan's dip test, $P < 10^{-5}$). These distributions separated at 210 μs, and cells with peak to trough times above this threshold were considered regular-spiking neurons and those with peak to trough times below this threshold were considered fast-spiking cells (Extended data Fig. 1g). The majority of cells (2,727) in PFC recordings were categorized as regular spiking, whereas approximately one-third (909) was categorized as fast spiking.

Histology. For histological verification of electrode position, drive-implanted mice were lightly anaesthetized using isoflorane and small electrolytic lesions were generated by passing current (10 μA for 20 s) through the electrodes. All mice were then deeply anaesthetized and transcardially perfused using phosphate-buffered saline (PBS) followed by 4% paraformaldehyde. Brains were dissected and post-fixed overnight at 4 °C. Brain sections (50 μm) were cut using a vibratome (LEICA) and fluorescent images were obtained on a confocal microscope (LSM800, Zeiss). Confocal images are shown as maximal projection of 10 confocal planes, 20 μm thick.

Analysis of firing rate. For all PFC and mediodorsal neurons, changes in firing rate associated task performance were assessed using peri-stimulus time histograms (PSTHs). PSTHs were computed using a 10-ms bin width for individual neurons in each recording session[4] convolved with a Gaussian kernel (25 ms full-width at half-maximum) to create a spike density function (SDF)[31,32], which was then converted to a *z* score by subtracting the mean firing rate in the baseline (500 ms before event onset) and dividing by the variance over the same period. For comparison of overall firing rates across conditions, trial number and window size were matched between groups. Homogeneity of variance for firing rates across conditions was determined using the Fligner–Killeen test for homoscedasticity[33]. For comparisons of multiple groups, a Kruskal–Wallis one-way ANOVA was used to assess variance across groups before pairwise comparisons.

Identification of peaks in task-modulated neurons. A total of 3,444 single units were recorded within the PFC and 974 single units were recorded in the mediodorsal across animals. Overall assessment of firing rates during the task delay period showed that individual regular-spiking PFC neurons did not exhibit sustained increases in spiking relative to baseline (population shown in Extended Data Fig. 1) and a comparison of variance homoscedasticity (Fligner–Killeen test) did not reveal changes in variance. In a subset of cells, however, a brief enhancement of spike-timing consistency at a defined moment in the delay period was observed (Fig. 1b). To formally identify these neurons we used the following steps.

First, periods of increased consistency in spike-timing across trials were identified using a matching-minimization algorithm[34]. This approach was used to determine the best moments of spike time alignment across trials (candidate tuning peaks). The number of these candidate tuning peaks (n) was based on firing rate values during the delay period for each neurons. n was obtained by minimizing the equation:

$$\sum_{k=1}^{N} |n_k - n| \quad (1)$$

Where n_k is the number of observed spikes in a trial k. As such, the initial (and maximum) number of candidate peaks is equal to the median number of spikes observed across trials.

With an initial number of candidate peaks in hand, their times were subsequently estimated. These times were initially placed randomly within the delay window, and iteratively adjusted to obtain the set of final candidate peak times. The result of this iterative process was the solution to the equation:

$$S = \arg_{C \in S} \min \sum_{k=1}^{N} d_2(S_k, C)^2 \quad (2)$$

Where the set of final candidate peak times S is obtained by iteratively minimizing the temporal distance between candidate peak times (in each iteration) C and the observed spike times across trials S_k on the basis of a penalty associated with increased temporal distance, computed across all trials k. In the first step, temporal adjustment for each candidate peak time was based on finding the local minimum of the temporal distance function, d_2 (as described in ref. 34) after which spikes were adjusted by linear interpolation. In brief, neighbouring spike times across trials were sorted by their temporal offset to a given candidate peak time, and their linear fit was computed. Each candidate peak time was then moved to the midpoint of that fitted line, to achieve a local minimum. In a second step, cost minimization was jointly computed for all putative peaks using the Lagrange multiplier solution to the global minimization equation[34] and intervals between peak times were adjusted on the basis of this global minimum. Both the local and the global minimization steps were iterated until the spike-time variance, defined as the sum of the squared distances between spikes across trials, converged and a set of final candidate peak times were determined.

Next, to identify genuine tuning peaks, we applied two further conditions. First, for 75% of the trials, at least one spike was required to fall within ±25 ms of each final candidate peak time. This conservative threshold was based on the median firing rates observed during the delay (around 10 Hz) predicting that inter-trial spike distances will be greater than 50 ms if spikes were randomly distributed, making it highly improbable to fulfill this condition by chance. Second, these candidate peaks needed to have z-score values of >1.5 (equivalent to a one-sided test of significance) to be considered genuine tuning peaks. The z score of spiking across trials during the delay was computed relative to the pre-delay 500-ms baseline (10-ms binning, convolved with a 25-ms full-width at half-maximum Gaussian kernel). Obtaining a genuine tuning peak identified a unit as task-modulated, which was subsequently used for most analyses in this study. The vast majority of units only showed a single tuning peak using this method. Independent validation of this method's validity is discussed in Supplementary Discussion 1.

Principle component and linear regression analysis of population code. To estimate the extent to which task modulated units differentially encode task rules, a PCA was first performed as described previously[10]. Next, linear regression was applied to define the two orthogonal, task-related axes of rule type and movement direction. These analyses were performed on neural z-core time-series, separately for each comparison (trials separated by rule type or movement direction). In brief, a data matrix X of size $N_{\text{unit}} \times (N_{\text{condition}} \times T)$, was constructed in which columns corresponded to the z-scored population response vectors for a given task rule or movement direction at a particular time (T) within the 1-s window following task initiation. This window size was chosen to provide sufficient samples for analysis, but only the delay period data were examined for this study.

The contribution of each principal component to the population response across time was quantified by projecting the trial-type-specific z-score time-series (for example, attend to vision rule) onto individual principal components and computing the variance. The first principal component was used for all subsequent analyses as subsequent principal components were found to be uninformative in the initial analysis.

Multi-variable linear regression was applied to determine the contribution of task rule and subsequent movement to principal component divergence across time for the corresponding trial-type comparisons. Specifically, linear analysis related the response of unit i at time t to a linear combination of these two task variables using the following equation:

$$r_{i,t}(k) = \beta_{i,t}(1)\text{movement}(k) + \beta_{i,t}(2)\text{rule}(k) + \beta_{i,t}(5) \quad (3)$$

Where $r_{i,t}(k)$ is the z-score response for a neuron in trial set (k) for each task variable; movement and rule. The regression coefficients (β) were used to describe the extent to which z-score time-series variation in the firing rate of the unit at a given time point describes a particular task variable. This analysis was generally only applied to correct responses.

Regression coefficients were then used to identify dimensions in state space corresponding to variance across neural response data for the two task variables. Vectors of these coefficients across z-score time-series matrices separated by trial types (for example, rule1 versus rule 2) were projected onto subspaces spanned by the previously identified principal component.

We next constructed task-variable axes ($\beta_v^{\perp}$) using QR-decomposition to identify principal component separation associated with each task variable (v). To identify movement along these axes for each population response, their associated z-score time-series were projected onto these axes across time as follows:

$$p_{v,c} = \beta_v^{\perp} \boldsymbol{X}_c \quad (4)$$

Where $\boldsymbol{X}_c$ is the population vector for trial type c. This projection resulted in two time-series vectors $\boldsymbol{p}_{v,c}$ for each task variable that compared movement across trial types (rule 1 versus rule 2; right versus left) on their corresponding axes. The difference between these two time-series was used as the main metric for information (task rule or movement) in this study. For evaluating rule information in error trials when their number permitted analysis (>20 error trials; based on empirical assessment of minimum trial numbers required for principal component divergence), trial type axes obtained from correct trials were multiplied by −1 to reverse directionality. The significance bounds for all time-series were obtained using random subsampling and bootstrapping (around 60% of total neurons per bootstrap, 200 replications). The 95% confidence bounds at each time point were then estimated on the basis of the resulting distribution. To determine whether our inference that rule information was related to tuning peaks, task-modulated spike times were randomly jittered by 500 ms and the PCA repeated. This resulted in loss of rule-information-related principal component divergence, validating our inference.

Peak strength analysis. To obtain a quantitative estimate of peak fidelity across multiple trials, an internal neural synchrony measurement[35] was modified for short-term synchrony, which was associated with identified peaks. This approach was applied to spike trains associated with differing task conditions and responses. Each spike within the train was convolved with a Gaussian kernel with a 9-ms half width. Trials were then summed and divided by the kernel peak size and trial number giving a maximum value (for perfect alignment) of one at any point. Convolution vector values around the tuning in the baseline condition were compared to the value within the same time window in the other condition.

Cross-correlation analysis. To compute cross-correlation histograms (cross-correlograms), the MATLAB function 'crosscorr' was applied to whole-session spike trains from pairs of cells. Continuous traces at a 1-kHz sampling rate were first generated on the basis of the spike times, with times at which spikes occurred set to one and all other times to zero. Crosscorr was then applied to trains from all possible cell pairs, using a maximum lag time of ±50 ms.

The significance of a cross-correlogram was determined by randomly jittering all spike times independently and re-computing the cross-correlogram. Jitter values were drawn from a Gaussian distribution centred at zero with a s.d. of 3 ms. This process was repeated 100 times for each pair, and if the observed peak cleared the 95% confidence bounds of all shuffled sets, the pair was determined to have a significant cross-correlation.

