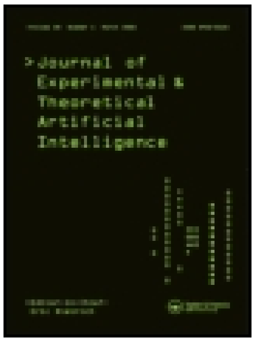

Taylor & Francis
Taylor & Francis Group

Journal of Experimental & Theoretical Artificial Intelligence

ISSN: 0952-813X (Print) 1362-3079 (Online) Journal homepage: http://www.tandfonline.com/loi/teta20

Why build a virtual brain? Large-scale neural simulations as jump start for cognitive computing

Matteo Colombo

To cite this article: Matteo Colombo (2016): Why build a virtual brain? Large-scale neural simulations as jump start for cognitive computing, Journal of Experimental & Theoretical Artificial Intelligence, DOI: 10.1080/0952813X.2016.1148076

To link to this article: http://dx.doi.org/10.1080/0952813X.2016.1148076

Published online: 22 Feb 2016.

JOURNAL OF EXPERIMENTAL & THEORETICAL ARTIFICIAL INTELLIGENCE, 2016
http://dx.doi.org/10.1080/0952813X.2016.1148076

Why build a virtual brain? Large-scale neural simulations as jump start for cognitive computing

Matteo Colombo

Tilburg center for Logic, Ethics and Philosophy of Science, Tilburg University, Tilburg, The Netherlands

ABSTRACT
Despite the impressive amount of financial resources recently invested in carrying out large-scale brain simulations, it is controversial what the pay-offs are of pursuing this project. One idea is that from designing, building, and running a large-scale neural simulation, scientists acquire knowledge about the computational performance of the simulating system, rather than about the neurobiological system represented in the simulation. It has been claimed that this knowledge may usher in a new era of neuromorphic, cognitive computing systems. This study elucidates this claim and argues that the main challenge this era is facing is not the lack of biological realism. The challenge lies in identifying general neurocomputational principles for the design of artificial systems, which could display the robust flexibility characteristic of biological intelligence.

ARTICLE HISTORY
Received 8 January 2015
Accepted 25 May 2015

KEYWORDS
Large-scale neural simulation; epistemology computer simulation; neuromorphic systems; biological realism; principles of neural computing

"If just reproducing the brain is the aim,

then there are better (presumably fun) ways to do that.

If we just want to simulate a real brain with all details,

maybe we should just reproduce." Anonymous

1. Introduction

Since the late 1980s, several research groups have been carrying out large-scale brain simulations. Carrying out large-scale brain simulations requires expertise from several different fields – including AI, machine learning, computational cognitive neuroscience, neurobiology and engineering – it also involves an impressive amount of financial resources. In the face of these efforts, it remains controversial what the pay-offs are of carrying out large-scale brain simulations. In particular, it is matter of heated debate whether any significant contribution to our understanding of cognitive behaviour could be made by simulating a large-scale model of the brain.

This study explores these issues, asking: currently, what do scientists learn from designing, building and running large-scale neural simulations? One plausible answer is that scientists learn at least what it takes to simulate these large-scale systems. By facing up to serious modelling and implementation

CONTACT Matteo Colombo m.colombo@uvt.nl

challenges, at least for some such simulations, scientists acquire new knowledge about the computational performance of their simulating systems.

Some neural simulations imitate some features of a real neural system not in order to serve as surrogates that are investigated to gain new knowledge about biological brains. Rather, these neural simulations imitate some features of a real neural system in order to gain useful knowledge about the simulating system itself (Colombo, 2015a).

Plausible as it sounds, the significance of this practice should not be downplayed, for at least two reasons. First, most work on the epistemology of computer simulation overlooks or downplays the computational and material aspects of computer simulation. But learning about the computational and material performance of a machine is in fact far from trivial (Sections 2 and 3). Second, the kinds of neural simulations examined in this study involve an interesting set of practices that are believed to usher in a new era of neuromorphic, cognitive computing systems (Section 4). Neuromorphic cognitive computing systems aim at solving complex, data-rich problems filled with uncertainties and ambiguity, by mimicking the function, power, volume and real-time performance of biological brains (Modha et al., 2011).

Many researchers believe that the major challenge in using large-scale neural simulations for developing neuromorphic cognitive computing devices lies in improving the biological realism of the simulations. Despite this widespread thought, the main challenge is not biological realism; rather, it lies in figuring out general neurocomputational principles that could enable artificial brains to display the robust flexibility characteristic of biological cognition (Section 5).

