

Neuroinformatics

Editors

Giorgio A. Ascoli

Erik De Schutter

David N. Kennedy

Special Issue:

Neurorobotic Models in Neuroscience and Neuroinformatics

Guest Editors: Anil Seth, Olaf Sporns, and Jeffrey Krichmar

Indexed and Abstracted in:
Medline/Pubmed/Index Medicus
Science Citation Index®

 HUMANA PRESS

HumanaJournals.com
Search, Read, and Download

Original Article

Spatial Navigation and Causal Analysis in a Brain-Based Device Modeling Cortical–Hippocampal Interactions

Jeffrey L. Krichmar,* Anil K. Seth, Douglas A. Nitz, Jason G. Fleischer,
and Gerald M. Edelman

The Neurosciences Institute, 10640 John J. Hopkins Drive, San Diego, CA 92121, USA

Abstract

We describe Darwin X, a physical device that interacts with a real environment, whose behavior is guided by a simulated nervous system incorporating aspects of the detailed anatomy and physiology of the hippocampus and its surrounding regions. This brain-based device integrates cues from its environment and solves a spatial memory task. The responses of simulated neuronal units in the hippocampal areas during its exploratory behavior are comparable to place cells in the rodent hippocampus and emerged by associating sensory cues during exploration. To identify different functional hippocampal pathways and their influence on behavior, we employed

a time series analysis that distinguishes causal interactions within and between simulated hippocampal and neocortical regions while the device is engaged in a spatial memory task. Our analysis identified different functional pathways within the neural simulation and prompts novel predictions about the influence of the perforant path, the trisynaptic loop and hippocampal–cortical interactions on place cell activity and behavior during navigation. Moreover, this causal time series analysis may be useful in analyzing networks in general.

Index Entries: Perforant pathway; episodic memory; Granger causality; place cell.

(Neuroinformatics DOI: 10.1385/NI:3:3:197)

Introduction

Successful learning of an appropriate path through an environment requires the coordination of several cognitive processes operating simultaneously in the brain. Integrating multi-

modal information over temporal sequences is not only necessary for navigation, but also for the development of episodic memory. Episodic memory puts together information about temporal events and provides information about the “what,” “when,” and “where” of events

*Author to whom all correspondence and reprint requests should be addressed.

E-mail: krichmar@nsi.edu

(Tulving, 1972; Griffiths et al., 1999). In humans, the medial temporal lobe, where the hippocampus resides, is necessary for the acquisition of episodic memories (Scoville and Milner, 1957; Vargha-Khadem et al., 1997). Recent studies, which have reported episodic-like memory responses in rodent hippocampal cells (Frank et al., 2000; Ferbinteanu and Shapiro, 2003; Battaglia et al., 2004), have raised the possibility that the rodent may be used as a model for episodic memory. However, these studies have been able to record from only one region of the hippocampus at a time and only from a limited number of cells simultaneously from within that region.

It is likely that the unique anatomy and connectivity of the hippocampal region and its surrounding areas are critical both for spatial navigation and for the formation of episodic memories. At a macroscopic level, highly processed neocortical information from all modalities converges onto the medial temporal lobe. After several levels of further processing within the medial temporal lobe, and specifically the hippocampus, information diverges in broad projections back to the neocortex (Lavenex and Amaral, 2000; Witter et al., 2000a). Within the hippocampus itself, there are several levels of looping (Amaral et al., 1990; Bernard and Wheal, 1994; Treves and Rolls, 1994; Witter et al., 2000b). Signals from the neocortex enter the hippocampus via the entorhinal cortex. Within the hippocampus, there are connections from the entorhinal cortex to the dentate gyrus (the perforant path), from the dentate gyrus to CA3 (mossy fibers), from CA3 to CA1 (Schaffer collaterals), and then from CA1 back to the entorhinal cortex. This pathway is referred to as the trisynaptic loop. Although the perforant pathway mainly projects from the entorhinal cortex to the dentate gyrus, perforant path connections from the entorhinal cortex to CA3 or CA1 can provide short-circuits of the trisynaptic loop. There are also strong recurrent connections within the dentate gyrus and CA3

regions. A possible function for this unique anatomy is that the multilevel looping integrates and registers inputs over different timescales. However, testing this hypothesis in animals would require simultaneous electrophysiological recordings from many neurons over multiple neural areas. This is not feasible using the present-day physiological techniques.

Our method explores these issues through synthetic neural modeling (Reeke et al., 1990) and brain-based devices (Edelman et al., 1992). Like other neurobotic systems (Pfeifer and Scheier, 1997; Floreano and Mondada, 1998; Arleo and Gerstner, 2000; Gaussier et al., 2002; Sporns and Alexander, 2002; Uchibe and Doya, 2004; Weng, 2004; Chavarriaga et al., 2005; Ijspeert et al., 2005), the behavior of a brain-based device (BBD) is guided by a simulated nervous system. BBDs can be considered a class of neurobotics that are based on features of vertebrate neuroanatomy and neurophysiology, emphasizing the organism's interaction with the environment, and strictly constrained by the following design principles:

1. The device needs to be situated in a physical environment.
2. The device needs to engage in a behavioral task.
3. The device's behavior must be controlled by a simulated nervous system having a design that reflects the brain's architecture and dynamics.
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 vertebrate brain architecture and dynamics, high performance computing to run the network in real time, and the engineering of specialized physical devices to embody the network. The power of this approach is that it allows for simultaneously recording the state of all components of its

simulated nervous system at all levels during a behavioral task in the real world.

Over the past 12 yr, we have successfully constructed BBDs to test theories of the nervous system having to do with perceptual categorization, primary and secondary conditioning, visual binding, and texture discrimination (Edelman et al., 1992; Almasy et al., 1998; Krichmar and Edelman, 2002; Seth et al., 2004a, b).

In this paper, we describe Darwin X, a BBD incorporating aspects of the detailed anatomy and physiology of the hippocampus and its surrounding regions. Darwin X has also been described in another study under a different behavioral condition and with a different analysis methodology (Krichmar et al., 2005). This BBD can integrate cues from its environment and provide flexible navigation solutions to spatial memory tasks. The responses of simulated neuronal units in the hippocampal areas during its exploratory behavior are comparable to those of neurons in the rodent hippocampus. By simultaneously sampling from the BBD's neural regions during a navigation task, we test the notion that the architecture of the hippocampus provides a means to organize multimodal information over different timescales and show that this organization is critical for the development of spatial and episodic memories.

In a previous study, we introduced an analysis that "traces back" from any reference neuronal unit in our simulated nervous system to identify all of the units that were related to the activity of this unit both anatomically and dynamically over short time spans (Krichmar et al., 2005). This study was able to characterize functional anatomical pathways that affected the device's behavior. In particular, it identified the importance of both the perforant pathway and the trisynaptic circuit in generating appropriate behavior.

In the present study, we apply a novel analysis that identifies different functional hippocampal pathways and their influence on behavior, based on statistical parameters.

Specifically, we introduce a time series analysis that distinguishes causal interactions within and between simulated hippocampal and neocortical regions during behavior. Unlike the "backtrace" method, this time series analysis does not rely on the knowledge of the underlying anatomy and it makes inferences based on extended temporal sequences of neural activities. As a result, the present analysis makes novel predictions about the influence of experience on the perforant path, the trisynaptic loop, and hippocampal-cortical interactions, and on place cell activity and behavior during navigation.

Methods

Some of the methods have been described elsewhere (*see* the supporting text of Krichmar, et al., 2005), but for completeness and convenience we include here the relevant details of the model and experimental design.

Experimental Protocol

Spatial memory in Darwin X was assessed by developing a dry variant of the Morris water maze (Morris, 1984) in which the device must find a hidden platform in its environment. Successful performance on this task is reflected by the device using only visual landmarks and self-movement cues to navigate to the hidden platform from any starting position.

Darwin X was allowed to explore a 16' × 14' enclosure in which there were visual landmarks on the walls (*see* Fig. 1). Paper strips of different colors and varying widths were hung on each of the walls. A circular "hidden platform" of reflective black construction paper, 24 inches in diameter, was placed on the floor in the center of one quadrant. The platform could not be detected visually by Darwin X, but was detectable at close range by an IR sensor on the front of the device.

During a training trial, Darwin X explored its enclosure until it encountered the hidden platform or until a time limit of 1000 s was reached. A training block was defined as a set

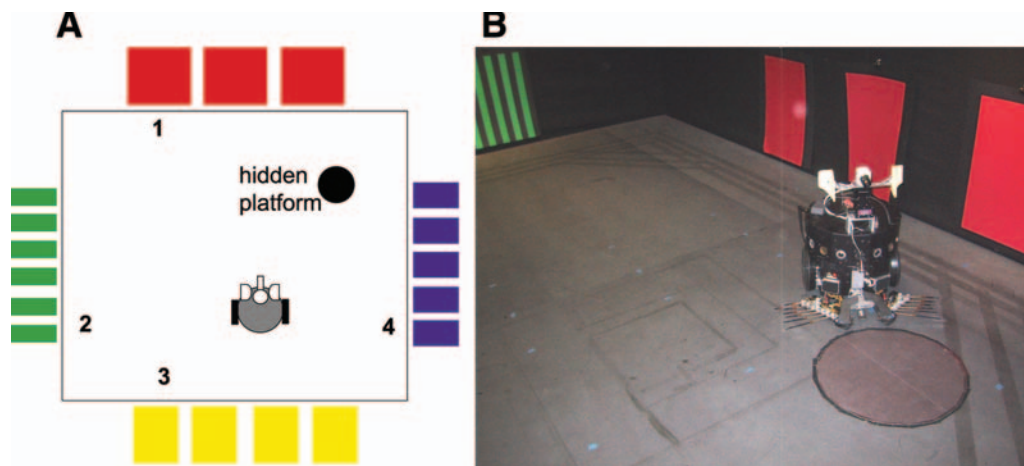


Fig. 1. Layout of the enclosure used for the hidden platform task. **(A)** Schematic of the environment. Enclosure is 16' \times 14' with black walls and flooring. Pieces of colored construction paper, of varying widths, were hung on each of the walls. A hidden platform, 24'' in diameter and made of reflective black construction paper, was placed in the center of a quadrant in the enclosure. The device could detect when it was over the platform by means of an IR sensor. Each trial began in one of four starting locations (see numbers 1–4 in the figure). **(B)** Snapshot of Darwin X in its environment. (Adapted from Krichmar et al., 2005.)

of four trials from each of four starting locations (see numbers 1–4 in Fig. 1A). At the start of each trial, Darwin X's initial heading was set such that it was perpendicular from the wall nearest to the starting location. Four blocks (16 trials) were completed by the device during training. After training, a probe trial, in which the hidden platform was removed, was used to assess the BBD's performance.

Device

The physical device portion of Darwin X consists of a wheeled mobile base equipped with a charge-coupled device (CCD) camera for vision, odometry for self-movement cues, infrared (IR) transceivers for obstacle avoidance, and a front-mounted, downward-pointing IR transceiver to detect the hidden platform (Fig. 1B). Light-emitting diodes on top of Darwin X, which were detectable by two cameras placed over the enclosure, were used to track Darwin X's position.

