Brain-Based Devices for the Study of Nervous Systems and the Development of Intelligent Machines

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

Abstract The simultaneous study of brain function at all levels of organization is difficult to undertake with current experimental tools. Present day electrophysiology only allows the recording of at most hundreds of neurons while an animal is performing a behavioral task. Because of this limitation and the sheer complexity of the nervous system, computational modeling has become essential in developing theories of brain function. Accordingly, our group has constructed a series of brain-based devices (BBDs), that is, physical devices with simulated nervous systems that guide behavior, to serve as a heuristic for testing theories of brain function. Unlike animal models, BBDs permit analysis of activity at all levels of the nervous system as the device behaves in its environment. Although the principal focus of developing BBDs has been to test theories of brain function, this type of modeling may also provide a basis for robotic design and practical applications.

Keywords
Adaptive behavior, learning, neuroscience, perception, robots

1 Introduction

Elucidation of brain mechanisms underlying behavior requires simultaneous measurements across multiple levels. The heuristic value of synthetic modeling using brain-based devices (BBDs), which will be described here, is supported by the fact that these types of measurements are difficult to obtain and compare in living animals. Given the construction of BBDs, we are able to observe their overall behavior while simultaneously recording the state of their simulated nervous systems at all levels. Since our purpose is to test theories of real nervous systems in order to arrive at a better understanding of brain function, we base the BBD's design on principles of neural anatomy and physiology.

We argue that a BBD should be constrained by the following design principles:

1. The device needs to engage in a behavioral task.
2. The device's behavior must be controlled by a simulated nervous system having a design that reflects the brain's architecture and dynamics.
3. The device needs to be situated in the real world [6, 7].
4. The behavior of the device and the activity of its simulated nervous system must allow comparisons with empirical data.

Because of these constraints, BBD simulations tend to require large-scale networks of neuronal elements that reflect the brain's anatomy and physiology, high-performance computing to run the network in real time, and the engineering of specialized physical devices to embody the network.

BBDs are not programmed by instructions like computers, but instead, like biological systems, they operate according to selectional principles that allow them to adapt to the environment [11]. This design, which possesses neuroanatomical structure and large-scale neural dynamics, differs fundamentally from that of robots. Robotic approaches using classical artificial intelligence are based on data representation, rule-driven algorithms, and the manipulation of formal symbol systems [24, 25]. Artificial intelligence has been somewhat successful in emulating logical aspects of human behavior, but has been less successful in dealing with perception, categorization, and movement in the world, which is a strength of synthetic neural models and BBDs [28, 29]. Purely reactive or behavior-based robots carry out actions that are controlled through arbitration of several primitive behavioral repertoires without neural architectures [2, 4]. Evolutionary robotics, in which control systems are selected after each trial or lifetime according to a fitness function [26], can evolve complex behaviors with very simple systems, but also do not emphasize neuronal responses. A recent hybrid between evolutionary algorithms and artificial neural network learning rules was designed to mutate learning rules between trials, allowing learning during the lifetime of the robot [14]. Typically, however, the artificial neural networks controlling the evolutionary robot's behavior are small (on the order of tens of artificial "neural units") and do not reflect neuroanatomical organization.

2 Examples of Brain-Based Devices

All BBDs that we have constructed contain the following attributes:

1. A morphology, or body plan, that allows for active exploration in a real environment
2. A brain simulation embedding detailed neuroanatomy, based on vertebrate nervous systems to control the BBD's behavior and shape its memory
3. A value system that signals the salience of environmental cues to the BBD's nervous system, causing change in the nervous system that results in modification of the device's behavior.

These features result in a system that generalizes signals from the environment into perceptual categories and adapts its behavior so that it becomes increasingly successful in coping with its environment. These BBDs are designated the Darwin series of automata. Various Darwin automata over the last 12 years have been shown to develop perceptual categorization, invariant visual object recognition, integration of scenes containing multiple visual shapes with overlapping features, fusion of different sensory modalities, and learning in the form of operant conditioning [1, 13, 19–22, 33].

In this section we describe two recent BBDs in the series, Darwin VII and Darwin VIII, with the aim of conveying the principles underlying these devices and the power of the BBD approach in testing theories of the brain.

2.1 Darwin VII—Perceptual Categorization and Operant Conditioning in a BBD

The behavior of Darwin VII showed that a synthetic brain-based device operating on biological principles and without prespecified instructions can carry out perceptual categorization and conditioned responses (for more details regarding Darwin VII experiments and methodology, see [19, 20]). Darwin VII's task was similar to a foraging task in which it learned to approach and sample *positive-value* blocks and avoided *negative-value* blocks. The arbitrarily assigned value was initially derived from the *taste*, or conductivity, of the metal blocks. After conditioning, Darwin VII assessed a block's value based on the block's visual and auditory cues. The successful performance of the device

rested on the selectional modulation of its neuronal activity by behavior interacting with constraints provided by its value system. The development of categorical responses required exploration of the environment and sensory-motor adaptation through specific and highly individual changes in connection strengths.

We observed Darwin VII's overall behavior while at the same time recording the state of every neuronal unit and synaptic connection in its simulated nervous system. By collecting these neuronal data, we were able to demonstrate the development of neuronal groups during categorization and recognition, to show that reliable classification of responses to visual stimuli could be based on the sampling of a small subpopulation of neuronal units, and to relate learning responses to functional changes in synaptic efficacy.

Darwin VII consists of a mobile base equipped with a CCD camera for vision, microphones for hearing, conductivity sensors for taste, and effectors for movement of its base, of its head, and of a gripping manipulator having one degree of freedom (Figure 1).