Pairs of cells were grouped as follows: the control group was composed of cells in which only the first cell was rule-tuned. The test group was composed of pairs in which both cells were tuned. This test group was further broken down into two subgroups: one in which both cells responded to the same rule and one in which the cells responded to different rules.

Within these groups, co-modulation was defined as the number of significant cross-correlograms divided by the total number of cross-correlograms. After overall group comparison using a χ^2 test, proportion differences were statistically evaluated in a post hoc pairwise fashion using binomial proportion tests.

To examine the effect of tuning to the same rule on co-modulation strength, the distributions of cross-correlogram peak heights were also compared for the groups of pairs described above. An empirical CDF (cumulative distribution function) was constructed using the peak heights of each group, and these distributions were compared using a signed-rank test.

Finally, the relationship between cross-correlogram peak height and inter-alignment time was explored. The inter-alignment times among neuronal pairs tuned to the same rule were calculated by taking difference in spike alignment times of each pair.

To more effectively assess putative monosynaptic connections, the significant cross-correlograms between tuned pairs were also re-computed at a 50-μs resolution. Significance thresholding at this resolution was repeated by determining whether a sequence of two or more successive bins of the adjusted trace, which exceeded two standard deviations of the overall trace, occurred within 10 ms of the centre bin[19]. Cross-correlograms containing such outliers were further characterized on the basis of their peak times. Those with peaks at 300 μs or later were categorized as putative monosynaptic connections[18,19]. Among these putative connections, the pairs were split into two groups: those that were tuned to the same rule, and those that were tuned to opposite rules. To compare peak strength, spike probability was estimated by subtracting a shuffled distribution of spike times with

the same average firing rate as the postsynaptic neuron and dividing by the number of spikes in the presynaptic neuron[17]. The distributions of the resulting peak strengths among same rule and opposite rule putative monosynaptic connections were compared using the Kolmogorov–Smirnov test. Finally, the peak strengths of these pairs were plotted against their inter-alignment time. As in the above analysis, only same rule pairs were included.

Nonlinear decoding analysis. To further assess the degree of rule representation in the PFC and mediodorsal thalamus, we applied two population decoding approaches, the maximum correlation coefficient (MCC) and Poisson naive Bayes (PNB) classifiers as implemented in the neural decoding toolbox[36]. These analyses were applied to all tuned neurons recorded from either structure, each of which were pooled into a pseudo-population for each structure ($n = 604$ neurons in the PFC and $n = 156$ neurons in the mediodorsal thalamus). For MCC decoding, firing rate response profiles in individual correct trials associated with each rule were preprocessed by converting them to a z score using the mean and variance in the corresponding trial to prevent baseline spike-rate differences from affecting classification[37]. For PNB classification, neuron spiking activity was modelled as a Poisson random variable with each neuron's activity assumed to be independent. Trial-specific z scores (MCC) or spike counts (PNB) from these pseudo-populations were then repeatedly and randomly subsampled (200 resampling runs) and divided into training and test subsets (six training and two test trials per recording session across $n = 360$ PFC and $n = 116$ mediodorsal sessions). For each subsampling, the classifier was trained using the training subset to produce a predictive mean response template ($\bar{x}$) for each rule (i). Templates were constructed separately for 100-ms overlapping windows across the trace (step size = 20 ms) and classifiers trained for each template. The windowed classifiers allowed us to estimate the temporal evolution of information in the population. In the cross-validation step, these templates were used to predict the class for each test trial in the test set (x^*) by maximizing the correlation decision function ($i^* = \arg\max_i \mathrm{corr}(x^*, \bar{x})$) in the case of MCC or the log-likelihood decision function ($i^* = \arg\max_i \mathrm{LL}(x^*, \bar{x})$) in the case of the PNB classifier[38]. Finally, we estimated the predictive strength of population activity at each time point, that is, the extent to which activity in that time bin predicts the trial type, as the average of the correct predictions in the test set. To determine the variability of this estimate, a bootstrapping procedure was applied in which 25% of neurons were subsampled from the overall population and the same procedure was repeated (50 resampling runs). The resulting traces were used to estimate the 95% confidence intervals of the initial estimate from the full population.

Granger causality analysis. To determine the degree of causal connectivity in the ensemble of recorded neurons within the PFC or their counterpart in our simulated network, we used the Weiner–Granger vector autoregressive (VAR) causality analysis as implemented in the multivariate Granger causality toolbox (MVGC)[25]. Spike train data from each recorded or simulated neuron within a session was converted to a continuous signal by binning in 1-ms increments[39,40] and convolving the resulting signal with a Gaussian filter (half width 5 ms). For all neurons in individual sessions, this analysis used 500-ms segments either within the delay period (delay) or just before (task engagement) along with an equal number of randomly selected segments recorded outside of the behavioural environment (out of task). For assessment of laser effects, a matched number of correct trials in the laser and non-laser condition were compared for each recording session across neurons. To improve stationarity in the signal, segments were adjusted by subtracting the mean and dividing by the s.d. of each segment[39,41] and stationarity was checked by determining whether the spectral radius of the estimated full model was less than one[25]. All models met this stationarity criteria. Model order was estimated empirically for each subset using Bayesian information criteria after which VAR model parameters were determined for the selected model order. On the basis of the resulting parameters, time-domain conditional Granger causality measurements were calculated for each cell pair across all trials. Causal density for a given condition in each session was taken as the mean pairwise-conditional causality[25].

Connectivity assay. To assess the effect of changes in thalamic excitability on cortical connection strength, we measured intra-cortical responses evoked by ChR2-mediated activation of the contralateral cortex for V1/LGN (94 neurons in two mice) and PFC/mediodorsal thalamus (96 neurons in three mice). Responses to either cortical stimulation alone (10 ms ChR2 activation to the contralateral cortex), thalamic activation alone (500 ms SSFO activation in ipsilateral LGN or mediodorsal thalamus) or the combination were recorded in V1 and PFC (100 interleaved trials per condition). For the combined condition, thalamic activation preceded cortical stimulation by 100 ms.

Computational modelling. *Network structure and dynamics.* We constructed a model that consisted of excitatory (regular-spiking) and inhibitory (fast-spiking) PFC neurons as well as mediodorsal neurons. Within the PFC, regular-spiking cells formed subnetworks representing each task rule consisting of multiple interconnected chains. Neurons in each of these chains were locally connected to their nearest neighbour within the chain as well as to other chains within the same subnetwork. While neurons representing different rules were connected, connections were made stronger within each subnetwork (for example, among neurons representing the same rule) on the basis of our cross-correlation experimental data. Regular-spiking neurons of either rule sent overlapping projections to mediodorsal neurons and received reciprocal inputs from the mediodorsal thalamus. Mediodorsal inputs were modulatory with a longer time constant than for the PFC (1 ms versus 10 ms), and resulted in increased spiking of fast-spiking neurons (direct synaptic drive, $w = 0.6$) while providing an amplifying input (factor, 1.6×) to connections between regular-spiking neurons (regardless of rule tuning).

During rule encoding, the arrival of input attributed to one rule simultaneously activated the starter neuron (first neuron in a chain) in chains encoding that rule, engaging mediodorsal neurons and enhancing their firing through synaptic convergence. In turn, mediodorsal neurons enabled signal propagation that was specific to that rule by amplifying currently active regular-spiking neuronal connections, while preventing irrelevant synchrony elsewhere through augmented inhibition.

Spiking neuron model. We employed the leaky integrate-and-fire (LIF) model to simulate both of the network paradigms described above. LIF is a simplified spiking neuron model that is frequently used to mathematically model the electrical activity of neurons. The evolution of the membrane voltage of neuron j using the LIF equation is as follows:

$$C\frac{dV_j}{dt} = -\alpha_j(V_j - E) + I_j^{\mathrm{ext}} + I_j^{\mathrm{syn}} \quad (5)$$

where C is the membrane capacitance, V_j is the jth neuron's membrane voltage, α is the leak conductance ($\alpha = 0.95$). I^{ext} is an externally applied current with amplitude taken independently for each neuron from a uniform distribution ($\mu = 0.825$, s.d. = 0.25 for PFC and mediodorsal neurons). I_j^{syn} is the synaptic input to cell j, and this is defined as follows:

$$I_j^{\mathrm{syn}} = \sum_i \omega_{ij} A_{ij} (H(t) - H(t - \tau)) \quad (6)$$

where ω_{ij} represents the strength of the connection between presynaptic cell i and the postsynaptic neuron j; A_{ij} is the connectivity matrix that denotes the connectivity map. τ is the spike duration (1 ms in our simulations) and the H(t) is a Heaviside function that is zero for negative values ($t < \tau$) and one for positive values ($t > \tau$). In this model the voltage across the cell membrane grows, and after it reaches a certain threshold ($V^{\mathrm{th}} = 1$), the cell fires an action potential, and its membrane potential is reset to the reset voltage. Here, the resting potential (E) and reset-potential V^{reset} are set to zero. The neuron enters a refractory period ($T^{\mathrm{ref}} = 1.5$ ms) immediately after it reaches the threshold ($V = V^{\mathrm{th}}$) and spikes. To integrate the LIF equation, we used the Euler method with a step size of $\Delta t = 0.01$ ms. To reproduce the spontaneous activity of the network, we introduced a noise that arrives randomly at each cell with a predefined probability ($f_N = 10$ Hz).

