Principles Underlying the Construction of Brain-Based Devices

Jeffrey L. Krichmar
The Neurosciences Institute
10640 John J. Hopkins Drive
San Diego, CA 92121 USA
krichmar@nsi.edu

Gerald M. Edelman
The Neurosciences Institute
10640 John J. Hopkins Drive
San Diego, CA 92121 USA
edelman@nsi.edu

Abstract

Without a doubt the most sophisticated behaviour seen in biological agents is demonstrated by organisms whose behaviour is guided by a nervous system. Thus, the construction of behaving devices based on principles of nervous systems may have much to offer. Our group has built series of brain-based devices (BBDs) over the last 14 years to provide a heuristic for studying brain function by embedding neurobiological principles on a physical platform capable of interacting with the real world. These BBDs have been used to study perception, operant conditioning, episodic and spatial memory, and motor control through the simulation of brain regions such as the visual cortex, the dopaminergic reward system, the hippocampus, and the cerebellum. Following the brain-based model, we argue that an intelligent machine should be constrained by the following design principles: (i) it should incorporate a simulated brain with detailed neuroanatomy and neural dynamics that controls behaviour and shapes memory, (ii) it should organize the unlabeled signals it receives from the environment into categories without a priori knowledge or instruction, (iii) it should have a physical instantiation, which allows for active sensing and autonomous movement in the environment, (iv) it should engage in a task that is initially constrained by minimal set of innate behaviours or reflexes, (v) it should have a means to adapt the device's behaviour, called value systems, when an important environmental event occurs, and (vi) it should allow comparisons with experimental data acquired from animal nervous systems. Like the brain, these devices operate according to selectional principles through which they form categorical memory, associate categories with innate value, and adapt to the environment. Moreover, this approach may provide the groundwork for the development of intelligent machines that follow neurobiological rather than computational principles in their construction.

1 Introduction

Although much progress has been made in the neurosciences over the last several decades, the study of the nervous system is still a wide open area of research. This is not due to a lack of first-rate research by the neuroscience community, but instead it reflects the complexity of the problem. Therefore, novel approaches to the problem, such as computational modelling and robotics, may be necessary to come to a better understanding of brain function. Moreover, as our models and devices become more sophisticated and more biologically realistic, the devices themselves may approach the complexity and adaptive behaviour that we associate with biological organisms and may find their way in practical applications. In this review, we will outline what we believe are the design principles necessary to achieve these goals (Krichmar and Edelman, 2005; Krichmar and Reeke, 2005). We will illustrate how these principles have been put into practice by describing two recent brain-based devices (BBDs) from our group.

2 Brain-Based Modelling Design Principles

2.1 Incorporate A Simulated Brain With Detailed Neuroanatomy And Neural Dynamics

Models of brain function should take into consideration the dynamics of the neuronal elements that make up different brain regions, the structure of these different brain regions, and the connectivity within and between these brain regions. The dynamics of the elements of the nervous system (e.g. neuronal activity and synaptic transmission) are important to brain function and have been modelled at the single neuron level (Borg-Graham, 1987; Bower and Beeman, 1994; Hines and Carnevale, 1997), network level (Izhikevich et al., 2004; Pinsky and Rinzel, 1994), and synapse level in models of plasticity (Bienenstock et al., 1982; Song et al., 2000; Worgotter and Porr, 2005). However, structure at the gross anatomical level is critical for function, and it has often been ignored in models of the nervous system. Brain function is more than the activity of disparate regions; it is the interaction between these areas that is crucial as we have shown in Darwins IV through X (Edelman et al., 1992; Krichmar and Edelman, 2005; Krichmar et al., 2005b; Seth et al., 2004). Brains are defined by a distinct neuroanatomy in which there are areas of special function, which are defined by their connectivity to sensory input, motor output, and to each other.

2.2 Organize the Signals from the Environment into Categories Without a priori Knowledge or Instruction

One essential property of BBDs, is that, like living organisms, they must organize the unlabeled signals they receive from the environment into categories. This organization of signals, which in general depends on a combination of sensory modalities (e.g. vision, sound, taste, or touch), is called *perceptual categorization*. Perceptual categorization in models (Edelman and Reeke, 1982) as well as living organisms makes object recognition possible based on experience, but without *a priori* knowledge or instruction. A BBD selects and generalizes the signals it receives with its sensors, puts these signals into categories without instruction, and learns the appropriate actions when confronted with objects under conditions that produce responses in value systems.