2.1.1 Darwin VII's Neural Architecture and Dynamics

Darwin VII's behavior is guided by a nervous system modeled after the vertebrate system, but obviously with far fewer neurons and simpler architecture. Six major systems make up the simulated brain: an auditory system, a visual system, a taste system, sets of motor neurons capable of triggering behavior, a visual tracking system, and a value system. The complete nervous system contained

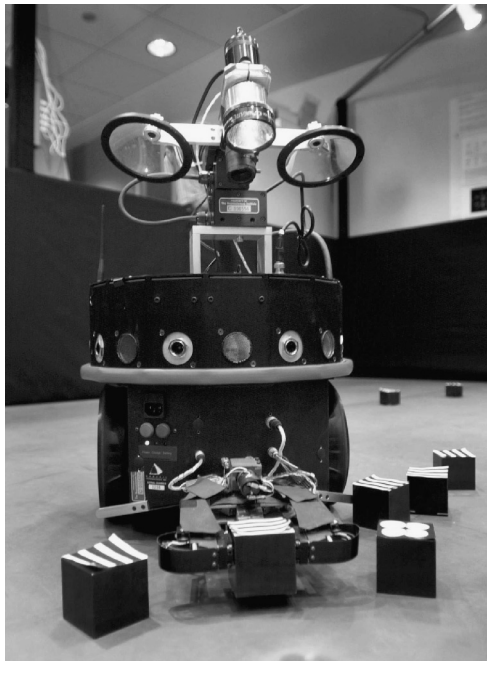

Figure 1. Darwin VII consists of a mobile base (developed by Nomadic Technologies Inc., Mountain View, CA) equipped with several sensors and effectors and a neural simulation running on a remote computer workstation. It contains a radio modem to transmit status and auditory information to the computer workstation carrying out the neural simulation and to receive motor commands from the simulation. Video output from a CCD camera mounted on Darwin VII is sent to the workstation via RF transmission. RF input and output to Darwin VII allows for untethered exploration. The CCD camera, two microphones on either side of the camera, and sensors embedded in the gripper that measure the surface conductivity of stimuli provide sensory input to the neuronal simulation. Eight infrared (IR) sensors are mounted at 45° intervals around the mobile platform. The IR sensors are responsive to the boundaries of the environment and were used to trigger reflexes for obstacle avoidance. All behavioral activity other than obstacle avoidance is triggered by signals received from the neural simulation. (From [19].)

18 neuronal areas, 19,556 neuronal units, and approximately 450,000 synaptic connections between neuronal units. Figure 2 shows a high-level diagram of the simulated nervous system and its neuroanatomical structure.

A neuronal unit in Darwin VII is simulated by a mean-firing-rate model; the activity of each unit corresponds roughly to the average activity of a group of neurons (approximately 100 neurons) over a time period of approximately 100 ms. The mean-firing-rate model was chosen because it captures many of the attributes of real neurons at a level of resolution consistent with electrophysiological measurements, while being sufficiently simple to allow computation in real time. In the model, updates were based on a simulation cycle: the period of time during which the current sensory input is processed, the activities of all neuronal units are computed, the connection strengths of all plastic connections are computed, and motor output is generated (see [19] for details). Each simulation cycle in Darwin VII took approximately 200 ms of real time, which is sufficiently fast in that the device does not wait for commands from the nervous system and can respond to stimuli without delays.

The strength of connections between two units can change based on the activity of the sending unit, called the presynaptic neuronal unit, and the receiving unit, called the postsynaptic unit. If the sending and receiving units are simultaneously strongly active, the connection between them is strengthened. If the sending and receiving units are simultaneously weakly active, the connection is weakened. Some connections are also affected by the value system. The connection strength between these units was amplified when the value system was active.

Activation of the simulated value system (area *Value*, Figure 2) signaled the occurrence of salient sensory events and contributed to the modulation, at that time, of connection strengths of all active synapses in the value-dependent pathways. For example, tasting a block picked up by Darwin VII's gripper is a salient event affecting subsequent behavior that is reinforced or weakened through synaptic change. The *Value* area is thus analogous to an ascending neuromodulatory value or reward system [31, 35].

In experiments where individual variation was examined, each Darwin VII subject shared the same physical device, but had an instantiation in which the simulated nervous system was unique due to different random initializations in both the connectivity between individual neuronal units and the initial connection strengths between those units. Because the connectivity between neuronal units was constrained by a common set of projections, large-scale connectivity (i.e., projections between neural areas) was similar between subjects.

2.1.2 Darwin VII's Environment

Darwin VII's environment consisted of an enclosed area with black walls and a floor covered with opaque black plastic panels, on which we distributed stimulus blocks (6 cm metallic cubes) in various arrangements (Figure 1). The top surfaces of the blocks were covered with black-and-white patterns; the other surfaces of the blocks were featureless and black. All experiments described here were carried out with multiple exemplars of two basic patterns: *blobs* (several white patches 2–3 cm in diameter) and *stripes* (width 0.6 cm, evenly spaced). Stripes on blocks in the gripper could be viewed in either horizontal or vertical orientations, yielding a total of three stimulus classes of visual patterns to be discriminated (*blob*, *horizontal*, and *vertical*). A flashlight mounted on Darwin VII and aligned with its gripper caused the blocks, which contained a photodetector, to emit a beeping tone when Darwin VII was in the vicinity. The sides of the stimulus blocks were metallic and could be rendered either strongly conductive (appetitive or "good taste") or weakly conductive (aversive or "bad taste"). Gripping of stimulus blocks activated the appropriate taste neuronal units (either area $Taste_{app}$ or $Taste_{ave}$) to a level sufficient to drive the motor areas above a behavioral threshold. In the experiments, strongly conductive blocks with a striped pattern and a 3.9 kHz tone were chosen arbitrarily to be positive value exemplars, whereas weakly conductive blocks with a blob pattern and a 3.3 kHz tone represented negative value exemplars.