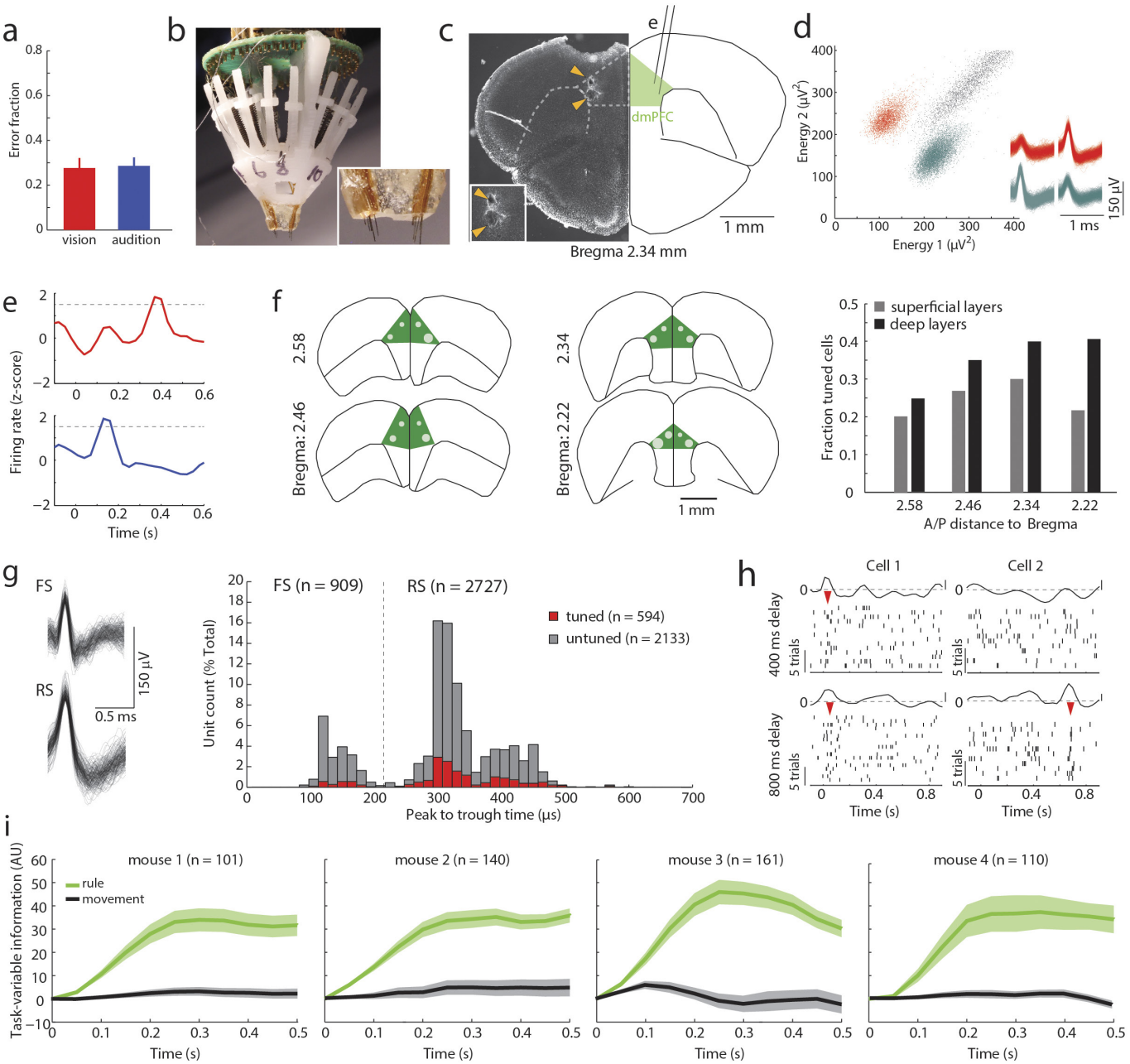
Statistical analysis. For each statistical analysis provided in the manuscript, an appropriate statistical comparison was performed. For large sample sets, the Kolmogorov–Smirnov normality test was first performed on the data to determine whether parametric or non-parametric tests were required. Variance testing for analysis involving comparisons of firing rates under differing behavioural conditions and following optogenetic manipulations was done using the Fligner–Killeen test of variance homoscedasticity. For small sample sizes ($n < 5$) non-parametric tests were used by default. Two different approaches were used to calculate the required sample size. For studies in which sufficient information on response variables could be estimated, power analyses were performed to determine the number of mice needed. For sample size estimation in which effect size could be estimated, the sample number needed was estimated using power analysis in MATLAB (sampsizepwr) with a β of 0.7 (70%). For studies in which the behavioural effect of the manipulation could not be prespecified, including optogenetic experiments, we used a sequential stopping rule[42]. This method enables null-hypothesis tests to be performed in sequential stages, by analysing the data at several experimental points using non-parametric pairwise testing. In these cases, the experiment initially uses a small cohort of mice which are tested over multiple behavioural sessions. If the P value for the trial comparison across mice falls below 0.05, the effect is considered significant and the cohort size is not increased. If the P value is greater than 0.36 following four sessions that met criteria, the investigator stopped the experiment and retained the null hypothesis. Using this strategy, the required number of animals was determined to be between three and five animals per cohort across testing conditions. For multiple comparisons, a non-parametric ANOVA (Kruskal–Wallis H-test) was performed followed by pairwise post hoc

analysis. All post hoc pairwise comparisons were two-sided. No randomization or investigator blinding was done for experiments involving electrophysiology. Blinding was used for experiments involving SSFO and behaviour (mediodorsal versus PFC).

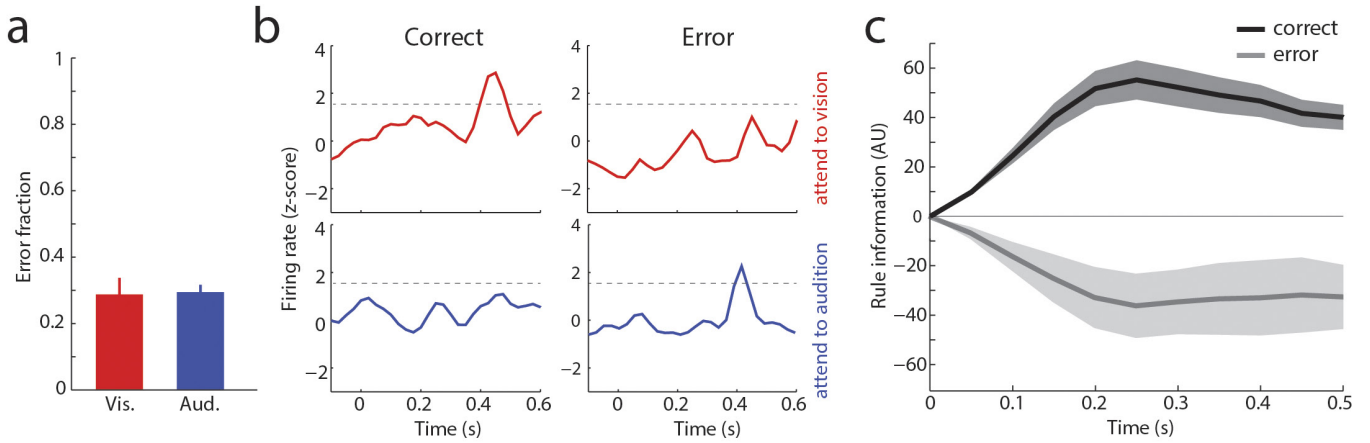
Code availability. All computer code used for analysis and simulation in this study was implemented in MATLAB computing software (MathWorks). Code will be made freely available to any party upon request. Requests should be directed to the corresponding author.

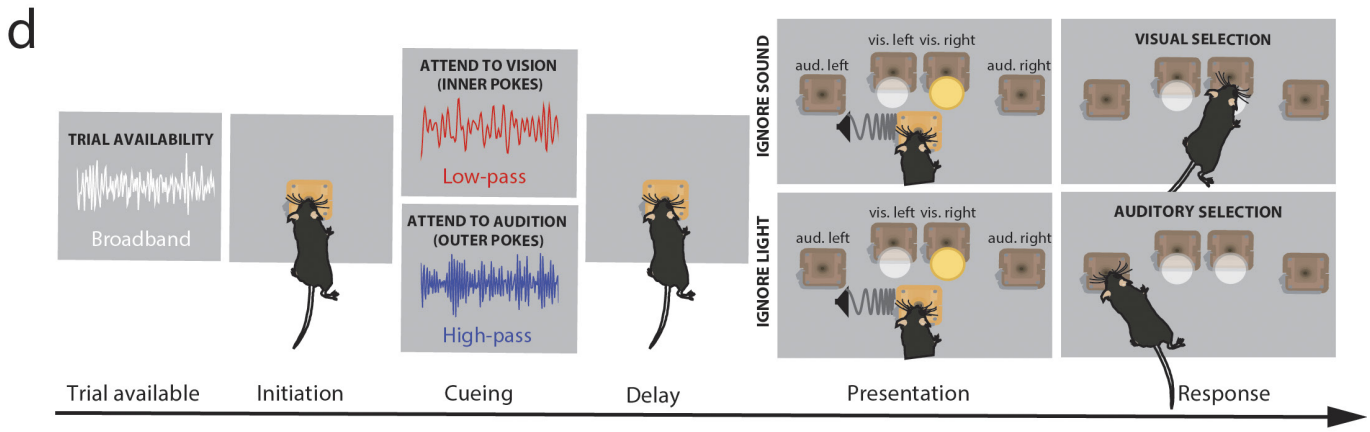
Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