2.3 Active Sensing and Autonomous Movement in the Environment

Brains do not function in isolation; they are tightly coupled with the organism's morphology and environment. In order to function properly, an agent, artificial or biological, needs to be situated in the real world (Chiel and Beer, 1997; Clark, 1997). Therefore, models of brain function should be embodied in a physical device and explore a real as opposed to a simulated environment. For our purposes, the real environment is required for two reasons. First, simulating an environment can introduce unwanted and unintentional biases to the model. For example, a computer generated object presented to a vision model has its shape and segmentation defined by the modeller and directly presented to the model, whereas a device that views an object hanging on a wall has to discern the shape and figure from ground segmentation based on its on active vision. Second, real environments are rich, multimodal, and noisy; an artificial design of such an environment would be computationally intensive and difficult to simulate. However, all these interesting features of the environment come for "free" when we place the BBD in the real world. The modeller is freed from simulating a world and need only concentrate on the development of a device that can actively explore the real world.

2.4 Engage in a Behavioural Task

It follows from the above principle that a situated agent needs to engage in some behavioural task. Similar to a biological organism, an agent or BBD needs a minimal set of innate behaviours or reflexes in order to explore and initially survive in its environmental niche. From this minimal set, the BBD can learn and adapt such that it optimizes its behaviour. How these devices adapt is the subject of the next principle, which describes value systems (see section 2.5). This approach is very different from the classic artificial intelligence or robotic control algorithms, where either rules or feedback controllers with pre-defined error signals need to be specified *a priori*. In the BBD approach, the agent selects what it needs to optimize its behaviour and thus adapts to its environment.

A second and important point with regard to behavioural tasks is that it gives the researcher a metric by which to score the BBD's performance. Moreover, these tasks should be made similar to experimental biology paradigms so that the behavioural performance of the BBD can be compared with that of real organisms (see section 2.6 below).

2.5 Adapt Behaviour when an Important Environmental Event Occurs

Biological organisms adapt their behaviour through value systems, which provide non-specific, modulatory signals to the rest of the brain that bias the outcome of local changes in synaptic efficacy in the direction needed to satisfy global needs. Stated in

the simplest possible terms, behaviour that evokes positive responses in value systems biases synaptic change to make production of the same behaviour more likely when the situation in the environment (and thus the local synaptic inputs) is similar; behaviour that evokes negative value biases synaptic change in the opposite direction. Examples of value systems in the brain include the dopaminergic, cholinergic, and noradrenergic systems (Aston-Jones and Bloom, 1981; Hasselmo et al., 2002; Schultz et al., 1997) which respond to environmental cues signalling reward prediction, uncertainty, and novelty. Theoretical models based of these systems and their effect on brain function have been developed (Doya, 2002; Friston et al., 1994; Montague et al., 1996; Yu and Dayan, 2005) and embedded in real world behaving devices (Arleo et al., 2004; Krichmar and Edelman, 2002; Sporns and Alexander, 2002).

2.6 Comparisons with Experimental Data Acquired from Animal Models

The behaviour of BBDs and the activity of their simulated nervous systems must be recorded to allow comparisons with experimental data acquired from animals. The comparison should be made at the behavioural level, the systems level, and the neuronal element level. These comparisons serve two purposes: First, BBDs are powerful tools to test theories of brain function. The construction of a complete behaving model forces the designer to specify theoretical and implementation details that are easy to overlook in a purely verbal description and it forces those details to be consistent among them. The level of analysis permitted by having a recording of the activity of every neuron and synapse in the simulated nervous system during its behaviour is just not possible with animal experiments. The results of such situated models have been compared with rodent hippocampal activity during navigation, basal ganglia activity during action selection, and attentional systems in primates (Burgess et al., 1997; Guazzelli et al., 2001; Itti, 2004; Prescott et al., 2006). Second, by using the animal nervous system as a metric, designers can continually make their simulated nervous systems closer to that of the model animal. This, in turn, allows the eventual creation of practical devices that may approach the sophistication of living organisms.