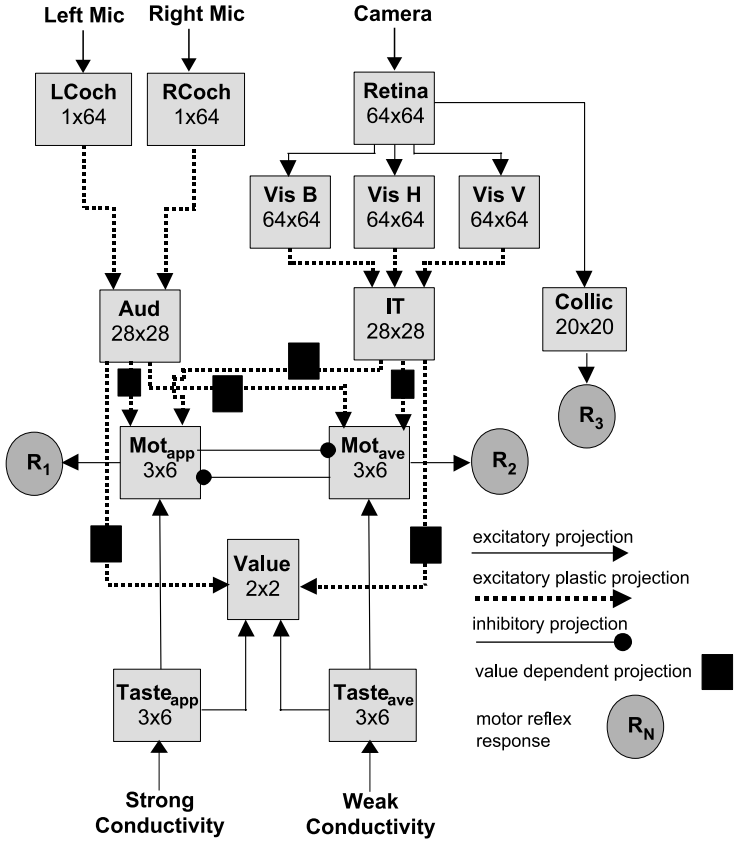

Figure 2. Schematic of the regional and functional neuroanatomy of Darwin VII. There are six major systems that make up the simulated nervous system: an auditory system, a visual system, a taste system, sets of motor neurons capable of triggering behavior, a visual tracking system, and a value system. In the version used in the present experiments, the simulated nervous system contained 18 neuronal areas, 19,556 neuronal units, and approximately 450,000 synaptic connections. A neuronal unit corresponded to the mean activity of a small group of real neurons over approximately 100 ms. The boxes in the diagram represent neural areas, and the numbers in the boxes denote the number of neuronal units in each area. Arrows denote synaptic projections between neural areas. The 64×64 gray level pixel image captured by the CCD camera was relayed to a retinal area *Retina* and transmitted via topographic connections to a primary visual area *Vis*. There were three subpartitions in *Vis*, selective for bloblike features, for short horizontal line segments, and for short vertical line segments (*Vis B*, *Vis H*, and *Vis V*). Responses within *Vis* closely followed stimulus onset and projected non-topographically via activity-dependent plastic connections to a secondary visual area, analogous to the inferotemporal cortex (*IT*). The frequency and amplitude information captured by Darwin VII's left and right microphones was relayed to simulated cochlear areas (*LCoch* and *RCoch*, respectively) and transmitted via mapped tonotopic and activity-dependent plastic connections to a primary auditory area *Aud*. No attempt was made in this study to use the difference between the two cochlear regions for sound localization. The activity of each cochlear neuronal unit was broadly tuned to a preferred frequency and scaled according to the signal amplitude. *Aud* and *IT* contained local excitatory and inhibitory interactions producing firing patterns that were characterized by focal regions of excitation surrounded by inhibition. *Aud* and *IT* sent plastic projections to the value system (*Value*) and to the motor areas Mot_{app} and Mot_{ave}. These two neuronal areas were capable of triggering two distinct behaviors, appetitive and aversive. Appetitive or aversive responses were triggered if the difference in instantaneous activity between the motor areas Mot_{app} and Mot_{ave} exceeded a behavioral threshold. Picking up and sampling the conductivity across the surface of stimulus objects, as measured by sensors in Darwin VII's gripper, is innate in Darwin VII's behavior. In all the experiments, strongly conductive blocks activated $Taste_{app}$ and weakly conductive blocks activated $Taste_{ave}$. The taste system sent information to the motor areas (Mot_{app} and Mot_{ave}) and the value system (*Value*). The *Value* area projected diffusely with long-lasting value-dependent activity to the auditory, visual, and motor behavior neuronal units. The visual tracking system controlled navigational movements, in particular the approach to objects identified by brightness contrast with respect to the background. To achieve tracking behavior, the retinal area *Retina* projected to the area *Collic* ("colliculus"). The activity in *Collic* was converted into wheel motor commands such that a bright object on the left (right) caused a turn towards the left (right). (Adapted from [19].)

The three visual categories and two auditory categories described above are not labeled or known a priori by Darwin VII. Therefore, Darwin VII must learn these perceptual categories based on experience-dependent plasticity in the connections between its primary sensory areas (*LCoch*/*RCoch* for auditory and *Vis* for visual in Figure 2) and its association areas (*Aud* and *IT* in Figure 2). Categorical memory in Darwin VII is elicited by patterns of activity in its association areas. The patterns that emerge in response to stimuli used in the experiments described here are a small subset of the possible activity patterns.

