SCALING UP NEUROSCIENCE

Blending computational and experimental neuroscience

Patricia S. Churchland[1] and Terrence J. Sejnowski[2,3]

The launch of the United States' BRAIN Initiative brings with it a new era in systems neuroscience that is being driven by innovative neurotechnologies, increases in computational power and network-style artificial intelligence. A new conceptual framework for understanding cognitive behaviours based on the dynamical patterns of activity in large populations of neurons is emerging.

The oft-heard grumblings that neuroscience is data rich and theory poor are half right. Neuroscience is theory poor: although piecemeal insights abound, it cannot yet provide integrated explanations of how we see, recall past experiences or swat a fly. Likewise, it has not yet given us molecule-to-behaviour explanations for neurological conditions such as schizophrenia or chronic depression. However, despite regular complaints from novice neuroscientists that there is just too much data, veterans ruefully realize that — at the level of neuronal networks — neuroscience is also decidedly data poor. Moreover, this dearth of data hobbles theoretical innovation: networks are the organizational matrix that links individual neurons to large-scale systems, and, without data that show how these networks operate, theories that aim to explain linkages across these levels are under-constrained.

The BRAIN (Brain Research through Advancing Innovative Neurotechnologies) Initiative, which was launched in the United States in 2013, was motivated by recognition of the urgent need to find ways to access the operations of neural networks (at the micro and macro scales) and to understand how their activity can be modulated by the activity of other networks. To make progress towards addressing this problem, we will require new technologies — new tools to obtain the data and new methods to analyse them. Accordingly, the BRAIN Initiative is directed towards inventing new technologies that will, it is hoped, foster the discovery of a unified set of principles that link all levels of nervous systems and through which brains perform their jobs.

At the same time that the BRAIN Initiative has hatched, several breakthroughs in artificial intelligence (AI) research (outlined below) reached public awareness. It seems possible that we are on the brink of a new era of collaboration between systems neuroscience and AI. This era is being driven, in part, by the development of innovative neurotechnologies, such as optogenetics and large-scale multi-electrode arrays, that allow for the generation of gigantic data sets. However, equally important — and our focus here — is the ever-increasing computational power that allows for hitherto impossible analyses of these gigantic data sets and the breakthroughs in machine learning that can be harnessed to find deep patterns and organizational features within them.

Over the past 50 years, much of our understanding of neuronal properties — and, consequently, our conceptual frameworks for brain function — was derived by recording from single neurons, one at a time. Although the technology was the best available at the time, this narrow scope led to problems similar to those faced when looking at a scene through a soda straw, sampling one pixel at a time from random locations. Without understanding the higher-order relationships between pixels (or neurons), it is difficult to extract their global properties and configuration. Today, recording from hundreds of neurons using optics rather than electricity has become routine, and the size of the data sets produced is likely only to grow in the wake of tools produced as part of the BRAIN Initiative. In light of these technical breakthroughs, what conceptual ideas are available to help us understand what the data mean in terms of how networks contribute to representations of the world and to making decisions about what to do next?

In 1992, we published *The Computational Brain*[1], which went beyond the single neuron and explored the ways in which populations of neurons could represent the properties of the world. Insights were derived from artificial neural network (ANN) models that used learning algorithms to update the strength of connectivity between network 'units' (representing neurons) in response to training. These models could solve difficult computational problems by developing distributed representations in 'hidden layers' located between input and output layers. Although the units were not very neuron-like, it was intriguing to note that their patterns

[1]*Department of Philosophy, University of California, San Diego.*
[2]*Howard Hughes Medical Institute, the Salk Institute for Biological Studies, La Jolla, California 92037, USA.*
[3]*Division of Biological Sciences, University of California, San Diego, La Jolla, California 92093, USA.*

Correspondence to T.J.S.
terry@salk.edu

doi:10.1038/nrn.2016.114
Published online 9 Sept 2016

of activity resembled those in populations of neurons recorded one at a time. Once trained, the output of the ANN could provide a categorization 'answer' from the input data it was given. So too, it was conjectured, could real neural networks.