31. Fries, P., Neuenschwander, S., Engel, A. K., Goebel, R. & Singer, W. Rapid feature selective neuronal synchronization through correlated latency shifting. *Nat. Neurosci.* **4,** 194–200 (2001).
32. Szűcs, A. Applications of the spike density function in analysis of neuronal firing patterns. *J. Neurosci. Methods* **81,** 159–167 (1998).
33. Wimmer, K., Nykamp, D. Q., Constantinidis, C. & Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* **17,** 431–439 (2014).
34. Wu, W. & Srivastava, A. Towards statistical summaries of spike train data. *J. Neurosci. Methods* **195,** 107–110 (2011).
35. Golomb, D. & Rinzel, J. Dynamics of globally coupled inhibitory neurons with heterogeneity. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **48,** 4810–4814 (1993).
36. Meyers, E. M. The neural decoding toolbox. *Front. Neuroinform.* **7,** 8 (2013).
37. Zhang, Y. *et al.* Object decoding with attention in inferior temporal cortex. *Proc. Natl Acad. Sci. USA* **108,** 8850–8855 (2011).
38. Duda, R. O., Hart, P. E. & Stork, D. G. *Pattern Classification* Vol. 18 (Wiley, 2001).
39. Cadotte, A. J., DeMarse, T. B., He, P. & Ding, M. Causal measures of structure and plasticity in simulated and living neural networks. *PLoS One* **3,** e3355 (2008).
40. Zagha, E., Ge, X. & McCormick, D. A. Competing neural ensembles in motor cortex gate goal-directed motor output. *Neuron* **88,** 565–577 (2015).
41. Ding, M., Bressler, S. L., Yang, W. & Liang, H. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biol. Cybern.* **83,** 35–45 (2000).
42. Dejean, C. *et al.* Prefrontal neuronal assemblies temporally control fear behaviour. *Nature* **535,** 420–424 (2016).
43. Gardner, R. J., Hughes, S. W. & Jones, M. W. Differential spike timing and phase dynamics of reticular thalamic and prefrontal cortical neuronal populations during sleep spindles. *J. Neurosci.* **33,** 18469–18480 (2013).



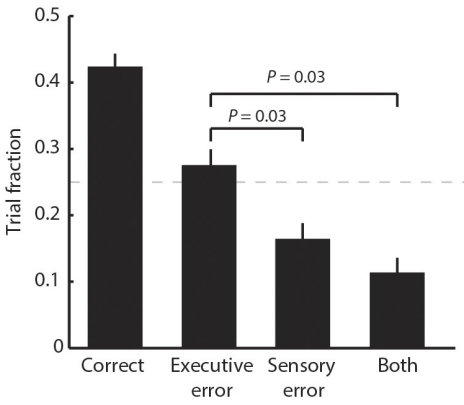
Extended Data Figure 1 | Behavioural and electrophysiological features of the 2AFC task. a, Mice display equal performance across trial types ($n = 4$ mice, $P = 0.52$, Wilcoxon rank-sum test). **b**, Multi-electrode implant used for PFC neural recordings. Inset, magnification showing electrodes. **c**, Post-mortem histology in an example brain showing electrode tip locations (arrowheads). **d**, Example of spike sorting in energy space to identify single units. Two identified clusters reflect two single units. Inset, corresponding spike-waveforms. **e**, In 17% of rule-tuned cells, tuning is observed for both task rules (example PSTHs shown), albeit with distinct temporal offsets during the delay. **f**, Schematic showing electrode locations from which rule-tuned neurons (dots) were recorded, illustrating that they are most frequently found in deeper layers. Dot sizes are scaled in proportion to the number of tuned neurons found at that location ($n = 594$ cells from four mice). **g**, Fast-spiking (FS) and regular-spiking (RS) neurons are identified on the basis of the peak to trough time of their spike waveform (left, example waveforms; right, peak to through time histogram, dashed line represents cut off for fast-spiking to regular-spiking classification[21,43]). **h**, Example rasters and PSTHs for two cells during delay periods of either 400 or 800 ms, randomized within the same recording session. In the first, an early peak is present in both conditions (left), while in the other a late peak is only evident in the 800-ms condition (right). **i**, Task-variable information for each mouse of our first cohort (manipulation free). Task-variable information is based on the PCA from the divergence of population activity of task-modulated PFC neurons on the axis associated with each variable (see Methods) and is highly informative for task rule (green), but contains no information about movement (side selection, grey). Shaded areas indicate the bootstrapped 95% CI.



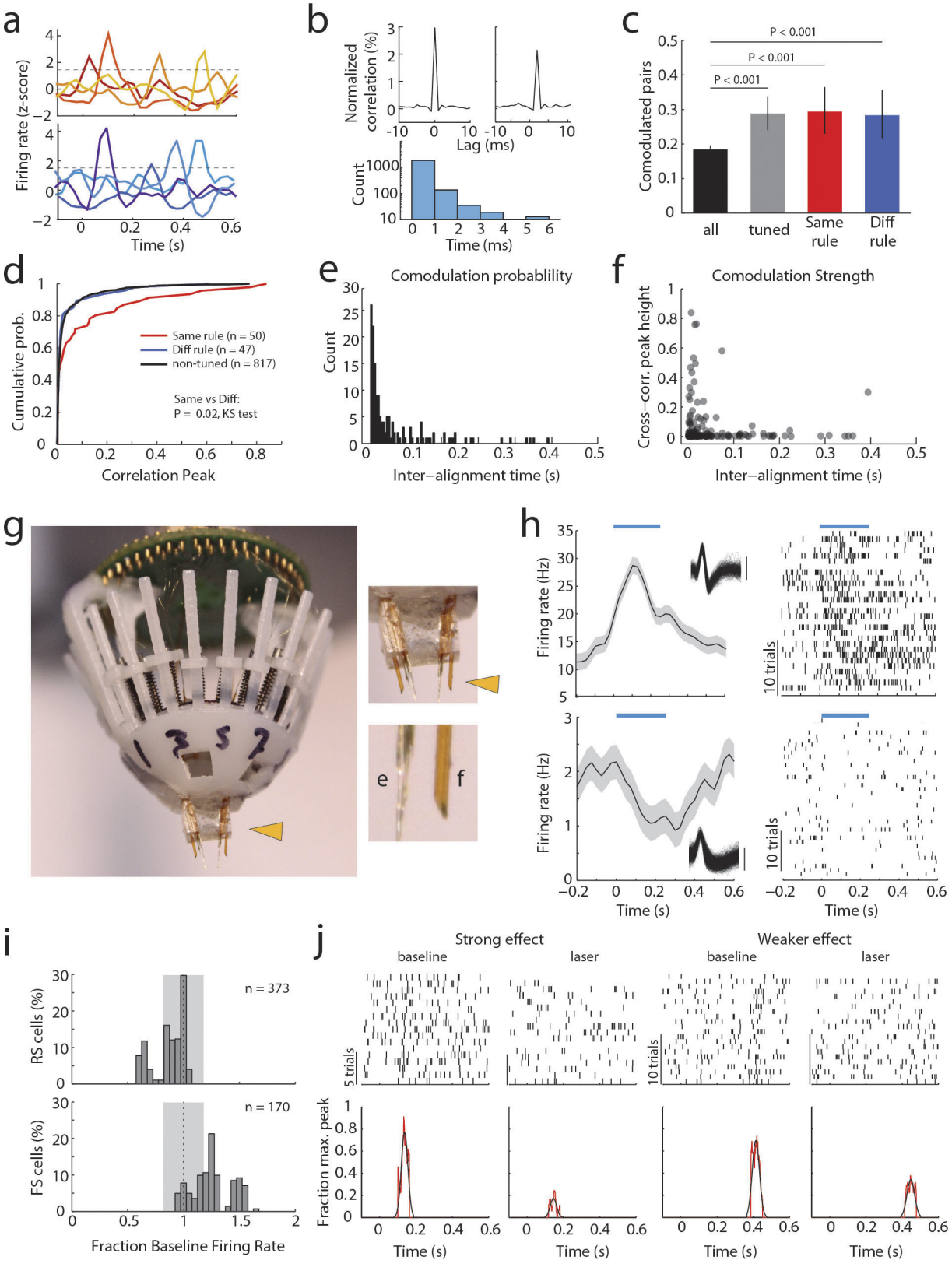


e

Response \ Target	audio left	audio right	visual left	visual right
audio left	correct	sensory	executive	both
audio right	sensory	correct	both	executive
visual left	executive	both	correct	sensory
visual right	both	executive	sensory	correct

