BUILDING A BEHAVING BRAIN

Chris Eliasmith

One of the grand challenges that the National Academy of Engineers identified is to reverse engineer the brain. Neuroscientists and psychologists would no doubt agree that this is, indeed, a grand challenge.

But what exactly does it mean to "reverse engineer" a brain? In general, reverse engineering is a method by which we take an already made product and systematically explore its behavior at many levels of description so as to synthesize (that is, *build*) a similar product. We attempt to identify its components and how they work, as well as how they are composed to give rise to the global behavior of the system. With systems as complex as the brain (or a competitor's silicon chip), the synthesizing step is usually carried out as a software simulation.

Reverse engineering the brain could bring many benefits. For instance, it would allow us to better understand the biological mechanisms that the brain employs and how they tend to fail in disease. At a more abstract level, reverse engineering the brain might allow us to discover effective information-processing strategies that we can import into our own engineered devices. Perhaps more surprisingly, our understanding of the basic properties of physical computation also stand to benefit from such research—neurons, after all, do not compute like a typical digital chip. In short, reverse engineering the brain will allow us to: (1) understand the healthy and unhealthy brain and develop new medical interventions, (2) develop new kinds of algorithms to improve existing machine intelligence, and (3) develop new technologies that exploit the physical principles exhibited by neural computation.

There are currently several large-scale brain simulations already being developed, each aiming to understand the actions of a million neurons or more. One project, supported by the Defense Advanced Research Projects Agency (DARPA), is IBM's SyNAPSE project, which aims to build a new kind of computer patterned after the brain. That

team recently announced a five-hundred-billion-neuron simulation (the human brain has about one hundred billion neurons). The individual neurons in SyNAPSE resemble actual neurons in that they generate neural action potentials (or "spikes") to communicate, and they incorporate some elements of individual neuron physiology (although they are much simpler than their biological counterparts in that they have no spatial extent and model only a few of the many currents found in a cell). A second high profile brain model is the €1 billion Human Brain Project, which grew out of the Swiss Blue Brain project, a simulation of (thus far) one million neurons. Although the total number of neurons simulated is small by comparison to the SyNAPSE project, the Human Brain Project aspires to model individual neurons in considerable detail, capturing neuron shape, hundreds of currents, and the dynamics of neural spiking for each cell. The trade-off for this increase in biological detail is that each neuron is far more computationally costly to simulate. Compared to the few equations per neuron in the SyNAPSE project, the Human Brain Project simulations have hundreds of equations per neuron. While this level of detail can be surpassed by adding more detailed molecular dynamics or including the important contributions of glial cells, at present this degree of biological fidelity is much higher than in other large-scale models.

From a reverse engineering perspective, large-scale simulations are an important step forward. They establish the computational feasibility of simulating large numbers of components. However, existing large-scale brain simulations like SyNAPSE and the Human Brain Project lack a key ingredient for successful reverse engineering: showing how the vast array of neural components relate to *behavior*. As yet, these models do not remember, see, move, or learn, so it is difficult to evaluate them in terms of what is, arguably, the purpose of the brain.

Behavior and the Brain

My group has taken a different approach, aimed at understanding the neural underpinnings of behavior. Our most recent model, called Spaun (Semantic Pointer Architecture Unified Network), has a single eye, which takes digital images as input, and a single, physically simulated

arm, which it moves to provide behavioral output (see figure 1a). Internally, its 2.5 million neurons generate neural spikes to process the input (for example, recognize and remember digits) and generate relevant output (for example, draw digits with its arm; see figure 1b). These neurons are organized to simulate about twenty out of the approximately one thousand different areas typically identified in the brain (see figure 2a). These areas were chosen to provide a suitably rich set of functions while remaining computationally tractable. The biophysical model of individual neurons that Spaun uses is quite simple. As in the SyNAPSE project, only a few equations are needed to describe each neuron. These neurons communicate using neural action potentials (spikes). When impacting a neighboring neuron's synapse, these spikes elicit a simulated version of one of four neurotransmitters (out of the tens or hundreds of different kinds) found in the brain. Again, this level of physiological and anatomical detail provides a practical compromise between computational simplicity and functionality.

