

and update our understanding of dopamine in action, motivation and learning.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Salamone, J.D. *et al. Curr. Top. Behav. Neurosci.* published online, doi:10.1007/7854_2015_383 (1 September 2015).
2. Berridge, K.C. *Eur. J. Neurosci.* **35**, 1124–1143 (2012).

3. Montague, P.R., Dayan, P. & Sejnowski, T.J. *J. Neurosci.* **16**, 1936–1947 (1996).
4. Robbins, T.W. & Everitt, B.J. *Psychopharmacology (Berl.)* **191**, 433–437 (2007).
5. Syed, E.C.J. *et al. Nat. Neurosci.* **19**, 34–36 (2016).
6. Hamid, A.A. *et al. Nat. Neurosci.* **19**, 117–126 (2016).
7. Guitart-Masip, M. *et al. Neuroimage* **62**, 154–166 (2012).
8. Lobb, C.J., Troyer, T.W., Wilson, C.J. & Paladini, C.A. *Front. Syst. Neurosci.* **5**, 25 (2011).
9. Cockburn, J., Collins, A.G. & Frank, M.J. *Neuron* **83**, 551–557 (2014).
10. Hart, A.S., Rutledge, R.B., Glimcher, P.W. & Phillips, P.E.M. *J. Neurosci.* **34**, 698–704 (2014).
11. Howe, M.W., Tierney, P.L., Sandberg, S.G., Phillips, P.E.M. & Graybiel, A.M. *Nature* **500**, 575–579 (2013).
12. Collins, A.G.E. & Frank, M.J. *Psychol. Rev.* **121**, 337–366 (2014).
13. Cragg, S.J. *Trends Neurosci.* **29**, 125–131 (2006).
14. Fiorillo, C.D. & Tobler, P.N. *Science* **299**, 1898–902 (2003).
15. Gershman, S.J. *Neural Comput.* **26**, 467–471 (2014).

Parietal and prefrontal: categorical differences?

Daniel Birman & Justin L Gardner

A working memory representation goes missing in monkey parietal cortex during categorization learning, but is still found in the prefrontal cortex.

If you can imagine reading to the end of this sentence and forgetting what was written at the beginning, you may start to appreciate how critical working memory is to much of what we take to be higher cognition. Indeed, a life without such short-term memory would lurch between disconnected events, threatening not just our cognitive abilities, but the core continuity of our conscious selves. The finding in the 1980s that, during delay periods in which monkeys remembered the location of an instructed eye movement, prefrontal¹ and parietal² neurons display persistent, spatially specific activity cemented the idea that these cortical areas are allied in serving this critical memory function. However, in this issue of *Nature Neuroscience*, Sarma *et al.*³ report parietal neurons in the lateral intraparietal area (LIP) to be unexpectedly and puzzlingly forgetful, whereas their counterparts in the prefrontal cortex are not.

This finding comes from experiments probing another hallmark cognitive function for which parietal and prefrontal neurons appear to share responsibility: categorization. Categorization is our ability to generalize properties of, say, an apple across many exemplars with incidental differences in size, color or shape. Without categorization, each and every apple might have to be individually memorized. Categorization is clearly a foundational cognitive capacity; we use it not just when we recognize an apple, but when we distinguish specific states of importance, such as whether it is edible or rotten. We might imagine that categorical decisions are just as critical for a monkey as for a human.

But how exactly does one get a monkey to make categorical decisions repeatedly and in a controlled way so that the neural representations can be systematically studied?

Continuing in the tradition of their laboratory⁴, Sarma *et al.*³ have developed formidable skills in training monkeys to do just this. In the current work, they shaped behavior incrementally, starting with a sequence of simpler tasks before making the leap into categorical decisions. The authors trained monkeys to perform a delayed match-to-sample task in which the monkeys were required to remember the direction of a patch of briefly presented moving dots and release a lever only if the direction of a second dot patch, presented after a short delay, exactly matched the remembered direction. After monkeys reached criterion performance, they were trained on the full categorization task, which was identical except for one crucial difference. In the categorization task, the monkeys were trained to release the lever not when the two directions were identical but when they were in the same experimenter chosen category (that is, moved in the same direction relative to an arbitrary category boundary). A series of studies has provided abundant support that, after such categorization training, neurons in LIP and prefrontal cortex show a beautifully simple and stable representation of the category, preferring stimuli in either one category or the other⁴.

The new insight concerning working memory came from a relatively simple proposition that led to an unexpected result: record the activity of neurons before as well as after the training of the categorization task. Given that, after training, LIP neurons encode category during the delay period, might one expect that a working memory representation encoding direction would be present for the delayed match-to-sample task? Oddly, the answer is no.