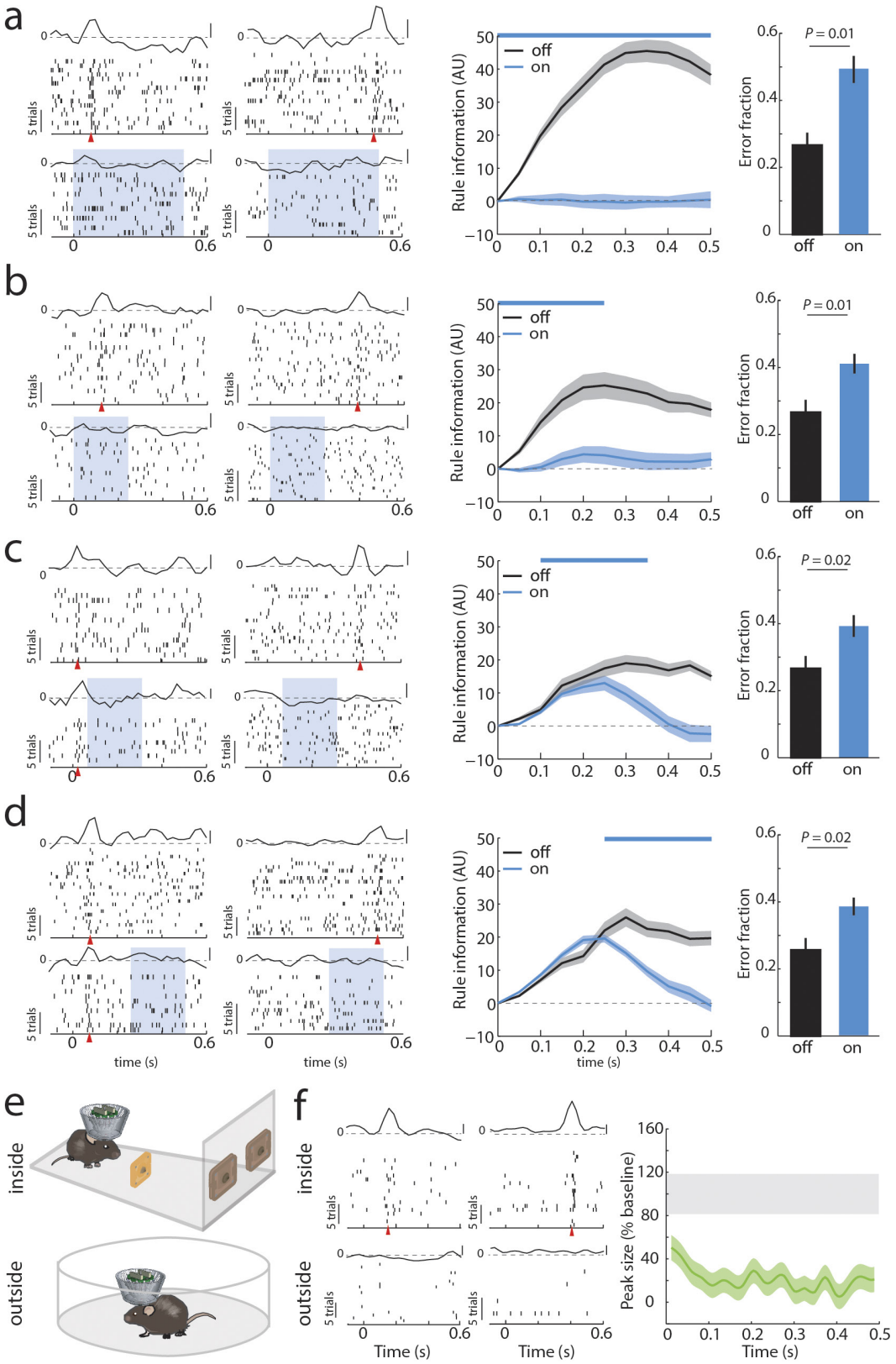


Extended Data Figure 2 | Behavioural errors are primarily driven by inappropriate rule encoding. a, Mice show comparable performance on trials with one target modality presented compared to performance in conflict trials ($n = 4$ mice, $P = 0.81$). **b**, Example PSTHs of a neuron whose appropriate tuning to the attend to vision rule is observed in error trials of the attend to audition rule. **c**, Rule information derived from PCA of sessions in which sufficient numbers of errors allowed for their analysis (93 neurons, 18 sessions from four mice) show that they contain information about the other rule; directionality of rule-related axis in error trials are along the same axis used in correct trials (see Methods). Shaded areas indicate bootstrapped 95% CI. **d**, Schematic of the 4AFC task developed to distinguish between errors related to rule encoding (executive) and those related to target cue perception (sensory, see Methods). Visual and auditory targets are reported at different response port pairs (inner versus outer), making it possible to distinguish between outcomes in which the animal makes a selection on the basis of prior cueing, the spatial location of sensory targets, both, only one or neither. **e**, 4AFC task outcomes illustrated in a confusion matrix showing outcomes conditional upon sensory target modality and location. Note that sensory conflict is not specified for these trials, as it can be either spatially congruent or in conflict with the appropriate target. **f**, Executive errors represent the majority of those observed, accounting for about 50% of all errors across mice ($n = 4$ mice). Dashed line represents chance performance (25%). All behavioural data was compared using a Wilcoxon rank-sum test.



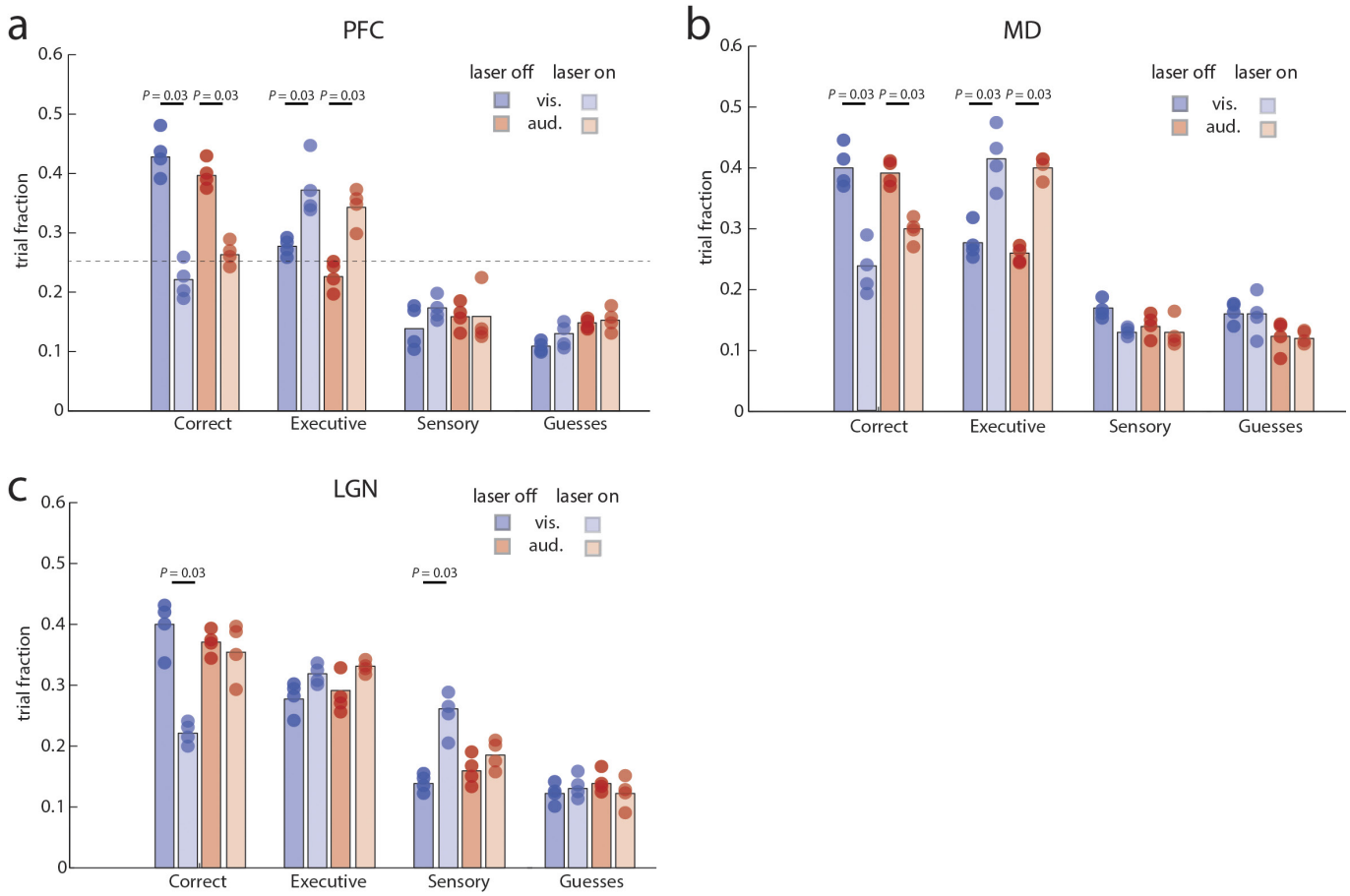
Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Combining PFC recordings with local optogenetic control of inhibitory interneurons. a, Tuning peak examples of multiple PFC neurons simultaneously recorded in a single recording session. Tuning peaks associated with either rule occur at multiple times across the delay period in different neurons suggesting precisely timed, sequential activation. **b**, Top, two examples of a short-latency cross-correlation (shuffle corrected; see Methods) observed between pairs of tuned neurons. Bottom, histogram of cross-correlation peak times ($n = 914$ pairs). **c**, Increased connection probability between tuned neurons (all, 914 pairs; tuned, 97 pairs; same rule, 50 pairs; opposite rule, 47 pairs; comparison with a binomial test). **d**, Cross-correlation strength is significantly higher for neurons representing the same rule. **e**, **f**, Co-modulation probability (**e**) and strength (**f**) show dependence on temporal distance between tuning peaks among same rule-representing pairs ($n = 138$). **g**, Photograph of a multi-electrode implant used to record from PFC with simultaneous optogenetic manipulation. Inset, enlargement of the drive component targeting bilateral PFC with optic fibres and electrodes (top) and enlargement showing electrodes and optic fibre for one hemisphere (bottom) (**e**, electrodes and **f**, optic fibre). **h**, Examples of a fast-spiking neuron that is driven (top) and a regular-spiking neuron that is inhibited (bottom) by exposure to blue light (blue bar, 473 nm). **i**, Quantification of laser effects on fast-spiking and regular-spiking cell firing rate shows that this holds true at the population level (albeit with the population mean of regular-spiking neurons being generally smaller than the example). Grey shading represents 95% CI of the no laser condition. **j**, Top, example task-modulated spike rasters showing laser effect on tuning peak. Bottom, visualization of peak strength measurement for these examples (see Methods).