One of Spaun's virtues, relative to SyNAPSE and the Human Brain Project, is its global, brain-like structure. Whereas the neurons in SyNAPSE form a largely undifferentiated, or statistically uniform mass, in Spaun they are organized to reflect the known anatomy and function of the brain. One set of neurons is modeled after those in the frontal cortices, playing important roles in working memory and the tracking of task context. Other neurons make up a simulated basal ganglia, where they help the model learn new behavioral strategies and control the flow of information throughout much of the cortex. Still others are modeled after the neurons in the occipital lobe, allowing Spaun to visually recognize handwritten digits it has never seen before. Neurons in Spaun are physiologically similar (that is, using the kinds of neurotransmitters found in that part of the brain, spiking at similar rates, and such), functionally similar (that is, are active in similar ways under similar behavioral circumstances as in the brain), and are connected in a similar manner (that is, receiving inputs from and projecting out to some of the same brain areas that a real neuron would) to neurons in the corresponding area of a biological brain (see figure 2a). For example, there are two different kinds of medium spiny neurons in the simulated basal ganglia that receive cortical projections and are inhibitory, but they have different kinds of dopamine receptors and project to different parts of the globus pallidus.

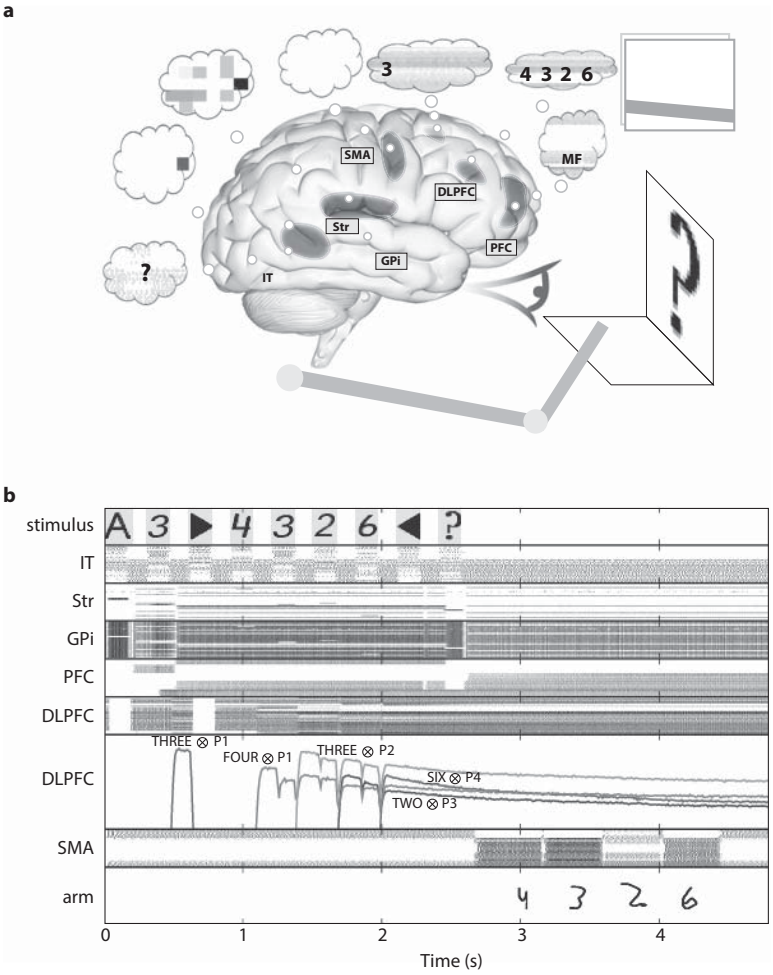

Figure 1. A serial working memory task in Spaun. a. A conceptual description of the processing Spaun performs. It is first shown a randomly chosen handwritten digit that it compresses through its visual system, allowing it to recognize the digit and map it to a conceptual representation (or "semantic pointer," SP). That representation is then further compressed by binding it to its position in the list and storing the result in working memory. Any number of digits can be shown in a row and will be processed in this manner. Once a question mark is shown, Spaun proceeds to decode its working memory representation by decompressing the items at each position and sending them to the motor system to be written out, until no digits remain. b. A screen capture from the simulation movie of this task, taken 2.5 s into the simulation time course shown in c. The input image is on the right; the output is drawn on the surface beside the arm. Spatially organized (neurons with similar tuning are near one another), low-pass-filtered model neuron activity is approximately mapped to the relevant cortical areas