3 Illustrative Examples of Brain-Based Devices

In this section, we will use our group's two most recent BBDs as illustrative examples of the above principles. The first example, Darwin X (Krichmar et al., 2005a; Krichmar et al., 2005b), is a BBD that develops spatial and episodic memory by incorporating a detailed model of the hippocampus and its surrounding regions. The second example is a BBD capable of predictive motor control based on a model of cerebellar learning (McKinstry et al., 2006).

3.1 An Embodied Model of Spatial and Episodic Memory

Darwin X was used to investigate the functional anatomy specific to the hippocampal region during a memory task. Darwin X incorporates aspects of the anatomy and physiology of the hippocampus and its surrounding regions, which are known to be necessary for the acquisition and recall of spatial and episodic memories. The simulated nervous system contained 50 neural areas, 90,000 neuronal units, and 1.4 million synaptic connections. It included a visual system, a head direction system, a hippocampal formation, a basal forebrain, a value or reward system, and an action selection system. Darwin X used camera input to recognize the category and position of distal visual objects and used odometry to develop head direction sensitivity.

Darwin X successfully demonstrated the acquisition and recall of spatial and episodic memories in a maze task similar to the Morris water maze (Morris, 1984) by associating places with actions. The association was facilitated by a dopaminergic value system based on the known connectivity between CA1 and nucleus accumbens and frontal areas (Thierry et al., 2000). The responses of simulated neuronal units in the hippocampal areas during its exploratory behaviour were comparable to neuronal responses in the rodent hippocampus; i.e., neuronal units responded to a particular location within Darwin X's environment (O'Keefe and Dostrovsky, 1971).

Darwin X took into consideration the macro- and micro-anatomy between the hippocampus and cortex, as well as the within the hippocampus. In order to identify different functional hippocampal pathways and their influence on behaviour, we developed two novel methods for analyzing large scale neuronal networks: 1) Backtrace - tracing functional pathways by choosing a unit at a specific time and recursively examining all neuronal units that led to the observed activity in this reference unit (Krichmar et al., 2005a), and 2) Causality - a time series analysis that distinguishes causal interactions within and between neural regions (Seth, 2005). These analyses allowed us to examine the information flow through the network and highlighted the importance of the perforant pathway from the en-

torhinal cortex to the hippocampal subfields in producing associations between the position of the agent in space and the appropriate action it needs to reach a goal. This functional pathway has recently been identified in the rodent (Brun et al., 2002).

As with other BBDs in the Darwin series, Darwin X follows the brain-based modelling principles. It is a physical device in a real world that carries out a task similar to that conducted with animal models. It adapts its behaviour based on its value system, and the dynamics of its nervous system were analyzed during its behaviour and compared with the responses of real nervous systems.

3.2   A Model of Predictive Motor Control Based On Cerebellar Learning and Visual Motion

Recently, our group constructed a BBD which included a detailed model of the cerebellum and cortical areas that respond to visual motion (McKinstry et al., 2006). One theory of cerebellar function proposes that the cerebellum learns to replace reflexes with a predictive controller (Wolpert et al., 1998). Synaptic eligibility traces in the cerebellum have recently been proposed as a specific mechanism for such motor learning (Medina et al., 2005). We tested whether a learning mechanism, called the delayed eligibility trace learning rule, could account for the predictive nature of the cerebellum in a real-world, robotic visuomotor task.

The BBD's visuomotor task was to navigate a path designated by orange traffic cones. The platform for this task was a Segway Robotic Mobility Platform modified to have a camera, a laser range finder, and infrared proximity detectors as inputs. The BBD's nervous system contained components simulating the cerebellar cortex, the deep cerebellar nuclei, the inferior olive, and a cortical area MT. The simulated cortical area MT, which responds to visual motion, was constructed based on the suggestion that the visual system makes use of visual blur for determining motion direction (Geisler, 1999; Krekelberg et al., 2003). The simulated nervous system contained 28 neural areas, 27,688 neuronal units, and 1.6 million synaptic connections. Using an embedded Beowulf computer cluster of six compact personal computers, it took roughly 40 ms to update all the neuronal units and plastic connections in the model each simulation cycle. Initially, path traversal relied on a reflexive movement away from obstacles that was triggered by infrared proximity sensors when the BBD was within 12 inches of a cone. This resulted in clumsy, crooked movement down the path. The infrared sensor input was also the motor error signal to the cerebellum via simulated climbing fibre input. Over time, the cerebellar circuit predicted the correct motor response based on visual motion cues preventing the activation of the reflex and resulting in smooth movement down the centre of the path. The system learned to slow down prior to a curve and to turn in the correct direction based on the flow of visual information. The system adapted to and generalized over different courses with both gentle and sharp angle bends.