However, scaling up the size of these early ANNs was a dead-in-the-water problem that was only solved by an exponential increase in computational power provided by hardware engineers. Although the pioneering ANNs had only a few hundred model neurons and a few thousand connections, current industrial-strength networks have millions of model neurons and billions of connections, organized into dozens of layers. 'Deep learning' in these ANNs produced unexpectedly high levels of performance in tasks that previously only humans could do well, such as speech recognition and object recognition in images[2]. This achievement helps us to conceptualize, and then test, how massive populations of neurons in a hierarchical architecture can support complex cognitive behaviours. The mathematical insights about the dynamics of these networks in state space are already helping to explain experimental recordings[3], and their influence is likely to grow in tandem with research from the BRAIN Initiative.

Making decisions to reach a goal is essential for survival, and modelling such decision making is an important challenge for computational neuroscience. The temporal differences (TD) algorithm models the ability of dopamine cells in the ventral tegmental area of the midbrain to learn when a particular sensory stimulus predicts a reward[4]. This type of trial-and-error learning can enable an animal to make a sequence of decisions to optimize future rewards. A dramatic example of the use of advanced versions of this learning algorithm to model category formation and decision making was the TD-Gammon ANN[5]. It learned to play backgammon— a game that depends on an uncertain role of a dice — and, after playing a million of games against itself, achieved world-champion performance. Similarly, the AlphaGo ANN learned to play Go, a famously complex game that is much more difficult than chess, at championship level[6]. In March 2016, AlphaGo beat Lee Sedol, the Korean Go champion, four games to one. AlphaGo displayed human-like creativity, demonstrating the power of reinforcement learning coupled with deep learning. One reason this result was surprising is that some psychologists, perhaps inspired by Chomsky's criticism of behaviourism, decried reinforcement learning as far too feeble to accomplish much in the cognitive domain and favoured models for cognition that involved rules, such as the rules of logic. The ANN results thus showed that reinforcement learning in multilayered networks may be vastly more powerful and more capable of generalizing to new cases and solving problems than the conventional wisdom ever imagined.

Because learning and adapting to the environment are fundamental functions of all nervous systems, it is tempting to lapse into the supposition that ANNs have unravelled the brain's mystery. This is a mistake. Powerful though ANN models are, they cannot yet accommodate many fundamental survival functions, including motivation, drive, sociality and aggression. These are constant contributors to decision making and behaviour. ANNs cannot run or hide from a predator, they cannot find energy sources, and they care not one whit about whether or not they survive. To capture the autonomy of biological creatures, and how they use knowledge to make a living in a causally and socially complex world, will require the broader warp and woof of goals, motivation, caring, wanting and deciding. Cognitive capacities are a product of biological evolution and thus are shaped by the abiding and indispensable scaffolding of movement and motivational organization into which they must fit. For example, curiosity and the desire to explore are not incidental to learning and using what is learned. To address these wider issues, it will be important for the next generation of AI models — developed in parallel with the BRAIN Initiative — to include many more brain systems to determine how they are integrated with each other.

Working out how nervous systems organize an action by integrating multiple current perceptions and a relevant portfolio of learned knowledge in the context of registered needs (both immediate and long-term) is a grand challenge now facing neuroscience. Our growing understanding of how large-scale ANNs can perform exceptionally complicated tasks provides a conceptual platform to tackle this challenge. The platform is useful because it not only embraces the mind-boggling size of neural populations but also allows us to analyse what roles these populations are performing in high-dimensional state spaces.

1. Churchland, P. S. & Sejnowski, T. J. *The Computational Brain* (The MIT Press, 1992).
2. LeCun, Y., Bengio, Y. & Hinton, G. E. Deep learning. *Nature* **521**, 436–444 (2015).
3. Shenoy, K. V., Sahani, M. & Churchland, M. M. Cortical control of arm movements: a dynamical systems approach. *Annu. Rev. Neurosci.* **36**, 337–359 (2013).
4. Montague, P. R., Dayan, P. & Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).
5. Tesauro, G. Temporal difference learning and TD-Gammon. *Commun. ACM* **38**, 58–68 (1995).
6. Silver, D. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature.* **529**, 484–489 (2016).

Competing interests statement
The authors declare no competing interests.