2.1.3 Darwin VII's Behavior

Basic modes of behavior built into Darwin VII included IR-sensor-dependent obstacle avoidance, visual exploration, visual approach and tracking, gripping and tasting, and two main classes of innate behavioral reflex responses (appetitive and aversive). With the exception of obstacle avoidance, selection among the above behaviors was under control of the simulated nervous system. Obstacle avoidance was initiated by the infrared proximity sensors around the device's base. Activation of an IR proximity detector by a wall, for example, triggered a reflexive movement away from the obstacle. The tracking system allows Darwin VII to orient towards visual stimuli. The activity of the area *Collic* (analogous to the superior colliculus) dictates where Darwin VII directs its camera gaze. Tracking in Darwin VII is achieved by signals to Darwin VII's wheels based on the vector summation of the activity of the neuronal units in *Collic*. Each neuronal unit in *Collic* has a receptive field that matches its preferred direction, and the area has a topographic arrangement such that if activity is predominately on the left side of *Collic*, signals to Darwin VII's wheels are issued that evoke a turn towards the left. Appetitive and aversive responses were triggered when the difference in activity between the motor areas Mot_{app} and Mot_{ave} exceeded a behavioral threshold ($\beta = 0.3$). These responses could be activated by taste (the unconditioned stimulus, US, triggering an unconditioned response, UR) or by auditory or visual stimuli (the conditioned stimulus, CS, triggering a conditioned response, CR).

2.1.4 Darwin VII's Conditioning

In a series of conditioning experiments, Darwin VII was trained to associate the taste value of objects with their visual and auditory characteristics. Seven individually different Darwin VII subjects participated in the experiments, in which each subject encountered at least 10 appetitive and 10 aversive blocks. Appetitive and aversive responses were triggered initially by taste, but after training or conditioning, these responses could be triggered by auditory or visual stimuli. After conditioning, Darwin VII continued to grip and taste appetitive blocks, but learned to back away without picking up aversive blocks, avoiding these blocks over 90% of the time (see Figure 3). Thus, during the conditioning experiments, in which many stimuli were encountered over an extended period of time, Darwin VII developed perceptual categories that modified its behavioral responses in an adaptive fashion.

Furthermore, we were able to carry out second-order conditioning experiments with Darwin VII. After associating the initially neutral visual pattern with an innate value-loaded taste, the visual pattern was paired with a tone that emitted from the block. Darwin VII successfully learned this association, and it avoided blocks with a tone predictive of bad taste and approached blocks with a tone predictive of good taste over 90% of the time (see [19] for details).

While performance improved with training, it never reached perfection, and occasional mistakes were made. Some of these mistakes were due to sensor noise, but some were due to experiential differences that naturally occur when sampling a noisy, unpredictable real-world environment (e.g., no two views of a striped block will be exactly alike). This unpredictability is a general property of selectionist systems, that is, systems consisting of a population of variant repertoires that are differentially amplified, thus yielding responses to unpredicted or novel events [10]. The unpredictability of behavioral responses in Darwin VII coupled with the variability of a complex environment

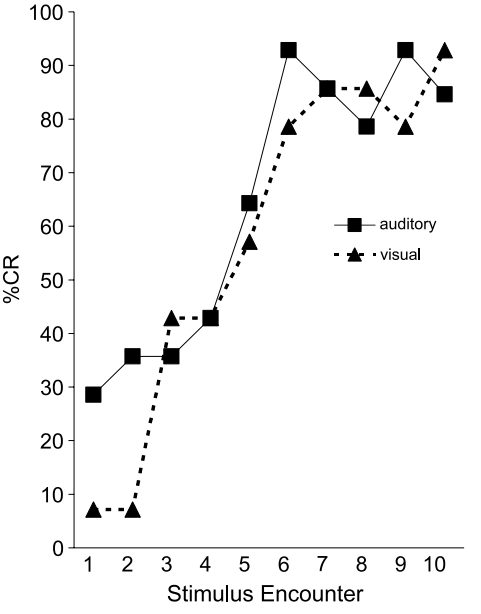

Figure 3. Percentage of conditioned responses (CR) during learning trials for both visual and auditory conditioning as a function of stimulus block encounters (seven subjects). The responses of aversive and appetitive stimulus encounters were pooled together.

allowed the device to learn after making mistakes, to generalize over sensory inputs, and to deal with novel situations.

2.1.5 Darwin VII's Perceptual Categorization

The development of perceptual categories was essential for success in conditioning tasks. Darwin VII's nervous system has three features that are critical for perceptual categorization:

1. Connectivity from a topographically mapped primary area with transient activity to a non-topographically-mapped higher area with more persistent activity (e.g., see *Vis*→*IT* in Figure 2)
2. Sensory input that is continuous and temporally correlated with self-generated movement
3. Activity-dependent learning in which competitive mechanisms categorize sensory information and select for appropriate behavioral repertoires.

These features allow Darwin VII to achieve invariant object recognition, that is, to recognize an object despite changes in orientation, size, and position in its visual field. Invariant object recognition is a system property that emerges dynamically from competitive neuronal group interactions within and between neural areas.

Darwin VII never displayed identical patterns of neural activity, even during repetitions of the same behavior (see Figure 4). The adaptive behaviors tend to remain similar, however, despite variation in system properties resulting from multiple interactions across circuitry, plastic synaptic connections, fluctuating value systems, and variable object encounters. In this respect, Darwin VII is an example of a degenerate system: different circuits and dynamics can yield similar behavior [12, 39]. Experiments comparing the development of responses to strongly biased samples of appetitive or aversive stimuli demonstrate, however, that even with identical starting architectures, changes in experiential sequences can have profound effects on subsequent behavior. While this has been