2. Large-scale neural simulations: aims and prospects

For many *large-scale neural simulations*, a simulating system implements some algorithm that finds solutions to mathematical equations that are believed to describe the dynamics and patterns of connectivity of a large number (e.g. over a million) of neurons and synapses (for reviews of different large-scale neural simulations, see Brette et al., 2007; de Garis, Shuo, Goertzel, & Ruiting, 2010; Eliasmith & Trujillo, 2014; Goertzel, Ruiting, Itamar, de Garis, & Chen, 2010; Sandberg & Bostrom, 2008).

A large-scale neural simulation is a type of computer simulation. Computer simulation can be characterised broadly as "a comprehensive method for studying systems," which "includes choosing a model; finding a way of implementing that model in a form that can be run on a computer; calculating the output of the algorithm; and visualizing and studying the resultant data" (Winsberg, 2013). Accordingly, some real-world system should be picked as the representational target of the computer simulation; some mathematical equations should be chosen, which are believed to model (some aspect of) the behaviour of the target system; and an appropriate simulating system, consisted of both hardware and software components, should be used to implement and run the mathematical model.

In line with much of the philosophical literature, where models and simulations are understood as serving as representations of some system about which one wants or hopes to gain knowledge (e.g. Grüne-Yanoff & Weirich, 2010, pp. 21–26; Hartmann, 1996, p. 83; Humphreys, 2004, p. 110), Winsberg (2013) claims that the entire process constituting computer simulation is "used to make inferences about the target system that one tries to model."

The claim also coheres with the stated goal of many large-scale neural simulations. For example, James M. Bower, who contributed to establishing *GENESIS*, one of the earliest neural simulators, in the early 1990s (Wilson, Bhalla, Uhley, & Bower, 1989), claims that understanding how biological brains compute will depend on "computer simulations that are very closely linked to the detailed anatomical and physiological structure" of the brain (Bower, 1998, p. 197).

More recently, the *Human Brain Project*, led by Henry Markram, set out to "simulate brains of mammals with a high level of biological accuracy and, ultimately, to study the steps involved in the emergence of biological intelligence" (Markram, 2006, p. 153).[1] The objective of carrying out large-scale neural simulations would be to understand why and how many different ion channels, receptors, neurons and synaptic pathways in the brain contribute to different brain functions and to emergent, intelligent behaviour (p. 158).

Similarly, Kwabena Boahen and collaborators have built *Neurogrid*, a system for simulating the behaviour of over a million neurons and their synaptic connections, with the ultimate objective of explaining "how intelligent behavior arises from bioelectrical processes at spatial and temporal scales six orders of magnitude smaller (from nanometers to millimeters and from microseconds to seconds)" (Benjamin et al., 2014, p. 699).

Despite significant differences, the goal shared by these and other large-scale neural simulation projects is to facilitate understanding of how brains' multi-scale, complex organisation contributes to generate intelligent behaviour. This goal may be reached. Yet, it is far from uncontroversial that, currently, a large-scale neural simulation is an especially fruitful or effective approach to addressing questions about how neurons and synapses' dynamics generate different brain functions, cognitive phenomena and intelligent behaviour (cf., Mainen & Pouget, 2014).

Commenting on this approach, Carandini (2012) argues that, currently, "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" (p. 509). If the underlying neural computations are not understood, there is little hope to learn how neural circuits generate different brain functions and integrated cognitive phenomena. A mechanistic understanding of how neural circuits generate intelligent behaviour requires the formulation and systematic testing of hypotheses about what is computed and by means of which algorithms. It cannot merely proceed from piecing together various, biophysical building blocks of neural circuitry.

In a similar vein, Sporns (2012) points out that the success of projects such as Markram's *Human Brain Project* "depends on knowledge about the organization of neurons and molecules into complex networks whose function underpins system dynamics" (p. 168). If the hope is to understand how intelligent behaviour is generated, then aggregating cells into circuits and circuits into systems in a neural simulation should be guided by knowledge of the computational architecture supporting brain function. Such knowledge is currently sparse and not easily incorporable into large-scale neural simulations, particularly into simulations that take a "bottom-up" approach, which is not driven by computational hypotheses about the function of different brain circuits.

Because of these issues, few large-scale neural simulations allow us to begin to bridge the gap between biophysical phenomena happening at the neural level and cognitive phenomena displayed by intelligent agents. Although, as we shall see in Section 5 below, a promising path has been opened up by Chris Eliasmith and colleagues' (2012) 2.5 million neuron simulation, most current large-scale neural simulations can at best fit existing biophysical data and display emergent properties that are not reducible to properties of individual brain components (e.g. Izhikevich & Edelman, 2008).