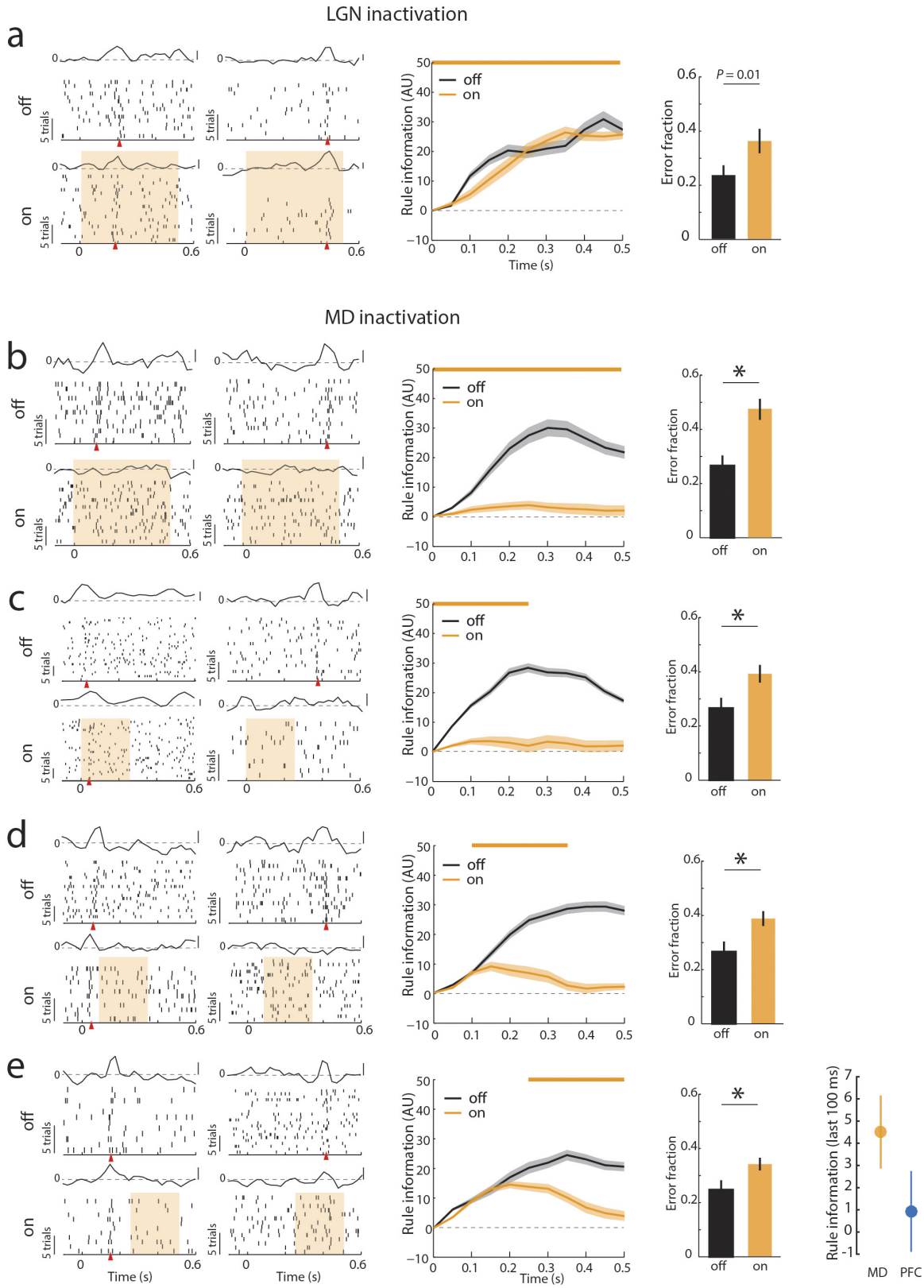


Extended Data Figure 4 | See next page for caption.

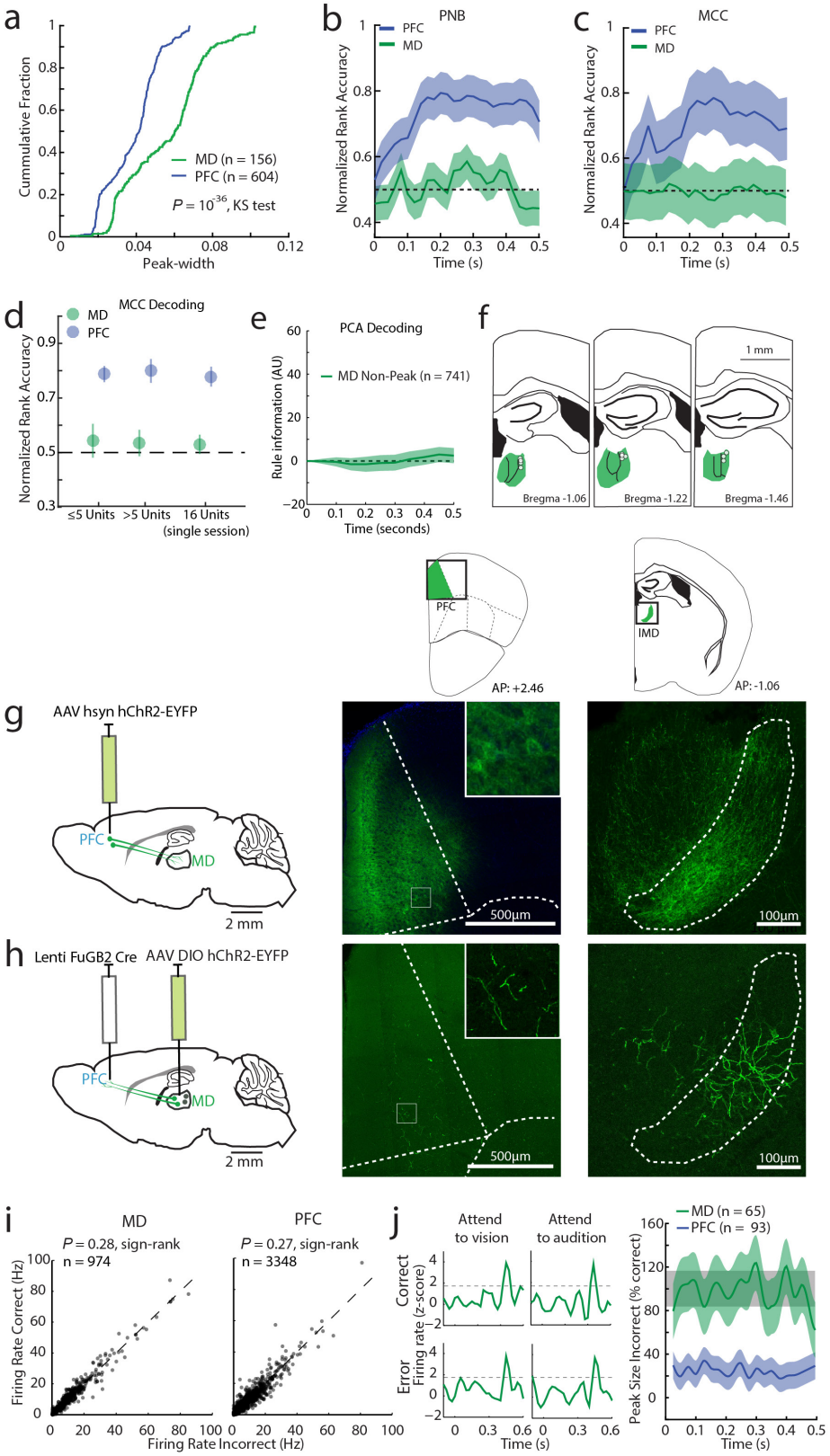
Extended Data Figure 4 | Causal evidence for task-specific sequential PFC activity maintaining rule representation. a, Effect of bilateral optogenetic enhancement of local inhibition on PFC rule tuning and behaviour. Left, raster and PSTH examples of neurons tuned either early or late in the delay (blue shading indicates laser presentation), shows loss of tuning with minimal impact on overall spiking (see Extended Data Fig. 3i for quantification of laser on spike rates). Middle, laser effect on population rule information ($n = 94$ neurons, three mice). Right, laser effect on behaviour ($n = 12$ sessions; four from each mouse). **b**–**d**, Temporally limited optogenetic manipulations show that later PFC tuning is dependent on early tuning. **b**, Manipulation limited to the first 250 ms is sufficient to mimic the effect of suppression during the full delay period on neuronal tuning with a smaller impact on behaviour probably owing to a smaller laser dose and being close to the optical fibres ($n = 52$ neurons, 12 sessions). **c**, The effect from **b** persists even when rule presentation period is spared ($n = 46$ neurons, 12 sessions). **d**, Late laser activation only impacts late activity ($n = 53$ neurons, 12 sessions). **e**, Cartoon of experimental comparison of the effect of sensory selection rule presentation inside and outside of the task. **f**, Left, example of two neurons that display tuning following rule-related cue presentation inside the task but not outside of it. Right, group quantification for population tuning shown on the left ($n = 283$ neurons from five mice). For peak size, shaded error regions show the 95% CI of the measurement, while the grey bar denotes the subsampled bootstrapped 95% CI for baseline error estimate. For rule information shaded areas indicate bootstrapped 95% CI. Wilcoxon rank-sum test was used for all behavioural comparisons.