Darwin X was equipped with a set of innate behavioral responses for exploration, obstacle avoidance, and platform detection. Its default

behavior was to proceed forward for approx 10 s, rotate to its left, then to its right, and then choose a new heading. If Darwin X detected a large obstacle, such as a wall with its IR sensors, it would initiate an obstacle avoidance response. If it detected the hidden platform with the downward-facing IR sensor, Darwin X would stop and rotate to the left and then to the right.

Neural Simulation

Darwin X's behavior is guided by a simulated nervous system modeled on the anatomy and physiology of the mammalian nervous system but, obviously, with far fewer neurons and a much less complex architecture. It consists of a number of areas labeled according to the analogous neocortical, hippocampal, and subcortical brain regions (see Table 1). Each area contains neuronal units that can be either excitatory or inhibitory, each of which represents a local population of neurons (Edelman, 1987), in which the mean firing rate variable of each unit corresponds to the average activity of a group of roughly 100 real neurons during a

time period of approx 200 ms. To distinguish modeled areas from corresponding regions in the mammalian nervous system, the simulated areas are indicated in italics (e.g., *IT*).

During each simulation cycle of Darwin X, sensory input is processed, the states of all neuronal units are computed, the connection strengths of all plastic connections are determined, and motor output is generated. The neural simulation was run on a Beowulf cluster containing 12 1.4-GHz Pentium IV computers running the Linux operating system. In our experiments, execution of each simulation cycle required approx 200 ms of real time, which was limited by our cluster's computing power. Although it would have been preferable to have shorter cycle times, a 200 ms cycle time was sufficient such that neuronal responses did not lag behind the dynamic real-world sensory input. All sensory input from the device and motor commands to the device were communicated through wireless links between the device and one of the cluster's workstations. During each simulation cycle, all neuronal activities were saved on a hard disk, and Darwin X's position was recorded.

In the present experiments, the simulated nervous system contained 50 neural areas, 90,000 neuronal units, and approx 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. Figure 2A shows a high-level diagram of the simulated nervous system including the various neural areas and the arrangement of synaptic connections. Figure 2B shows the detailed neuroanatomy of the simulated hippocampus. Specific parameters relating to each area and to patterns of connectivity are given in Tables 1 and 2.

The visual system was modeled on the primate occipitotemporal or ventral cortical pathway and a dorsal cortical pathway. The ventral cortical pathway, in our model (*V1-color* → *V2/4-color* → *IT*), contained neuronal units in succes-

sive areas having progressively larger receptive fields until, in inferotemporal cortex (*IT*), receptive fields cover nearly the entire visual field (Ungerleider and Haxby, 1994). The dorsal cortical pathway in our model (*V1-width* → *V2/4-width* → *PR*) contained neuronal units that responded to the size and position of objects (Ungerleider and Mishkin, 1982).

Visual images from Darwin X's CCD camera were filtered for color and edges and the filtered output directly affected neural activity in area *V1*, which is composed of functionally segregated subareas for color and shape. The CCD camera sends 320 × 240 pixel red-green-blue video images, via a radiofrequency transmitter, to a frame grabber attached to one of the workstations running the neural simulation. The image was spatially averaged to produce an 80 × 60 pixel image. Gabor filters of different sizes (8 × 8, 16 × 16, 32 × 32, and 64 × 64) were used to detect vertical edges of varying widths. The output of the Gabor function mapped directly onto the neuronal units of the corresponding *V1* subarea (*V1-width8*, *V1-width16*, *V1-width32*, and *V1-width64*). Color filters (red-positive center with a green-negative surrounding, red-negative center with a green-positive surrounding, blue-positive with red-green negative, and blue-negative with red-green positive) were applied to the image. The outputs of the color filters were mapped directly onto the neuronal units of *V1-red*, *V1-green*, *V1-blue*, and *V1-yellow*. *V1* neuronal units projected retinotopically to neuronal units in *V2/V4* (see Fig. 2 and Table 2).

A head direction system was modeled after areas of the rodent nervous system (e.g., anterior thalamic nuclei) that respond selectively to the animal's heading (Muller et al., 1996; Taube, 1998). Neurons in these areas are often called head direction cells. Odometer information obtained from Darwin X's wheels was used to estimate current heading. This information was input into the head direction neural area (*HD*). Each of the 360 *HD* neuronal units had a cosine tuning curve, which

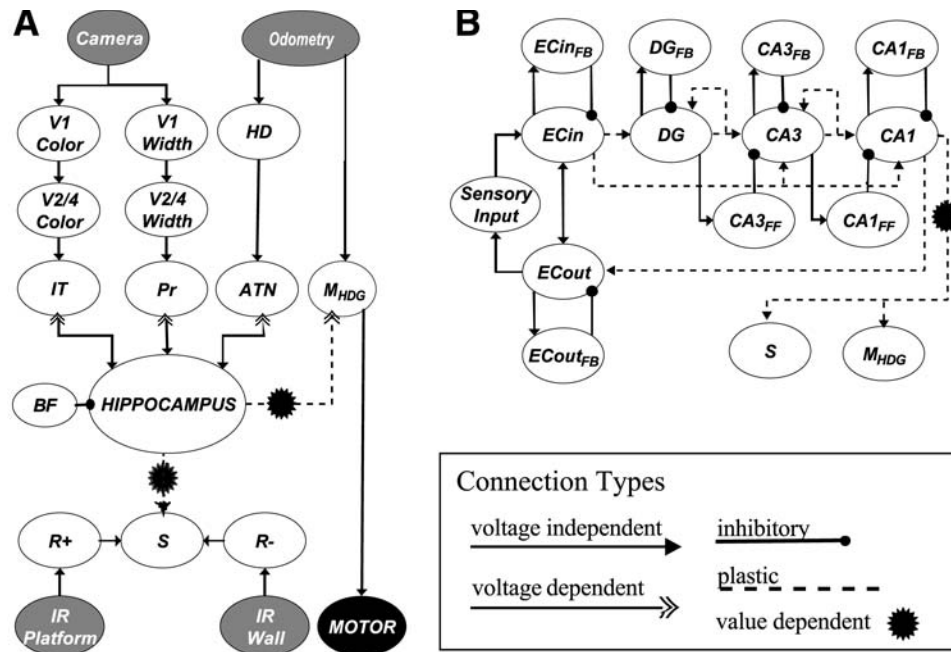


Fig. 2. Schematic of the regional and functional neuroanatomy of Darwin X. The simulated nervous system contained 50 neuronal areas, 90,000 neuronal units, and approx 1.4×10^6 synaptic connections. Ellipses denote different neural areas. Arrows denote projections from one area to another. **(A)** High-level diagram of the neural simulation connectivity. Input to the neural simulation came from a CCD camera, wheel odometry, and IR sensors for wall and platform detection. The simulation contained neural areas analogous to visual cortex (*V1*, *V2/4*), the inferotemporal cortex (*IT*), parietal cortex (*PR*), the head direction system (*HD*), anterior thalamic nuclei (*ATN*), motor areas for egocentric heading (M_{HDG}), a value system (*S*), and positive and negative reward areas (R^+ , R^-). The hippocampus received input from three major sensory streams (*IT*, *PR*, and *ATN*). The hippocampus projected to a motor area (M_{HDG}), and the value system (*S*). The hippocampus received rhythmic inhibition from a simulated basal forebrain (*BF*) region. For clarity, the intrinsic connections within neural areas are omitted. **(B)** Detailed connectivity within the hippocampal region. *Sensory input* (see *ATN*, *IT*, and *PR* in Fig. 2A) enters the hippocampus via EC_{IN} . The modeled hippocampus contained areas analogous to entorhinal cortex (EC_{IN} , EC_{OUT}), dentate gyrus (*DG*), and the *CA3* and *CA1* subfields. These areas contained interneurons that implemented feedback inhibition (e.g., $CA3 \rightarrow CA3_{FB} \rightarrow CA3$) and feedforward inhibition (e.g., $DG \rightarrow CA3_{FF} \rightarrow CA3$). Recurrent activity in hippocampal subregions and the *BF* to hippocampal region connectivity was omitted for clarity. See text, and Tables 1 and 2 for details. (Reprinted with the permission of Krichmar et al., 2005.)

responded maximally to a preferred heading with a tuning width of π rad:

$$HD_i = \cos\left(\frac{i}{360}2\pi - \text{curr_heading}\right)^5 \quad (1)$$

where HD_i is a head direction unit with a preferred direction of $[(i/360)2\pi]$ and i ranges from 0 to 359. HD_i was not allowed to have values less than zero.

The head direction units projected topographically to an area analogous to the anterior thalamic nucleus (see $HD \rightarrow ATN$ in Table 2 and Fig. 2) and to a motor area (see $HD \rightarrow M_{HDG}$ in Table 2 and Fig. 2) used for selecting a new heading (see Eq. 5).

The architecture of the simulated hippocampal formation was based on rodent neuroanatomy. In this article, we restrict our model

Table 1
Values of Parameters Defining Properties of Neuronal Units in Darwin X

Neural Area Name	Area	Size	σ -fire	σ -vdep	ω	g
Primary visual cortex	<i>V1</i> (8)	60 × 80	–	–	–	–
Head direction system	<i>HD</i>	1 × 360	–	–	–	–
Positive value	<i>R</i> ⁺	1 × 1	–	–	–	–
Negative value	<i>R</i> [–]	1 × 1	–	–	–	–
Basal forebrain	<i>BF</i>	1 × 1	–	–	–	–
Visual cortex—color	<i>V2/4-color</i> (4)	6 × 8	0.20	0.10	0.0	1.0
Visual cortex—spatial tuning	<i>V2/4-width</i> (4)	15 × 20	0.20	0.10	0.0	1.0
Inferotemporal cortex	<i>IT</i>	30 × 30	0.20	0.10	0.0	1.0
IT inhibitory interneurons	<i>IT</i> _{<i>i</i>}	15 × 15	0.20	0.10	0.15	1.0
Parietal cortex	<i>PR</i>	30 × 30	0.20	0.10	0.0	1.0
Anterior thalamic nucleus	<i>ATN</i>	30 × 30	0.10	0.10	0.50	1.0
ATN inhibitory interneurons	<i>ATN</i> _{<i>i</i>}	30 × 30	0.10	0.10	0.15	1.0
Motor area for heading	<i>M</i> _{<i>HDG</i>}	1 × 60	0.0	0.10	0.0	1.0
<i>M</i> _{<i>HDG</i>} inhibitory interneurons	<i>M</i> _{<i>HDG</i>} _{<i>i</i>}	1 × 60	0.0	0.10	0.0	1.0
Value system	<i>S</i>	4 × 4	0.0	0.10	0.0	1.0
Entorhinal cortex—input	<i>EC</i> _{<i>IN</i>}	30 × 30	0.10	0.10	0.50	1.0
<i>EC</i> _{<i>IN</i>} inhibitory interneurons	<i>EC</i> _{<i>IN</i>} _{<i>i</i>}	15 × 15	0.02	0.10	0.0	1.0
Entorhinal cortex—output	<i>EC</i> _{<i>OUT</i>}	30 × 30	0.10	0.10	0.50	1.0
<i>EC</i> _{<i>OUT</i>} inhibitory interneurons	<i>EC</i> _{<i>OUT</i>} _{<i>i</i>}	15 × 15	0.02	0.10	0.0	1.0
Dentate gyrus	<i>DG</i>	30 × 30	0.10	0.10	0.50	0.75
DG feedback inhibitory interneurons	<i>DG</i> _{<i>FB</i>} _{<i>i</i>}	15 × 15	0.02	0.10	0.0	1.0
CA3 hippocampal subfield	<i>CA3</i>	15 × 15	0.05	0.10	0.50	0.75
CA3 feedback inhibitory interneurons	<i>CA3</i> _{<i>FB</i>} _{<i>i</i>}	8 × 8	0.02	0.10	0.0	1.0
CA3 feedforward inhibitory interneurons	<i>CA3</i> _{<i>FF</i>} _{<i>i</i>}	15 × 15	0.02	0.10	0.0	1.0
CA1 hippocampal subfield	<i>CA1</i>	20 × 20	0.05	0.10	0.50	0.75
CA1 feedback inhibitory interneurons	<i>CA1</i> _{<i>FB</i>} _{<i>i</i>}	10 × 10	0.02	0.10	0.0	1.0
CA1 feedforward inhibitory interneurons	<i>CA1</i> _{<i>FF</i>} _{<i>i</i>}	10 × 10	0.02	0.10	0.0	1.0

