nature neuroscience

SUPPLEMENTARY INFORMATION

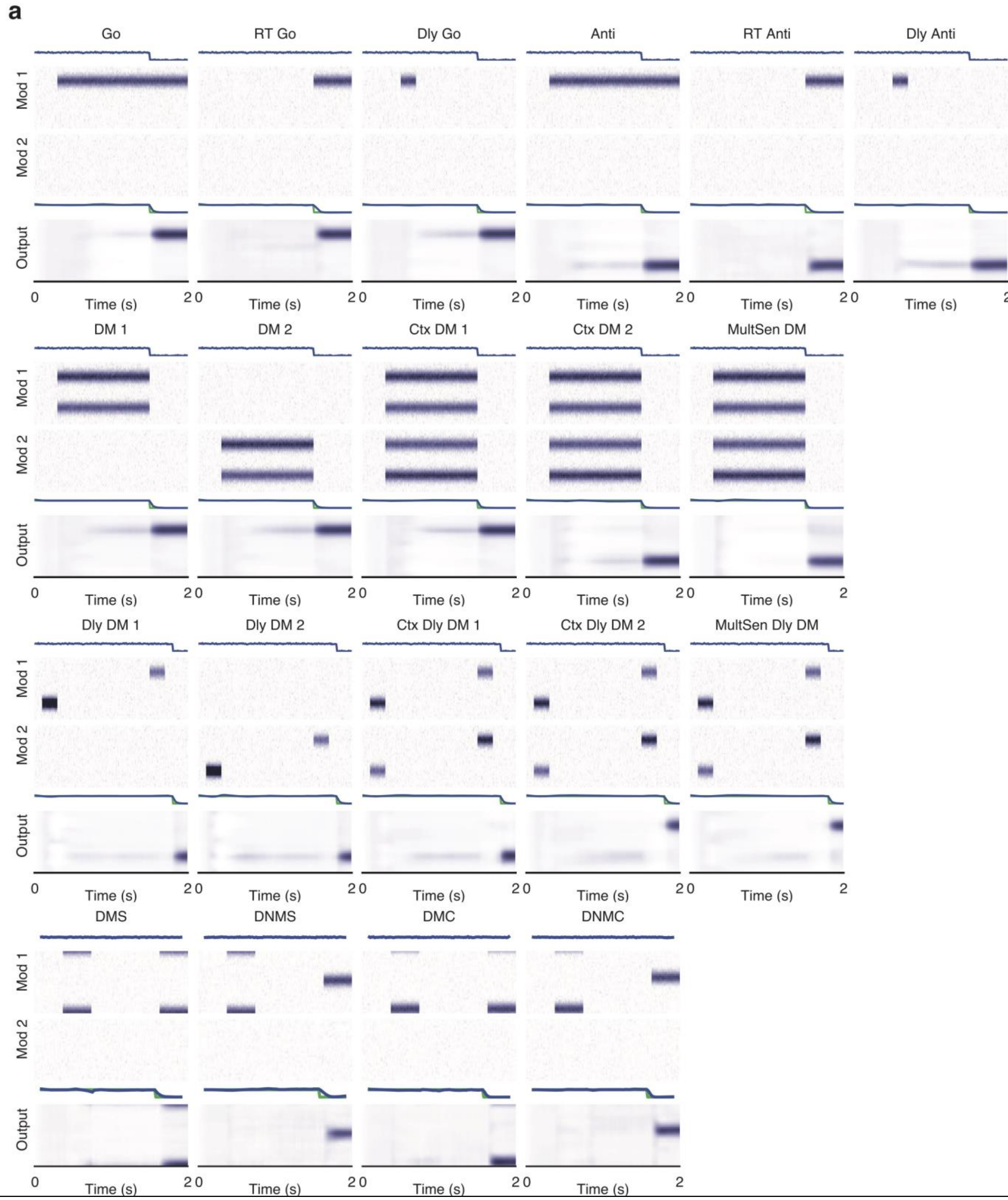
https://doi.org/10.1038/s41593-018-0310-2

In the format provided by the authors and unedited.

