From circuits to behavior: a bridge too far?

Matteo Carandini

Neuroscience seeks to understand how neural circuits lead to behavior. However, the gap between circuits and behavior is too wide. An intermediate level is one of neural computations, which occur in individual neurons and populations of neurons. Some computations seem to be canonical: repeated and combined in different ways across the brain. To understand neural computations, we must record from a myriad of neurons in multiple brain regions. Understanding computation guides research in the underlying circuits and provides a language for theories of behavior.

One of the fundamental mandates of neuroscience is to reveal how neural circuits lead to perception, thought and, ultimately, behavior. The general public might think that this goal has already been achieved; when they read that a behavior is associated with some part of the brain, they take that statement as an explanation[1]. But most neuroscientists would agree that, with a few notable exceptions, the relationship between neural circuits and behavior has yet to be established.

We clearly need to do more work, and institutions are aware of this. For instance, the University of California San Diego has a Center for Neural Circuits and Behavior (**Fig. 1**), and my university is forming a Centre for Neural Circuits and Behaviour. These institutions and their funders are right to invest in this, as it is an exciting and not unreasonable goal. But how shall we proceed? Can we go directly from circuits to behavior or is it a bridge too far?

Let's imagine that, instead of the brain, we were trying to understand a laptop computer (**Fig. 2a**) with the knowledge and tools available a hundred years ago. Physiologists might discover and characterize transistors, chips, buses, clocks and hard drives. Anatomists might strive for a 'connectome' of the wires across and in the chips. A furious debate, however, might divide them, as the details of wiring would differ across models (older versus newer) and across brands (different microprocessors). Psychologists might concentrate on general input and output properties of software applications, but those who study a business application would disagree with those studying videogames.

Matteo Carandini is at University College London, London, UK.
e-mail: matteo@carandinilab.net

No theories, at this stage, would likely connect the hardware to the operation of the computer.

What discovery would bridge this gap between circuits and behavior? It would be the realization that there is an intermediate level: the level of computer languages and operating systems. This level decouples the hardware from the software. Different models and brands of chips have different circuits but perform exactly the same computations. Understanding these computations would allow the researchers to ask the right questions about the circuits and understand how they work. Theories about software applications, in turn, would lie on a foundation of computer algorithms, without needing to speak of wires and electrical charge. Grasping this intermediate level of description would explain how computers work.

In some ways, this is a tired analogy. Each generation tends to compare the brain to a complex technology of their time: a loom, a telephone exchange, a chemical plant or a hologram[2]. These comparisons elicit smiles a few years afterwards. Moreover, the brain may be more of a special purpose machine: the circuits for vision, olfaction or body movement might be more tightly linked to the resulting function than in a general-purpose computer. Even so, the brain is undeniably an information processing device, so it may serve to compare it to the information processing devices that we build.

Notably, the computer analogy illustrates a general rule in science, which is to seek an appropriate level of description[3]. This level is intermediate between detailed mechanism (too much reductionism) and overall function (too much holism). In physics, for instance, the equations for particle interactions become impossible to solve or even simulate once a

Figure 1 Before and after. (**a**) The Center for Neural Circuits and Behavior at University of California, San Diego (photo by Bassam Atallah). (**b**) An exceedingly literal interpretation of this article's viewpoint (rendering by Anita Horn).

system involves more than ten particles[4]. So, to describe what a decent-sized piece of matter does, solid-state physicists have developed remarkably successful theories operating at mesoscopic levels[5]. Similar examples abound in biology. For instance, it is much preferable to describe proteins in terms of a handful of domains rather than of thousands of amino acids. Protein domains are fairly independent of precise amino acid sequence. They constitute an intermediate level that decouples the level of structure from that of overall function.

It is reasonable to suspect that a similar approach will lead to success in understanding the brain. We might be able to identify an intermediate stage between circuits and behavior, the equivalent of computer languages for brain operation (**Fig. 2b**). This is a stage of computations that occur in the activity of individual neurons and especially of populations of neurons.

Research in recent decades has indeed started to reveal some elements of these computations. There is evidence that the brain relies on a core set of standard (canonical) neural computations: combined and repeated across brain regions and modalities to apply similar operations to different problems. As examples, consider two computations that are close to my own expertise: linear filtering and divisive normalization.

Linear filtering is a widespread computation in sensory systems, in which neurons operate on sensory inputs by weighted summation in linear receptive fields. It is performed, at least approximately, at various stages in the visual system[6], in the auditory system[7] and in the somatosensory system[8]. It may also be involved in motor systems, where neural activity can specify force fields obeying linear superposition[9].