So, currently, carrying out large-scale neural simulations may be fruitful to explaining some biophysical phenomena displayed by networks of neurons. However, it is more problematic to claim that, currently, a large-scale neural simulation yields any novel insight into how the neurobiological systems represented in the simulation produce cognitive phenomena and contribute to intelligent behaviour.

3. Computational performance: from brains to computers

More plausible is that, currently, from at least *some* large-scale neural simulations, scientists gain novel knowledge about the computational performance of the simulating system itself, rather than about the neural system that the simulation represents.

Simulating systems are computing systems comprising both software and hardware components. They include a computational architecture and a set of algorithms appropriately formulated as computer programs that can be executed on a concrete computing machine made of specific materials and chips. The computational performance of the simulating system depends on a complex combination of properties of its architecture, of the algorithms it uses, the programs it executes and of the materials and technological devices of which it is made.

Three common dimensions on which computational performance can be assessed are as follows: the time it takes for the computing system to carry out a given task, the maximum number of tasks that can be completed by the system in a given time interval and the electrical power it takes for the system to carry out a task.

The total time required for a computing system to complete a task is called *execution time*. One way to measure the execution time of a program is in terms of clock period, which is the time length (in nanoseconds) of a cycle of the clock built into the system that determines when events take place in the hardware. The clock rate (in hertz) is the inverse of the clock period. Increasing computational performance for a given program requires decreasing its execution time, which may be tackled as an engineering problem – viz. as the problem of reducing the clock period (i.e. increasing the clock rate) – or as a computational problem – viz. as the problem of designing a more efficient computational architecture or more efficient algorithms and programs.

The number of tasks that can be completed per unit time by a computing system is called *throughput*. If we focus on the communication channels of a computing system, then the maximum throughput of a channel is often called *bandwidth* (measured in bits of data/second). The amount of time it takes for a communication channel to become unoccupied so that it can allow for data transfer is called *latency*. The available bandwidth of a communication channel is a limited resource and should be used sparingly. The greater the bandwidth capacity, or the lower the latency of the communication channels, the more likely it is that the system displays better computational performance. The throughput, bandwidth and latency of a computing system are a complex function of the physical medium being used for communications, the system's wiring architecture and the type of code used for programming.

Power consumption is a major constraint on computational performance. The microprocessors of computing systems dissipate heat. Heat must be removed from a computing system; else, its hardware components will overheat. Conserving power and avoiding overheating, while improving computational performance, have led computer scientists and engineers to explore novel architectures, hardware technologies, software solutions and programming languages for highly efficient computing systems.

There are two reasons why carrying out a computer simulation of a large number of neurons and synapses can yield non-trivial knowledge of the computational performance of the simulating system. The first reason is that brains can be understood as computational systems, which can be used to set a real biological benchmark for artificial computing systems' performance. The second reason has to do with scale: the *scalability* (or *scaling efficiency*) of a computing system indicates how efficient an application is when using increasing numbers of parallel processing units or amount of computational resources.

If the brain is a computing system, then it displays high performance in the face of low power consumption and small size. On average, the human brain weighs around 1.3–1.5 kg is constituted by about 100 billion neurons and around 100 trillion synapses, and its volume is about 1400 ml. For carrying out its computations, it consumes energy at a rate of about 20 W. Brains' computational architecture and style of computing are very different from those of modern artificial computing systems. Modern artificial computing systems possess von Neumann architecture and have stored programs, which are typically implemented in digital, serial, synchronous, centralised and fast microcircuits. By contrast, biological brains possess a non-von Neumann, multiscale, network architecture; they have distributed computational units, which carry out mixed-mode analogue–digital, parallel, asynchronous, slow, noisy, computations (Montague, 2007, Chapter 2; Piccinini & Bahar, 2013; von Neumann, 1958).

If the brain is a computing machine, then there is a set of properties possessed by both biological brains and artificial computing systems such that specific instantiations of these properties determine the computational performance that the computing machine – biological or otherwise – can reach. Available information about computational features of biological brains can provide *one* basis for benchmarking the performance of artificial computational systems along some dimension of interest such as power consumption or scalability. Comparing the computational performance of the simulating system in a large-scale neural simulation to that of its neurobiological target along some dimension of interest allows scientists to learn about why and how certain features of the simulating system (e.g. its network architecture, its physical materials) impact its performance relative to that dimension.