# Task representations in neural networks trained to perform many cognitive tasks

**Guangyu Robert Yang[1,2], Madhura R. Joglekar[1,6], H. Francis Song[1,7], William T. Newsome[3,4] and Xiao-Jing Wang[1,5*]**

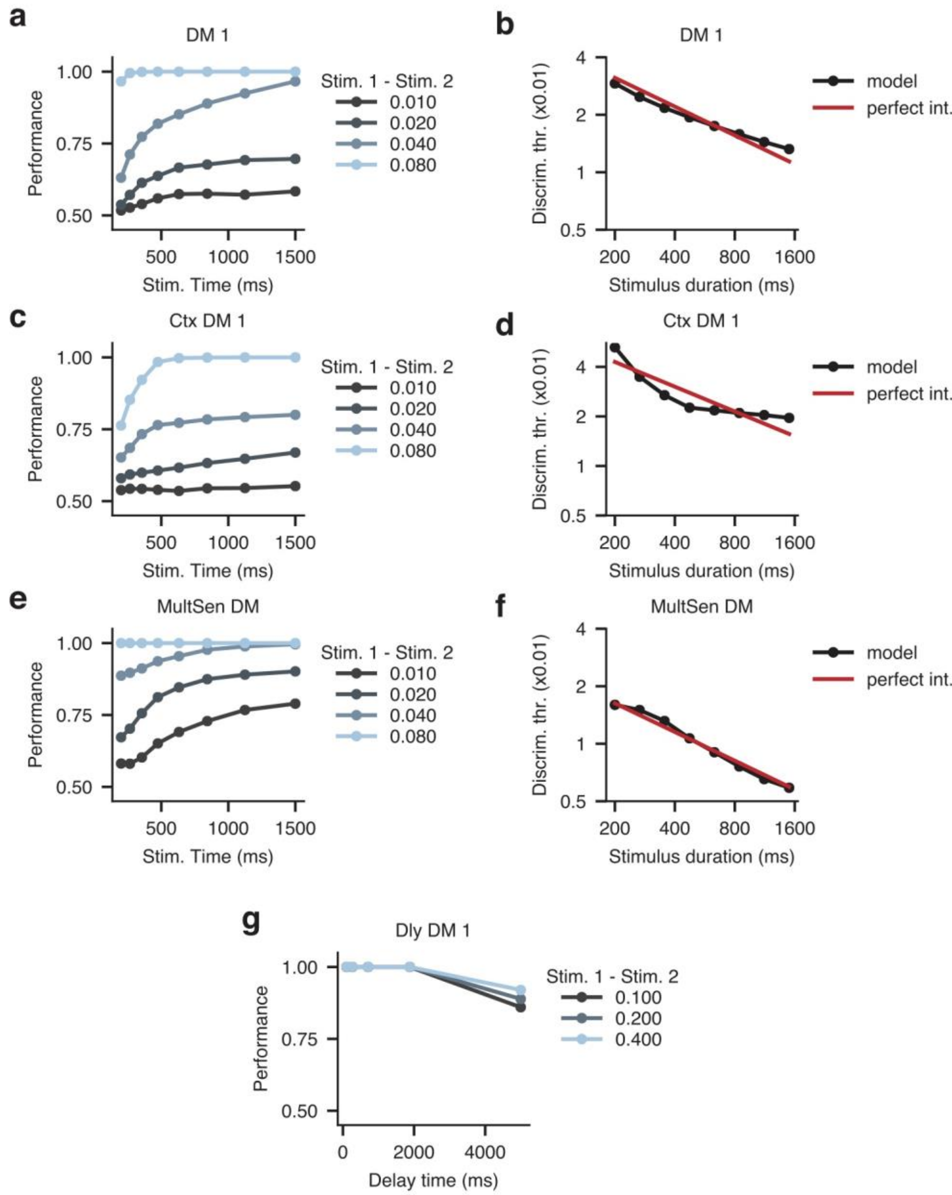
[1]Center for Neural Science, New York University, New York, NY, USA. [2]Mortimer B. Zuckerman Mind Brain Behavior Institute, Department of Neuroscience, Columbia University, New York, NY, USA. [3]Department of Neurobiology, Stanford University, Stanford, CA, USA. [4]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. [5]Shanghai Research Center for Brain Science and Brain-Inspired Intelligence, Shanghai, China. [6]Present address: Courant Institute of Mathematical Sciences, New York University, New York, NY, USA. [7]Present address: DeepMind, London, UK. *e-mail: xjwang@nyu.edu