The experiments, which depend both on the dynamics of the delayed trace eligibility learning and on the architecture of the cerebellum, demonstrated how the cerebellum can predict impending errors and adapt its movements. Moreover, by analyzing the responses of the cerebellum and the inputs from the simulated area MT during its behaviour, we were able to predict the types of signals the nervous system might select to adapt to such a motor task. The BBD's nervous system categorized the motion cues that were predictive of different collisions and associated those categories with the appropriate movements. The neurobiologically inspired model described here prompts several hypotheses about the relationship between perception and motor control and may be useful in the development of general-purpose motor learning systems for machines.

4   Conclusions

Higher brain functions depend on the cooperative activity of an entire nervous system, reflecting its morphology, its dynamics, and its interaction with its phenotype and the environment. BBDs are designed to incorporate these attributes such that they can test theories of brain function. Like the brain, they operate according to selectional principles through which they form categorical memory, associate categories with innate value, and adapt to the environment. These BBDs also provide the groundwork for the development of intelligent machines that follow neurobiological rather than computational principles in their construction.

Acknowledgements

References

Arleo, A., Smeraldi, F., and Gerstner, W. (2004). Cognitive navigation based on nonuniform Gabor space sampling, unsupervised growing networks, and reinforcement learning. IEEE Trans Neural Netw *15*, 639-652.

Aston-Jones, G., and Bloom, F. E. (1981). Nonrepi-nephrine-containing locus coeruleus neurons in behaving rats exhibit pronounced responses to non-noxious environmental stimuli. J Neurosci *1*, 887-900.

Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. J Neurosci *2*, 32-48.

Borg-Graham, L. (1987). Modelling the electrical behavior of cortical neurons - simulations of hippocampal pyramidal cells., In Computer Simulation in Brain Science, R. M. J. Cotterill, ed. (Cambridge: Cambridge University Press).

Bower, J. M., and Beeman, D. (1994). The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.: TELOS/Springer-Verlag).

Brun, V. H., Otnass, M. K., Molden, S., Steffenach, H. A., Witter, M. P., Moser, M. B., and Moser, E. I. (2002). Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry. Science *296*, 2243-2246.

Burgess, N., Donnett, J. G., Jeffery, K. J., and O'Keefe, J. (1997). Robotic and neuronal simulation of the hippocampus and rat navigation. Philos Trans R Soc Lond B Biol Sci *352*, 1535-1543.

Chiel, H. J., and Beer, R. D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. Trends Neurosci *20*, 553-557.

Clark, A. (1997). Being there. Putting brain, body, and world together again. (Cambridge, MA: MIT Press).

Doya, K. (2002). Metalearning and neuromodulation. Neural Netw *15*, 495-506.

Edelman, G. M., Reeke, G. N., Gall, W. E., Tononi, G., Williams, D., and Sporns, O. (1992). Synthetic neural modeling applied to a real-world artifact. Proc Natl Acad Sci U S A *89*, 7267-7271.

Edelman, G. M., and Reeke, G. N., Jr. (1982). Selective networks capable of representative transformations, limited generalizations, and associative memory. Proc Natl Acad Sci U S A *79*, 2091-2095.

Friston, K. J., Tononi, G., Reeke, G. N., Sporns, O., and Edelman, G. M. (1994). Value-dependent

selection in the brain: simulation in a synthetic neural model. Neuroscience *59*, 229-243.

Geisler, W. S. (1999). Motion streaks provide a spatial code for motion direction. Nature *400*, 65-69.

Guazzelli, A., Bota, M., and Arbib, M. A. (2001). Competitive Hebbian learning and the hippocampal place cell system: modeling the interaction of visual and path integration cues. Hippocampus *11*, 216-239.