Divisive normalization, in turn, is an operation in which neuronal responses are divided by a common factor, the summed activity of a pool of neurons. Normalization was developed to explain responses in primary visual cortex and is now thought to operate throughout the visual system and in multiple other sensory modalities and brain regions[10]. It is thought to underlie operations as diverse as the representation of odors, the deployment of visual attention, the encoding of value and the integration of multisensory information.

Both computations are examples of bridges between circuits and behavior. For instance, a standard model of human visual detection starts with a front end made of linear filters[11,12] and is typically followed by a stage of divisive normalization[10]. Linear filtering and divisive normalization, moreover, summarize the activity of large populations of neurons and of individual neurons in the early visual system[13,14]. As such, they have guided a multitude of experiments aimed at the underlying circuits[10].

Linear filtering and divisive normalization are just two instances of plausible candidates for canonical neural computations. Other examples include thresholding and exponentiation, recurrent amplification, associative learning rules, cognitive spatial maps, coincidence detection, gain changes resulting from input history and cognitive demands, population vectors, and constrained trajectories in dynamical systems. Of course, one hopes that further research will identify new computations and tell us about the various ways that the computations are combined in different brain regions and modalities.

Crucially, research in neural computation does not need to rest on an understanding of the underlying biophysics. Some computations, such as thresholding, are closely related to underlying biophysical mechanisms. Others, however, such as divisive normalization, are less likely to map one-to-one onto a biophysical circuit. These computations depend on multiple circuits and mechanisms acting in combination, which may vary from region to region and species to species. In this respect, they resemble a set of instructions in a computer language, which does not map uniquely onto a specific set of transistors or serve uniquely the needs of a specific software application.

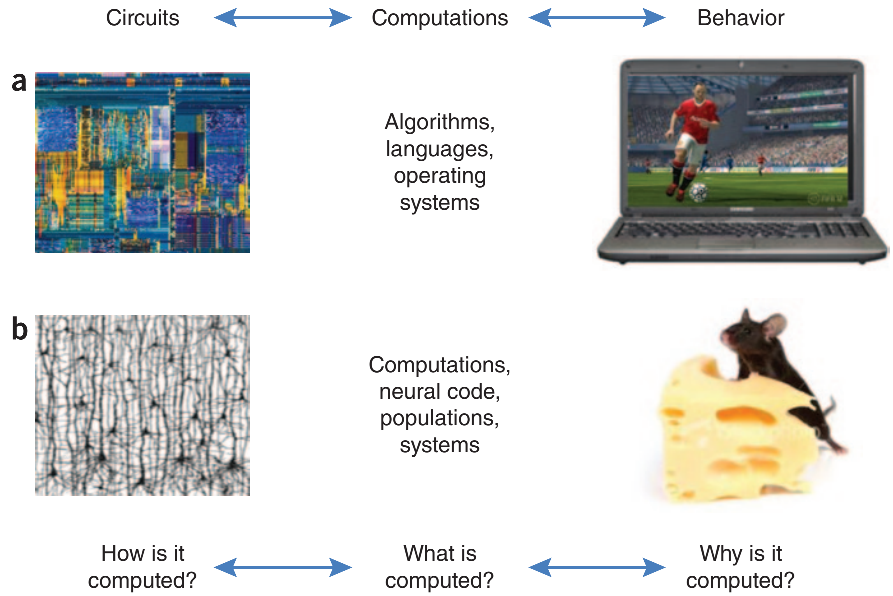

Figure 2 Between circuits and behavior: the Marr approach applied to computers and brains. (**a**) The wiring of a fraction of an Intel microprocessor and a laptop playing a popular videogame (FIFA 12). (**b**) Pyramidal neurons in cortex (detail of a drawing by Ramon y Cajal) and a mouse engaged in a pleasant behavior.

Nonetheless, once they are discovered, neural computations can serve as a powerful guide for research into the underlying circuits and mechanisms. It is hard to understand a circuit without knowing what it is computing, be it linear filtering with thresholding and divisive normalization[15] or the detection of time differences between two sets of inputs[16].

On the other hand, developing a *Biophysics of Computation*[17] can occasionally work the other way. For instance, studying recurrent excitation in a vertical column of cortex leads to the suggestion that it may act as an amplifier and to proposals as to why amplification may be a useful computation[18].

The basic idea that one should concentrate on computation was laid out in the 1980s by Marr in his influential book *Vision*. Marr argued that "any particular biological neuron or network should be thought of as just one implementation of a more general computational algorithm"[19]. He suggested that "the specific details of the nervous system might not matter". This may seem extreme, but it is useful as it firmly distinguishes between the question of what is computed and the questions of how and why it is computed (**Fig. 2**).