Supplementary Figure 1

**Sample trials from the 20 tasks trained.**

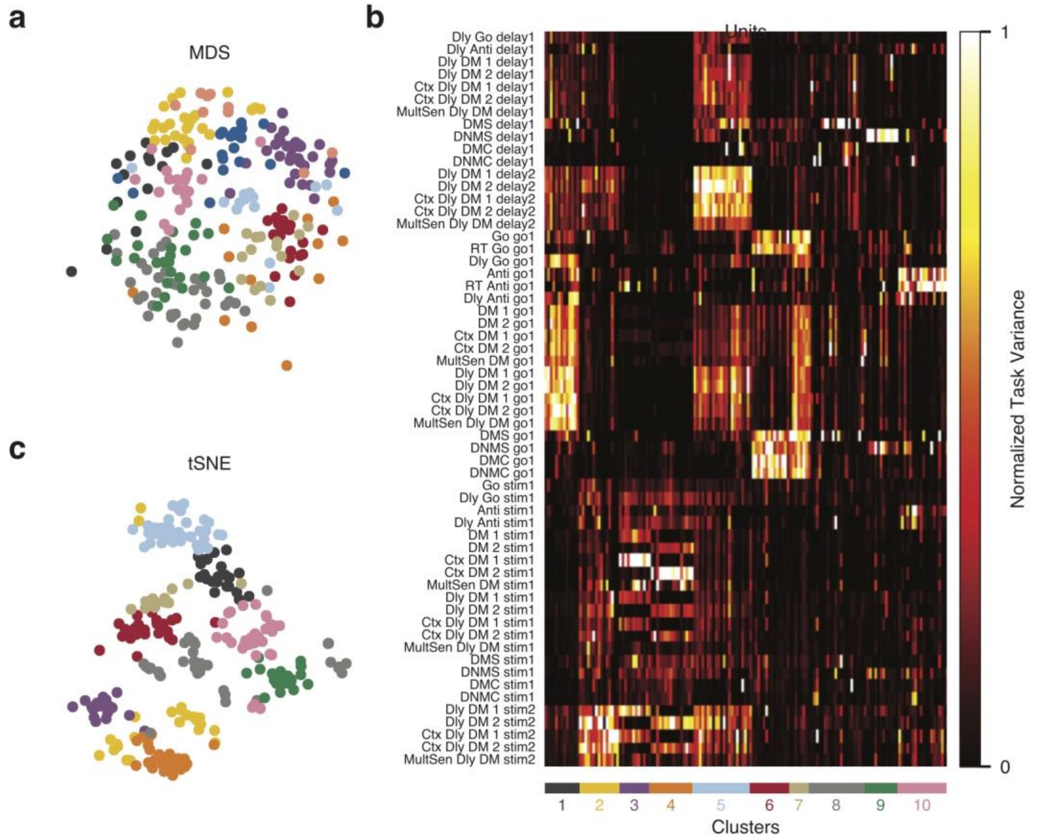
(**a**) Convention is the same as Fig. 1a. Output activities are obtained from a sample network after training. Green lines are the target activities for the fixation output unit.




**Supplementary Figure 2**

**Psychometric tests for a range of tasks.**

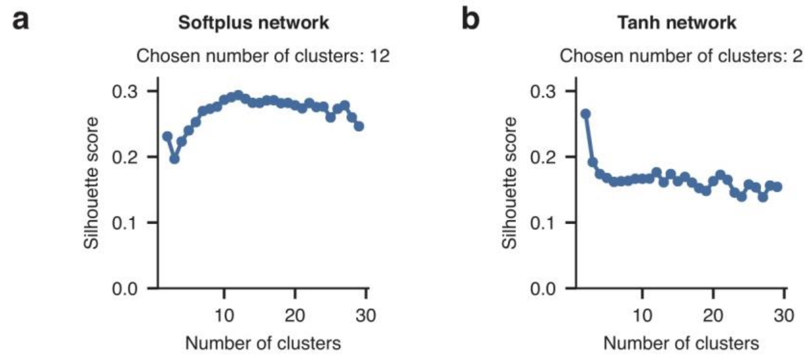
(**a**) Decision making performances improve with longer stimulus presentation time and stronger stimulus coherence in the DM 1 task in a sample reference network. (**b**) Discrimination thresholds decrease with longer stimulus presentation time in the DM 1 task. The discrimination thresholds are estimated by fitting cumulative Weibull functions. (**c-f**) Same analyses as (**a,b**) for the Ctx DM 1 (**c,d**) and MultSen DM (**e,f**) task. In all n=20 independent networks studied, performance improves with longer stimulus presentation time. However, in many networks the improvement is different from that expected of perfect integration (red line). This variation has no impact on other results. (**g**) A sample network is able to perform well above chance in the Dly DM 1 task for a delay period of up to five seconds.