Areas *V1*, *HD*, *R*⁺, and *R*[–] are input areas and their activity is based on the camera image, odometry, and IR sensors, respectively. Areas *V1* and *V2/4* have four subareas tuned for color (red, green, blue, and yellow) and four subareas tuned for varying width vertical bars. An area with a lowercase *i* in its name (e.g., *EC*_{*IN*}_{*i*}) contains inhibitory neuronal units. The table indicates the number of neuronal units in each area or subarea (size). Neuronal units in each area have a specific firing threshold (σ -fire), a threshold above which voltage-dependent connections can have an effect (σ -vdep), a persistence parameter (ω), and a scaling factor (g).

to a single hemisphere of the hippocampus and do not take into consideration the connectivity between hemispheres. The input streams into the hippocampus are from the associative cortical areas in the simulation (see *ATN*→*EC*_{*IN*}, *IT*→*EC*_{*IN*}, and *PR*→*EC*_{*IN*} in Table 2 and *sensory input* in Fig. 2B). Parameter values for the neu-

ronal units and connections in these areas were tuned such that each cortical area (*ATN*, *PR*, and *IT*) had an equivalent synaptic influence on *EC*_{*IN*} (see Tables 1 and 2). The relative numbers of neuronal units in each area, the relative number of synapses, and the intrinsic and extrinsic synaptic pathways of the hippocampus

Table 2
Properties of Anatomical Projections and Connection Types in Darwin X

Projection	Arbor	p	$c_{ij}(0)$	Type	φ	η	$k1$	$k2$
V1-color→V2/4-color	[] 1 × 1	1.00	003,0.050	VI	–	–	–	–
V2/4-color→V2/4-color(intra)	[] 2 × 2	0.40	0.5,0.6	VD	–	–	–	–
V2/4-color→V2/4-color(inter)	[] 3 × 3	1.00	–0.0012, –0.28	VI	–	–	–	–
V1-width→V2/4-width	[] 1 × 1, 2 × 2, 3 × 3, 4 × 4	1.00	0.008,0.009	VI	–	–	–	–
V2/4-width→V2/4-width(intra)	[] 2 × 2	0.40	0.5,0.6	VD	–	–	–	–
V2/4-width→V2/4-width(inter)	[] 3 × 3	1.00	–0.012, –0.014	VI	–	–	–	–
V2/V4→IT	Nontopo	0.05	0.03,0.04	VI	–	–	–	–
IT→IT	[] 1 × 1	1.00	0.08,0.14	VI	–	–	–	–
IT→Iti	⊖ 2,3	1.00	0.06,0.08	VI	–	–	–	–
IT ⁱ →IT	[] 1 × 1	1.00	–0.36, –0.50	VI	–	–	–	–
V2/V4→PR	[] 1 × 1	0.25	0.25,0.30	VD	–	–	–	–
PR→PR	⊖ 4,6	1.00	–0.06, –0.08	VI	–	–	–	–
HD→ATN	[] 30 × 2	0.20	0.01,0.02	VI	–	–	–	–
ATN→ATN ⁱ	⊖ 10,15	0.25	0.01,0.02	VI	–	–	–	–
ATN ⁱ →ATN	[] 1 × 1	1.00	–0.36, –0.50	VI	–	–	–	–
HD→M _{H_{HDG}}	[] 1 × 1	1.00	0.01,0.01	VI	–	–	–	–
M _{H_{HDG}} →M _{H_{HDG}} ⁱ	⊖ 20,30	0.50	0.10,0.20	VI	–	–	–	–
M _{H_{HDG}} ⁱ →M _{H_{HDG}}	[] 1 × 1	1.00	–0.36, –0.50	VI	–	–	–	–
ATN→EC _{IN}	Nontopo	0.001	0.40,0.50	VI	–	–	–	–
IT→EC _{IN}	Nontopo	0.001	0.40,0.50	VI	–	–	–	–
PR→EC _{IN}	Nontopo	0.001	0.40,0.50	VI	–	–	–	–
EC _{IN} →EC _{OUT}	Nontopo	0.05	0.04,0.08	VI	–	–	–	–
EC _{IN} →EC _{IN} ⁱ	⊖ 2,3	0.10	0.45,0.60	VI	–	–	–	–
EC _{IN} ⁱ →EC _{IN}	[] 1 × 1	1.00	–0.90, –1.20	VI	–	–	–	–
EC _{IN} →DG	[] 3 × 3	0.10	0.45,0.60	VI	0.75	0.05	0.90	0.45
EC _{IN} →CA3	[] 3 × 3	0.05	0.15,0.20	VI	0.75	0.05	0.90	0.45
EC _{IN} →CA1	[] 3 × 3	0.04	0.30,0.40	VI	0.75	0.05	0.90	0.45
EC _{OUT} →ATN	Nontopo	0.01	0.40,0.45	VD	–	–	–	–
EC _{OUT} →IT	Nontopo	0.01	0.40,0.45	VD	–	–	–	–
EC _{OUT} →PR	Nontopo	0.01	0.40,0.45	VD	–	–	–	–
EC _{OUT} →EC _{IN}	Nontopo	0.05	0.04,0.08	VI	–	–	–	–
EC _{OUT} →EC _{OUT} ⁱ	⊖ 2,3	0.10	0.45,0.60	VI	–	–	–	–
EC _{OUT} ⁱ →EC _{OUT}	[] 1 × 1	1.00	–0.90, –1.20	VI	–	–	–	–
DG→CA3	[] 3 × 3	0.03	0.45,0.60	VI	–	–	–	–

(Continued)

Table 2 (Continued)

Projection	Arbor	p	$c_{ij}(0)$	Type	φ	η	k_1	k_2
$DG \rightarrow DG$	$[] 1 \times 1$	1.00	0.10,0.14	VI	–	–	–	–
$DG \rightarrow DG_{FB}^i$	$\Theta 2,3$	0.10	0.45,0.60	VI	–	–	–	–
$DG_{FB}^i \rightarrow DG$	$[] 1 \times 1$	1.00	-0.90, -1.20	VI	–	–	–	–
$DG \rightarrow CA3_{FF}^i$	$\Theta 2,3$	0.10	0.45,0.60	VI	–	–	–	–
$CA3_{FF}^i \rightarrow CA3$	$[] 1 \times 1$	1.00	-0.90, -1.20	VI	–	–	–	–
$CA3 \rightarrow CA1$	$[] 3 \times 3$	0.08	0.45,0.60	VI	0.75	0.05	0.90	0.45
$CA3 \rightarrow CA3$	Nontopo	0.10	0.15,0.20	VI	0.75	0.05	0.90	0.45
$CA3 \rightarrow CA3_{FB}^i$	$\Theta 2,3$	0.10	0.45,0.60	VI	–	–	–	–
$CA3_{FB}^i \rightarrow CA3$	$[] 1 \times 1$	1.00	-0.90, -1.20	VI	–	–	–	–
$CA3 \rightarrow CA1_{FF}^i$	$\Theta 2,3$	0.10	0.45,0.60	VI	–	–	–	–
$CA1_{FF}^i \rightarrow CA1$	$[] 1 \times 1$	1.00	-0.90, -1.20	VI	–	–	–	–
$CA1 \rightarrow EC_{OUT}$	$[] 3 \times 3$	0.25	0.60,0.75	VI	0.75	0.05	0.90	0.45
$CA1 \rightarrow M_{HDG} \#$	$[] 3 \times 3$	1.00	0.01,0.02	VD	–	0.05	0.90	0.45
$CA1 \rightarrow S \#$	$[] 3 \times 3$	1.00	0.01,0.02	VD	–	0.005	0.90	0.45
$R^+ \rightarrow S$	Nontopo	1.00	0.25,0.25	VI	–	–	–	–
$R^- \rightarrow S$	Nontopo	1.00	0.25,0.25	VI	–	–	–	–
$BF \rightarrow EC_{IN}, EC_{OUT},$ $DG, CA3, CA1$	Nontopo	0.05	-0.01, -0.02	VI	–	–	–	–

A presynaptic neuronal unit connects to a postsynaptic neuronal unit with a given probability (p) and given projection shape (Arbor). This arborization shape can be rectangular “[]” with a height and width ($h \times w$), doughnut-shaped or a ring of lateral connections “ Θ ” with the size of the ring constrained by an inner and outer radius (r_1, r_2), or nontopographical “Nontopo” in which any pairs of presynaptic and postsynaptic neuronal units have a given probability of being connected. The initial connection strengths $c_{ij}(0)$ are set randomly, with a uniform distribution, within the range given by a minimum and maximum value (min, max). A negative value for $c_{ij}(0)$ indicates inhibitory connections. Connections marked with “intra” denote those within a visual sub-area and connections marked with “inter” denote those between visual subareas. Projections marked # are value-dependent. A connection type can be voltage-independent (VI) or voltage-dependent (VD). φ denotes the persistence of the connection. Nonzero values for η, k_1 , and k_2 signify plastic connections. Expansion of neural area abbreviations can be found in the Fig. 2 caption.

were implemented based on known anatomical measurements (Amaral et al., 1990; Bernard and Wheal, 1994; Treves and Rolls, 1994). The perforant path projects mainly from entorhinal cortex to the dentate gyrus but also to the CA3 and CA1 subfields (see $EC_{IN} \rightarrow DG, EC_{IN} \rightarrow CA3$, and $EC_{IN} \rightarrow CA3$ in Table 2 and Fig. 2). The mossy fibers (see $DG \rightarrow CA3$ in Table 2 and Fig. 2), Schaffer collaterals (see $CA3 \rightarrow CA1$ in Table 2 and Fig. 2), and divergent projections from the hippocampus back to cortex (see $CA1 \rightarrow EC_{OUT} \rightarrow ATN, IT, PR$) in Table 2 and Fig. 2)

were also reflected in the neural simulation. Moreover, the prevalent recurrent connectivity found in the hippocampal formation was included in the model (see $EC_{IN} \leftrightarrow EC_{OUT}, DG \rightarrow DG$, and $CA3 \rightarrow CA3$ in Table 2 and Fig. 2).