To evaluate the model we compared it to a range of empirical data, drawn from both neurophysiological and behavioral studies. For instance, a common reinforcement learning task asks rats to figure out which of several actions is the best one, given some probabilistic reward (as if it were choosing between better- and worse-paying tables in a casino). Single neuron spike patterns can be recorded from the animals while they are performing this task. Spaun matches the behavioral choice patterns of the rat, but in addition, the firing patterns of neurons in the ventral striatum of both the model and the rodent exhibit similar changes during delay, approach, and reward phases of this task.

There are several examples of Spaun's neural firing patterns reproducing those found in real brains. In comparing to spiking data gathered from monkeys performing a simple working memory task, Spaun exhibits the same spectral power changes of populations of neurons (and of single neurons) while performing the same task. Similarly, by comparing to data from a monkey visual task, we have shown that the tuning of neurons in the primary visual area of the model matches those recorded in monkeys. In each case, the spiking data from the model and the animal were analyzed using exactly the same methods, to generate appropriate comparisons.

While matches to single neuron data can help build confidence in the basic mechanisms of the model, if we want to understand *human* cognition, it is often the case that such data is unavailable. As a result, in studying humans we must often rely more on behavioral comparisons. Here again, Spaun provides a good fit in many cases. For example,

---

and shown in gray scale (dark is high activity, light is low). Thought bubbles show example spike trains, and the results of decoding those spikes are in the overlaid text. For striatum (Str), the thought bubble shows decoded utilities of possible actions, and in globus pallidus internus (Gpi) the selected action is darkest. c. Time course of a single trial of the serial working memory task for four digits. The stimulus row shows input images. "A3" indicates it is performing task 3 (serial working memory), the triangles provide structure to the input, and the question mark indicates a response is expected. The arm row shows digits drawn by Spaun. Other rows are labeled by their corresponding anatomical area. Similarity plots (solid gray lines) show the dot product (i.e., similarity) between the decoded representation from the spike raster plot and concepts in Spaun's vocabulary. Raster plots in this figure are generated by randomly selecting 2,000 neurons from the relevant population and discarding any neurons with a variance of less than 10 percent over the run. Adapted from Eliasmith (2013).