**Supplementary Figure 3**

**Task and epoch variances.**

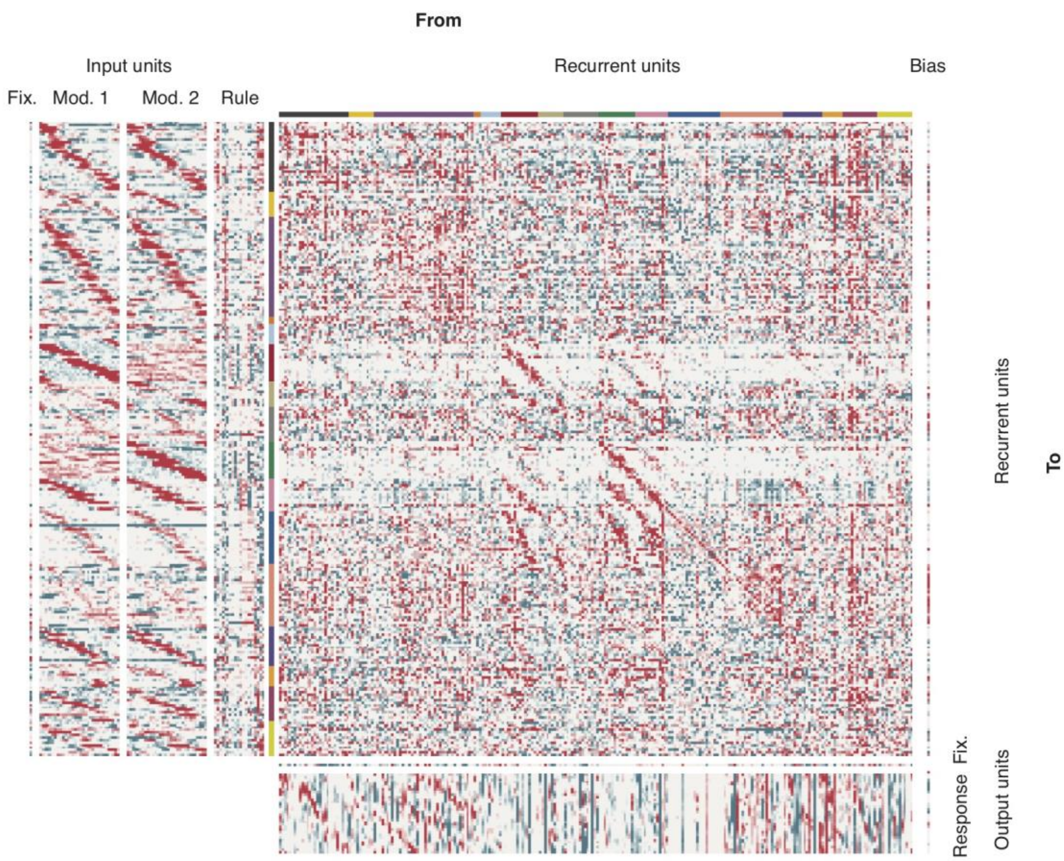
(**a**) Visualization of the task variance map using classical multi-dimensional scaling (MDS). MDS tends to preserve global structures, while tSNE tends to emphasize local structures (e.g., clustering). (**b**) Epoch variance is computed in a similar way to task variance, except that it is computed for individual task epochs instead of tasks. There are clusters of units that are selective in specific epochs. (**c**) Visualization of the epoch variance map in the same style as Fig. 2d.