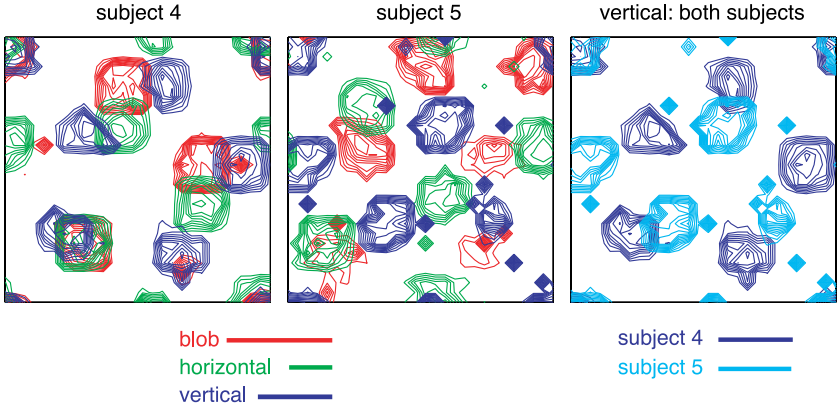

Figure 4. Comparison of patterns of activity in the area *IT* for the three visual stimuli across different Darwin VII subjects. The first two contour plots on the left show, for two representative subjects, the borders of neuronal group activity in response to blobs (red lines), horizontal stripes (green lines), and vertical stripes (blue lines). The contours on the far right show the activity for two different subjects in the same plot, in response to vertical stimuli for subject 4 (dark blue) and for subject 5 (light blue). Across all subjects tested (n = 7), the mean overlap of activity within each subject but between stimuli was 26.6% (σ = 0.10). The mean overlap of activity between different subjects was 22.1% (σ = 0.09) in response to blob patterns, 26.9% (σ = 0.11) in response to horizontal striped patterns, and 20.0% (σ = 0.10) in response to vertical striped patterns. (Adapted from [19].)

documented phenomenologically with living organisms, the experiments described here may suggest possible mechanisms underlying such epigenetic biases.

Differences in an individual's perceptual history can have a profound effect on the organization and response of the nervous system. Using Darwin VII, we performed experiments concerned with experience-dependent effects on categorization during the development of perceptual categories as well as after such development [21]. In these experiments, Darwin VII started with an identical simulated nervous system operating on the same physical platform, ensuring that any differences would only be due to experiential history. One set of experiments investigated the effect of variations in stimulus order and frequency on early development. The experiments began with a visually naive Darwin VII exploring an environment that was partially divided into two subareas with disparate distributions of stimuli. The number of neuronal units in *IT* selectively responding to a given stimulus (whether blob, horizontal stripe, or vertical stripe) increased with an increase in the frequency of presentation of that stimulus class [19, 21]. These findings are similar to the results of neuronal recordings in the monkey inferotemporal cortex in that more neurons responded to familiar than to unfamiliar objects in recognition tasks [18].

Another set of experiments investigated the effect of stimulus frequency on a BBD that had previous visual experience. The experience consisted of exemplars of the three stimulus classes presented in equal proportion until categorization was achieved with high accuracy. This "adult" Darwin VII was then presented with exemplars of only two out of the three stimulus classes. In contrast to the previous experiments on early development, after extensive experience, the number of neuronal units in *IT* responding to more frequently sampled stimuli did not change significantly, suggesting that responses in *IT* had become saturated with respect to the familiar stimuli. However, in the experiments in which Darwin VII responded to the less frequently sampled stimulus, the number of *IT* neuronal units was significantly less than that in the controls (Table 1). In essence, Darwin VII had forgotten these perceptual categories.

2.2 Darwin VIII—Visual Binding through Reentrant Connectivity and Synchronization

A more recent BBD, Darwin VIII, demonstrated the ability to bind the attributes of stimuli in a scene to form coherent perceptual categories [22, 33]. Thus, Darwin VIII solves the so-called

Table 1. Role of history in perceptual categorization. After visual categories had been developed (eight presentations of each stimulus class), the activity in *IT* stayed relatively constant for stimuli presented in higher proportions, but decreased for stimuli presented in lower proportions. Each row in the table shows the median neural activity in *IT* in response to stimuli for a control group (eight additional presentations of blob, horizontal, and vertical stimuli) and three experimental groups in which eight additional stimuli from two out of the three classes were presented. There were 10 trials for each group, with identical initial conditions in the simulated nervous system. The asterisks denote a significant difference ($p < 0.05$) in medians between the control group and an experimental group using the Wilcoxon rank sum test of medians. (Adapted from [19].)

	Effect of stimulus presentation frequency on *IT* activity after development of visual categorization			
Test stimuli	Control	Stripes only	Horizontal and blob	Vertical and blob
Blob	88.8	72.9*	89.9	91.6
Horizontal	54	55.6	57.7	31.5*
Vertical	24.2	24.4	17.9*	23.8

* $p < 0.05$; Wilcoxon rank sum test.

binding problem, that is, it shows how different brain areas and modalities can yield a coherent perceptual response in the absence of any superordinate control. We had previously presented a computational model to account for visual binding. This model emphasized the interaction between neural areas, and showed that reentrant connections between different visual areas were sufficient for recognizing and selecting multiple objects in a scene [38]. The model included nine simulated cortical areas of the visual system, as well as a value system. It was trained to prefer a particular visual object via simulated saccades, and its performance showed that binding and discriminative behavior could be achieved through the interaction of different neural areas via reentrant connections. Despite its success in showing the capabilities of reentrant circuits, the model had several limitations. The stimuli used were taken from a limited set and were of uniform scale. Furthermore, its behavior did not emerge in a rich and noisy environment of the kind confronted by behaving organisms.

To address these limitations, we constructed a BBD, Darwin VIII, to test the necessity and sufficiency of synchronous activity brought about by reentrant connections between neural areas to bind together and discriminate objects in a visual scene [22, 33]. Darwin VIII, which had a neural model similar to the reentrant cortical model described above (see [38] for details), autonomously approached and viewed multiple visual shapes containing overlapping features. It was trained to prefer one of these shapes, by pairing the shape with an auditory cue, and it demonstrated this preference by orienting toward the preferred object. When confronted by a pair of these shapes, Darwin VIII learned to successfully track towards the preferred object, designated the target; and avoid the other objects, designated the distracters. Figure 5 shows that Darwin VIII subjects with intact reentrant connections tracked each of four different targets over 80% of the time, but subjects with severed reentrant connections had significant degradation in their tracking performance. These observations indicate that reentrant connectivity was necessary for the reliable discrimination of targets from visually similar distracters. In contrast to previous models of target selection, which required external intervention or an artificial environment [16, 17, 27, 38], Darwin VIII autonomously solved the binding problem in a rich environment in which there was self-movement that generated changes in the size and location of visual stimuli.