The task ahead is to discover and characterize more neural computations and to find out how these work in concert to produce behavior. How shall we proceed? The known neural computations were discovered by measuring the responses of single neurons and neuronal populations and relating these responses quantitatively to known factors (for example, sensory inputs, perceptual responses, cognitive states or motor outputs). This approach clearly indicates a way forward, which is to record the spikes of many neurons concurrently in multiple brain regions in the context of a well-defined behavior. How many neurons? Currently, we can record from hundreds of neurons[20,21], and new technology will hopefully soon grow this to thousands. And which neurons shall we aim for? This will likely depend on the methods for establishing functional connectivity. There are exciting improvements in these methods[22] and we are likely to see further improvements in coming years.

To guide these experiments and to interpret the resulting flood of data, we will need new theories. Ideally, these theories will establish new metaphors for the concerted activity of large neuronal populations. Great models can do that. Consider, for example, the highest success of computational neuroscience: Hodgkin and Huxley's model of the action potential. This model bridged structure and function by relying not on a chemical description, but on a metaphor: the equivalent electrical circuit. By extending this metaphor beyond passive membranes, it captured vast amounts of data and guided decades of research into the underlying biological hardware (voltage-sensitive ion channels).

There are, of course, alternatives to Marr's approach, and a notable one is the quest for the full diagram of the circuits of the brain, the connectome[23]. This diagram will undoubtedly prove useful to understand how circuits give rise to computations (**Fig. 2**). For instance, a tiny piece of connectome was recently obtained[24] for a piece of retina (a circuit) and it answered a longstanding question about direction selectivity (a computation). However, this approach will do little to explain how various computations are used together to produce behavior (**Fig. 2**).

More generally, knowing a map of connections may not be as useful as one expects, especially if this map comes with no information about connection strength. For instance, we have long known the full connectome for the worm *C. elegans*, detailing the more than 7,000 connections between its 302 neurons[25], and yet we are hardly in a position to predict its behavior, let alone the way that this behavior is modified by learning. Similarly, the scientists that were trying to understand the computer in our opening metaphor would benefit more from a manual of a programming language than from a blueprint of a microprocessor (**Fig. 2a**).

Another alternative to Marr's approach is the effort to simulate brain circuits in all their glorious complexity, to obtain a 'simulome' (apologies for the neologism). This approach was championed in the 1990s with the neural simulator Genesis[26] and had a revival in the BlueBrain project[27] and possibly the Human Brain Project[28]. Its central hypothesis is that an "understanding of the way nervous systems compute will be very closely dependent on understanding the full details of their structure"[29]. According to this hypothesis, one should seek "computer simulations that are very closely linked to the detailed anatomical and physiological structure" of the brain, in hopes of "generating unanticipated functional insights based on emergent properties of neuronal structure"[26].

This quest for the simulome has been a bit of a disappointment. Two decades since the idea was put forward, we have not discovered much by putting together highly detailed simulations of vast neural systems. Where Genesis and other detailed neural simulators have succeeded is when they have concentrated on a more microscopic scale: detailed simulations of myriad items as tiny as ion channels can be necessary for understanding computation in single neurons or dendrites. However, putting all of the subcellular details (most of which we don't even know) into a simulation of a vast circuit is not likely to shed light on the underlying computations[30].

Indeed, although we have good examples of the reductionist approach working well (from behavior to computations to circuits), the case still needs to be made for the constructivist approach (from circuits to computations to behavior). A similar situation is seen in other sciences: "the ability to reduce everything to simple fundamental laws does not imply the ability to start from these laws and reconstruct the universe"[3].

Luckily, there is a strong sense that the level of the subcellular and the level of the network are decoupled. For instance, very similar patterns of cellular and network responses (and therefore very similar computations) can be obtained with wide differences in biophysical details[31]. Conversely, small changes in biophysical details can lead to wide differences in cellular properties[32] (and therefore in computations). This decoupling of levels gives us hope that we will indeed understand the relationships between circuits and behavior. Conversely, if understanding behavior requires understanding a myriad of inter-relationships between molecules, channels, receptors, synapses, dendrites, neurons and so forth, then we have little hope of success.

To conclude, the gap between circuits and behavior is too wide to be bridged without an intermediate stage. Following on the basis laid by Marr, it seems evident that this stage is one of computation. Neuroscientists have already identified some computations that appear to be canonical: repeated and combined in different ways across the brain. Hopefully new experiments, new technologies and new theories will soon identify an even wider array of computations, and give us more concrete examples of how these are combined to determine behavior. Subscribing to this view does not mean arguing for a separation of those who study circuits from those who study behavior (**Fig. 1b**). Rather, it means arguing that researchers of circuits and of behavior go furthest when they speak a common language of computation.