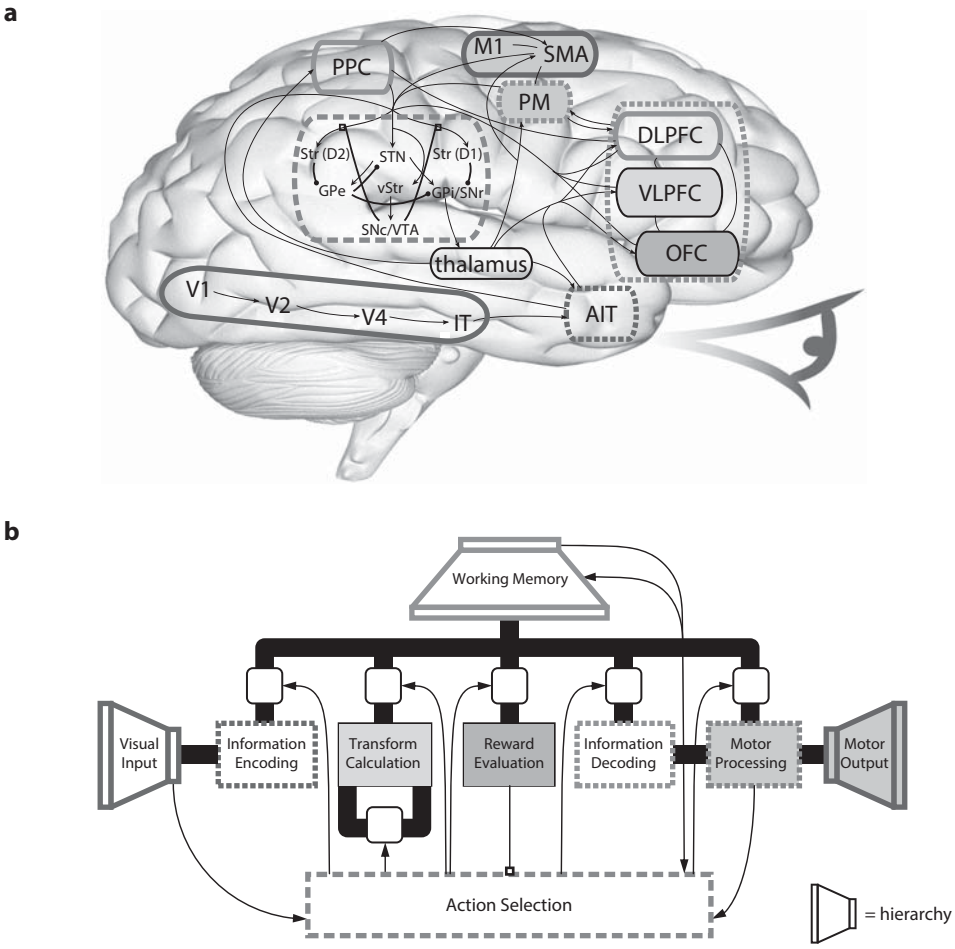

Figure 2. The architecture of the Spaun model. a. The anatomical architecture of the model (using standard anatomical abbreviations) drawn on the outline of a brain to indicate correspondences between model components and brain areas. Lines with circular endings indicate inhibitory projections. Lines with square boxes indicate modulatory connections exploited during learning. Other connections are excitatory. b. The functional organization of the model showing information flow between components. Thick lines indicate information flow between elements of model cortex, thin lines indicate information flow between the action selection mechanism (basal ganglia) and model cortex, and rounded boxes indicate elements that can be manipulated to control the flow of information within and between subsystems. The circular end of the line connecting reward evaluation and action selection indicates that this connection modulates connection weights. Line styles and fills indicate the mapping to the anatomical architecture in a. Adapted from Eliasmith (2013).

Spaun makes the same kinds and frequency of errors as humans during a serial working memory task (this task requires remembering and repeating back a list of numbers). This suggests that the neural mechanisms in the model are plausible, although evidence is indirect. Similarly, Spaun takes the same length of time per count as people do when internally counting numbers. Moreover, it also parallels people in showing an increase in the variance of the reaction of time for longer counts, reproducing Weber's famous law from psychophysics. There are many tests yet to be run, but as we continue to test the model in a variety of ways—both neurally and behaviorally—we strengthen our case that the principles we have used for reverse engineering the brain are on the right track.

This case is made significantly stronger by noting that it is the exact same model being used in each of these comparisons. Mathematical models, like Spaun, often have parameters that are tuned to match specific experimental results. This leads to the common worry that a model is "overfit" to a particular experiment or type of experiment. However, we have made significant efforts to allay this concern. For example, the decay rate of working memory is set using data from human experiments that are not included in any of the eight tasks that Spaun does. Many of the other parameters are set automatically using three principles of neural implementation that we have developed over a decade of research (Eliasmith and Anderson 2003). But, most importantly, no matter how they are set, they remain constant across all of the tasks that Spaun performs (or, more accurately, only the model can change them itself, through learning). By leaving these parameters untouched across experiments, and by testing the model against a wide variety of experiments, such concerns of overfitting become less plausible because the parameters are clearly not picked to work only under one or a few experimental conditions.

One of the central reasons for constructing such a model is to determine what it can teach us about how the brain functions. Interestingly, Spaun has generated several specific predictions that are currently being tested. For instance, the model exhibits a particular pattern of errors on question answering tasks, despite a constant reaction time in responding to questions. In particular, questions about either the identity or position of an item are more likely to be incorrectly answered the closer

they are to the middle of the list. To the best of our knowledge this task has not yet been run on people, consequently it is an ideal prediction to test. Spaun has also given rise to specific neural predictions. For example, it suggests a particular pattern of similarity between the neural activity during encoding of a single item in working memory, versus encoding that same item along with other items. Specifically, the similarity of neural firing in Spaun drops off exponentially as items are added. This prediction contradicts that from other models of working memory in which the similarity stays constant. As a result, this particular prediction is an excellent test of the mechanisms and assumptions of Spaun.