What about scalability? Although it is problematic to precisely define "scalability," the term is used in computer science to denote the capacity of a multiprocessor parallel computing system to accommodate a growing number of processing units or to carry out a growing volume of work gracefully (Hill, 1990). Scalability is a desirable feature of a computing system because it allows for hardware or software components to be added in the system without outgrowing it.

Two more specific notions, helpful to assess the performance of a large-scale simulation, are those of *strong scaling* and *weak scaling*, which denote ,respectively, the capacity of a system to reduce execution time for solving a fixed-size problem by adding processors, and the capacity to keep execution time constant by adding processors so as to accommodate additional workload. Assessing strong scaling is particularly relevant to learning about why some program takes a long time to run (something that is CPU-bound). Assessing weak scaling is particularly relevant to learning why some program takes a lot of memory to run (something that is memory-bound).

The scaling properties of a system provide scientists with useful criteria to assess the cognitive performance of a computational architecture. However, it is hard to predict whether the architecture underlying a large-scale neural simulation will scale. Predictions about scaling depend on background knowledge about the system of interest. "Scaling can seldom be fully characterized as 'more of the same,' because we may not know which 'same' is most relevant until we actually scale" (Eliasmith, 2013, 307). Thus, knowledge about the dimension along which scaling is carried out is crucial to determining whether the system will break down or will exhibit novel phenomena.[2]

While lack of scalability in a large-scale neural simulation may depend on a wrong choice in the relevant dimension to scale, it can indicate that the architecture of the simulating system cannot effectively solve problems of a certain size that biological brains can solve quickly. It can indicate that adding more simulated neurons and synapses to the simulating system is not an efficient strategy to execute a certain program more quickly, as the communication costs would increase as a function of the number of processors added to the system. It can also indicate that the power consumption required by a system that grows larger is too costly.

By taxing an artificial computing system by simulating millions of neurons and synapses, scientists can learn about trade-offs between memory, computation and communication in a certain computational architecture. Although acquiring this sort of knowledge is seldom the explicit goal of carrying out a large-scale neural simulation, it is useful knowledge that is in fact employed to design of neuromorphic, cognitive computing systems.

4. Brains, simulating systems and neuromorphic devices

Learning about the computational performance of a computing system can be important for developing *neuromorphic technologies for cognitive computing*. Neuromorphic technologies are devices for information processing and data analysis that aim to approximate the computational architecture and style of computing of biological brains in complementary metal-oxide semiconductor (CMOS) very large-scale integration systems. Such technologies include vision systems, auditory processors, multi-sensor integrators, autonomous robots and tools for handling and analysing large amount of data (Boahen, 2005; Choudhary et al., 2012; Indiveri & Horiuchi, 2011).

Systems of Neuromorphic Adaptive Plastic Scalable Electronics (SyNAPSE) is an ongoing research program funded by the US Defense Advanced Research Projects Agency (DARPA). "The vision for the SyNAPSE program is to develop electronic neuromorphic machine technology that scales to biological levels" (DARPA BAA08-28). This research program aims to develop electronic technology with similar computational performance to the mammalian brain in terms of size, speed and energy consumption.

Under the SyNAPSE program, Preissl and colleagues (2012) carried out a computer simulation of a very large neural circuit with the ultimate goal of exploring how closely one can "approximate the function, power, volume and real-time performance of the brain within the limits of modern technology" (p. 10). The representational target system of their simulation was a network comprising 65 billion neurons and 16 trillion synapses, which imitated the largest known wiring diagram in the macaque monkey's brain. This biological target was modelled as a network of neurosynaptic cores containing digital integrate-leak-and-fire neurons.

The simulating system involved a 16-rack Blue Gene/Q supercomputer of 16,384 to 262,144 CPUs and 256 TB of main memory, and *Compass*, a multi-threaded, massively parallel software, which enabled the simulation of billions of neurosynaptic cores operating in a parallel, distributed and semi-synchronous fashion.

The modelling choices of Preissl and colleagues were congenial to the pursuit of an engineering goal. The neurons, synapses and axons in their simulation were modelled as event-driven (asynchronous), digital, integrate-leak-and-fire circuits. The leaky integrate-and-fire model is one of the simplest models of spiking neurons. Given its lack of biophysical detail, the range of phenomena that this model can address is limited. Nonetheless, the model is analytically solvable and relatively easy to implement in a computer simulation. For many integrate-and-fire neurons models, the model fits nicely with an event-driven simulation, whereby all operations in the simulation are driven by neural spike events, which is generally well suited to decrease computational time and minimise memory load. The inter-core pattern of connections embodied in *Compass* imitated the macaque's neural wiring. That is, the relationship between the model-network and its neurobiological target was not isomorphic; it was a *similarity* relation, which can be sufficient to allow scientists to learn from computer simulations, especially when, like in this case, some relevant aspects and degrees of similarity are specified based on the research question at hand, on available background knowledge and on the larger scientific context (cf., Teller, 2001; Weisberg, 2013, Chapter. 8).