**Supplementary Figure 4**

**Determining number of clusters.**

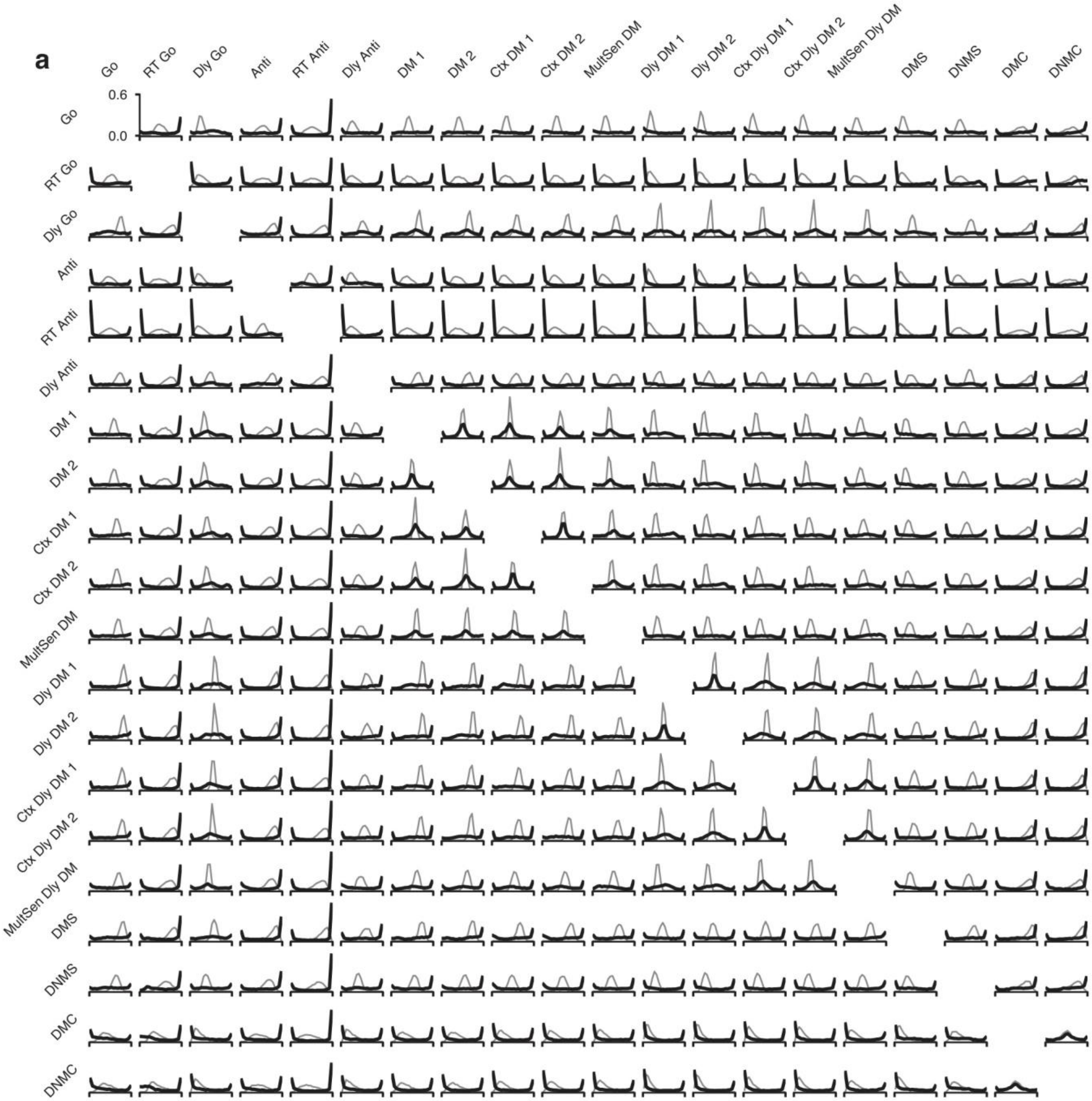
The silhouette score as a function of the number of clusters for an example network with the Softplus activation function (**a**) and one with the Tanh activation function (**b**). The silhouette score assesses the quality of a clustering scheme (see Online Methods). The "optimal" or natural number of clusters is chosen to be the one with the highest silhouette score.




**Supplementary Figure 5**

**Connectivity matrix.**

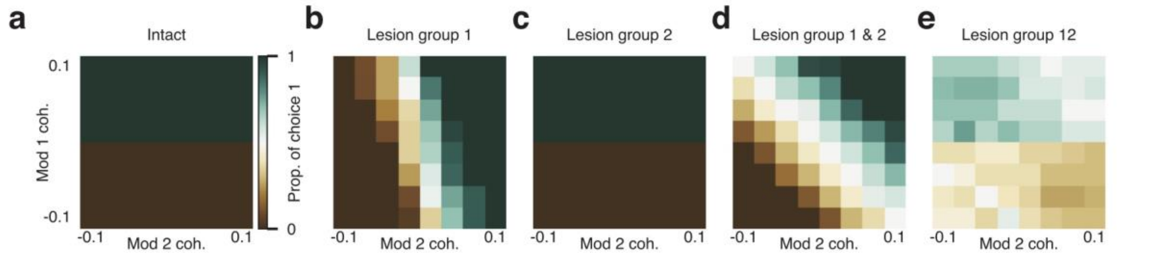
The full connectivity matrix for an example reference network. The network units are first sorted according to their cluster identity. Within each cluster, the units are sorted according to their preferred input directions, as defined by the input direction making the strongest connection weights to each unit (summed across modality 1 and 2). Color range is determined separately for each sub-matrix for better visualization. Red means more excitatory and blue means more inhibitory.