While performing this task, temporally coherent neuronal circuits were activated for each object in the visual scene. These circuits were composed of active neuronal groups distributed throughout different areas in the simulated nervous system. Circuits associated with objects having different combinations of features were distinguishable by their temporal characteristics, specifically the relative phase of their neural activity.

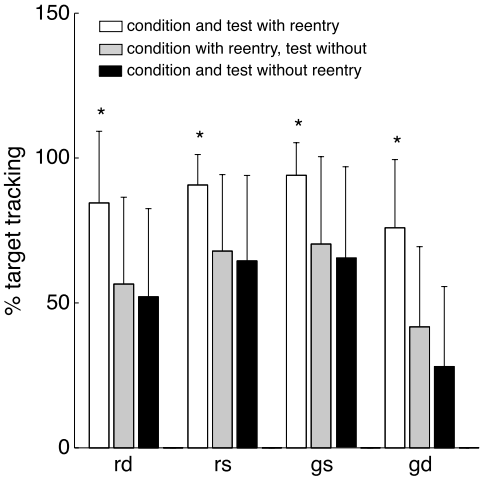

Figure 5. Darwin VIII's behavior following conditioning. Three separate Darwin VIII subjects were conditioned to prefer one of four target shapes (rd = red diamond, rs = red square, gs = green square, gd = green diamond). Bars represent the mean percentage of the time Darwin VIII centered its gaze on the target shape, with error bars denoting the standard deviation. Subjects with reentrant connections intact (white bars) tracked targets significantly better than subjects with lesions, where the reentrant connections between neural areas were cut during testing (light gray bars), and subjects with the reentrant connections cut during both training and testing (black bars). Asterisks denote $p < 0.01$ using the Wilcoxon rank sum test.

The results of these experiments demonstrated: (i) the importance of synchronization of neuronal activity in perceptual categorization, (ii) the role of reentrant connectivity (i.e., the activity across reciprocal excitatory connections between neuronal units) in sustaining this synchrony, and (iii) that the interaction between local processes (i.e., activity in each neural area) and global processes (i.e., emergent synchronously active circuits distributed throughout the nervous system) was necessary for successful discriminatory behavior.

3 Discussion

By correlating neural activity with behavioral responses during perceptual and conditioning tasks performed by our brain-based devices (BBDs), we have obtained new insights into the organization of autonomous behavior and into its underlying mechanisms. We have shown that BBDs, such as Darwin VII and Darwin VIII, can be used to gain insight into how nervous systems may categorize an unlabeled world and shape behavior, and can test specific theories of brain function, such as the visual binding of objects through reentrant synchronous circuits.

The development of adaptive and autonomous behavior by BBDs is novel in its neurally based approach and has implications for the construction of intelligent machines. Without a doubt the most sophisticated behavior seen in either biological or artificial agents is demonstrated by organisms whose behavior is guided by a nervous system. Thus, the construction of behaving devices based on principles of nervous systems may have much to offer.

However, this raises a question concerning which aspects of the nervous system should be incorporated into these devices and at what level of organization the nervous system should be modeled.

3.1 Choosing the Right Level of Abstraction

One of the most fundamental purposes of a biological model is to achieve explanatory power by facilitating direct comparisons with real biological function and experimental data. For example,

Thelen and colleagues have developed a dynamical system model of arm reaching and decision making by an infant in the A not B task [37]. This model has provided an explanation of the factors influencing a reach decision and has made novel predictions, such as the effect of salience of hidden objects on reaching decisions, which in turn have suggested further experiments that could be tested with infants.

However, there is a tendency among modelers to apply Occam's razor (i.e., the best theory is the simplest theory with the least assumptions) to their designs. For example, Beer and colleagues have introduced the notion of simple idealized models that show minimally cognitive behavior [3, 34]. These agents are developed through an evolutionary algorithm and contain on the order of tens of artificial neurons. Their agents show sensitivity to the future location of an object and can act is if they are able to switch attention between objects. Because their model is simple enough that a rigorous dynamical systems analysis can be applied, they can make interesting predictions about the cognitive mechanisms underlying attention and working memory.

In neuroscience, however, the simplest explanation is rarely the best explanation. Brains are large in scale, complex in dynamics, and have multiple levels of control (e.g., from molecular events to gross anatomical structure). What should be the proper level of abstraction and complexity for a brain-based device?

We designed BBDs to test theories of the brain that could not presently be tested with animals in the laboratory. These BBDs were designed so that they generate data, in the form of neuronal activities in different brain regions, that can be directly compared with experimental data. Equally important in the design is that a BBD must display adaptive behavior and this behavior must be measurable by an observer. The BBD's neural model, by necessity, is developed at a systems level, in which the structure of the brain and its different regions gives rise to adaptive behavior. Although these models are still too simple to make direct comparisons with neurophysiology, they can make predictions about the neuroanatomical and dynamical constraints that subserve adaptive behavior.

Preserving this structure tends to make the model very large in terms of the number of distinct neural areas and the number of neuronal units within each of these areas. For example, Darwin VII's early visual processing region contained as few neuronal units as possible to resolve the different shapes in its environment (see *Retina* in Figure 2). Because its nervous system was based on the anatomy of the visual stream, this early region, with its 4096 neuronal units, had projections to "higher" regions of the same scale (see *Vis B*, *Vis H, Vis V*, and *IT* in Figure 2), resulting in a large-scale simulation.