Implementing the macaque's wiring diagram "challenges the communication and computational capabilities of *Compass* in a manner consistent with supporting brain-like networks" (p. 11). The performance of the simulating system could then be compared with that of the real neurobiological system represented in the computer simulation. A quantitative characterisation of the deviations between the real neural system and the simulating system allowed scientists to identify which features of architectural and communication-design contributed to computational efficiency.

Preissl and colleagues' computer simulation could be used as a test bed for learning about the performance of hardware and software components of a simulating system put under serious computational stress. Simulating a neural network at that scale poses major challenges for computation, memory and communication, even with current supercomputers. If we consider N neurons, whose average firing rate is H, and whose average number of synapses is S, and we take account of all spike transmissions, then a real-time simulation of 1 s of biological time should process $N \times H \times S$ spike transmissions. This minimal number of operations set a benchmark to assess the computational performance of a neural simulation (Brette et al., 2007, pp. 350–351).

Preissl et al.'s (2012) simulation yielded two main results. First, as the average spiking rate of neurons was 8.1 Hz, the simulation was 388× slower than real time. Second, simulating the pattern of structural connectivity of the macaque's brain, the simulating system displayed near-perfect weak and strong scaling. While acquiring this type of information does not obviously yield novel insight about phenomena produced by biological brains, it is relevant to the development of more efficient neuromorphic, artificial computing systems. As Preissl and colleagues put it:

> *Compass* is a harbinger of an emerging use of today's modern supercomputers for midwifing the next generation of application-specific processors that are increasingly proliferating to satisfy a world that is hungering for increased performance and lower power while facing the projected end of CMOS scaling and increasing obstacles in pushing clock rates ever higher. (p. 11)

In using the brain as a template for machine intelligence, work such as Preissl and colleagues' work demonstrates that *some* large-scale neural simulations provide useful information for developing neuromorphic systems for cognitive computing, and *not* for understanding how biological brains work and produce intelligent behaviour.

5. Challenges for a new era of cognitive computing

Neuromorphic, cognitive computing systems aim at solving complex, data-rich problems, characterised by uncertainty and ambiguity by mimicking the function, power, volume and real-time performance of biological brains. Cognitive computing systems have three characteristic capacities: the capacity to learn from experience, the capacity to deal effectively with uncertainty and the capacity to extract useful information from sensory data and stored knowledge. These capacities rely on algorithms for unsupervised and reinforcement learning, for data mining, pattern recognition and natural language processing, which aim at mimicking the way biological brains work, and whose computational performance can be assessed courtesy of large-scale neural simulations (Kelly & Hamm, 2013; Modha et al., 2011).

When large-scale neural simulations aim at mimicking neural algorithms that sustain cognitive capacities and that may also be implemented in neuromorphic systems, several researchers believe that the main challenge lies in improving the realism of the simulations. They believe that neural simulations should build in as much biological detail as possible. Markram, for example, harshly criticised Modha and collaborators' work because of the lack of biological realism of their model neurons, which are "missing 99.999% of the brain" (Adee, 2010; see also de Garis et al., 2010, p. 3.2). The argument underlying this criticism is simple: If you use biological brains as templates for designing and building cognitive computing devices, then you should take account of as much biological detail as possible.

This argument is also misguided, however. For it assumes a generic ideal of biological realism that cannot inform large-scale neural simulations and related work in cognitive computing. The degree of realism (or descriptive accuracy) of a neural simulation is a function of the features included in the simulation that *matter* to the phenomena exhibited by the system. What matters and what does not is jointly determined by the causal structure of the real-world system under investigation, the modellers' varying epistemic interests and purposes in relation to that system and the modellers' audience (see also Colombo, 2015b).

Preissl et al. (2012) carried out their neural simulation in order to understand specific scaling properties of a particular simulator. Their neural simulation imitated some feature of the brain not in order to serve as a surrogate investigated in its stead. Rather, some features of the brain were imitated because the brain offers a biological benchmark against which a simulating system's design and performance can be assessed. Information about how certain properties determine the computational performance of biological brains can then be used not only to try and instantiate those properties in the design of artificial systems, but also to characterise the discrepancy between the brain's and the simulating system's performance. This characterisation provides insight about what types of constraints and what computational properties an artificial computing system need to instantiate for carrying out some task of interest more efficiently.