Despite the fact that, to perform the task, the monkey must have some working memory of the direction of the dots, neurons in LIP showed near chance-level encoding of motion direction during the delay period. This lack of delay-period selectivity is not a result of information about motion direction not entering LIP; during stimulus presentation, the motion direction was clearly represented. The neurons just seem to forget the direction when it most matters. Notably, the nearby medial superior temporal area, MST, which also receives direction-selective information from the medial temporal motion selective area MT, has by contrast been shown to carry working memory representations of motion direction⁵.

Although clearly a provocative finding, task and training are intertwined in the experiments as outlined above, making it unclear which accounts for the difference in LIP activity. Parietal working memory activity might arise only after sufficient training. Given that the delayed match-to-sample task was trained first, it is possible that lack of working memory representation comes not from a difference in task, but from a lack of training. The authors have provided a detailed view into their training structure, which helps to allay this concern. In particular, the monkeys were extensively trained on the delayed match-to-sample task (hundreds of daily sessions and hundreds of thousands of trials), and their performance had plateaued. Although sorting out training and task by design awaits future replication studies, clearly the monkeys had substantial experience with the delayed match-to-sample task, therefore suggesting that task, and not training, accounts for the difference.

Potential alternative explanations aside, pause for a second to consider the extensive training regime. Why does it take monkeys so long to learn so little? It seems intuitive that a

Daniel Birman and Justin L. Gardner are in the Department of Psychology, Stanford University, Stanford, California, USA.
e-mail: jlg@stanford.edu

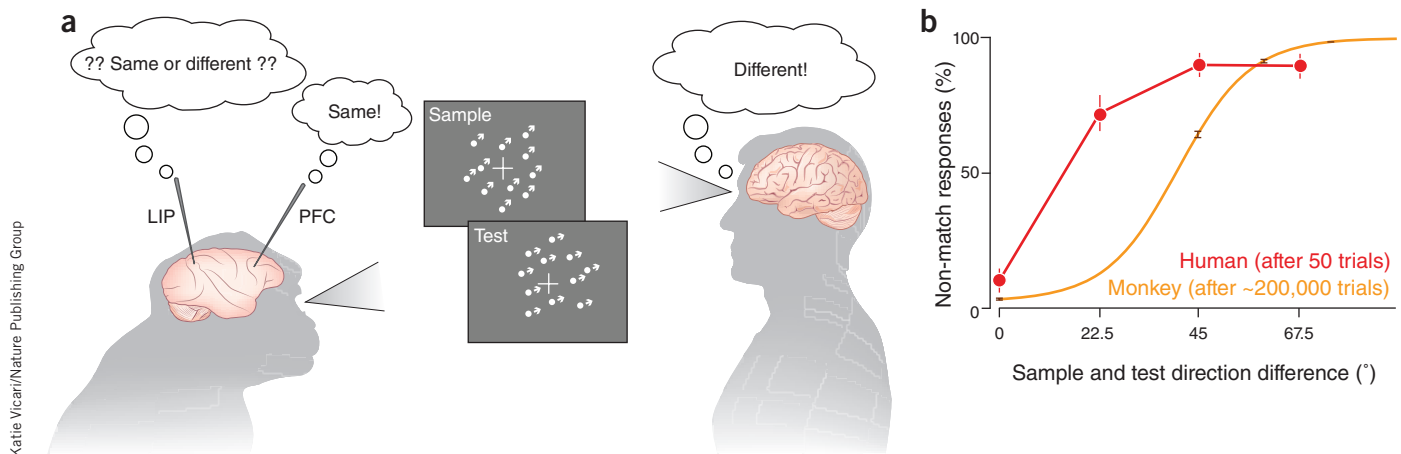


Figure 1 Humans learn a delayed match-to-sample task much faster and more thoroughly than monkeys. **(a)** Sarma *et al.*³ found that LIP is distinguished from prefrontal cortex (PFC) in that, surprisingly, it does not show a working memory representation (left) for the direction of dots in a delayed match-to-sample task (center). Humans whom we tested on Mechanical Turk (right) quickly and easily learned this task through feedback, even without being verbally told the rule. **(b)** A 22.5° difference in direction between sample and task would typically result in humans reporting “different,” whereas monkeys would typically report “the same” even after performance had plateaued after hundreds of days of training. Monkey data adapted from Sarma *et al.*³, Nature Publishing Group; error bars represent s.e.m. Human data are presented as mean and s.e.m. over 45 subjects.

human could learn both tasks extremely rapidly and attain perfect performance in a single leap once they understand the rule. If this distinction exists, what does that mean for extrapolating the results of these animal experiments to human cognition? Verbally explaining a rule gives a distinct advantage, but how easy is it actually to learn the match-to-sample and category rules given that they are not explicitly instructed, but implicitly instructed through feedback? Out of curiosity, we tested our intuitions about human category learning by having human participants on Mechanical Turk, an online platform that enables recruitment of subjects for scientific experiments, learn these two tasks implicitly. The details of the stimulus were kept as similar as possible to those used for the monkey experiments. We stress that these results have not been peer-reviewed and are provided here simply to stimulate thought and discussion. All data and experimental code are accessible from <http://gru.stanford.edu/doku.php/categorization/>.

