

Closing the gate to distractors during decision-making

Edmund Chong, Athena Akrami

You are on your way to the grocery store when your partner calls and asks you to buy a few items. While mentally constructing a shopping list and planning your path through the store, you bump into a friend, and start chatting—while still trying to hold the shopping list in your mind. How the short-term maintenance of information or planned actions can be robust or susceptible to distracting information is regarded as one hallmark of working memory [Oberauer et al., 2018 Psych Bulletin]. What goes on in the brain during working memory under the presence of irrelevant information?

In this issue of *Nature Neuroscience*, Finkelstein et al. reveal the possible computations performed by brain circuits which may explain how decisions can be robust to conflicting or distracting information, when planning actions. Prior research from the same lab has already shown that a brain region known as Anterior Lateral Motor cortex (ALM) is responsible for maintaining planned actions across a short delay of several seconds [Li et al., 2015 Nature] and, further, that ALM activity can be described as the coordinated activity of networks of neurons that evolves over time during the delay, and converges to one of several discrete states (‘attractors’) that correspond to the upcoming actions [Inagaki et al., 2019, Nature].

Key to the concept of attractors is that small perturbations of network activity do not affect its ultimate convergence onto the original attractor—the network converges onto the same attractor as if the perturbation had not occurred. On the other hand, large or targeted perturbations can lead to the network converging onto a different attractor, resulting in a different action of the animal from the one that was originally planned. Finkelstein et al. specifically investigate how new or distracting information perturbs the network sufficiently to converge onto different attractors and produce different actions, and conditions under which network activity is robust to such distracting information.

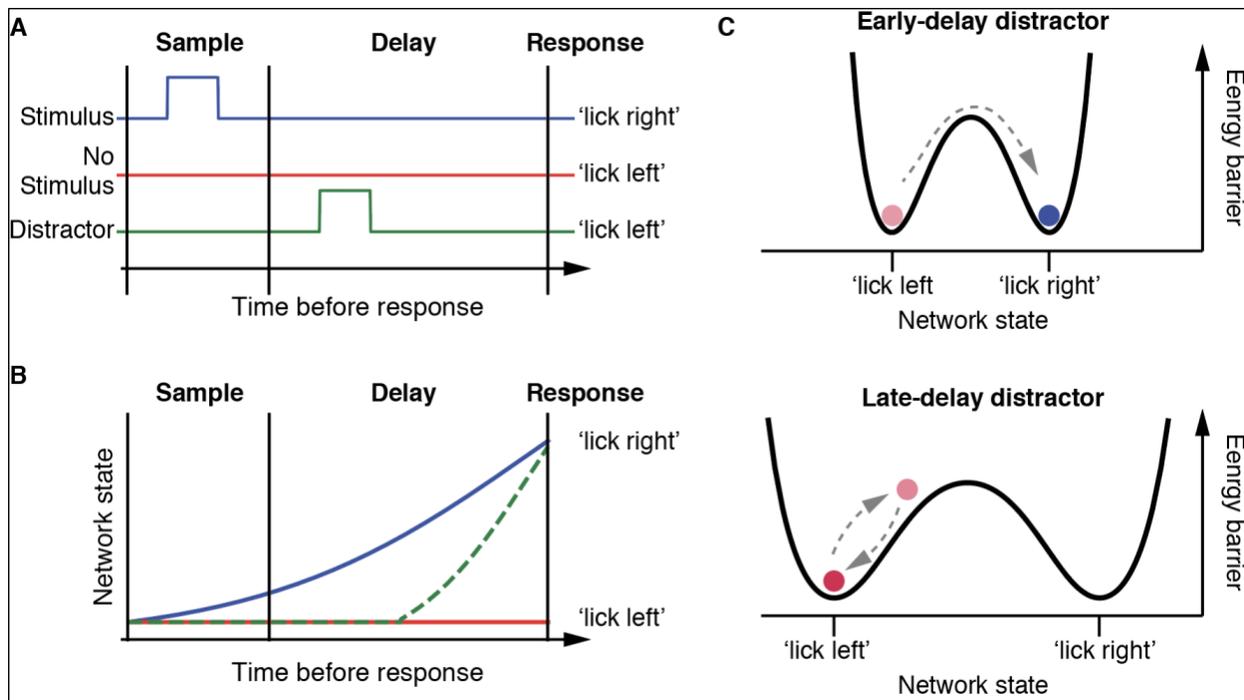


Figure 1. Experimental design and representations of ALM activity in Finkelstein et al. A.