Extended Data Figure 5 | Optogenetic dissection of error types in the 4AFC task. a, Inactivation of PFC in four VGAT–ChR2 mice during the delay period specifically increases executive errors, whereas the sensory errors remain comparable. **b**, Mediodorsal inhibition leads to a similar increase in executive errors. **c**, By contrast, LGN inactivation specifically increases sensory errors in attend to vision trials, whereas executive errors remain comparable. Coloured bars show median values and dots represent average performance of each mouse (4–5 sessions per mouse). For visual clarity, error bars were not included. Wilcoxon rank-sum test was used for all comparisons.

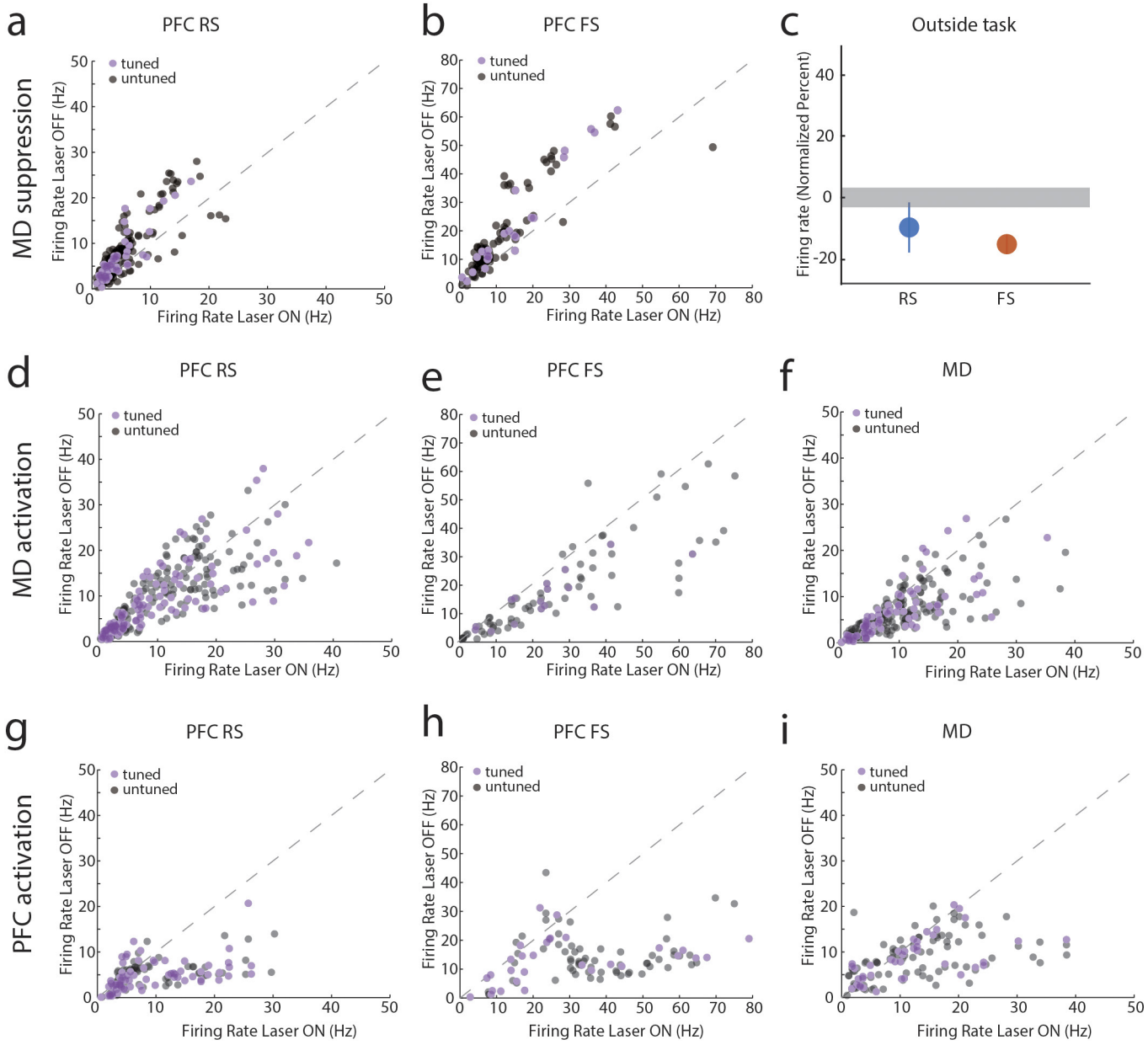


Extended Data Figure 6 | Mediodorsal recruitment by the PFC is related to delay period length in the 2AFC task. a, Bilateral optogenetic LGN suppression through activation of NpHR3.0 (orange bar) had no effect on PFC tuning during the delay period, but did increase errors in the 2AFC task. Left, raster and PSTH examples of neurons tuned either early or late in the delay (shading indicates laser presentation), shows that rule tuning persists during LGN inactivation. Middle, laser effect on population rule information over the delay ($n = 33$ cells, two mice). Right, laser effect on behaviour ($n = 2$ mice, three sessions each). **b**–**e**, Mediodorsal suppression using the same approach as in the LGN leads to loss of tuning and disrupts behavioural performance. Data are presented as example units (**b**–**e**, left), followed by the PCA for the laser on verus off conditions (second left) and behavioural impact (**b**–**d**, right; **e**, third graph). **e**, The right most graph shows the direct comparison between late PFC (Extended Data Fig. 4d (middle)) suppression and late mediodorsal suppression (second left) on PFC rule information in the last 100 ms (mean ± 95% CI).