**Supplementary Figure 6**

**Fractional variance distributions for all pairs of tasks.**

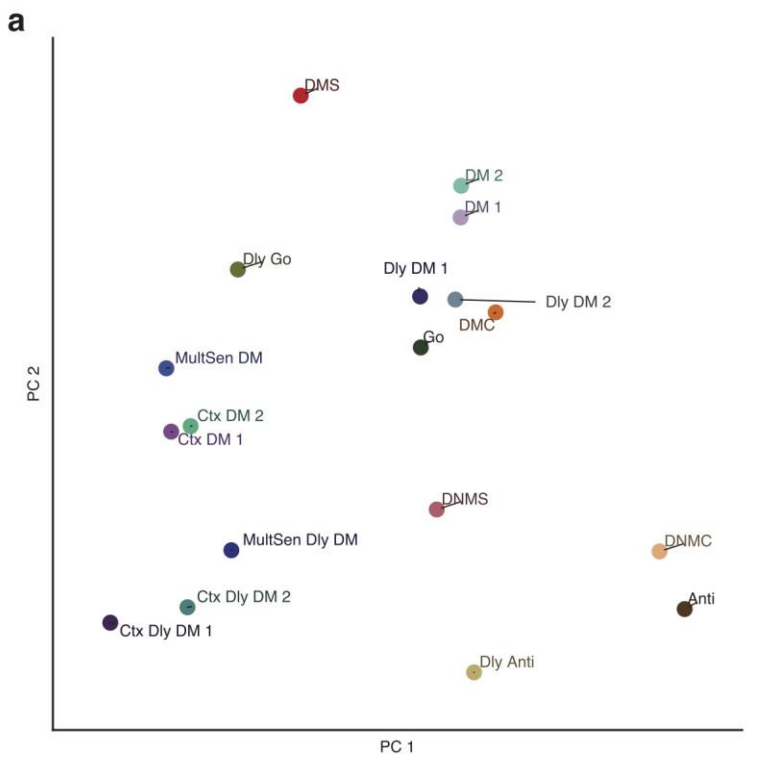
(**a**) There is a total of 190 unique pairs of tasks from all 20 tasks trained. Each fractional variance distribution (black) shown here is averaged across 20 independently trained networks. As a control, we also computed fractional variance distributions (gray) from activities of surrogate units that are generated by randomly mixing activities of the original network units (see Online Methods). The y-axis range is shared across all plots.




**Supplementary Figure 7**

**Detailed behavioral effect of lesioning on the Ctx DM 1 task.**

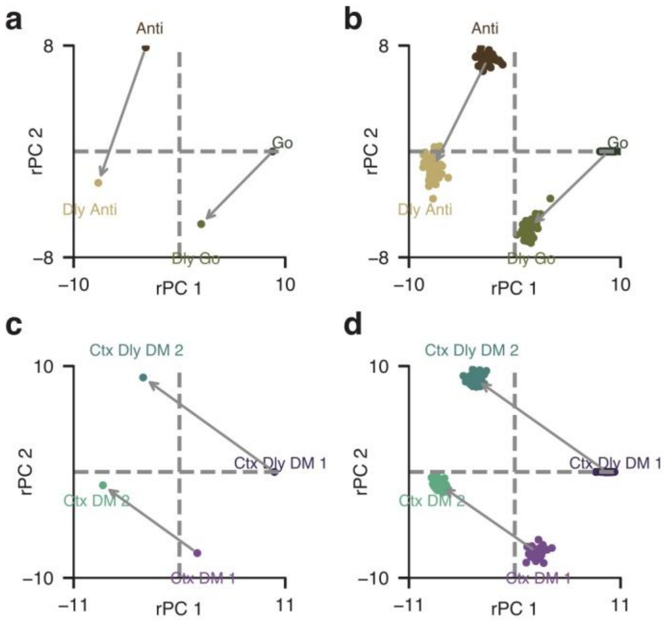
(**a-e**) The network choice in the Ctx DM 1 task for different combinations of modality 1 and modality 2 coherence in various networks. (**a**) The intact network's choice only depends on the coherence of modality 1. (**b**) Lesioning group 1 makes the network more dependent on the coherence of modality 2. (**c**) Lesioning group 2 has no impact for the Ctx DM 1 task. (**d**) Lesioning both group 1 and 2 allow the network to weigh both modalities equally. (**e**) Lesioning group 12 led to failure in making decisions. Although some preference towards modality 1 is preserved, the network is largely unable to choose decisively.