3.2 Importance of Neuroanatomy

The brain can be studied at many different levels (e.g., molecular, synaptic, cellular, and network), but structure at the gross anatomical level is critical for function and often ignored in models of the nervous system. Thus, chemical changes in the nervous system, such as the effect of Prozac on the serotonin system, serve to modulate brain function. A loss of neuroanatomical structure due to stroke, trauma, or injury can cause serious deficits. For example, an injury to the hippocampal and subhippocampal structures can cause anterograde amnesia, that is, the loss of the ability to form new memories [23, 32]. Lesions of the prefrontal cortex can cause a loss in the ability to plan for the future, make rational decisions, and process emotion [8]. Additionally, in cases where a stroke or some other injury causes lesions to one side of the parietal cortex, subjects are reported to neglect the side opposite the lesion [9]. In general, damage to the neuroanatomical structure of the brain, whether due to stroke, disease, or trauma, can lead to dramatic deficits in function.

Any model of brain function must not only take into consideration the structure of different brain regions, but also pay attention to the connectivity within and between these brain areas. 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 Darwin VII and Darwin VIII. Synaptic connections, which occur within and between neural areas, help define the function of a region. For example, the auditory cortex can

be induced to carry out visual processing by rerouting the projection from the optic nerve to the auditory cortex [36]. Additionally, special connectivity that occurs during a critical period of the developing animal allows the visual system to be responsive to orientation [5]. Thus, brains are defined by a distinct neuroanatomy in which there are areas of special function, which are defined by their connectivity to sensory input, to motor output, and to each other.

3.3 Brain-Based Devices and Embodiment

Brains do not function in isolation; they are tightly coupled with the organism's morphology and environment. Therefore, our models are 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 into the model. For example, a computer-generated object presented to a vision model has its shape and segmentation defined by the modeler 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 own 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 modeler is freed from simulating a world and need only concentrate on the development of a device that can actively explore the real world.

Interesting and unexpected results, which would not be apparent using a simulated agent in a virtual environment, have emerged by placing BBDs in the real world. For example, invariant object recognition in Darwin VII and Darwin VIII arose mainly as a result of plasticity between topographic feature detectors (e.g., *Vis* in Figure 2) and non-topographic association areas (e.g., *IT* in Figure 2) that were reinforced and expanded by subsequent inputs from stimuli during the BBD's movements. When the temporal sequence of the images leading to invariance was artificially shuffled [1], invariant object recognition did not occur. Another example of the influence of the environment was seen in our experience-dependent perceptual categorization experiments in which we altered the environment so that Darwin VII tended to sample different sequences and occurrences of stimuli from one trial to the next [21]. Differences in an individual's perceptual history had an effect on the organization and response of the nervous system. Similarly to results with primates [18], more neurons in the area *IT* responded to familiar than to unfamiliar stimuli.

3.4 Analysis of the Brain-Based Device's Behavior

While a central goal of research in the neurosciences is to understand the relationships between brain structure, function, and behavior, several related factors make this a challenging task. Presently, it is not possible simultaneously to analyze all levels of control from molecular events to motor responses. Using current electrophysiological techniques, one can simultaneously record only from a few hundred neurons in a behaving animal. To confront these issues and complement these approaches, we have adopted synthetic neural modeling [13, 30].

The advantage of a synthetic model is that these measurements can be carried out at all levels and in every neuron and synapse of the BBD's nervous system during the acquisition and recall of a behavior. However, researchers using synthetic models need to analyze their data in such a way that they can compare their results with empirical data. By analyzing the neural dynamics of the model (i.e., spike rates, correlations between areas, neural decoding, and prediction) and choosing a behavioral paradigm similar to those used when studying behaving animals (mazes, conditioning paradigms, decision-making tasks, etc.), the modeler can directly compare the BBD's results with the results of psychological and neurophysiological experiments. However, this does put the burden on modelers of including complexity in their models sufficient for these psychological and physiological metrics to be obtainable.

We have made a number of insights and predictions based on results of experiments with BBDs. For example, in Darwin VII, the activity of neuronal units in the area *IT* was sufficient to predict the

visual stimulus seen by Darwin VII's camera. We showed that this prediction held even when the number sampled was very small (see Figure 6 in [19]), which is in accord with results in behaving animals [15, 40]. In agreement with experimental results with primates, our experiments with Darwin VII showed that experience can have an effect on the number of neurons that respond to a category similar to results shown in the inferotemporal cortex of the primate [18]. Furthermore, by controlling the conditions in a way that would be impossible with a living organism, we demonstrated that the effect is altered if the organism is first exposed to a rich environment (see Table 1). In Darwin VII, we showed that the value system can shift from being activated by an innate unconditioned stimulus to being activated by an initially neutral stimulus, as do dopaminergic cells in the primate [35]. Thus, the activity in the value system becomes predictive of a stimulus' value and can shape behavior. In Darwin VIII, our model suggests that synchrony between widely separated neural areas may play a key role in solving the binding problem and demonstrates the importance of reentrant connections in facilitating binding through synchrony. It demonstrates that binding through synchrony is feasible in a real-world environment where objects are constantly changing in size and position.

4 Summary

Higher brain functions depend on the cooperative activity of an entire nervous system, reflecting its morphology, its dynamics, and its interaction with the phenotype and the environment. Brain-based devices are designed to incorporate these attributes so that they can test such theories of brain function. We have demonstrated that BBDs can address many difficult and unsolved problems, without instruction or intervention, such as object recognition, visual binding of objects in a scene, and operant conditioning. 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 devices also provide the groundwork for the development of intelligent machines that follow neurobiological rather than computational principles in their construction.

Acknowledgments

This work was supported by The W. M. Keck Foundation and by the Neurosciences Research Foundation, which supports The Neurosciences Institute. We thank Drs. J. A. Gally and A. K. Seth for detailed criticism.