There are distinct patterns of intrinsic and extrinsic, feedback and feedforward inhibitory connections in the hippocampal circuitry (Bernard and Wheal, 1994; Freund and Buzsaki, 1996). Feedback inhibitory connections (see $EC \rightarrow EC_{FB} \rightarrow EC, DG \rightarrow DG_{FB} \rightarrow DG, CA3 \rightarrow CA3_{FB} \rightarrow CA3$, and $CA1 \rightarrow CA1_{FB} \rightarrow CA1$ in Table

2 and Fig. 2) and feedforward inhibitory connections (see $EC \rightarrow DG_{FF} \rightarrow DG$, $DG \rightarrow CA3_{FF} \rightarrow CA3$, and $CA3 \rightarrow CA1_{FF} \rightarrow CA1$ in Table 2 and Fig. 2) were included in the model. These connections were important for separating inputs and maintaining network stability.

A simplified model of the basal forebrain provided an extrinsic rhythm for the neural simulation. The function of the simulated basal forebrain area was to gate input into the hippocampus and keep activity levels stable. Our simulated extrinsic rhythm is based on the theta rhythm activity, which appears to be regulated by inhibitory GABA-ergic (where GABA stands for g-aminobutyric acid) input from the medial septum or basal forebrain (Stewart and Fox, 1990; Hasselmo et al., 2002b). In order to simulate this rhythm, BF activity followed a sinusoidal curve:

$$BF(t) = \text{phasic rhythm}(t \bmod n) \quad (2)$$

where $\text{phasic rhythm} = \{0.01, 0.165, 0.33, 0.495, 0.66, 0.825, 1.00, 0.825, 0.66, 0.495, 0.33, 0.165, 0.01\}$, t is the time step, and n is the number of simulation cycles for a complete sine wave. To allow this discretized sine wave to be smooth and also allow sufficient time for new input to propagate through the network, n was set to 13. Owing to computational limitations, the duration of the cycle was not comparable to that of real nervous systems. BF projected to all hippocampal areas with inhibitory connections (see $BF \rightarrow EC_{IN}, EC_{OUT}, DG, CA3, CA1$ in Fig. 2 and Table 2). In order to keep the level of activity of a hippocampal region within specific ranges, BF inhibition to each hippocampal region was adjusted every simulation cycle based on the region's current activity:

$$\begin{aligned} \Delta sf_r(t) &= [s_r(t) - tgt_r] \\ BF_r(t) &= BF(t) + \Delta sf_r(t) \end{aligned} \quad (3)$$

where r denotes the region (i.e., EC_{IN} , EC_{OUT} , DG , $CA3$, and $CA1$), $sf_r(t)$ is the scale factor at time t , $s_r(t)$ is the percentage of active neuronal units in region r at time t , tgt_r is the desired percentage of active units in area r ($EC_{IN}=10\%$,

$EC_{OUT}=10\%$, $DG=20\%$, $CA3=5\%$, and $CA1=10\%$), and $BF_r(t)$ is the presynaptic neuronal unit activity for a BF to hippocampus region r connection.

Activity in the simulated value system (Area S, Fig. 2) signals the occurrence of salient sensory events and this activity contributes to the modulation of value-dependent connection strengths in synaptic pathways ($CA1 \rightarrow S$ and $CA1 \rightarrow M_{HDC}$). Activity in S is analogous to that of ascending neuromodulatory systems in that it influences large regions (see the subsection "Neuronal Dynamics and Synaptic Plasticity") of the simulated nervous system and persists for several cycles (Aston-Jones and Bloom, 1981; Schultz et al., 1997; Sporns et al., 2000). The projection from our simulated CA1 to the value and goal decision areas is consistent with the connectivity between CA1 and nucleus accumbens and frontal areas (Mogenson and Nielsen, 1984; Thierry et al., 2000). Initially, S is activated by the hidden platform IR detector (see $R^+ \rightarrow S$ in Table 2 and Fig. 2), causing potentiation of value-dependent connections, or by obstacle avoidance IR detectors (see $R^- \rightarrow S$ in Table 2 and Fig. 2), causing depression of value-dependent connections. After experience, the value system could be activated by CA1. The magnitude of potentiation or depression is based on a neural implementation of temporal difference learning rule (Sutton and Barto, 1990; Montague et al., 1996).

$$TD(t) = \begin{cases} R^+(t) - \overline{S(t-\tau)}, & R^+ > 0 \\ \overline{S(t-\tau)} - R^-(t), & R^- > 0 \\ \overline{S(t)} - \overline{S(t-\tau)}, & \text{otherwise} \end{cases} \quad (4)$$

where $\overline{S(t)}$ is the average activity of the value system at time t , τ is one phasic rhythm cycle (13 simulation cycles), R^+ is positive reward and equal to 1 if the BBD is over the hidden platform, and R^- is negative reward and equal to 1 if the BBD is too close to a wall. The basic idea of the temporal difference rule is that learning is based

on the difference between temporally successive predictions of rewards. The goal of learning is to make the learner's current prediction of expected reward match more closely the actual expected reward at the next time interval (τ). If the expected reward value increases over τ , temporal difference (TD) is positive and the affected synaptic connections are potentiated, and if the change in value decreases, TD is negative and the affected synaptic connections are depressed. Further details on how the temporal difference is applied to individual synaptic connections are given in the subsection "Neuronal Dynamics and Synaptic Plasticity" below.

Darwin X selected a new heading every three phasic rhythm cycles (39 simulation cycles or 7.8 s), based on activity in its motor area (M_{HDG}). From its original heading, Darwin X would first turn counterclockwise 60° and wait for 3 s, then turn clockwise for 60° and wait 3 s, then another 60° clockwise turn and wait 3 s, and finally turn counterclockwise returning to its original heading. The average activity of M_{HDG} was calculated during the wait periods. A softmax algorithm was used to create a probability distribution for choosing a new heading:

$$p(\text{newhdg}) = \frac{\exp\left\{\beta\left[\overline{M_{HDG}(\text{newhdg})}\right]\right\}}{\sum_{h=\text{hdg}-60, \text{hdg}, \text{hdg}+60} \exp\left\{\beta\left[\overline{M_{HDG}(h)}\right]\right\}} \quad (5)$$

where newhdg is a possible new heading for Darwin X, $\overline{M_{HDG}(\text{newhdg})}$ is the average activity of M_{HDG} at a possible new heading, β is the exploration-exploitation parameter set at 40, hdg is the current heading, and h has three positions (current heading, current heading less 60° , and current plus 60°).

Neuronal Dynamics and Synaptic Plasticity

Neuronal units in Darwin X are simulated by a mean firing rate model and synaptic connections between neuronal units, both within and between neural areas, are set to be either

voltage-independent or voltage-dependent, and either plastic or nonplastic (see Fig. 2 and Table 2). Voltage-independent connections provide synaptic input regardless of postsynaptic state. Voltage-dependent connections represent the contribution of receptor types (e.g., NMDA receptors) that require postsynaptic depolarization to be activated and tend to play a modulatory role in neuronal dynamics (Wray and Edelman, 1996; Grossberg, 1999).

The mean firing rate of each neuronal unit ranges continuously from 0 (quiescent) to 1 (maximal firing). The state of a neuronal unit is updated as a function of its current state and contributions from voltage-independent and voltage-dependent inputs (see Fig. 2). The voltage-independent input to unit i from unit j is

$$A_{ij}^{VI}(t) = c_{ij}s_j(t) \quad (6)$$

where $s_j(t)$ is the activity of unit j and c_{ij} is the connection strength from unit j to unit i . The voltage-independent postsynaptic influence POST_i^{VI} on unit i is calculated by summing over all the inputs onto unit i :

$$\text{POST}_i^{VI}(t) = \varphi\left[\text{POST}_i^{VI}(t-1)\right] + (1-\varphi)\left(\sum_{l=1}^M \sum_{j=1}^{N_l} [A_{ij}^{VI}(t)]\right) \quad (7)$$

where M is the number of different anatomically defined connection types (see Table 2), N_l is the number of connections of type M projecting to unit i , and φ is the persistence of synaptic input.

The voltage-dependent input to unit i from unit j is

$$A_{ij}^{VD}(t) = \Phi\left[\text{POST}_i^{VI}(t)\right]c_{ij}s_j(t),$$

where

$$\Phi(x) = \begin{cases} 0, & x < \sigma_i^{\text{vdep}} \\ x, & \text{otherwise} \end{cases} \quad (8)$$

where σ_i^{vdep} threshold for the postsynaptic activity below which voltage-dependent connections have no effect (see Table 1).

The voltage-dependent postsynaptic influence on unit i , $POST_i^{VD}$, is given by

$$POST_i^{VD}(t) = \varphi [POST_i^{VD}(t-1)] + (1-\varphi) \left(\sum_{l=1}^M \sum_{j=1}^{N_j} [A_{ij}^{VD}(t)] \right) \quad (9)$$

The total postsynaptic influence on neuronal unit i is given by

$$POST_i = \sum_{j=1}^{N_{VI}} POST_j^{VI} + \sum_{k=1}^{N_{VP}} POST_k^{VD} \quad (10)$$

The new activity is determined by the following activation function:

$$s_i(t+1) = \phi \left(\tanh \left\{ g_i [POST_i + \omega s_i(t)] \right\} \right),$$

where

$$\phi(x) = \begin{cases} 0, & x < \sigma_i^{\text{fire}} \\ x, & \text{otherwise} \end{cases} \quad (11)$$

where ω determines the persistence of unit activity from one cycle to the next, g_i is a scaling factor, and σ_i^{fire} is a unit-specific firing threshold. Specific parameter values for neuronal units are given in Table 1, and synaptic connections are specified in Table 2.

Synaptic strengths are subject to modification according to a synaptic rule that depends on the pre- and postsynaptic neuronal unit activities. Plastic synaptic connections are either value-independent (see $EC_{IN} \rightarrow DG, CA3, CA1; DG \rightarrow CA3; CA3 \rightarrow CA1$; and $CA1 \rightarrow EC_{OUT}$ in Fig. 2 and Table 2) or value-dependent (see $CA1 \rightarrow S$ and $CA1 \rightarrow M_{HDC}$ in Fig. 2 and Table 2). Both of these rules are based on a modified BCM learning rule (Bienenstock et al., 1982). Synapses between neuronal units with strongly correlated firing rates are potentiated and synapses between neuronal units with weakly correlated firing rates are depressed; the magnitude of change is determined by pre- and postsynaptic activities. The specific parameter settings for fine-scale synaptic connections are given in the equations below and Table 2.