Counterintuitively, we found that human learning of categorization resembled monkey learning, whereas delayed match to sample performance suggested stark differences. Of 63 subjects trained on the category task (without prior training on the match-to-sample task), only 5 (8%) successfully learned the rule and could verbalize it within 150 trials. Subjects in general showed incremental improvements. Although many demonstrated performance that was clearly above chance, they came away with various superstitions that did not match the actual category rule. In contradistinction, the delayed match-to-sample task was dead easy for humans. Performance after 75 trials was nearly perfect, and 42 of the 45 subjects

(93%) could verbalize the rule exactly. Of course, there are differences in the amount of evidence each trial of these two tasks provides for potential rules, but the match to sample rule nonetheless appeared to be highly intuitive for human subjects.

These informal results suggest that humans are more like monkeys when learning the categorization task and differentiate themselves dramatically from monkeys in the delayed match-to-sample task. Indeed, after extensive training of the direction match-to-sample task, the monkeys persistently make large percentages of errors when directions are as different as 22.5°, whereas humans rarely made such errors after 75 trials (**Fig. 1**). The monkey behavior is puzzling, as monkeys are able to tell that even a few degrees of difference in motion direction are different (at least around vertical⁶) and 22.5° differences should be easily visually differentiable. It is perhaps unknowable what monkeys have conceptualized about the tasks, but it suggests that the notion of sameness, so easy for humans to learn, is somehow much less natural for monkeys. One can only speculate about what could be different for the monkeys about matching a direction to sample such that this task would fail to result in stable working memory representations in LIP. Perhaps potential task strategy differences may also help to sort out the human literature, which has reported conflicting observations on whether stimulus-specific working memory representations are represented in parietal cortex^{7,8}.

An intriguing possibility is that lability and flexibility in how monkeys perform the task is associated with dynamic rather than stable representations. A hint, in the results of Sarma

*et al.*³, that this might be the case is that, in prefrontal cortex during the delayed match-to-sample task, working memory representations appear to not be completely stationary. That is, the representation of direction for neurons changes dynamically as a function of time, such that classifiers built at one time point are not as predictive of responses at other time points. If true, this suggests a potential link with views that cognitive functions are represented in dynamically changing activity^{9,10}.

LIP has been something of a field of dreams for neuroscientists looking for the neural basis of various cognitive functions. A war of ideas broke out in the 1990s over whether this associative area should be thought of in sensory or action terms; that is, whether we should use labels such as attention or intention. A field filled with different ideas grew up from this debate as it was discovered that responses in LIP are not just spatial, but are biased by factors reflecting many of the conceptual revolutions in systems neuroscience¹¹. Different groups sought and found evidence for cognitive functions in LIP, from value-based computations¹², evidence accumulation in decision making¹³, categorization⁴, to perceptual learning¹⁴ and more. Perhaps all of these different representations are multiplexed^{10,15}. However, the findings of Sarma *et al.*³ stand out not for what was found in LIP, but for what was not found. Though perhaps different from the spatial representation that brought LIP to the forefront of systems neuroscience, it is nonetheless somehow ironic and perplexing that it is a working memory representation that has gone missing.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

- Funahashi, S., Bruce, C.J. & Goldman-Rakic, P.S. *J. Neurophysiol.* **61**, 331–349 (1989).
- Gnadt, J.W. & Andersen, R.A. *Exp. Brain Res.* **70**, 216–220 (1988).
- Sarma, A., Masse, N.Y., Wang, X.-J. & Freedman, D.J. *Nat. Neurosci.* **19**, 143–149 (2016).
- Freedman, D.J. & Assad, J.A. *Nature* **443**, 85–88 (2006).
- Mendoza-Halliday, D., Torres, S. & Martinez-Trujillo, J.C. *Nat. Neurosci.* **17**, 1255–1262 (2014).
- Purushothaman, G. & Bradley, D.C. *Nat. Neurosci.* **8**, 99–106 (2005).
- Riggall, A.C. & Postle, B.R. *J. Neurosci.* **32**, 12990–12998 (2012).
- Ester, E.F., Sprague, T.C. & Serences, J.T. *Neuron* **87**, 893–905 (2015).
- Mante, V., Sussillo, D., Shenoy, K.V. & Newsome, W.T. *Nature* **503**, 78–84 (2013).
- Raposo, D., Kaufman, M.T. & Churchland, A.K. *Nat. Neurosci.* **17**, 1784–1792 (2014).
- Gottlieb, J. & Snyder, L.H. *Curr. Opin. Neurobiol.* **20**, 731–740 (2010).
- Platt, M.L. & Glimcher, P.W. *Nature* **400**, 233–238 (1999).
- Roitman, J.D. & Shadlen, M.N. *J. Neurosci.* **22**, 9475–9489 (2002).
- Law, C.-T. & Gold, J.I. *Nat. Neurosci.* **11**, 505–513 (2008).
- Meister, M.L.R., Hennig, J.A. & Huk, A.C. *J. Neurosci.* **33**, 2254–2267 (2013).