ACKNOWLEDGMENTS
Most of these thoughts have arisen from conversations with colleagues, among them D.J. Heeger, J. Schmidhuber, R.I. Wilson, J.A. Movshon and the attendees of the 2009 meeting on Canonical Neural Computation (http://www.carandinilab.net/canonicalneuralcomputation2009). The author's research is supported by the UK Medical Research Council (grant G0800791) and by the European Research Council (project CORTEX). The author holds the GlaxoSmithKline/Fight for Sight Chair in Visual Neuroscience.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

1. Weisberg, D.S., Keil, F.C., Goodstein, J., Rawson, E. & Gray, J.R. The seductive allure of neuroscience explanations. *J. Cogn. Neurosci.* **20**, 470–477 (2008).
2. Blakemore, C. *The Mind Machine*. (BBC Books, 1988).
3. Anderson, P.W. More is different. *Science* **177**, 393–396 (1972).
4. Laughlin, R.B. & Pines, D. The theory of everything. *Proc. Natl. Acad. Sci. USA* **97**, 28–31 (2000).
5. Laughlin, R.B., Pines, D., Schmalian, J., Stojkovic, B.P. & Wolynes, P. The middle way. *Proc. Natl. Acad. Sci. USA* **97**, 32–37 (2000).
6. Carandini, M. *et al.* Do we know what the early visual system does? *J. Neurosci.* **25**, 10577–10597 (2005).
7. Depireux, D.A., Simon, J.Z., Klein, D.J. & Shamma, S.A. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J. Neurophysiol.* **85**, 1220–1234 (2001).
8. DiCarlo, J.J. & Johnson, K.O. Receptive field structure in cortical area 3b of the alert monkey. *Behav. Brain Res.* **135**, 167–178 (2002).
9. Bizzi, E., Giszter, S.F., Loeb, E., Mussa-Ivaldi, F.A. & Saltiel, P. Modular organization of motor behavior in the frog's spinal cord. *Trends Neurosci.* **18**, 442–446 (1995).
10. Carandini, M. & Heeger, D.J. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62 (2011).
11. Watson, A.B. & Ahumada, A.J. Jr. A standard model for foveal detection of spatial contrast. *J. Vis.* **5**, 717–740 (2005).
12. Graham, N.V.S. *Visual Pattern Analyzers* (Oxford University Press, 1989).
13. Benucci, A., Ringach, D.L. & Carandini, M. Coding of stimulus sequences by population responses in visual cortex. *Nat. Neurosci.* **12**, 1317–1324 (2009).
14. Busse, L., Wade, A.R. & Carandini, M. Representation of concurrent stimuli by population activity in visual cortex. *Neuron* **64**, 931–942 (2009).
15. Priebe, N.J. & Ferster, D. Inhibition, spike threshold, and stimulus selectivity in primary visual cortex. *Neuron* **57**, 482–497 (2008).
16. Grothe, B. New roles for synaptic inhibition in sound localization. *Nat. Rev. Neurosci.* **4**, 540–550 (2003).
17. Koch, C. *Biophysics of Computation* (Oxford University Press, 1999).
18. Douglas, R.J., Koch, C., Mahowald, M., Martin, K.A.C. & Suarez, H.H. Recurrent excitation in neocortical circuits. *Science* **269**, 981–985 (1995).
19. Marr, D. *Vision* (W.H. Freeman & Co, 1982).
20. Buzsáki, G. Large-scale recording of neuronal ensembles. *Nat. Neurosci.* **7**, 446–451 (2004).
21. Harris, K.D. *et al.* How do neurons work together? Lessons from auditory cortex. *Hear. Res.* **271**, 37–53 (2010).
22. Marshel, J.H., Mori, T., Nielsen, K.J. & Callaway, E.M. Targeting single neuronal networks for gene expression and cell labeling *in vivo*. *Neuron* **67**, 562–574 (2010).
23. Seung, H.S. Neuroscience: towards functional connectomics. *Nature* **471**, 170–172 (2011).
24. Briggman, K.L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471**, 183–188 (2011).
25. White, J.G., Southgate, E., Thomson, J.N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **314**, 1–340 (1986).
26. Bower, J.M. & Beeman, D. *The book of GENESIS* (Springer, 1993).
27. Markram, H. The blue brain project. *Nat. Rev. Neurosci.* **7**, 153–160 (2006).
28. Waldrop, M.M. Computer modeling: brain in a box. *Nature* **482**, 456–458 (2012).
29. Bower, J.M. in *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System* (eds. J.M. Bower & D. Beeman) Ch. 11 (Springer-Verlag, 1998).
30. Sejnowski, T.J., Churchland, P.S. & Koch, C. Computational neuroscience. *Science* **241**, 1299–1306 (1988).
31. Prinz, A.A., Bucher, D. & Marder, E. Similar network activity from disparate circuit parameters. *Nat. Neurosci.* **7**, 1345–1352 (2004).
32. Mainen, Z.F. & Sejnowski, T.J. Influence of dendritic structure on firing pattern in model neocortical neurons. *Nature* **382**, 363–366 (1996).