Value-independent synaptic changes in c_{ij} are given by

$$\Delta c_{ij}(t+1) = \eta s_i(t) s_j(t) BCM(s_i) \quad (12)$$

where $s_i(t)$ and $s_j(t)$ are activities of post- and presynaptic units, respectively, and η is a fixed learning rate. The function BCM is implemented as a piecewise linear function, taking postsynaptic activity as input, which is defined by a sliding threshold, θ , two inclinations (k_1, k_2), and a saturation parameter ρ ($\rho = 6$ throughout):

$$BCM(s) = \begin{cases} -k_1 s, & s \leq \theta/2 \\ k_1 (s - \theta), & \theta/2 < s \leq \theta \\ k_2 \tanh[\rho(s - \theta)]/\rho, & \text{otherwise} \end{cases} \quad (13)$$

The threshold is adjusted based on the postsynaptic activity:

$$\Delta \theta = 0.25 (s^2 - \theta) \quad (14)$$

Value-independent plasticity was subject to weight normalization to prevent unbounded potentiation:

$$c_{ij} = \frac{C_{ij}}{\text{sqrt} \left(\sum_{k=1}^K c_{ik}^2 \right)} \quad (15)$$

where c_{ij} is a particular connection and K is the total number of connections onto unit i .

The rule for value-dependent plasticity differs from the value-independent rule in that synaptic change is governed by the presynaptic activity, postsynaptic activity, and temporal difference from the value system. The synaptic change for value-dependent synaptic plasticity is given by:

$$\Delta c_{ij}(t+1) = \eta s_i(t) s_j(t) TD(t) \quad (16)$$

where $TD(t)$ is the temporal difference value at time t [see Eq. (4)].

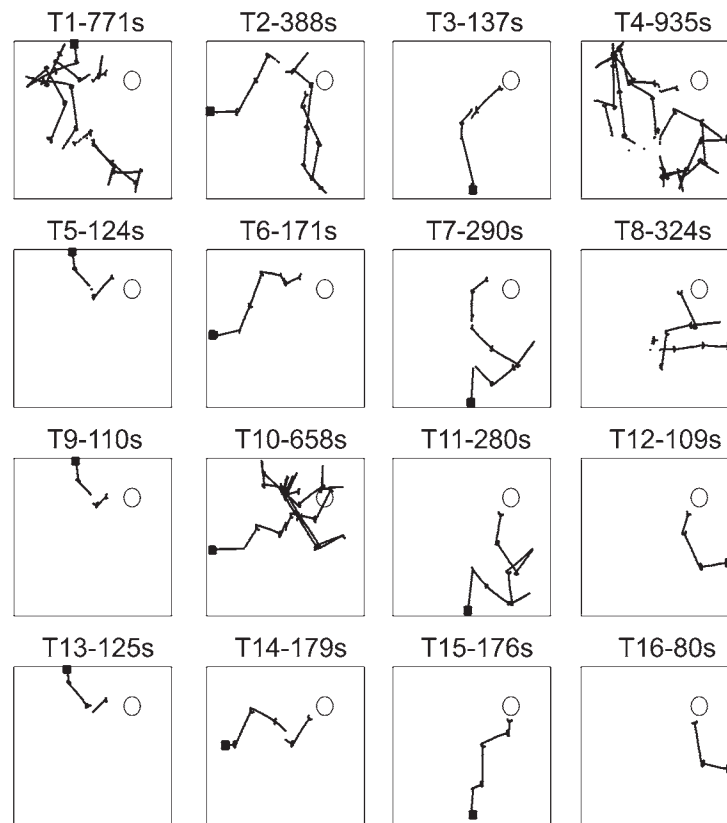


Fig. 3. Trajectories during training for a single subject. Each chart shows Darwin X's trajectory on a single trial. The title above the chart denotes the trial number (T) and the search time in seconds (s). The small black square along the wall signifies the start location and the circle indicates the hidden platform location.

Results

Training and testing in the task was repeated with nine different Darwin X "subjects." Each subject consisted of the same physical device, but each possessed a unique simulated nervous system. This variability among subjects was a consequence of random initialization in the probability of connectivity between individual neuronal units (p in Table 2) and the initial connection strengths between those units [$c_{ij}(0)$ in Table 2]. However, the overall connectivity among neuronal units, which was constrained by the synaptic pathways and arborization patterns, remained similar among different subjects (see Fig. 2 and Table 2 for specifics). Each simulation cycle, the position

of Darwin X, its heading, and the state of all neuronal units were recorded for analysis.

Behavioral Response

The group of Darwin X subjects showed significant improvement in the hidden platform task, as measured by the time to find the hidden platform (search time), as training progressed. The median search times of the last four trials were significantly shorter than the first four trials (223.5 s in trials 13–16 and 532.1 s in trials 1–4, $p < 0.01$, Wilcoxon sign rank test). Figure 3 shows a set of trajectories for a single subject. In general, after the second block (trials 9–16), Darwin X traversed directly to the hidden platform from multiple starting points.

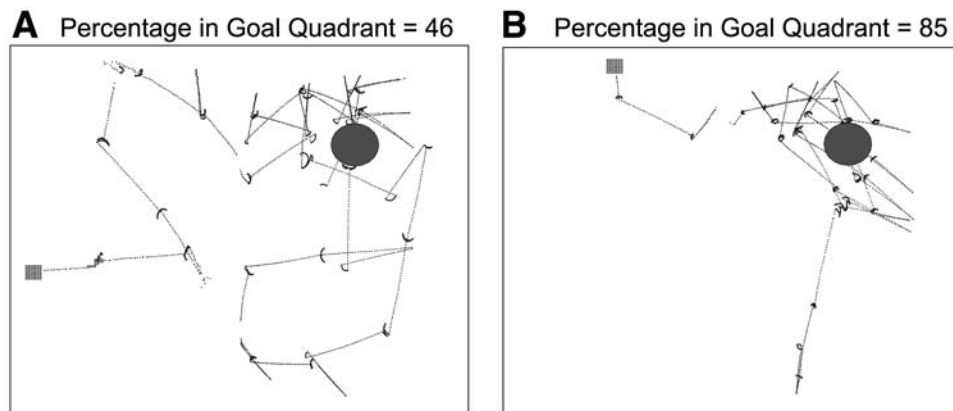


Fig. 4. Trajectories of subjects during the probe trial in which the hidden platform was removed. The gray circle denotes the location of the platform during training trials. Gray square denotes starting location. *Left:* A subject with average performance spent 46% of its time in the quadrant where the hidden platform was during training. *Right:* The subject that showed the best performance during the probe trial spent 85% of the time in the hidden platform quadrant.

During a probe trial, in which the hidden platform was removed and Darwin X explored the environment for 5000 simulation cycles (approx 17 min), the subjects spent a significant amount of time searching in the region where the hidden platform would have been located (*see* Fig. 4). Seven subjects were tested in the probe trial and each spent approximately half their time ($\mu=0.50$, $\sigma=0.23$) in the quadrant of the enclosure that contained the hidden platform (circle in Fig. 4).

Neuronal Response

As previously noted (Krichmar et al., 2005), many of the neuronal units in the hippocampal areas of the simulation showed responses typical of place cells (O'Keefe and Dostrovsky, 1971) where the neuronal unit was active exclusively in a specific region of the environment (*see* Fig. 5).

Metrics used to characterize hippocampal place cells, such as information, sparsity, selectivity, and coherence, were used to measure the spatial response of the simulated hippocampal neuronal units (Kubie et al., 1990; Skaggs et al., 1996). The spatial information is derived by considering a cell as a communication channel

whose input is the rat's location. Sparsity measures the fraction of the explored environment in which the cell was active. Selectivity is equal to the spatial maximum firing rate divided by the mean firing rate of the cell. The more tightly concentrated the cell's activity, the higher the selectivity. Coherence is a nearest-neighbor two-dimensional autocorrelation that measures the local smoothness of a positional firing pattern. The relationship between place metrics and firing characteristics can be seen in Fig. 5. For example, the place neuronal unit in the upper right of Fig. 5, which has relatively high coherence and sparsity, is active smoothly across a large area of the environment. In contrast, the place neuronal unit in the lower right of Fig. 5, which has high information and selectivity with low sparsity, is active in one specific area of the environment.

Table 3 lists the place metrics for simulated CA1 units that were active during the first trial (seven subjects, 1398 neuronal units) and the probe trial (seven subjects, 1635 neuronal units). These metrics were comparable to those reported for CA1 in the rodent (Kubie et al., 1990; Skaggs et al., 1996) and established that neuronal units in Darwin X's CA1 region

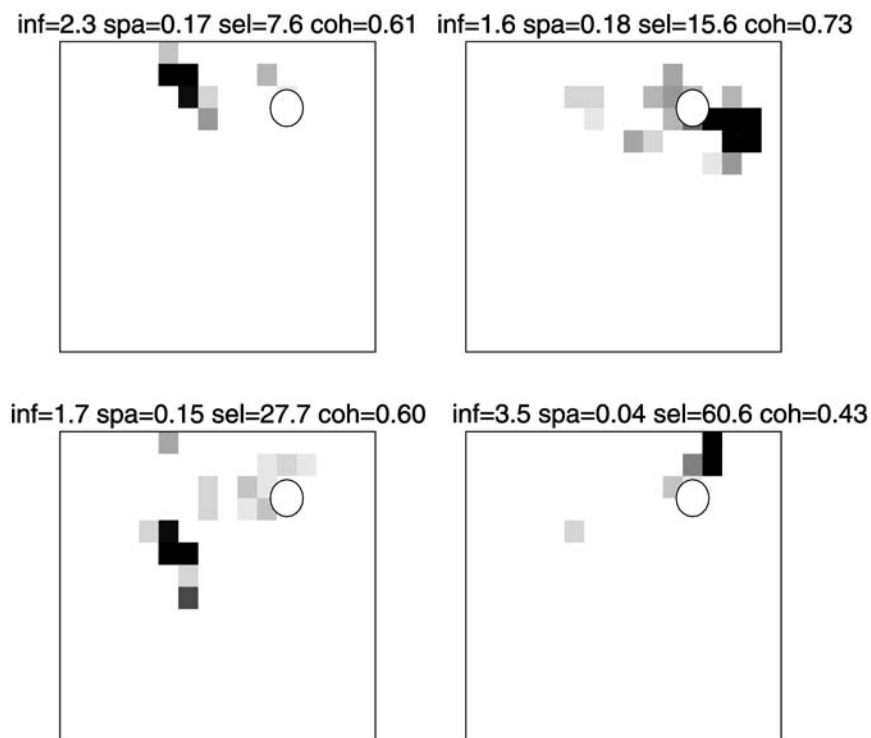


Fig. 5. Four representative place neuronal units recorded in simulated CA1 during a probe trial. Each chart represents Darwin X's enclosure and each pixel represents one square foot in the enclosure. The gray scale represents the activity of a given CA1 neuronal unit and is normalized from quiescent (white) to maximal firing rate (black). The circle denotes the location of the hidden platform. Above each chart are the spatial metrics for the unit: spatial information (inf), sparsity (spa), selectivity (sel), and coherence (coh).