References

1. Almassy, N., Edelman, G. M., & Sporns, O. (1998). Behavioral constraints in the development of neuronal properties: A cortical model embedded in a real-world device. *Cerebral Cortex*, *8*, 346–361.
2. Arkin, R. C. (1993). Modeling neural function at the schema level: Implications and results for robotic control In T. McKenna (Ed.), *Biological neural networks in invertebrate neuroethology and robotics* (pp. 383–409). Boston: Academic Press.
3. Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Science*, *4*, 91–99.
4. Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, *2*, 14–23.
5. Chapman, B., & Stryker, M. P. (1992). Origin of orientation tuning in the visual cortex. *Current Opinion in Neurobiology*, *2*, 498–501.
6. Chiel, H. J., & Beer, R. D. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neuroscience*, *20*, 553–557.
7. Clark, A. (1997). *Being there. Putting brain, body, and world together again*. Cambridge, MA: MIT Press.
8. Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., & Damasio, A. R. (1994). The return of Phineas Gage: Clues about the brain from the skull of a famous patient. *Science*, *264*, 1102–1105.
9. Driver, J., & Mattingley, J. B. (1998). Parietal neglect and visual awareness. *Nature Neuroscience*, *1*, 17–22.

10. Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection.* New York: Basic Books.

11. Edelman, G. M. (1993). Neural Darwinism: Selection and reentrant signaling in higher brain function. *Neuron*, *10*, 115–125.

12. Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences of the USA, 98,* 13763–13768.

13. Edelman, G. M., Reeke, G. N., Gall, W. E., Tononi, G., Williams, D., & Sporns, O. (1992). Synthetic neural modeling applied to a real-world artifact. *Proceedings of the National Academy of Sciences of the USA, 89,* 7267–7271.

14. Floreano, D., & Mondala, F. (1998). Evolutionary neurocontrollers for autonomous mobile robots. *Neural Networks*, *11*, 1461–1478.

15. Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, *233*, 1416–1419.

16. Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts in attention. *Vision Research*, *40*(10–12), 1489–1506.

17. Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transaction on Pattern Analysis and Machine Intellegence*, *20*, 1254–1259.

18. Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology*, *80*, 324–330.

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

20. Krichmar, J. L., & Snook, J. A. (2002). A neural approach to adaptive behavior and multi-sensor action selection in a mobile device. In *IEEE Conference on Robotics and Automation* (pp. 3864–3869).

21. Krichmar, J. L., Snook, J. A., Edelman, G. M., & Sporns, O. (2000). Experience-dependent perceptual categorization in a behaving real-world device. In J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, and S. W. Wilson (Eds.), *Animals to Animats 6: Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior* (pp. 41–50). Cambridge, MA: MIT Press.

22. McKinstry, J. L., Seth, A. K., Edelman, G. M., & Krichmar, J. L. (2002). Synchronous activity binds visual stimulus properties viewed by a brain-based device. *Society for Neuroscience Abstracts*, *558*, 12.

23. Mishkin, M., Vargha-Khadem, F., & Gadian, D. G. (1998). Amnesia and the organization of the hippocampal system. *Hippocampus*, *8*, 212–216.

24. Moravec, H. P. (1983). The Stanford cart and the CMU rover. *Proceedings of the IEEE, 71,* 872–884.

25. Nilsson, N. (1984). *Shakey the robot.* Menlo Park, CA: SRI International.

26. Nolfi, S., & Floreano, D. (2000). *Evolutionary robotics: The biology, intelligence, and technology of self-organizing machines.* Cambridge, MA: MIT Press.

27. Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, *13*, 4700–4719.

28. Pfeifer, R., & Schier, C. (1997). Sensory-motor coordination: The metaphor and beyond. *Robotics and Autonomous Systems*, *20*, 157–178.

29. Reeke, G. N., & Edelman, G. M. (1988). Real brains and artificial intelligence. *Daedalus: Proceedings of the American Academy of Arts and Sciences*, 117, 143–173.

30. Reeke, G. N., Sporns, O., & Edelman, G. M. (1990). Synthetic neural modeling: The "Darwin" series of recognition automata. *Proceedings of the IEEE, 78,* 1498–1530.

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

32. Scoville, W. B., & Milner, B. (2000). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neuropsychiatry and Clinical Neuroscience*, *12*, 103–113.

33. Seth, A. K., McKinstry, J. L., Edelman, G. M., & Krichmar, J. L. (2004). Visual binding through reentrant connectivity and dynamic synchronization in a brain-based device. *Cerebral Cortex*, *14*, 1185–1199.

34. Slocum, A. C., Downey, D. C., & Beer, R. D. (2000). Further experiments in the evolution of minimally cognitive behavior: From perceiving affordances to selective attention. In S. W. Wilson (Ed.), *From Animals to Animats 6: Sixth International Conference on Simulation of Adaptive Behavior* (pp. 430–439). Cambridge, MA: MIT Press.

35. Sporns, O., Almassy, N., & Edelman, G. M. (2000). Plasticity in value systems and its role in adaptive behavior. *Adaptive Behavior*, *8*, 129–148.

36. Sur, M. (1988). Visual projections induced into auditory thalamus and cortex: Implications for thalamic and cortical information processing. *Progress in Brain Research*, *75*, 129–136.

37. Thelen, E., Schoner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral Brain Sciences*, *24*, 1–34; discussion, 34–86.

38. Tononi, G., Sporns, O., & Edelman, G. M. (1992). Reentry and the problem of integrating multiple cortical areas: Simulation of dynamic integration in the visual system. *Cerebral Cortex*, *2*, 310–335.

39. Tononi, G., Sporns, O., & Edelman, G. M. (1999). Measures of degeneracy and redundancy in biological networks. *Proceedings of the National Academy of Sciences of the USA, 96*, 3257–3262.

40. Wilson, M. A., & McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, *261*, 1055–1058.