In particular, Preissl et al.'s (2012) neural simulation imitated some features of the biological brain in order to draw inferences about how closely the function, power, volume and real-time performance of the brain can be approximated within the limits of current technology. *Compass* incorporated "several innovations in communication, computation, and memory" based on available knowledge of some aspects of the function, power and volume of organic brains (p. 10). The neural scale and pattern of connectivity embodied in *Compass* challenged its communication, memory and computational capabilities. Given these challenges, the simulating system performance could be compared to that of a biological brain along dimensions such as neural spiking rates, latency and bandwidth. Running on the IBM Blue Gene/Q supercomputer, *Compass* was found to be 388× slower than real-time performance of the brain; importantly, *Compass* was found to have near-perfect weak and strong scaling when a model was run of the neural dynamics of a large circuit of the macaque's brain.

By themselves, these results do not yield novel information about some set of computational properties instantiated by biological brains. Instead, these results offered the basis for developing a novel, efficient, computational architecture that can support a host of neuromorphic applications (Kelly & Hamm, 2013). The biological details incorporated in Preissl et al's (2012) simulation did not serve to formulate an explicit hypothesis about how the brain generates intelligent behaviour. Choice about which biological features to include in the simulation (e.g. a macaque's wiring diagram) and which biological features to abstract away or distort (e.g. much biophysical features of real neurons) was functional to identifying and characterising certain computational properties of the simulator itself.

It is misguided to believe that the major challenge in using large-scale neural simulations for developing neuromorphic cognitive computing devices lies in improving the biological realism of the simulations. Large-scale neural simulations, and related work in cognitive computing, should not be evaluated in terms of a generic ideal of biological realism. The scientific or engineering purposes of a large-scale neural simulation should always inform judgements about its biological realism.

More serious is the challenge of identifying general neurocomputational principles that could enable the orchestrated functioning of learning, reasoning under uncertainty, and concept formation, characteristic of biological cognition, in a single, integrated, artificial system.

Currently, no unified theory of brain functioning and cognition is available that could guide AI efforts towards genuine cognitive computing. Many successful models and theories in current cognitive neurosciences are piecemeal, task-specific and sometimes mutually inconsistent. This makes it challenging to design artificial systems that can display flexible behaviour by relying on common computational resources that can be redeployed in order to tackle novel tasks. As a consequence, essentially all current neuromorphic technology can display only a narrow range of cognitive capacities.

Despite the apparent disunified status of current cognitive neuroscience, one theoretical development of the past twenty year has been the realisation that distinct neural processes, cognitive functions, algorithms and machine learning techniques can be understood in terms of *prediction* and *prediction error correction*. The idea is that brains are homeostatic prediction-testing mechanisms, the central activity of which is to minimise the errors of their predictions about the sensory inputs they receive from their local environment. The mechanistic activity of minimising prediction error would be constituted by various monitoring- and manipulation-operations on hierarchical, probabilistic, dynamic models of the causal structure of the world within a bidirectional cascade of cortical processing. Such kind of activity would give rise to perception, action, attention and a host of other cognitive capacities (see e.g. Clark, 2013; Friston, 2010; Tenenbaum, Kemp, Griffiths, & Goodman, 2011).

Unfortunately, this predictive processing theory of brains and cognition lacks implementational detail and unambiguous architectural commitments. Furthermore, this theory posits that brains represent probability distributions and carry out Bayesian inference on probabilistic models, which, for many real-life problem domains, would just be computationally intractable. Partly because of these outstanding issues, no attempt has been made thus far to build complex, unified, large-scale neural simulations or neuromorphic technologies grounded in predictive processing.

A related but distinct idea is that of *semantic pointers*. Chris Eliasmith and colleagues (2012) relied on this idea to build *Spaun*, a large-scale neural simulation that showed how a unified set of neurocomputational mechanisms can display aspects of the robust and rapid flexibility of biological systems. *Spaun* could display a variety of cognitive skills, including low-level perceptual and motor abilities, reward-based learning, high-level reasoning under uncertainty and concept formation.

"Semantic pointers are neurally realized representations of a vector space generated through a compression method" (Eliasmith et al., 2012, Supplementary Sections S.1). A semantic pointer is neurally realised because it consists of various spiking patterns in a population of biological or artificial neurons; if neural activity is represented with vectors, then semantic pointers correspond to vectorial representations. Semantic pointers are constructed by compressing sensory, motor, emotional, conceptual or symbolic representations, which are also patterns of neural firing captured by highly dimensional vectors. Analogously to JPEG picture files, semantic pointers provide "lighter" representations, whose dimensionality is lower than its constituents. Analogously to "pointers" in computer science, semantic pointers are the address of large data structures and can function as proxies for these data structures. Unlike pointers in computer science, semantic pointers systematically refer to the compressed data structures from which they were generated. Hence, they possess a "semantic."