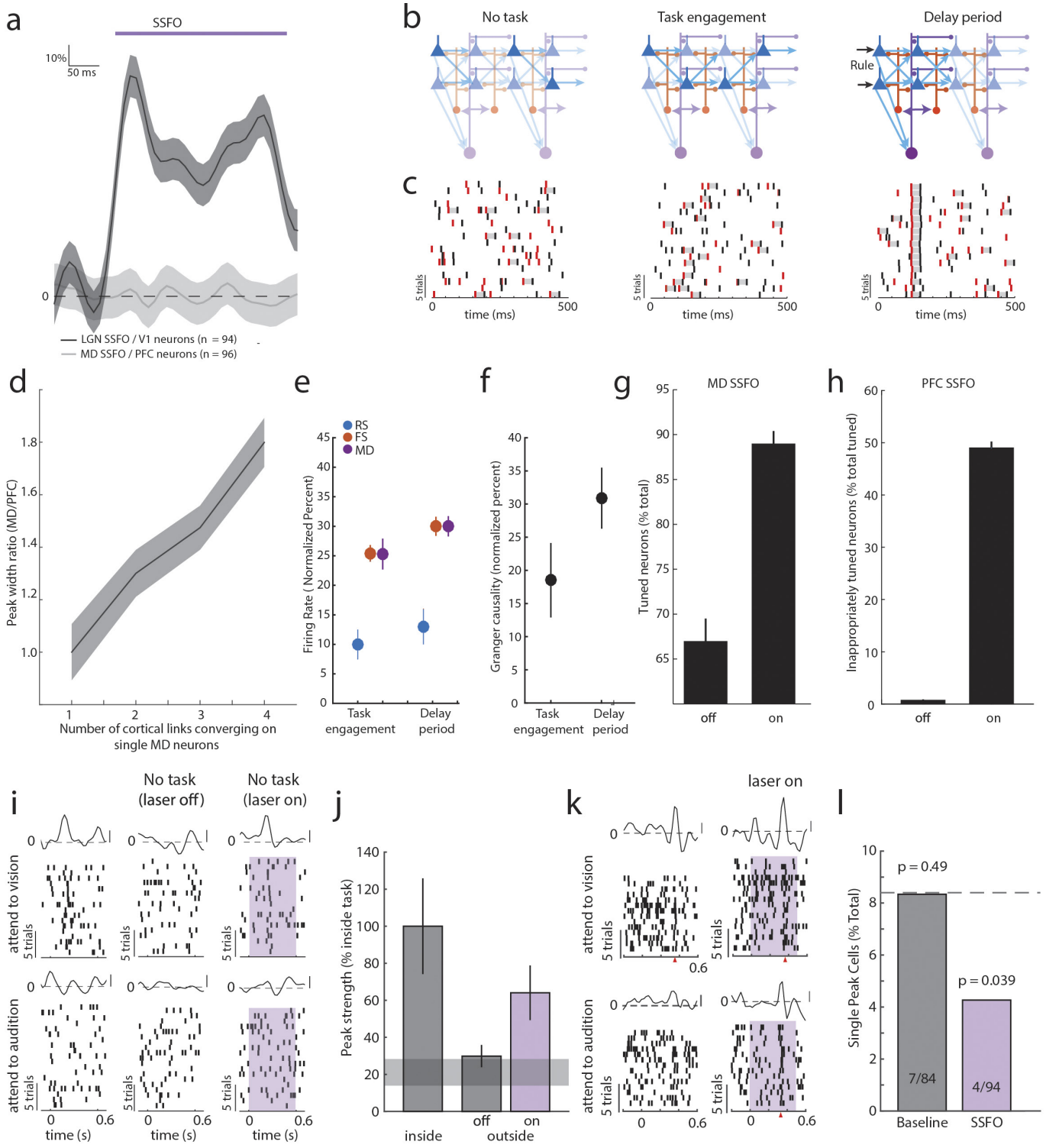


Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Connectivity pattern and response profile of mediodorsal and PFC neurons. a, Cumulative distributions of neuronal peak widths (measured as full-width at half-maximum) for mediodorsal thalamus and PFC. **b**, **c**, Two nonlinear decoding methods, PNB (**b**) and MCC (**c**), fail to reveal rule information among tuned mediodorsal neurons (PFC, $n = 604$ neurons, six mice; mediodorsal thalamus, $n = 156$ neurons, three mice). **d**, Rule information obtained from nonlinear decoding does not depend on the number of simultaneously recorded neurons. Decoding of rule information among tuned mediodorsal and PFC neurons is similar in sessions containing 1–5 neurons (PFC, $n = 318$ neurons, six mice; mediodorsal thalamus, $n = 73$ neurons, three mice) and sessions containing more than 5 neurons (PFC, $n = 286$ neurons, six mice; mediodorsal thalamus, $n = 83$ neurons, three mice). Similar results were obtained in single sessions with the highest population of simultaneously recorded mediodorsal neurons containing temporal peaks ($n = 16$), and an equivalent session containing the same number of simultaneously recorded tuned PFC neurons. Error bars are 95% CI. **e**, Rule information is not encoded by mediodorsal neurons that do not show peaks. **f**, Schematic diagram showing that tetrodes yielding mediodorsal neurons with peaks were located exclusively in the lateral mediodorsal thalamus (dots). **g**, Anterograde labelling of the PFC shows that their terminals are located in the lateral mediodorsal thalamus. **h**, Retrograde labelling from PFC identifies cells in the lateral mediodorsal thalamus. Insets show enlarged view. **i**, Firing rates are comparable in correct and incorrect trials for mediodorsal (left) or PFC (right) neurons. **j**, Left, example PSTH of a single mediodorsal neuron in correct or error trials showing similar peaks across all conditions. Right, quantification of peak size for the same rule in incorrect trials shows that mediodorsal peaks are retained, whereas PFC peaks are diminished. Shade indicates bootstrapped 95% CI.

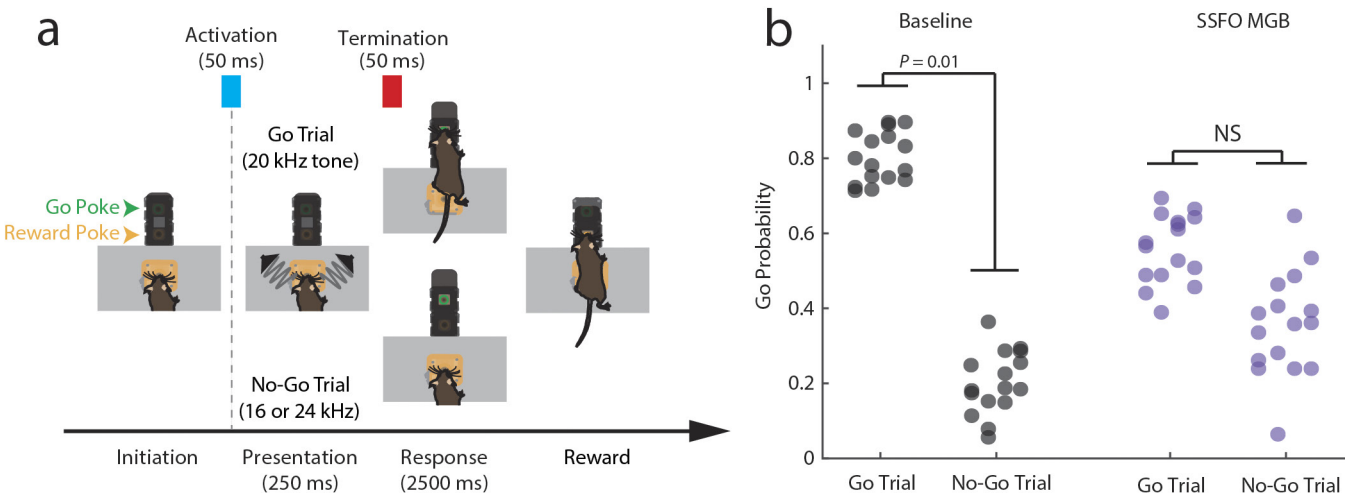


Extended Data Figure 8 | Mediodorsal thalamus to PFC and PFC to mediodorsal thalamus pathways are functionally asymmetric. **a**, **b**, Scatter plots comparing firing rates of PFC neurons during the delay period with and without mediodorsal suppression (each data point represents a neuron). Regular-spiking (**a**) and fast-spiking (**b**) cells show significantly reduced firing when the mediodorsal thalamus is optogenetically suppressed during the task delay period ($P < 0.001$). **c**, By contrast, mediodorsal suppression outside of the task only reduces fast-spiking firing rates (regular-spiking, $n = 245$ neurons fast-spiking, $n = 114$ neurons; data are presented as mean ± 95% CI; grey shading indicates 95% CI of null distribution). **d**–**f**, Increasing excitability in the mediodorsal thalamus through activation of SSFOs has no effect on the firing rate of regular-spiking neurons (**d**, $n = 303$ neurons, $P > 0.05$), but significantly increases spiking of fast-spiking (**e**, $n = 131$ neurons, $P < 0.001$) and mediodorsal spiking neurons (**f**, $n = 254$ neurons, $P = 0.001$). **g**–**i**, The same manipulation in the PFC increased firing rates in cortical regular-spiking (**g**, $n = 140$ neurons, $P < 0.001$), fast-spiking (**h**, $n = 91$ neurons, $P < 0.001$) and mediodorsal (**i**, $n = 111$ neurons, $P = 0.004$) neurons. **j**, All scatter plot data were compared using Wilcoxon signed-rank tests. Other than **c**, the scatter plots are the raw data used for the normalized values in Fig. 3.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Experimental and modelling results clarifying the attributes of the mediodorsal thalamus–PFC network. a, While LGN activation drives spiking in the V1, mediodorsal activation does not drive PFC spiking. **b**, Data-based schematic of the conceptual model showing mediodorsal, cortical fast-spiking and regular-spiking neurons in three different conditions. Triangles represent PFC regular-spiking cells tuned to a single rule which send convergent input to mediodorsal neurons (purple). The mediodorsal thalamus sends a modulatory like signal that enhances spiking in fast-spiking cells (orange) and amplifies connections among regular-spiking cells. Task engagement enhances mediodorsal activity and in turn fast-spiking neural activity. Rule information is simulated as synchronized input to starter regular-spiking neurons. **c**, Example data from the model in **b**, rasters of two regular-spiking cells (red, cell 1; black, cell 2) at different positions within a chain showing changes in activity across conditions. Overall spike rates do not change, but coordinated spiking (grey shading) increases. **d**, Systematically exploring the degree of convergence in the mediodorsal thalamus–PFC model suggests that 3–4 links in the PFC chain converge onto individual mediodorsal neurons ($n = 250$ neurons, three simulations per condition). **e**, **f**, The model captures firing rate (**e**) and connectivity changes (**f**) observed experimentally. **g**, Enhancing excitability in mediodorsal neurons by 10% significantly increases the number of rule-tuned cells in the PFC. **h**, Enhancing excitability in the PFC population by 8% markedly increased the proportion of neurons that show inappropriate tuning. **g**, **h**, Data are mean ± s.e.m.; $n = 250$ neurons, 10 simulations; $P = 0.002$, Wilcoxon rank-sum test. **i**, Example raster from a neuron tuned to one rule, showing that mediodorsal activation is sufficient to generate appropriate tuning outside the task. **j**, Population data shows that mediodorsal activation is sufficient to partially generate tuning outside the task ($n = 2$ mice, 31 tuned neurons). Grey shading indicates 95% CI of null distribution. All data are presented as mean ± 95% CI. **k**, Example of the effect of SSFO-based activation on a mediodorsal neuron containing only one peak, showing the addition of a second peak at the same time point in the opposite trial type. **l**, Relative to the population average (8.4%, dotted line), mediodorsal neurons showed significantly fewer single peaks in the SSFO condition despite the presence of an average number of single peaks in the same neurons without SSFO (cumulative binomial test versus population average).



Extended Data Figure 10 | Behavioural effects of excitability changes in MGB. a, Diagram showing task design and SSFO activation/termination timing in a Go/No-go auditory discrimination task (see Methods for task description). **b**, Comparison showing the probability of a Go response after either Go or No-go stimuli were presented across sessions (points) and mice (columns). NS, non-significant ($P = 0.52$), Wilcoxon signed-rank test.