**Supplementary Figure 8**

**Representation of all tasks in state space.**

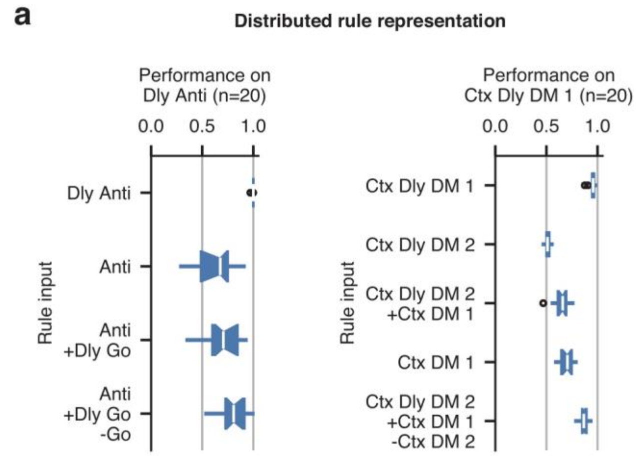
(**a**) The representation of each task is computed the same way as in Fig. 6. Here showing the representation of all tasks in the top two principal components. RT Go and RT Anti tasks are not shown here because there is no well-defined stimulus epoch in these tasks.




**Supplementary Figure 9**

**Visualization of connection weights of rule inputs.**

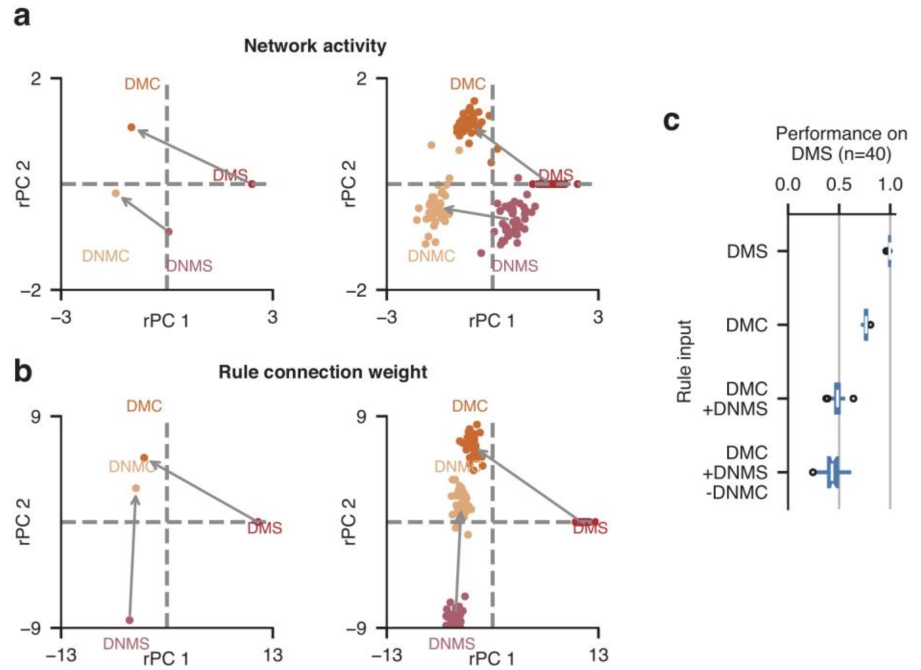
(**a**) Connection weights from rule input units representing Go, Dly Go, Anti, Dly Anti tasks visualized in the space spanned by the top two principal components (PCs) for a sample network. Similar to Fig. 6, the top two PCs are rotated and reflected (rPCs) to form the two axes. (**b**) The same analysis as in (a) is performed for 40 networks, and the results are overlaid. (**c**) Connection weights from rule input units representing Ctx DM 1, Ctx DM 2, Ctx Dly DM 1, and Ctx Dly DM 2 tasks visualized in the top two PCs for a sample network. (**d**) The same analysis as in (c) for 40 networks.




**Supplementary Figure 10**

**Distributed rule representation.**

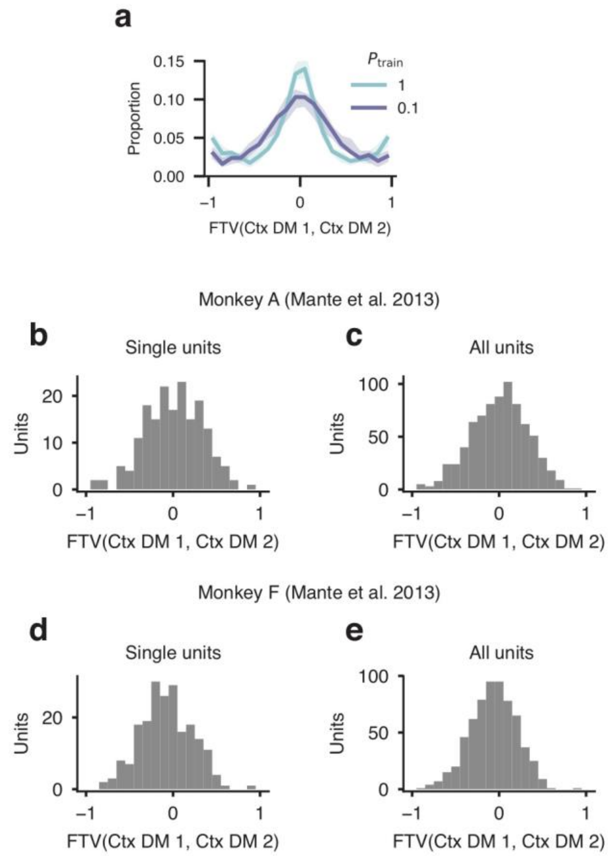
(**a**) The same analysis and box-plot convention as Fig. 7b,c, except that the networks are trained using distributed, instead of one-hot, rule representations.