Linking cognition to brain connectivity

Stephen Smith

Predicting an individual's behavior is a formidable challenge for neuroimaging. A study now finds a strong link between an individual's ability to sustain attention and an extended, but specific, set of brain connections.

The possibility of using neuroimaging data to predict an individual's behavior is of great interest because it may eventually lead to understanding how processing in the brain gives rise to cognition. Over recent decades, neuroimaging has been evolving from its original primary application of population-average brain mapping toward understanding how brain function varies across subjects and how these differences relate to the subjects' differences in behavioral performance. Moreover, whereas early neuroimaging studies concentrated largely on localizing functional areas, it is now the connectivities between these areas that are increasingly being probed. Indeed, investigating the relationship between connectivity and cognition, and how this varies across subjects, is a primary goal of major recent neuroimaging endeavors such as the Human Connectome Project¹.

In a study in this issue of *Nature Neuroscience*, Rosenberg *et al.*² investigated in depth one particular mind-brain relationship: whether indicators of an individual's ability to sustain attention can be found in brain connections. They found that the strengths of a specific set of brain connections can be used to predict a subject's attention ability with high accuracy. This held not just for connectivity estimates made from the attention task imaging data, but even when estimated from resting state data, collected when the subjects were not carrying out any explicit task. Rosenberg *et al.*² found that a large number of brain connections are involved in sustained attention and that, despite this being a highly extended network, the set of connections is specific to attention and does not successfully

predict other cognitive measures, such as IQ. Furthermore, there is evidence that this extended network of connections is modulated *en masse*, with the network as a whole being weaker or stronger in subjects with different abilities to sustain attention.

To study this link between attention and connectivity, Rosenberg *et al.*² used a recently developed protocol^{3,4} in which the subject watches a slowly varying slideshow of different scenes and is asked to respond when a particular scene type is shown. Over the course of the experiment, attention generally falls, and the average accuracy in correct reporting of scene types is found to be a sensitive indicator of sustained attention ability—one that correlates with numbers of attention-related errors in normal life, as well as level of 'mindfulness' (for example, awareness of one's present situation and thoughts). Notably, a person's level of mindfulness also correlates positively with mean reaction time in this attentional task, with slower responses relating to being more generally thoughtful.

To test whether a subject's attention score could be predicted by specific brain connections, Rosenberg *et al.*² started by analyzing the functional magnetic resonance imaging (fMRI) data acquired during the attention task. They parcellated the brain into 268 distinct functional regions and estimated a representative average activation time series for each region. Then they estimated the correlation between every pair of regions, generating data for ~36,000 region pairs (Fig. 1). Correlation is a simple measure of the functional connectivity between regions, based on the notion that functionally linked regions will co-fluctuate. These temporal fluctuations may be caused by a region altering its behavior in distinct task conditions. Alternatively, in resting state experiments, fluctuations would be spontaneous changes in the states of sensory,

motor or cognitive processes. Next the authors considered the strength of each region pair's connection and tested whether this varied across subjects in the same way that the attention score varied. They found that 1,000–2,000 connections' strengths varied across subjects in a way that correlated with subjects' variations in attentional ability. These connection strengths were then averaged together within each subject, resulting in a single measure representing average connection strength, and this simplified measure of brain connectivity was found to correlate very highly with sustained attention ability.

This result still held when the approach was applied to subjects not used in this prediction-training process; such evaluations (for example, leave-one-out testing) are crucial in situations such as this, where the possibility of over-fitting the data is high. Even more convincingly, when this same set of brain connections was combined together using separate fMRI data with the subjects at rest, the connection strengths were still strongly predictive of the subjects' separately measured attention ability. This tells us that this large set of brain connections is highly related to sustained attention ability as an innate property of the individuals' brains, and not just during the explicit attention task. That is consistent with another recent study from the same authors, indicating that connectivities estimated during rest correspond closely to those estimated during a range of tasks⁵. Finally, the authors conducted a fully independent replication test, investigating whether averaging connection strength across this same set of putative attentional brain connections was predictive of attentional problems in a cohort of Chinese children with ADHD. Again, they found highly significant predictions of attentional ability, in this case relating to a

Stephen Smith is at the Oxford University Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), Oxford, UK.
e-mail: steve@fmrib.ox.ac.uk