Table 3
Place Metrics for Simulated CA1 Neuronal Units During the First Training Trial and the Probe Trials

	Information			Sparsity			Selectivity			Coherence		
	Mean (std)	Min	Max	Mean (std)	Min	Max	Mean (std)	Min	Max	Mean (std)	Min	Max
Trial 1	1.72 (0.85)	0.13	5.28	0.26 (0.16)	0.01	0.86	18.9 (25.3)	1.7	476.5	0.15 (0.34)	-0.92	0.99
Trial 17	1.77 (0.96)	0.08	6.12	0.26 (0.17)	0.00	0.89	22.1 (30.5)	2.3	637.0	0.09 (0.33)	-0.99	0.99

The mean(standard deviation), minimum, and maximum is given for each metric. Metrics were calculated from all the CA1 neuronal units that were active during the first trial ($n = 1398$ across seven subjects) and the probe trial ($n = 1635$ across seven subjects).

showed spatial activity consistent with its navigation behavior. Note that the place metrics did not differ significantly between the first trial and the probe trial, but more neuronal units in the probe trial showed place activity than in the first trial. The population of CA1

neuronal units demonstrated place activity across the entire environment explored by Darwin X. Some of these units were very specific to a particular place and others were active for much of the environment (e.g., see minimum and maximum sparsity in Table 3).

Knowing the spatial activities of neuronal units in the model is not sufficient for understanding how the hippocampus contributes to successful navigation behavior. It is important to show how the system architecture is related to neural activity and how this correlates with the organism's experience. In the following section, we use a novel time-series analysis to show how the above place metrics may relate to the functional neuroanatomy of the hippocampus and its surrounding regions.

Causality Analysis of Neural Dynamics and Pathways

One way to characterize functional neuroanatomy is to identify the causal interactions that are generated by a neural system during behavior. To identify causal interactions in neuronal dynamics that depend on Darwin X's hippocampal anatomy, we used techniques based on linear regression modeling; specifically, Granger causality (Granger, 1969). Although similar techniques have been used previously to identify causal interactions in neurobiological data (Bernasconi and Konig, 1999; Liang et al., 2000; Kaminski et al., 2001; Brovelli et al., 2004), their application to detailed neural models is novel. Here, we present brief details of the method before applying it to Darwin X; a fuller account of the method itself appears in the work by Seth (in press).

The concept of Granger causality is based on prediction: if a signal A causes a signal B in the Granger sense, then past values of A should contain information that helps predict B, above and beyond the information contained in past values of B alone. Granger causality is usually tested in the context of linear autoregressive models that predict the evolution of a time series or of a set of time series. Consider two time series representing neural activities of two neuronal units X_1 and X_2 . Suppose that the temporal dynamics of these neuronal units can be described by a bivariate autoregressive model:

$$\begin{aligned} X_1(t) &= \sum_{j=1}^p A_{11,j} X_1(t-j) + \sum_{j=1}^p A_{12,j} X_2(t-j) + E_1(t) \\ X_2(t) &= \sum_{j=1}^p A_{21,j} X_1(t-j) + \sum_{j=1}^p A_{22,j} X_2(t-j) + E_2(t) \end{aligned} \quad (17)$$

where p is the maximum number of lagged observations included in the model (the model order), A contains the estimated coefficients of the model, and E_1 and E_2 are the residuals (prediction errors) for each time series. If the variance of the prediction error E_1 (or E_2) is reduced by the inclusion of X_2 (or X_1) terms in the first (or second) equation, then we say that X_2 (or X_1) Granger-causes X_1 (or X_2). In other words, X_2 Granger-causes X_1 if all the coefficients in A_{12} are jointly significantly different from zero. This can be tested by performing an F-test on the null hypothesis that $A_{12} = 0$, given assumptions of covariance stationarity on X_1 and X_2 (i.e., their means and variances are constant over time). The magnitude of a given Granger-causality interaction can be estimated by the logarithm of the corresponding F-statistic. Importantly, the concept of Granger causality can be readily extended to the multivariate case (Boudjellaba et al., 1992), allowing characterization of causal connectivity in a network.

Because we were interested in the differences in neuronal dynamics between exploration of a novel environment and exploration of a familiar environment, we analyzed the causal interactions in neural dynamics generated by a representative Darwin X subject during its first training trial (trial 1) and during the corresponding probe trial (trial 17). This analysis was based on a set of functional circuits identified as follows. For each neuronal unit in CA1, which we refer to as the reference unit, we found the neuronal unit in each remaining area of the hippocampus and entorhinal cortex (CA3, DG, EC_{IN} , EC_{OUT}) that showed the highest covariance in activity with the reference

unit. We call the resulting set of neuronal units a functional circuit.

Granger causality analysis was only applied to a subset of functional circuits. We first excluded circuits for which the reference unit was active on less than 0.5% of the simulation cycles. We then applied first-order differencing to the activity vector of each neuronal unit in each remaining functional circuit, after which circuits were excluded if they did not show covariance-stationarity in the differenced activity of all neuronal units ($p < 0.05$, Dickey–Fuller unit-root test). For the remaining circuits, linear regression models were estimated for the differenced data, using the method of ordinary least-squares, with model order $p=3$. This model order was selected using the Bayesian information criterion (Schwartz, 1978). Each functional circuit was also analyzed using model order $p = 9$, as suggested by the Akaike information criterion (Akaike, 1974), with similar results (data not shown). A final set of exclusions were made if the regression models for each circuit showed (1) correlated residuals ($p < 0.01$; Ljung–Box Q-statistic), or (2) insufficient capture of variance in the data (adjusted $R^2 < 0.3$). Out of a total of 400 CA1 neuronal units, 231 functional circuits satisfied all the above conditions in trial 1, and 162 in trial 17. Significant Granger-causality interactions were calculated, within each of these functional circuits, using an F-test corrected for multiple comparisons ($p < 0.01$, Bonferroni correction).

Figure 6 shows patterns of causal connectivity for a representative functional circuit in the probe trial (trial 17), together with the corresponding place field of the CA1 reference unit. Each arrow represents a statistically significant causal interaction and the width of the arrow and the size of the arrowhead represent the magnitude of the interaction (as determined by the logarithm of the corresponding F-statistic). A gray arrow denotes a unidirectional interaction and a black arrow denotes a bidirectional interaction.

In this case, the pattern shows a reciprocal causal link from EC_{IN} to CA1, which skips the trisynaptic loop by excluding DG from the set of significant causal interactions. It can be seen from the place activity of this CA1 neuronal unit that the unit fired specifically in a small region of the environment resulting in high spatial information and low sparsity, but had two discontinuous place fields resulting in moderate coherence.

Experience affected the response of units in the network as well as the causal interactions between areas. Figure 7 shows patterns of causal connectivity and the corresponding place fields for an example CA1 reference unit in trials 1 and 17. After experience, not only is the place field of the CA1 neuronal unit much sharper, but the web of causal interactions that involve the unit is also much denser. On trial 1, the neuronal unit shows little or no place activity and there is little causal interaction within the hippocampus. By trial 17, the CA1 reference unit shows place activity and its causal pattern has developed into a tri-synaptic loop, from cortex through the hippocampus, and back again to cortex. In the Granger-causality sense, signals from EC_{IN} cause activity in DG, which cause activity in CA3, which cause activity in CA1, which, finally, cause activity in EC_{OUT} . In contrast to the neuronal unit in Fig. 6, the place activity for this CA1 unit showed continuous firing over a large portion of the environment resulting in moderate spatial information, but high coherence and high sparsity. It is important to reiterate that a finding of significant causal influence from one unit to another must be interpreted as a statistical relation, that is, the activity of the latter can be better predicted by including past observations of the former in a multivariate model of the dynamics of all units in the selected functional circuit.

Figures 6 and 7 together suggest that there may be a relationship between place neuronal unit metrics, such as spatial information and sparsity, with causal interaction patterns. To

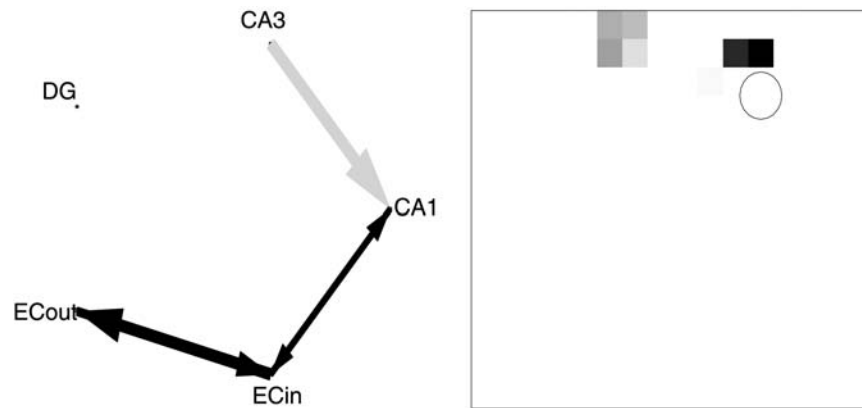


Fig. 6. Representative place neuronal unit and its causal connectivity pattern during a probe trial (trial 17). *Left:* The chart shows patterns of causal connectivity between the hippocampal regions. Each arrow represents a statistically significant causal interaction and the width of the arrow and the size of the arrowhead represent the magnitude of the interaction (logarithm of the corresponding F-statistic). A gray arrow denotes a unidirectional interaction and a black arrow denotes a bidirectional interaction. The causal connectivity pattern is sparse and shows a significant causal interaction from EC_{IN} to $CA1$ that skips the trisynaptic pathway. *Right:* The corresponding $CA1$ place neuronal unit activity had spatial information equal to 2.63, coherence equal to 0.40, sparsity equal to 0.13, and selectivity equal to 12.

better understand this relationship, we calculated the “causal flow” and the “unit causal density” for each $CA1$ reference unit. The unit causal density is the total number of significant incoming and outgoing causal connections involving a given neuronal unit and causal flow is the difference between the number of significant outgoing and incoming causal projections involving a given neuronal unit. Units with highly positive causal flow can be described as causal “sources,” that is, they causally influence network dynamics more than they are themselves influenced. Units with highly negative causal flows can be considered causal “sinks,” that is, they are caused more than they cause (for related terminology, see Kotter and Stephan, 2003; Brovelli et al., 2004). By correlating the place metrics with the causal metrics, we found that information and selectivity were negatively correlated with unit causal density, and that sparsity was positively correlated with unit causal density (see Table 4). In other words, $CA1$ neuronal units that fired very specifically with respect to a given place

would tend to require less integration between neural areas. Moreover, there was a positive correlation between information and causal flow, which suggests that high information neuronal units may serve as a source or drive for other areas.