Hasselmo, M. E., Hay, J., Ilyn, M., and Gorchetchnikov, A. (2002). Neuromodulation, theta rhythm and rat spatial navigation. Neural Netw *15*, 689-707.

Hines, M. L., and Carnevale, N. T. (1997). The NEURON simulation environment. Neural Comput *9*, 1179-1209.

Itti, L. (2004). Automatic foveation for video compression using a neurobiological model of visual attention. IEEE Trans Image Process *13*, 1304-1318.

Izhikevich, E. M., Gally, J. A., and Edelman, G. M. (2004). Spike-timing dynamics of neuronal groups. Cereb Cortex *14*, 933-944.

Krekelberg, B., Dannenberg, S., Hoffmann, K. P., Bremmer, F., and Ross, J. (2003). Neural correlates of implied motion. Nature *424*, 674-677.

Krichmar, J. L., and Edelman, G. M. (2002). Machine psychology: autonomous behavior, perceptual categorization and conditioning in a brain-based device. Cereb Cortex *12*, 818-830.

Krichmar, J. L., and Edelman, G. M. (2005). Brain-based devices for the study of nervous systems and the development of intelligent machines. Artif Life *11*, 63-77.

Krichmar, J. L., Nitz, D. A., Gally, J. A., and Edelman, G. M. (2005a). Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task. Proc Natl Acad Sci U S A *102*, 2111-2116.

Krichmar, J. L., and Reeke, G. N. (2005). The Darwin Brain-Based Automata: Synthetic Neural Models and Real-World Devices, In Modeling in the Neurosciences: From Biological Systems to Neuromimetic Robotics, G. N. Reeke, R. R. Poznanski, K. A. Lindsay, J. R. Rosenberg, and O. Sporns, eds. (Boca Raton: Taylor & Francis), pp. 613-638.

Krichmar, J. L., Seth, A. K., Nitz, D. A., Fleischer, J. G., and Edelman, G. M. (2005b). Spatial

navigation and causal analysis in a brain-based device modeling cortical-hippocampal interactions. Neuroinformatics *3*, 197-221.

McKinstry, J. L., Edelman, G. M., and Krichmar, J. L. (2006). A cerebellar model for predictive motor control tested in a brain-based device. Proc Natl Acad Sci U S A.

Medina, J. F., Carey, M. R., and Lisberger, S. G. (2005). The representation of time for motor learning. Neuron *45*, 157-167.

Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci *16*, 1936-1947.

Morris, R. (1984). Developments of a water-maze procedure for studying spatial learning in the rat. J Neurosci Methods *11*, 47-60.

O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res *34*, 171-175.

Pinsky, P. F., and Rinzel, J. (1994). Intrinsic and network rhythmogenesis in a reduced Traub model for CA3 neurons. J Comput Neurosci *1*, 39-60.

Prescott, T. J., Montes Gonzalez, F. M., Gurney, K., Humphries, M. D., and Redgrave, P. (2006). A robot model of the basal ganglia: Behavior and intrinsic processing. Neural Netw *19*, 31-61.

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. Science *275*, 1593-1599.

Seth, A. K. (2005). Causal connectivity of evolved neural networks during behavior. Network *16*, 35-54.

Seth, A. K., McKinstry, J. L., Edelman, G. M., and Krichmar, J. L. (2004). Active sensing of visual and tactile stimuli by brain-based devices. International Journal of Robotics and Automation *19*, 222-238.

Song, S., Miller, K. D., and Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nat Neurosci *3*, 919-926.

Sporns, O., and Alexander, W. H. (2002). Neuromodulation and plasticity in an autonomous robot. Neural Netw *15*, 761-774.

Thierry, A. M., Gioanni, Y., Degenetais, E., and Glowinski, J. (2000). Hippocampo-prefrontal cortex pathway: anatomical and electrophysiological characteristics. Hippocampus *10*, 411-419.

Wolpert, D., Miall, R., and Kawato, M. (1998). Internal models in the cerebellum. Trends in Cognitive Sciences *2*, 338-347.

Worgotter, F., and Porr, B. (2005). Temporal sequence learning, prediction, and control: a review of different models and their relation to biological mechanisms. Neural Comput *17*, 245-319.

Yu, A. J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. Neuron *46*, 681-692.