In Eliasmith and colleagues' (2012) simulation, semantic pointers supported complex syntactic operations, regulated the flow of information and facilitated the orchestrated functioning of perception, learning, concept formation and reasoning under uncertainty. Implemented along with biologically plausible learning algorithms, like TD-learning (Sutton & Barto 1998; see also Colombo, 2014), semantic pointers allowed *Spaun* to get much of its cognitive powers from being able to deal with both the statistical properties and symbolic regularities underlying different psychological tasks.

In the light of this work and of current theoretical advances in the cognitive sciences, large-scale neural simulations that are grounded in *predictions*, *prediction error correction* and *semantic pointers* are most likely to yield results that will reliably guide us in the quest for cognitive computing.

6. Conclusion

It is not obvious what sort of new information scientists can currently gain from designing, building and running a large-scale neural simulation. This study has argued that for some large-scale neural simulation, what they learn concerns the computational performance of the simulating system itself. Learning about the computational performance of a computing machine is far from trivial and can afford knowledge useful for designing novel neuromorphic, cognitive computing technologies.

Once this role is recognised of some large-scale neural simulations, it should be clear that the main challenge a new era of cognitive is facing is not the lack of biological realism – as many believe. Instead, the challenge lies in figuring out how the computational components of an artificial brain should be arranged to produce the robust flexibility of biological cognition. *Prediction*, *prediction correction* and *semantic pointers* are three ingredients that will help researchers to face up to this exciting challenge.

Notes

1. On the website of the project, we read: "Reconstructing the brain piece by piece and building a virtual brain in a supercomputer – these are some of the goals of the *Blue Brain Project*. The virtual brain will be an exceptional tool giving neuroscientists a new understanding of the brain and a better understanding of neurological diseases." http://bluebrain.epfl.ch/cms/lang/en/pid/56882.
2. Eliasmith (2013, pp. 306–307) illustrates this difficulty with the case of Tusko. In 1962, a male Asiatic elephant, named Tusko, was injected with LSD in order to study the phenomenon of "musth," a remarkable change in violent behaviour displayed by adult male elephants. Researchers tried to induce musth by administering Tusko .1 mg of LSD per each kg of its body weight, for a total of 297 mg of LSD. The decision to scale the dosage to Tusko by body weight was based on known effects of LSD on cats and monkeys. Body weight turned out to be the wrong dimension for scaling. In fact, within five minutes, Tusko collapsed to the ground and one hour and forty minutes later he died (West, Pierce, & Thomas, 1962).

Acknowledgement

This was work supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program "New Frameworks of Rationality" ([SPP 1516]). I am also grateful to Lasha Abzianidze, Luigi Acerbi, one anonymous referee and to the editor of Jetai, Eric Dietrich, for their constructive criticisms and helpful suggestions.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Adee, S. (2010). The Markram Modha Controversy. *IEEE Spectrum, 1* (47), 16–17.

Benjamin, B. V., Gao, P., McQuinn, E., Choudhary, S., Chandrasekaran, A. R., Bussat, J., … Boahen, K. (2014). Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. *Proceedings of the IEEE, 102*, 699–716.

Boahen, K. (2005). Neuromorphic microchips. *Scientific American, 292*, 56–63.

Bower, J. M. (1998). Constructing new models. In James M. Bower & D. Beeman (Eds.), *The book of GENESIS: Exploring realistic neural models with the GEneral NEural SImulation System* (2nd ed., Chapter 11, pp. 195–201). New York, NY: Springer-Verlag.

Brette, R. M., Carnevale, T., Hines, M., Beeman, D., Bower, J., Diesmann, M., … Destexhe, A. (2007). Simulation of networks of spiking neurons: A review of tools and strategies. *Journal of Computational Neuroscience, 23*, 349–398.

Carandini, M. (2012). From circuits to behavior: A bridge too far? *Nature Neuroscience, 15*, 507–509.

Choudhary, S., Sloan, S., Fok, S., Neckar, A., Trautmann, E., Gao, P., … Boahen, K. (2012). Silicon neurons that compute. In *Artificial neural networks and machine learning – ICANN 2012* (pp. 121–128). Berlin Heidelberg: Springer.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*, 181–253.