We also observed that the proportion of trisynaptic causal pathways decreased with experience and the proportion of pathways short-circuiting the trisynaptic loop, via perforant pathways from EC_{IN} to $CA3$ or $CA1$, increased with experience (see Table 5). A possible explanation for this may be that early in Darwin X’s experience more integration (e.g., $EC_{IN} \rightarrow DG \rightarrow CA3 \rightarrow CA1$) is necessary for navigation, but after learning a more direct pathway can be exploited (e.g., $EC_{IN} \rightarrow CA1$).

We extended the above analysis by including sensory input areas ATN , IT , and PR , and identifying causal interactions in the corresponding functional circuits. As before, neuronal units from each area were selected on the basis of high covariance in activity with the $CA1$ reference unit.

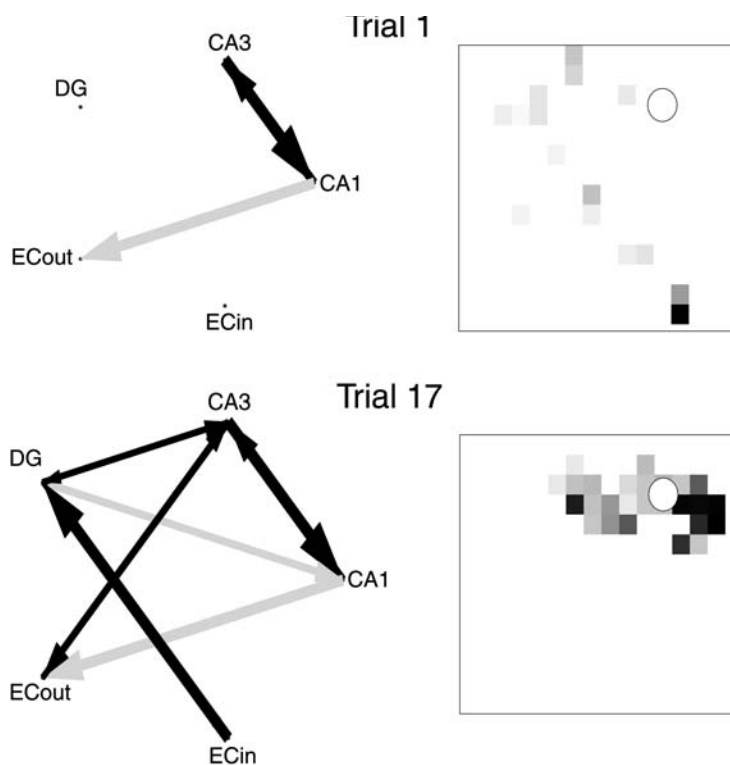


Fig. 7. Representative place neuronal unit and its causal connectivity pattern during the first trial and the probe trial (trial 17). *Top left:* The chart shows patterns of causal connectivity between the hippocampal regions for trial 1. *Top right:* The corresponding place neuronal unit activity for trial 1. *Bottom left:* On trial 17, the pattern of causal connectivity has increased in causal density showing a trisynaptic loop. *Bottom right:* The corresponding place neuronal unit activity for trial 17 (spatial information equal to 0.68, coherence equal to 0.59, sparsity equal to 0.50, and selectivity equal to 5).

Table 4
Correlation Between Place and Causal Metrics for CA1 Reference Units During a Probe Trial ($n = 155$)

	Causal Flow	Unit Causal Density
Information	0.166 ^a	-0.476 ^b
Coherence	-0.037	0.128
Sparsity	-0.124	0.501 ^b
Selectivity	0.019	-0.310 ^b

^a $p < 0.05$.
^b $p < 0.0001$.

Figure 8 shows causal flow profiles for causal connectivity graphs from trials 1 (black) and 17 (gray). In both trials, neuronal units in CA3

Table 5
Change in the Percentages of Causal Pathways with Experience

	Percentage of Causal Pathways	
	Trial 1 ($n = 231$)	Trial 17 ($n = 162$)
Trisynaptic	42.0	29.0
Short-circuit	14.8	21.6
Both	21.6	19.1
Neither	21.6	30.3

The proportion of trisynaptic pathways decreased from the first trial to the probe trial, and the proportion of pathways that short-circuited the trisynaptic loop with significant causal connections from EC_{IN} to CA3 or CA1 increased with experience. These changes were significant ($p < 0.0001$) using a chi-square goodness-of-fit test.

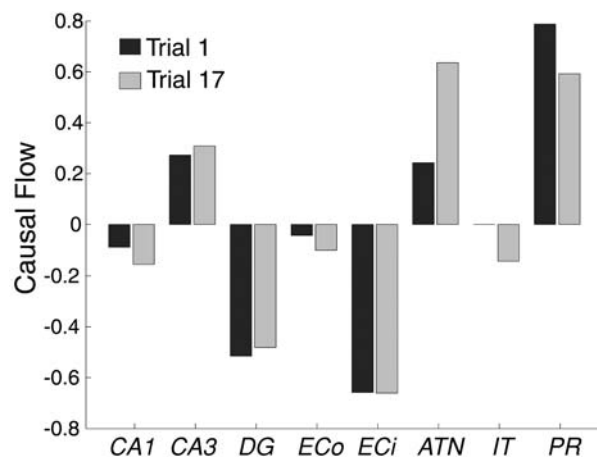


Fig. 8. Causal flow in the different regions of the neural simulation, based on functional circuits derived by using *CA1* neuronal units as reference units. *CA3*, *ATN*, and *PR* have significant positive causal flow, whereas *DG* and *EC_{IN}* have significant negative causal flow ($p < 0.01$ using a Wilcoxon sign rank test). Comparing causal flow from trial 1 with trial 17, region *ATN*, which responds to head direction, became a stronger causal source with experience ($p < 0.01$ using a Wilcoxon rank sum test).

and in certain cortical areas (*ATN*, *PR*) are causal sources (i.e., they drive circuit dynamics more than they are driven). Areas *DG* and *EC_{IN}* contain causal sinks, and neuronal units in *IT*, *EC_{OUT}*, and *CA1* are balanced (i.e., they cause as much as they are caused). It is striking that cortical areas *ATN* and *PR* have a considerably greater influence on circuit dynamics than *IT*, despite starting out with well-balanced synaptic input from each of these areas to *EC_{IN}*. This observation predicts that, *in this task*, head-direction signals and the “where” visual pathway from the simulated parietal region may have greater functional significance for spatial memory than the “what” visual signals filtered through inferotemporal region. Figure 8 also shows that causal flow profiles change as a result of experience. In particular, by the probe trial, the causal influence of *ATN* on functional circuit dynamics has grown markedly.

Discussion

The BBD methodology in this study was critical for investigation of the bases for episodic and spatial memory. Unlike other computational models of the hippocampus, Darwin X was embodied in the environment and made minimal assumptions about the environment inputs into the model or the behavioral actions required to solve a spatial memory task (Redish et al., 1993; Shapiro and Hetherington, 1993; Blum and Abbott, 1996; Recce and Harris, 1996; Samsonovich and McNaughton, 1997; Foster et al., 2000). No prior assumptions with respect to spatial tuning or underlying functions were built into the neuronal units of the simulated hippocampus regions. Moreover, unlike robotic systems, which abstract features of the hippocampus to build spatial memories (Mataric, 1991; Bachelder and Waxman, 1995; Burgess et al., 1997; Arleo and Gerstner, 2000; Gaussier et al., 2002; Milford et al., 2004), Darwin X had neural dynamics coupled with detailed neuroanatomy at the systems neuroscience level.

Darwin X takes into consideration the macro- and microanatomy linking the hippocampus with the cortex, as well as the anatomy within the hippocampus itself. Darwin X developed mappings from the neural responses in the hippocampus leading to purposeful behavior (see Figs. 3 and 4). It developed place neuronal units, comparable to place cells seen in the rodent, as a result of autonomous exploration and the integration of three input streams (“visual what” or *IT*, “visual where” or *PR*, and “self-movement” or *ATN*) over time (see Fig. 5 and Table 3).

We proposed techniques based on linear autoregressive modeling and “Granger causality” to identify causal interactions in neural dynamics generated by Darwin X during behavior. Unlike alternative methods for determining causality (Pearl, 1999; Tononi and Sporns, 2003; Keinan et al., 2004), the Granger approach does not require perturbation or lesioning of the neural system, and hence is

well suited to the analysis of “normal” neural dynamics generated by intact nervous systems during behavior. Lesion approaches do not generally allow inferences of causal interactions in intact systems since the nature of an impairment following a lesion does not necessarily correspond to the functional role of the lesioned element in the intact system (Farah, 1994).

It is important to emphasize again that Granger’s concept of causality is statistical rather than physical. If a system is only partially observed, there are several situations in which Granger causality may not correspond to physical causal chains. For example, A may enhance the predictability of B if both A and B are driven by a common (unobserved) input C, or if A physically causes an intermediate process D, which itself physically causes B. Such sensitivity to hidden variables is true in general for dynamic analyses and does not impact the method’s usefulness in showing directed dynamical interactions in a system without external perturbations or lesions.

Causal connectivity analysis of DarwinX generated novel insights of how the hippocampal region may shift its functional anatomy based on experience and context. In both the backtrace study (Krichmar et al., 2005) and the present causal analysis, we were able to identify functional analogs of both trisynaptic and perforant pathways in neural circuits spanning the hippocampus and entorhinal cortex. In the present study, we showed that the proportion of functional trisynaptic pathways decreased with experience. However, pathways that circumvented the trisynaptic loop via perforant pathways from EC to CA1 or CA3 became increasingly prevalent as a result of experience (*see* Table 5). One explanation for this phenomenon is that the trisynaptic pathway, which integrates more areas of the hippocampus, may be necessary when the organism is in unfamiliar environments. Moreover, because of the multiple intrinsic and extrinsic hippocampal loops, highly processed sensory input obtained at different times would

be integrated during the retrieval of a memory. In contrast, the more direct path from cortical information to CA1 and then motor response may be prevalent when the organism is familiar with its surroundings and is more “confident” in its motor actions, and high levels of integration may not be necessary for the retrieval of memories. Interestingly, this experiential shift was not seen in the backtrace analysis and this difference may be owing to the limited number of cycles examined.

Recent experimental and modeling results are consistent with the notion that there is a direct entorhinal to CA1 pathway that supports spatial memory (Brun et al., 2002; Hasselmo et al., 2002a). By isolating CA1, Brun and colleagues were able to show that this pathway, which short-circuited dentate gyrus and CA3, was sufficient for the acquisition of spatial memory and recognition memory. Based on experimental and simulation studies, Hasselmo and colleagues have proposed a model in which direct connections from entorhinal cortex to CA3 and CA1 are important for encoding memories. Moreover, Treves has hypothesized that the function of the CA3 region is to complete patterns and the function of the CA1 region is to predict the next pattern in a sequence (Treves, 2004). He suggests that differentiation between hippocampal subfield connectivity and firing rate adaptation affects the balance between prediction and retrieval. Although our model does not explicitly separate acquisition, recall, and recognition, it does suggest that this direct pathway from entorhinal cortex to CA1 may be exploited when converting familiar cues into motor actions.