**Supplementary Figure 11**

**Lack of compositionality for the family of matching tasks.**

(**a**) Visualization of task-based network activity for the DMS, DNMS, DMC, and DNMC tasks, for an example network (left) and for 40 networks (right). These plots have the same style as Fig. 6. (**b**) Visualization of connection weights for the same set of tasks in an example network (left) and for 40 networks (right). The rule weights are not compositional. These plots have the same style as Supplementary Fig. 9. (**c**) The DMS task can not be performed with a compositional rule input. The box plot convention is the same as the one in Fig. 7b.




**Supplementary Figure 12**

**Partially plastic networks and experimental data.**

(**a**) Networks where only 10% of connection weights are trained show a mixed FTV distribution for the Ctx DM 1 and Ctx DM 2 tasks. Solid lines are median over 60 networks. Shaded areas indicate the 95% confidence interval of the median estimated from bootstrapping. (**b-e**) FTV distributions derived from experimental data (reference 11). (**b**) Monkey A, single units. (**c**) Monkey A, all units. (**d**) Monkey F, single units. (**e**) Monkey F, all units.

| Task name | Abbreviation | Task family | Potential cognitive processes involved | Reference |
|---|---|---|---|---|
| Go | Go | Go | Pro-response | N/A |
| Reaction-time go | RT Go | Go | Pro-response | N/A |
| Delayed go | Dly Go | Go | Pro-response, Working memory | Funahashi 89 |
| Anti-response | Anti | Anti | Anti-response | Munoz 04 |
| Reaction-time anti-response | RT Anti | Anti | Anti-response | Munoz 04 |
| Delayed anti-response | Dly Anti | Anti | Anti-response, Working memory | Munoz 04 |
| Decision making 1 | DM 1 | DM | Decision making | Gold 07 |
| Decision making 2 | DM 2 | DM | Decision making | Gold 07 |
| Context-dependent decision making 1 | Ctx DM 1 | DM | Decision making, Gating mod 1 | Mante 13 |
| Context-dependent decision making 2 | Ctx DM 2 | DM | Decision making, Gating mod 2 | Mante 13 |
| Multi-sensory decision making | MultSen DM | DM | Decision making, Integrating mod 1 and 2 | Raposo 14 |
| Delayed decision making 1 | Dly DM 1 | Dly DM | Working memory | Romo 99 |
| Delayed decision making 2 | Dly DM 2 | Dly DM | Working memory | Romo 99 |
| Context-dependent delayed decision making 1 | Ctx Dly DM 1 | Dly DM | Working memory, Gating mod 1 | N/A |
| Context-dependent delayed decision making 2 | Ctx Dly DM 2 | Dly DM | Working memory, Gating mod 2 | N/A |
| Multi-sensory delayed decision making | MultSen Dly DM | Dly DM | Working memory, Integrating mod 1 and 2 | N/A |
| Delayed match-to-sample | DMS | Matching | Working memory, Comparison | Miller 96 |
| Delayed non-match-to-sample | DNMS | Matching | Working memory, Comparison | Miller 96 |
| Delayed match-to-category | DMC | Matching | Categorization, Working memory, Comparison | Freedman 16 |
| Delayed non-match-to-category | DNMC | Matching | Categorization, Working memory, Comparison | Freedman 16 |

Supplementary Table 1: Names and abbreviations of all tasks trained in the networks. Most of the trained tasks are derived from archetypal cognitive tasks used in non-human animal experiments. We grouped our tasks into five task families. We list cognitive processes that are potentially involved in each task. We are not aware of experimental studies that investigated the Ctx Dly DM 1, Ctx Dly DM 2, or MultSen Dly DM tasks in non-human animals.