Colombo, M. (2014). Deep and beautiful: The reward prediction error hypothesis of dopamine. *Studies in history and philosophy of science part C: Studies in history and philosophy of biological and biomedical sciences, 45*, 57–67.

Colombo, M. (2015a). Why Build a Virtual Brain? Large-scale Neural Simulations as Test-bed for Artificial Computing Systems. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 429–434). Austin, TX: Cognitive Science Society.

Colombo, M. (2015b). For a Few Neurons More. Tractability and Neurally-Informed Economic Modelling. *The British Journal for Philosophy of Science, 66*, 713–736.

de Garis, H., Shuo, C., Goertzel, B., & Ruiting, L. (2010). A world survey of artificial brain projects, Part I: Large-scale brain simulations. *Neurocomputing, 74*, 3–29.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. New York, NY: Oxford University Press.

Eliasmith, C., & Trujillo, O. (2014). The use and abuse of large-scale brain models. *Current Opinion in Neurobiology, 25*, 1–6.

Eliasmith, C., Terrence C., Stewart T. C., Choo, X., Bekolay, T., DeWolf, T., … Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science, 338*, 1202–1205.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*, 127–138.

Goertzel, B., Ruiting, L., Itamar, A., de Garis, H., & Chen, S. (2010). A world survey of artificial brain projects, Part ii. *Biologically inspired cognitive architectures. Neurocomputing, 74*, 30–49.

Grüne-Yanoff, T., & Weirich, P. (2010). Philosophy of simulation. *Simulation and Gaming: An Interdisciplinary Journal, 41*(1), 1–31.

Hartmann, S. (1996). The world as a process: Simulations in the natural and social sciences. In R. Hegselmann, U. Mueller, & K. G. Troitzsch (Eds.), *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View* (pp. 77–100). Dordrecht: Kluwer.

Hill, M. D. (1990). What is scalability? *ACM SIGARCH Computer Architecture News, 18*, 18–21.

Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*. New York, NY: Oxford University Press.

Indiveri, G., & Horiuchi, T. K. (2011). Frontiers in neuromorphic engineering. *Frontiers in Neuroscience, 5*, 118. doi: http://dx.doi.org/10.3389/fnins.2011.00118.

Izhikevich, E. M., & Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proceedings of the national academy of sciences, 105*(9), 3593–3598.

Kelly, J. E., & Hamm, S. (2013). *Smart machines: IBM's watson and the era of cognitive computing*. New York, NY: Columbia Business School Publishing.

Mainen, Z. F., & Pouget, A. (2014). European Commission: Put brain project back on course. *Nature, 511*, 534–534.

Markram, H. (2006). The blue brain project. *Nature Reviews Neuroscience, 7*, 153–160.

Modha, D. S., Ananthanarayanan, R., Esser, S. K., Ndirango, A., Sherbondy, A. J., & Singh, R. (2011). Cognitive computing. *Communications of the ACM, 54*, 62–71.

Montague, P. R. (2007). *Your brain is almost perfect*. New York, NY: Plume.

Piccinini, G., & Bahar, S. (2013). Neural computation and the computational theory of cognition. *Cognitive Science, 37*, 453–488.

Preissl, R., Wong, T. M., Datta, P., Myron, F., Raghavendra, S., Esser, S. K., … Modha, D. S. (2012, November 10–16). Compass: A scalable simulator for an architecture for cognitive computing. In *Proceedings of the international conference for high performance computing, networking, storage, and analysis. SC'12*. Salt Lake City, UT: IEEE Computer Society Press.

Sandberg, A., & Bostrom, N. (2008). *Whole brain emulation: A roadmap*. Oxford: Future of Humanity Institute, Oxford University.

Sporns, O. (2012). *Discovering the human connectome*. Cambridge, MA: MIT Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT press.

Teller, P. (2001). Twilight of the perfect model. *Erkenntnis, 55*, 393–415.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331*, 1279–1285.

von Neumann, J. (1958). *The computer and the brain*. New Haven, CT: Yale University Press.

Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. New York, NY: Oxford University Press.

West, L. J., Pierce, C. M., & Thomas, W. D. (1962). Lysergic acid diethylamide: Its effects on a male Asiatic elephant. *Science., 138*, 1100–1103.

Wilson, M., Bhalla, U., Uhley, J., & Bower, J. (1989). GENESIS: A system for simulating neural networks. In D. Anderson (Ed.), *Advances in neural information processing systems* (pp. 485–492). New York, NY: American Institute of Physics.

Winsberg, E. (2013). Computer simulations in science. In Edward N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2013 Edition). Retrieved from http://plato.stanford.edu/archives/sum2013/entries/simulations-science/