In contrast to large-scale simulations that produce a lot of neural activity but little observable behavior, I would argue that Spaun is providing detailed, quantifiable insights into the organization *and function* of the brain. (Videos of many of the experiments run on Spaun can be found at http://nengo.ca/build-a-brain/spaunvideos.)

Coordinated for Flexibility

One key contribution of Spaun relative to many competing architectures is that Spaun can perform a variety of *different* behaviors, much like an actual brain. For example, Spaun can use its visual system to recognize numbers that it then organizes into a list and stores in its working memory. It can later recall this list and draw the numbers, in order, using its arm. Furthermore, Spaun can use this same visual system to parse more complex input and recognize patterns in digits it hasn't seen before. To do so, it uses the same memory system, but in a slightly different way. As well, it uses other brain areas that it didn't use in the list recall task. That is, Spaun can deploy the same brain areas in different ways depending on what task it needs to perform (see figure 2b).

This kind of "flexible coordination" is something that sets animal cognition apart from most current artificial intelligence. Animals can determine what kinds of information processing needs to be brought "online" in order to solve a given challenging problem. In other words, different, specialized brain areas are *coordinated* in a task-specific—that is, *flexible*—way to meet a challenge presented by the environment. As people, this ability comes very naturally to us, so it is an ability that we

often overlook. When I switch from composing an e-mail to reading a book, to making a drink, to chasing my cat, I have coordinated many different parts of my brain in many different ways, and often with little delay in between. Because animals have evolved in a dynamic, challenging environment, this kind of behavioral flexibility is critical. In fact, Merlin Donald and others have suggested that humans are incredibly evolutionarily successful because they exhibit this kind of adaptability better than almost any other species.

One of the central goals of the Spaun project is to develop a preliminary understanding of how this kind of flexible coordination occurs in the mammalian brain. As a result, there is an important distinction in the model between midbrain and cortical regions. The midbrain regions, dominated by the basal ganglia, play a crucial role in coordinating information processing largely carried out in the cortex. So the architecture of Spaun essentially consists of an "action selector" (the basal ganglia), which monitors the current state of the cortex and determines how information needs to flow through the cortex to accomplish a given goal. However, the basal ganglia itself doesn't perform complex actions. Instead, it helps organize the cortex, so the massive computing power available there can be directed at the current problem in the right way. This allows Spaun to perform any of eight very different tasks in any order, while remaining robust to unexpected input and noise. Spaun determines what task to do by understanding its input. When it sees the letter "A" followed by a number, Spaun determines how to interpret subsequent input (for example, "A3" means that it should memorize the list of numbers it is shown next; see figure 1b).

The Benefits of Reverse Engineering

It is perhaps not surprising that in the mammalian brain, the basal ganglia have been found to be important for selecting what to do next. Problems including damage and neurodegeneration in the basal ganglia result in behaviors related to addiction, anxiety, and obsessive compulsive disorder. As well, the tremors associated with Parkinson's disease find their root in a malfunction of these areas. Consequently, understanding the mechanisms that underwrite flexible coordination have significant consequences for health.

In a similar vein, Spaun has allowed us to cast light on the cognitive decline associated with aging. There is currently a long-standing debate about whether or not the known reduction in brain cells with aging is related to the measured decline in performance on cognitive tests. The Raven's Progressive Matrices (RPM) test is a standard IQ test that has often been used to track this kind of change. The RPM test asks subjects to figure out how to complete a visual pattern of some kind. In fact, one of the tasks that Spaun performs is modeled after this test (and Spaun has been shown to perform about as well as a human of average intelligence). More recently, my lab has developed a model using the same architecture as Spaun that is able to perform the exact same test as is used on human subjects. Again, it performs about as well as average humans. Because the model has neurons, we can, for the first time, explore the causal relation between damaging those cells (as happens naturally during aging) and performance on the RPM. By running hundreds of versions of this kind of model, we can show that the performance of the models reproduces the standard "bell curve" of human populations, and that neuron loss due to aging can cause a uniform shift downward in that distribution. In short, we have been able to show how the cognitive decline due to aging could be a direct result of neuron loss.

