NEURAL NETWORKS

Linking task structure and neural network dynamics

The solutions found by neural networks to solve a task are often inscrutable. We have little insight into why a particular structure emerges in a network. By reverse engineering neural networks from dynamical principles, Dubreuil, Valente et al. show how neural population structure enables computational flexibility.

Christian David Márton, Siyan Zhou and Kanaka Rajan

Artificial neural networks can be trained to achieve human-level performance in games and brain-like performance on neuroscience tasks[1,2], but we still do not understand how the magic happens. In 'backprop' — the go-to learning algorithm for training deep networks — the error gradient gradually pushes each neuron away from the unfavorable activity it is generating and the network toward better task performance. Slowly edging away from the mistakes, ex negativo, seems to work well. But on the flipside, we still do not understand what units are actually doing once they have converged to a particular solution; it still feels a bit like pulling a rabbit out of a hat.

Currently, there are two opposing perspectives. One view is that each individual neuron has a certain computational function as evidenced by a particular response profile, and a network can be understood as the aggregate of all the individual neurons' response profiles. An example of this would be a neuron that responds only to a particular stimulus, but not to others, or fires only at a particular moment in time. Thus, at an aggregate level, we can have different groups of neurons tuned to particular things. The other view is that the computation is not expressed at the level of individual neurons, but only at the level of an entire population, where it can be understood as tracing out a low-dimensional manifold[3,4]. In the latter case, when looking at single-neuron response profiles, information may appear 'mixed' or 'multiplexed' — that is, a single neuron may respond to many different stimuli, and appear non-selective in its response profile.

Two recent studies exemplify this dichotomy well. Raposo, Kaufman and Churchland[5] found no non-random population structure in the posterior parietal cortex when mice performed multi-sensory decision-making tasks; in other words, all the individual neurons appeared mixed in
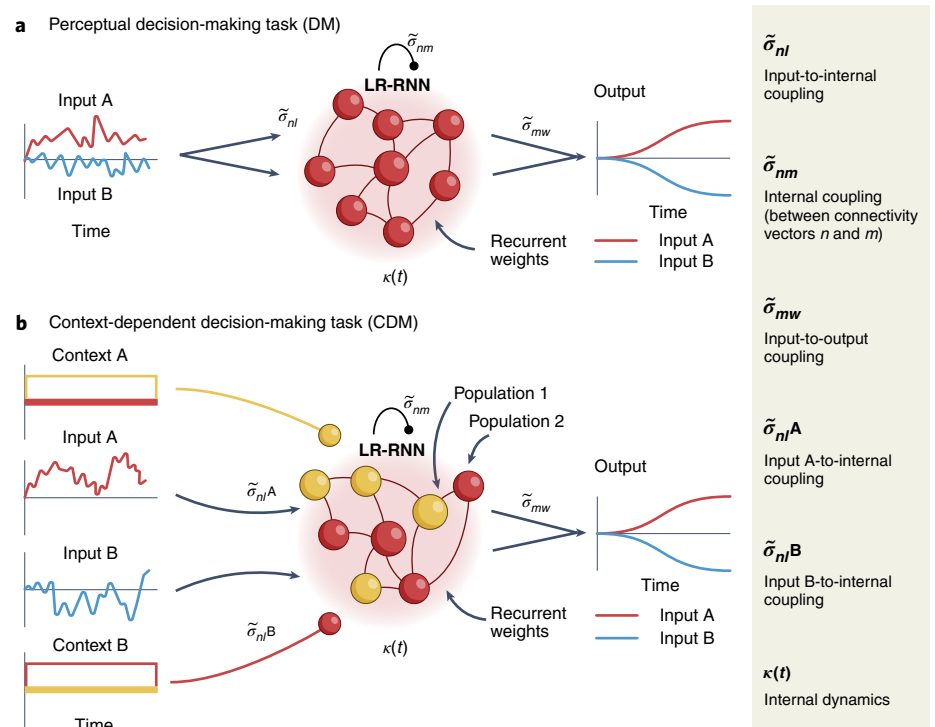