Animals were trained to lick right when vS1 stimulation occurred during the sample period ('Stimulus', blue), and lick left there was no stimulation ('No Stimulus', red) or if stimulation occurred after the sample period ('Distractor', green). Animals were permitted to respond only after a delay. **B.** ALM activity on any given trial, described by single value ('Network state') evolving over the trial. During Stimulus trials (blue), activity evolves across the sample and delay periods to a point (attractor) which is also predictive of the animal's lick left/right choice. For No Stimulus trials, activity evolves towards another discrete attractor (red). During distractor trials, especially for early-occurring distractors, the occurrence of the distractor may shift the network away from the path of one attractor to the other (green). **C.** Model of how ALM activity is robust to late but not early distractors. Network states converging towards attractors can be visualized as a ball moving along a hill and valleys--it is easy to roll downhill, but not uphill. The position of the ball represents the state of the network, and attractors are the lowest points of the valleys. Early on in the delay ('Early-delay distractor'), the network can be sufficiently perturbed to move from one attractor to the other. In contrast, later in the delay ('Late-delay distractor'), the two attractors are further apart and it becomes harder for the network to be perturbed towards the other attractor.

To carry out the task, Finkelstein et al. trained mice to report the presence versus absence of stimuli within demarcated time points in a trial (Fig. 1A). Stimuli that occurred later, should be treated as irrelevant ('distractors'). The authors used as stimuli, the optogenetic activation of vibrissal somatosensory cortex (vS1). As vS1 is known to be associated with the whisker sensing, and stimulating the whiskers activates vS1, directly stimulating vS1 neurons is a proxy for whisker stimulation. The choice of direct vS1 stimulation allows for fine control over

stimulus magnitude, while knowing the exact origin of cortical responses. Additionally, directly stimulating vS1 bypasses thalamic effects which have been previously implicated [Guo et al., 2017 Nature], allowing the authors to examine how ALM gates sensory information independently of thalamus.

The authors found that if distractors occurred early during the delay period, closer to the stimulus window, animals would be more likely to misreport that stimulation occurred during the demarcated period, meaning the distractor successfully interfered with the animal's choice. In contrast, late-occurring distractors were far less likely to influence the animal's behavior. The authors then recorded neural activity in vS1 and ALM to understand how cortical activity can support behavioral robustness to distractors.

One possibility is that ALM gates incoming sensory inputs by simple suppression. For example, in the task, ALM may inhibit sensory input after an initial period, especially later in the delay. If so, one would find weaker ALM responses to distractor stimuli that happen late in the delay interval. Or, that even at the sensory input level, within vS1, responses to later inputs would be weaker. However, the authors found that this was not the case—the average response of ALM or vS1 to distractor stimuli was no different whether the distractor occurred early or late. Instead, the authors found that ALM activity, similar to previous work [Inagaki et al.], was best explained by network dynamics converging onto discrete attractor states (Fig. 1B). In particular, the present study found that these dynamics were most susceptible to perturbation from distractor inputs which occurred early. On trials where the early distractor successfully biased the animals into making the wrong choice, ALM network activity would converge onto the attractor representing that choice.

The authors also used a recurrent neural network (RNN) to model ALM activity, training the RNN to match delay-period activity of ALM (without including responses to distractors during training). The authors found that the RNN could recapitulate network dynamics to distractors and produce output choices that mimicked animals' choices on the task. One important component of the RNN was a slow-ramping of activity that was non-selective for the planned choice of the animal. This slow ramping activity was observed experimentally in the present and also previous recordings of ALM [Inagaki et al.]. Critically, further analysis of the RNN revealed that this ramping activity can produce attractor dynamics, which potentially explains how ALM becomes more robust against distractors over time. The ramping activity produces attractor states that diverge over time during the delay period, such that as the network settles towards one attractor (Fig. 1C), it becomes more difficult for distractors to perturb the network towards the other attractor.

One difficult issue facing the field in general, is the precise operationalization of animal behavior during tasks aimed at measuring complex cognitive behaviors: animals may use task strategies

that differ from the experimenter's intent. In Finkelstein et al., during the delay period, have animals already decided upon, and are actively maintaining some planned action? Or, are they still open to encoding subsequent information? This is a subtle but important difference: going back to our initial example of grocery shopping: while mentally rehearsing a shopping list, instead of choosing to block out distracting chatter from a friend (maintenance), it could be continuing to listen as the friend mentions discounted items in the store (encoding).