Less obviously, understanding brain mechanisms is likely to provide us with new insights into how to build intelligent artificial systems. At the moment, most successes in machine intelligence master a single ability: machines are good at playing chess, or answering *Jeopardy!* questions, or driving a car. People, of course, can be quite good at all of these tasks. I believe this is because people can flexibly coordinate their skills in ways not currently available to machines. While most of the specific tasks that Spaun performs can be reproduced by artificial intelligence algorithms, the variety of tasks that Spaun performs is atypical of the field. Interestingly, Spaun also exhibits a nascent ability to learn new behaviors on its own (specifically, it can learn to choose different actions based on rewards in a limited manner), while preserving abilities it already has. One focus of future research on Spaun is to expand this ability to allow it to learn much more sophisticated tasks on its own, either through explicit instruction or through trial and error learning.

Building a *Physical* Brain

Even if we did understand the algorithms of the brain, it is not clear that we could usefully implement them on the computers of today. This is because the physical strategies the brain adopts for processing information lie in stark contrast to those we currently use in our computers. Silicon chips in our computing devices are engineered to eliminate uncertainty: transistors are either "on" or "off." This precision comes at the price of high power usage. Desktop computers of today typically use hundreds of watts. The brain, in contrast, uses only about 25 watts, and it performs far more sophisticated computations. And, it seems, the brain relies on highly unreliable, noisy devices: synapses fail much of the time, neurotransmitters are packaged in variable amounts, and the length of time it takes an action potential to travel down an axon can change.

Through reverse engineering, researchers have noticed these fundamental differences and have been motivated to develop "neuromorphic" silicon chips. Several of these chips arrange basic analog components of silicon chips in a manner that models of the behavior of cells in the cortex; these models have voltages with dynamics like those of neurons, and even communicate using spikes and synapses the way neurons do. Millions of such neurons can be arranged into a space smaller than a deck of cards and use less than 3 watts of power. In addition, they run in real time. This is important, since Spaun, for example, takes about 2.5 hours of real time to simulate 1 second of behavior using a digital supercomputer and kilowatts of power.

One reason these chips are promising is that many are currently fabricated with decades-old digital technology. Consequently, as they are moved to newer, already available, fabrication facilities they will be able to exploit the exponential improvement in component density. Furthermore, the limits on size affecting digital technology may not apply in the same way to neuromorphic approaches. This is because these limits are often a consequence of noise resulting from unexpected behaviors when devices get very small. Neuromorphic technology, being modeled after the noisy, stochastic brain, has faced such problems throughout its development: like the brain, neuromorphic hardware tends to be low power, analog, and asynchronous. These features tend to make the

effects of noise very salient—effects usually "engineered away" in digital hardware. Consequently, the improvements in computing power and efficiency we tend to expect of digital technology may now be more readily realized by brain-based approaches.

However, one challenge in usefully employing such neuromorphic hardware has historically been a lack of methods for systematically programming noisy, low-power, highly variable hardware of this type. But, as we continue to reverse engineer neural algorithms to build large-scale brain models, we have been concurrently developing such methods. Indeed, the same techniques used to build Spaun (called the Neural Engineering Framework, or NEF; Eliasmith and Anderson 2003) have been used to program several different kinds of neuromorphic chips. Consequently, the future of both neuromorphic programming and large-scale brain modeling are intimately tied. Together I believe they will usher in a new era of low-powered, robust, flexible, and adaptive computing.

In conclusion, efforts to address the grand challenge of reverse engineering the brain are clearly underway. Large-scale models at various levels of biological detail are being developed around the world. Models like Spaun—models that connect the activity of individual neurons to behavior—are an important part of that effort, as they provide fertile, specific hypotheses that stand to significantly improve our understanding of how the brain works. While Spaun has forty thousand times fewer neurons than are in the human brain, it nevertheless provides testable and predictive ideas about neural organization and function. As such models improve—and they are likely to do so exponentially in the coming years—they will have far-reaching consequences for the development of new treatments and new technologies. These models will begin to shed light on one of the most complex physical systems we have ever encountered, and, in so doing, change our basic understanding of who we are.

References

Eliasmith, C. 2013. *How to Build a Brain: A Neural Architecture for Biological Cognition*. Oxford: Oxford University Press.

Eliasmith, C., and C. Anderson. 2003. *Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems*. Cambridge, MA: MIT Press.