How to write a "How to Build a Brain" book*

Shimon Edelman

October 30, 2013

These are exciting times for those of us interested in comprehensive theories of cognition and the brain. A side by side comparison of Chris Eliasmith's new book, *How to Build a Brain* (2013), and W. Ross Ashby's *Design for a Brain* (1952) (to pick a similar earlier title) reveals just how much progress has been achieved in cognitive and brain sciences in the intervening six decades. In *HBB*, Eliasmith aims to translate this progress into constructing a broad-coverage, unified model of how the brain works, inviting the reader to play along by testing the accompanying code on a selection of problems, sampled from perception/action, memory, reasoning, and planning.

The conceptual vehicle that Eliasmith constructs for this purpose (clearly after a serious deliberation: many pages are devoted to explaining his design choices) seems, on the first reading, to lack a means of propulsion. Part I ("How to build a brain") is painstakingly detailed with regard to implementational niceties, such as the properties of the model neurons' spike trains and the items on pull-down menus in the accompanying simulation environment that must be clicked to make those neurons spike as intended. Discovering finally, in Part II ("Is that how you build a brain?"), an attempt to engage with the computational/functional issues that motivate the choice of algorithms and implementations back in Part I, I realized that the cart in this rig has been placed firmly before the horse, which, moreover, may not be easily persuaded to push where it needs to, or, indeed, to make the cart budge at all.

One would imagine that thirty years and more after a proper methodology for cognitive science has been formulated (Marr and Poggio, 1977; see (Edelman, 2008a, ch.4) for a textbook treatment and (Poggio, 2012) for a retrospective and an update), any attempt to explain the brain would take heed of it. The cart of explanation, especially one that carries the entire brain, will only move in the right direction if we hitch it to adequately explicit *computational* theories of whatever it is that the brain is up to.

Eliasmith does consider Marr's framework, only to set it aside (p.64). The result is that the book's many interesting ideas and contributions — e.g., semantic pointers, the importance of control, dealing with systematicity — are lost in the methodological maze of Part I. Theoretical projects of a similar scope typically paint a "big picture" of the brain's function. For instance, Clark (2013) posits and then defends the hypothesis that the task of the brain is prediction, which may or may not be right, but which does in any case connect with a crucially important category of questions at the right level of abstraction. In comparison, *HBB* does not put goal- or problem-level hypotheses front-and-center. It also tends to mix computational-level concepts (e.g., Bayes theory) with implementation-level ones (connectionism), while entertaining some false dichotomies (e.g., symbolic vs. dynamical systems; cf. Edelman, 2008b).

One manifestation of this methodological mishmash is a list of "core cognitive criteria" by which success

*A review of *How to Build a Brain*, C. Eliasmith (2013).

in brain-building is to be judged. These are introduced on p.16 and revisited on p.364, where we are offered a table that evaluates his approach alongside what he considers to be the main competitors. Tantalizingly, we're still not told what it is that the brain does, whether according to his model or any of the others.

A potential sticking point that recurs throughout the book is Eliasmith's use of the concept of semantics, which is central to his approach. The claim, found on p.79, that "in more connectionist approaches, semantics has been the focus" would surely surprise many of my colleagues (and not just the linguists). It turns out that what he means by "semantics" is similarity — a philosophically and empirically arguable reduction, which is never discussed beyond a brief passage on p.370, where the reader is referred to Eliasmith's earlier publications on "neurosemantics."

Indeed, combining cues from the introduction and the later chapters, I gathered that what Eliasmith thinks the brain does is similarity-based categorization (pp.19,88,247). The space allotted for this review doesn't allow me to discuss the notion that a single overarching idea may do justice to the question of brain function, or to examine, specifically, the viability of the familiar postulate of the centrality of similarity (which I quite like, albeit not without reservations, and about which I have written in the past). I wish that it were brought out and treated more explicitly, right from the start.

From the engineering standpoint, Eliasmith's modeling framework is very sensible. It uses networks of spiking neurons to implement the holographic reduced representation principle, adopted from Plate (1995), which is known for its ability to support compositional representations (Jones and Mewhort, 2007). Given his declared goal of building a brain, I am not convinced, however, that enough attention is given either to the global or circuit-level or to the local or cortical column-level anatomy of real brains (as, for instance, in the models of O'Reilly (2006) and Maass, Natschläger, and Markram (2003), respectively). This design choice detracts from what could have been a major contribution of this book: its attempt to deal with disparate problems in the same architecture. In the end, the circuits hand-constructed for the different problems (e.g., the Towers of Hanoi and the Wason card sorting task) look quite different.

The bespoke nature of those problem-specific incarnations of Eliasmith's general model and the excessive focus on the level of spiking neurons results in some important computational-level aspects of those problems being missed. For instance, a circuit tailored to solve Raven's matrix analogy is described as a success, because it decides correctly how many triangles should go into the empty cell in the matrix. And yet, the real problem here is for the model to realize without being hard-wired for it that the real challenge it faces is to figure out what a triangle is and that counting triangles is what it needs to do in the first place (cf. the discussion of the Bongard problems in Hofstadter, 1979).

Part II of the book consists of three chapters dedicated to coaxing the pony of computational-level understanding into giving a nudge to the cart of implementational and performance details, which, however, squarely blocks off its field of view. My impression is that the arguments marshaled in these chapters are repetitive (witness how many paragraphs begin with "Again") and that too much effort is spent on issues that have little bearing on the kind of big-picture understanding of the brain that one expects from this book. Upon finishing *HBB*, I was left with a feeling that chunks of the big picture are nevertheless there, fragmented and scrambled. Perhaps a better reader than I can piece the explanation together and set the pony free.

References

Ashby, W. R. (1952). *Design for a brain*. London: Chapman & Hall.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences 36*, 181–204.

Edelman, S. (2008a). *Computing the mind: how the mind really works*. New York: Oxford University Press.

Edelman, S. (2008b). On the nature of minds, or: Truth and consequences. *Journal of Experimental and Theoretical AI 20*, 181–196.

Eliasmith, C. (2013). *How to build a brain*. New York, NY: Oxford University Press.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: an Eternal Golden Braid*. New York: Basic Books.

Jones, M. N. and D. J. K. Mewhort (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review 114*, 137.

Maass, W., T. Natschläger, and H. Markram (2003). Computational models for generic cortical microcircuits. In J. Feng (Ed.), *Computational Neuroscience: A Comprehensive Approach*, Chapter 18, pp. 575–605. Boca Raton, FL: CRC-Press.

Marr, D. and T. Poggio (1977). From understanding computation to understanding neural circuitry. *Neurosciences Res. Prog. Bull. 15*, 470–488.

O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition. *Science 314*, 91–94.

Plate, T. A. (1995). Holographic Reduced Representations. *IEEE Transactions on Neural Networks 6*, 623–641.

Poggio, T. (2012). The levels of understanding framework, revised. *Perception 41*, 1017–1023.