Fig. 1 | Reverse engineering the solution to the perceptual decision-making (DM) and context-dependent decision-making (CDM) tasks with low-rank recurrent neural networks (LR-RNNs). a, In the DM task, the input signal (which can take the shape of A or B, for example) needs to be integrated over time. When neural networks were trained on this task, only one undifferentiated neural population was found to be sufficient to implement the task. In low-rank networks, adjusting coupling factors $\widetilde{\sigma}_{nl}$ and $\widetilde{\sigma}_{nm}$ was sufficient to reproduce the results. **b**, In the CDM task, depending on context (A or B), either input A or input B needs to be integrated. When neural networks were trained on this task, two different subpopulations with different response profiles (yellow and red) were found to be necessary. The results can be reproduced in low-rank networks by letting inputs A and B act on the internal dynamics through two different input-to-internal coupling factors, $\widetilde{\sigma}_{nlA}$ and $\widetilde{\sigma}_{nlB}$. The contextual inputs (A and B) act on the input-to-internal couplings, allowing context-dependent responses.

their response profiles. On the other hand, using a different behavioral task (odor discrimination), Hirokawa, Vaughan et al.[6] observed that neurons in rat prefrontal cortex could be arranged into different groupings with specific response profiles.

It seems that in certain cases looking at aggregate dynamics makes sense, but in others we can also learn something from individual response profiles. There has been very little attempt to try to align these views — until now.

Could it be that the particular choice of tasks leads to the differences observed in the studies above? If so, then when do we actually need different groups of neurons with particular response profiles rather than a population in which single units show a mixed profile? A similar confusion arises in the case of learning multiple tasks, or 'multi-task learning'[7–10]. When Yang et al.[7] trained artificial neural network models simultaneously on multiple cognitive tasks, several groups of neurons emerged, some of which were engaged across multiple tasks. Some tasks were seemingly more similar to each other and leveraged the same group of neurons. But it is not clear at what point distinct groups of neurons with different functionalities need to emerge to facilitate multi-task learning.

These are precisely the questions addressed in the work by Dubreuil, Valente et al.[11], where the authors reverse engineered artificial neural networks for a variety of tasks. The authors considered five commonly used cognitive neuroscience tasks and found that only in context-dependent decision-making (CDM; Fig. 1b) and delayed match-to-sample (DMS) tasks did trained networks exhibit groupings of neurons with different response profiles. In the other tasks (such as simple perceptual decision-making (DM); Fig. 1a), the information in the artificial networks was found to be fully mixed at the single-neuron level. To confirm that different groupings are indeed necessary in one case but not the other, the authors built surrogate networks by randomly resampling the weights from a Gaussian distribution that best fit the learned weights. The surrogate networks performed worse than the original networks on the CDM and DMS tasks (Fig. 1b), but not the others (such as DM; Fig. 1a). This confirmed that population structure is necessary for task performance and suggested that more complex tasks (for example, modifying outputs based on context, or matching inputs to outputs after a delay) require more complex population structures.

What about those tasks requires separate neural populations? Dubreuil, Valente et al. propose that this is best answered by viewing network activity through the lens of neural dynamics[12]. In accompanying work, the same team describes how to leverage low-rank networks to better understand neural dynamics[13,14]. In such low-rank neural networks, the evolution of neural activity can be fully determined by a small number of parameters. Dubreuil, Valente et al.[11] show that the activity of a particular neuron can be understood in terms of its projection onto a handful of parameters in low-rank networks (Fig. 1), and this projection can be used to group neurons into different populations (on the basis of their coupling to input, internal, and output dynamics; Fig. 1). The connectivity structure (or coupling structure, in low-rank parlance), in turn, controls the type of dynamics that emerges in the network. Thus, computational functionality can be achieved by controlling the dynamical structure in the network. To demonstrate this, the authors reverse engineered the solution found by artificial networks by re-constructing the same dynamical landscape from a small number of parameters in low-rank networks — without the learning algorithm. Low-rank networks were able to solve simpler tasks, such as perceptual decision-making (DM; Fig. 1a) without the need to tune any of the connectivity-related parameters, unlike more complex tasks such as CDM or DMS (Fig. 1b).

Putting all the pieces together, we now have a better handle on why more complex tasks, such as CDM, require a more complex population structure. More distinct groupings with different functionality in the network means more degrees of freedom in terms of network dynamics. This, in turn, allows more flexible input–output mappings, such as those required in the CDM task, where outputs need to be modulated in a context-dependent manner. In a non-differentiated, fully mixed network, joint dynamics would be modulated in unison by the inputs. By contrast, in the case of a more differentiated population with multiple groupings, different groupings could be modulated differently by a combination of inputs, allowing for more complex dynamics. This idea can be readily extrapolated to multi-tasking, where a model may need to process the same inputs differently in different tasks.