In the present study, it is unclear if animals are still actively trying to encode information during the delay period when distractors occur. Animals were familiarized with distractors during training, and distractors were informative of the correct response to make on each trial (lick left responses during distractor trials, i.e. late stimulations, were rewarded). Hence, even though one way to perform the task perfectly is to completely ignore the presence of distractors, animals can still be rewarded for encoding and utilizing information during the delay period. The authors also recorded ALM activity from animals which only experience distractors for the first time during ALM recordings. In this case, it is possible that animals will try to learn the value of these novel stimuli occurring during the delay, and encode them at least during the initial exposures. To overcome this issue, one could render the distractor stimuli completely uninformative about the correct response, by presenting them randomly across left-correct or right-correct trials. Another possible modification of Finkelstein et al.'s task design, could be the stimulation of other vS1 sites at random time periods, in all trials, such that the occurrence of any distractor is not predictive of the correct choice.

In human working memory experiments where task instructions explicitly divide trials into encoding and maintenance phases, and distractors are uninformative of the correct choice participants must make [Sakai, Rowe, Passingham, 2002, *Nature Neuroscience*], different neural correlates underlie the robustness against distractor- interference depending on whether the distractors occur during encoding [Liesefeld et al., 2020, *Visual Cognition*] or maintenance [Lorenc et al., 2021, *TICS*]. If for example animals are still actively encoding stimuli during the delay period, ALM dynamics which show a switching between attractor states, may actually reflect "changes of mind" in decision-making given additional information [Resulaj et al. 2009, *Nature*].

Another broader point of interest for future studies is the question of how ALM interacts with other brain regions in the presence of distractors. In work from the same and other labs, delay-period activity that predicts upcoming actions has previously been observed in thalamus [Guo et al., 2017 *Nature*] and cerebellar nuclei [Gao et al., 2018 *Nature*; Chabrol, Blot, Mrsic-Flogel, 2019 *Neuron*], and the activity in these other brain regions reciprocally influences ALM activity.

Finkelstein et al. represents an important advance within an already-impressive body of work from the present lab [Svoboda & Li, 2018 *Curr Opin Neurobiol*] related to the study of short-

term memory. By combining behavioral experiments, precise input manipulations, recordings, and detailed modelling of network activity, the authors shed light on how complex network dynamics can support maintenance of information or planned actions over a short delay period that is robust to interference.

References

- Chabrol, F. P., Blot, A., & Mrcic-Flogel, T. D. (2019). Cerebellar Contribution to Preparatory Activity in Motor Neocortex. *Neuron*, *103*(3), 506-519.e504.
- Gao, Z., Davis, C., Thomas, A. M., Economo, M. N., Abrego, A. M., Svoboda, K., . . . Li, N. (2018). A cortico-cerebellar loop for motor planning. *Nature*, *563*(7729), 113-116.
- Guo, Z. V., Inagaki, H. K., Daie, K., Druckmann, S., Gerfen, C. R., & Svoboda, K. (2017). Maintenance of persistent activity in a frontal thalamocortical loop. *Nature*, *545*(7653), 181-186.
- Inagaki, H. K., Fontolan, L., Romani, S., & Svoboda, K. (2019). Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature*, *566*(7743), 212-217.
- Li, N., Chen, T.-W., Guo, Z. V., Gerfen, C. R., & Svoboda, K. (2015). A motor cortex circuit for motor planning and movement. *Nature*, *519*(7541), 51-56.
- Liesefeld, H. R., Liesefeld, A. M., Sauseng, P., Jacob, S. N., & Müller, H. J. (2020). How visual working memory handles distraction: cognitive mechanisms and electrophysiological correlates. *Visual Cognition*, *28*(5-8), 372-387.
- Lorenc, E. S., Mallett, R., & Lewis-Peacock, J. A. (2021). Distraction in Visual Working Memory: Resistance is Not Futile. *Trends in Cognitive Sciences*, *25*(3), 228-239.
- Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., . . . Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885-958.
- Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263-266.
- Sakai, K., Rowe, J. B., & Passingham, R. E. (2002). Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nature Neuroscience*, *5*(5), 479-484.
- Svoboda, K., & Li, N. (2018). Neural mechanisms of movement planning: motor cortex and beyond. *Current Opinion in Neurobiology*, *49*, 33-41.