Dubreuil, Valente et al. built a bridge across the seemingly opposing views of single-unit response profiles and low-dimensional neural population dynamics. Now we realize that this has been a false dichotomy: population structure (evident at the level of single neurons or groups of neurons) emerges from the need for more flexible neural dynamics to fulfill particular, task-dependent, computational roles. So from the standpoint of neural dynamics, the two views can be complementary. Further, this work suggests that it is essential to consider them jointly when reporting results in order to characterize how a neural system implements a task.

This is the boon and bane of the insights generated by this work: it makes predictions for particular task-dependent effects to be observed experimentally. It is unfortunate that we are not (yet) at the stage of making more general predictions about what neural dynamics should look like for particular computational functions. But this work draws our attention to the fact that our observations are, in fact, task-dependent. If we are to get a handle on overarching computational principles, though, we need to be meticulous and systematic about recording task-related effects. Perhaps then we can start to group tasks on the basis of their dynamical profiles and recognize common building blocks[8].

From there, we may be able to build on this work to crystalize a better notion of task complexity, which scientists often play fast and loose with. Perhaps complexity can be defined in terms of the minimum number of adjustable levers needed to implement (solve) a task, as recent work suggests[15]. This is reminiscent of Kolmogorov complexity, which is still equally inscrutable. The work by Dubreuil, Valente et al. suggests that what matters may be not just the number of levers (the rank of the network), but also the diversity of configurations possible (the connectivity structure). In any case, these findings imply that we need to think carefully about the kinds of tasks we use and how 'complex' they may really be.

The authors also show how we can build up dynamical capacity in neural networks by adjusting a few levers, entirely without using a learning algorithm. It would be interesting to push this approach further and see whether we can build a lexicon of dynamical motifs that can be re-used to solve new tasks, as some recent work suggests[9,10].

Overall, this work lifts the veil on the magic of neural networks, at least partially, so we can sneak a peek into the hat. Depending on the task, one or more rabbit(s) may be necessary. ❏

Christian David Márton[1], Siyan Zhou[1,2] and Kanaka Rajan [iD][1]✉

[1]Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]Department of Neurobiology, Harvard Medical School, Boston, MA, USA.
✉e-mail: kanaka.rajan@mssm.edu

References
1. Botvinick, M. et al. Trends Cogn. Sci. 23, 408–422 (2019).
2. Chaisangmongkon, W., Swaminathan, S. K., Freedman, D. J. & Wang, X.-J. Neuron 93, 1504–1517.e4 (2017).
3. Cunningham, J. P. & Yu, B. M. Nat. Neurosci. 17, 1500–1509 (2014).
4. Gallego, J. A. et al. Nat. Commun. 9, 4233 (2018).
5. Raposo, D., Kaufman, M. T. & Churchland, A. K. Nat. Neurosci. 17, 1784–1792 (2014).

6. Hirokawa, J., Vaughan, A., Masset, P., Ott, T. & Kepecs, A. *Nature* **576**, 446–451 (2019).
7. Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T. & Wang, X.-J. *Nat. Neurosci.* **22**, 297–306 (2019).
8. Yang, G. R., Cole, M. W. & Rajan, K. *Curr. Opin. Behav. Sci.* **29**, 134–143 (2019).
9. Márton, C. D., Lajoie, G. & Rajan, K. Preprint at https://doi.org/10.48550/arXiv.2105.14108 (2021).
10. Duncker, L., Driscoll, L., Shenoy, K. V., Sahani, M. & Sussillo, D. *Adv. Neural Inf. Process. Syst.* **33**, 14387–14397 (2020).
11. Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F. & Ostojic, S. *Nat. Neurosci.* https://doi.org/10.1038/s41593-022-01088-4 (2022).
12. Vyas, S., Golub, M. D., Sussillo, D. & Shenoy, K. V. *Annu. Rev. Neurosci.* **43**, 249–275 (2020).
13. Beiran, M., Dubreuil, A., Valente, A., Mastrogiuseppe, F. & Ostojic, S. *Neural Comput.* **33**, 1572–1615 (2021).
14. Mastrogiuseppe, F. & Ostojic, S. *Neuron* **99**, 609–623.e29 (2018).
15. Stroud, J. P., Porter, M. A., Hennequin, G. & Vogels, T. P. *Nat. Neurosci.* **21**, 1774–1783 (2018).

Competing interests

The authors declare no competing interests.