Another interesting prediction is the relation between place neuronal unit firing and functional pathways. Highly coherent place neuronal units tended to have higher unit causal density and be more integrative, whereas high spatial information units tended to have lower unit causal density and be a part of more direct pathways (*see* Figs. 6 and 7, and Table 4). Sparsity, and coherence to some degree, go

together with integration; a high sparsity neuronal unit should respond to many combinations of inputs that relate to similar places in an environment; such integration is consistent with high unit causal density in that the neuronal unit can be affected by (and affect) units in multiple regions. Alternatively, a high information or highly selective neuronal unit is likely to have low unit causal density since it should only fire, given highly specific input statistics. Interestingly, the backtrace method identified a trend between functional anatomical pathways and place field metrics that contrasts the correlation between place metrics and causal density shown here (compare Table 1 in Krichmar et al., 2005 with Table 4). *CA1* place fields that had high information and selectivity tended to involve *CA3* (i.e., $EC_{IN} \rightarrow CA3 \rightarrow CA1$ and $DG \rightarrow CA3$), whereas low information and low selectivity *CA1* place fields tended to be driven by the entorhinal cortex ($EC_{IN} \rightarrow CA1$).

By extending our analysis to incorporate neuronal units from cortical areas *ATN*, *IT*, and *PR*, we were able to show differential causal roles of these sensory streams on the hippocampus and its corresponding functional circuits. Specifically, cortical neuronal units in which self-movement (HD) or visuospatial information (PR) acted as causal sources in the functional circuit had a stronger influence on the circuit than object recognition (IT) neuronal units (see Fig. 8). Moreover, the influence of the head direction system grew stronger as a result of experience. This analysis showed that *DG* is a causal sink that is influenced by other areas, *CA3* is a causal source that influences other areas, and *CA1* shows a balance of causal influences suggesting different functional roles for these subregions of the hippocampus.

The incorporation of substantial neuroanatomical detail into Darwin X offers new opportunities for the analysis of simulation models that may have implications for systems neuroscience. The present Granger causality method,

and the backtrace method (Krichmar et al., 2005) provide two such analysis tools. In comparing the findings from each analysis, the following distinctions for each analysis should be borne in mind. The backtrace method takes into account knowledge of the anatomy, and examines a specific time period. Tracing back for more than several simulation cycles leads to an explosion in network size that rapidly becomes computationally infeasible. In contrast, the Granger causality method does not rely on knowledge of the anatomy, but instead can identify causal pathways between any neuronal units, based on their dynamics. Also, the Granger causality method takes into account (and indeed requires) an extended series of data points covering many simulation cycles. Its inferences therefore relate to extended periods of behavior, and not to specific time points. Finally, whereas the backtrace is physical by nature, the Granger method is statistical. For these reasons, the two analysis methods can be seen as complementary, and both may be generally applicable in the examination of nervous systems.

Although this modeling and analysis approach is at an early stage, it allows for the investigation of the multiple loops from cortex to hippocampus and within the hippocampus proper. We believe this functional architecture is crucial for integrating inputs over time and building spatial memory that depends on context. These context-dependent responses are critical for the acquisition and recall of sequences of multimodal memories and are the hallmark of episodic memory, a subject that we are now in a position to explore.

Acknowledgments

This work was supported by the Neurosciences Research Foundation, W. M. Keck Foundation, the Defense Advance Research Projects Agency (DARPA), and the Office of Naval Research. We thank J. A. Snook, D. Moore, and D. Hutson for their contribution to the design of Darwin X.

References

- Akaike, H. (1974) A new look at the statistical model identification. *IEEE Trans. Automat. Control* 19, 716–723.
- Almassy, N., Edelman, G. M., and Sporns, O. (1998) Behavioral constraints in the development of neuronal properties: a cortical model embedded in a real-world device. *Cereb. Cortex* 8, 346–361.
- Amaral, D. G., Ishizuka, N., and Claiborne, B. (1990) Neurons, numbers and the hippocampal network. In: *Progress in Brain Research*. Ottersen, O. P. (ed.) Elsevier Science, Amsterdam, pp. 1–11.
- Arleo, A. and Gerstner, W. (2000) Modeling rodent head-direction cells and place cells for spatial learning in bio-mimetic robotics. Paper presented at: From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior, MIT Press, Paris, France.
- Aston-Jones, G. and Bloom, F. E. (1981) Non-repinephrine-containing locus coeruleus neurons in behaving rats exhibit pronounced responses to non-noxious environmental stimuli. *J. Neurosci.* 1, 887–900.
- Bachelder, I. A. and Waxman, A. M. (1995) A view-based neurocomputational system for relational map-making and navigation in visual environments. *Auton. Syst.* 16, 267–289.
- Battaglia, F. P., Sutherland, G. R., and McNaughton, B. L. (2004) Local sensory cues and place cell directionality: Additional evidence of prospective coding in the hippocampus. *J. Neurosci.* 24, 4541–4550.
- Bernard, C. and Wheal, H. V. (1994) Model of local connectivity patterns in CA3 and CA1 areas of the hippocampus. *Hippocampus* 4, 497–529.
- Bernasconi, C. and Konig, P. (1999) On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. *Biol. Cybern.* 81, 199–210.
- 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.
- Blum, K. I. and Abbott, L. F. (1996) A model of spatial map formation in the hippocampus of the rat. *Neural Comput.* 8, 85–93.
- Boudjellaba, B., Dufour, J.-M., and Roy, R. (1992) Testing causality between two vectors in multivariate ARMA models. *J. Am. Stat. Assoc.* 87, 1082–1090.
- Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., and Bressler, S. L. (2004) Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *Proc. Natl. Acad. Sci. USA* 101, 9849–9854.
- Brun, V. H., Otnass, M. K., Molden, S., et al. (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 Neural Simulation of the Hippocampus and Rat Navigation. *Biol. Sci.* 352, 1535–1543.
- Chavarriga, R., Strösslin, T., Sheynikhovich, D., and Gerstner, W. (2005) A computational model of parallel navigation systems in rodents. *Neuroinformatics* 3(3), in press.
- Edelman, G. M. (1987) *Neural Darwinism: The Theory of Neuronal Group Selection*. Basic Books, Inc., New York.
- 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. USA* 89, 7267–7271.
- Farah, M. (1994) Neuropsychological inference with an interactive brain: A critique of the locality assumption. *Behav. Brain Sci.* 17, 43–61.
- Ferbinteanu, J. and Shapiro, M. L. (2003) Prospective and retrospective memory coding in the hippocampus. *Neuron* 40, 1227–1239.
- Floreano, D. and Mondada, F. (1998) Evolutionary neurocontrollers for autonomous mobile robots. *Neural Netw.* 11, 1461–1478.
- Foster, D. J., Morris, R. G., and Dayan, P. (2000) A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* 10, 1–16.
- Frank, L. M., Brown, E. N., and Wilson, M. (2000) Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* 27, 169–178.
- Freund, T. F. and Buzsáki, G. (1996) Interneurons of the hippocampus. *Hippocampus* 6, 347–470.
- Gaussier, P., Revel, A., Banquet, J. P., and Babeau, V. (2002) From view cells and place cells to cognitive map learning: processing stages of the hippocampal system. *Biol. Cybern.* 86, 15–28.
- Granger, C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.
- Griffiths, D., Dickinson, A., and Clayton, N. (1999) Episodic memory: what can animals remember about their past? *Trends Cogn. Sci.* 3, 74–80.
- Grossberg, S. (1999) The link between brain learning, attention, and consciousness. *Conscious Cogn.* 8, 1–44.

- Hasselmo, M. E., Bodelon, C., and Wyble, B. P. (2002a). A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Comput.* 14, 793–817.
- Hasselmo, M. E., Hay, J., Ilyn, M., and Gorchetchnikov, A. (2002b). Neuromodulation, theta rhythm and rat spatial navigation. *Neural Netw.* 15, 689–707.
- Ijspeert, A. J., Crespi, A., and Cabelguen, J.-M. (2005) Towards a salamander robot: applying neurobiological principles to the control of locomotion in robots. *Neuroinformatics* 3(3), in press.
- Kaminski, M., Ding, M., Truccolo, W. A., and Bressler, S. L. (2001) Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance. *Biol. Cybern.* 85, 145–157.
- Keinan, A., Sandbank, B., Hilgetag, C. C., Meilijson, I., and Ruppin, E. (2004) Fair attribution of functional contribution in artificial and biological networks. *Neural Comput.* 16, 1887–1915.
- Kotter, R. and Stephan, K. E. (2003) Network participation indices: characterizing component roles for information processing in neural networks. *Neural Netw.* 16, 1261–1275.
- 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., Nitz, D. A., Gally, J. A., and Edelman, G. M. (2005) Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task. *Proc. Natl. Acad. Sci. USA* 102, 2111–2116.
- Kubie, J. L., Muller, R. U., and Bostock, E. (1990) Spatial firing properties of hippocampal theta cells. *J. Neurosci.* 10, 1110–1123.
- Lavenex, P. and Amaral, D. G. (2000) Hippocampal-neocortical interaction: a hierarchy of associativity. *Hippocampus* 10, 420–430.
- Liang, H., Ding, M., Nakamura, R., and Bressler, S. L. (2000) Causal influences in primate cerebral cortex during visual pattern discrimination. *Neuroreport* 11, 2875–2880.
- Mataric, M. J. (1991) Navigating with a rat brain: A neurobiologically-inspired model for robot spatial representation. In: *From Animals to Animats*. Meyer, J. A. and Wilson, S. W. (eds.) MIT Press, Cambridge, MA, pp. 169–175.
- Milford, M. J., Wyeth, G. F., and Prasser, D. (2004) RatSLAM: A Hippocampal Model for Simultaneous Localization and Mapping. Paper presented at: Proceedings of the 2004 IEEE International Conference on Robotics & Automation, New Orleans, LA.
- Mogenson, G. and Nielsen, M. (1984) Neuropharmacological evidence to suggest that the nucleus accumbens and subplidial region contributes to exploratory locomotion. *Behav. Neural Biol.* 42, 52–60.
- 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.
- Muller, R. U., Ranck, J. B., Jr., and Taube, J. S. (1996) Head direction cells: properties and functional significance. *Curr. Opin. Neurobiol.* 6, 196–206.
- 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.
- Pearl, J. (1999) *Causality: Models, reasoning, and inference*. Cambridge University Press, Cambridge.
- Pfeifer, R. and Scheier, C. (1997) Sensory-motor coordination: The metaphor and beyond. *Robot. Auton. Syst.* 20, 157–178.
- Recce, M. and Harris, K. D. (1996) Memory for places: a navigational model in support of Marr’s theory of hippocampal function. *Hippocampus* 6, 735–748.
- Redish, A. D., Touretzky, D., and Wan, H. S. (1993) Neural representation of space using sinusoidal arrays. *Neural Comput.* 5, 869–884.
- Reeke, G. N., Sporns, O., and Edelman, G. M. (1990) Synthetic neural modeling: The “Darwin” series of recognition automata. *Proc. IEEE* 78, 1498–1530.
- Samsonovich, A. and McNaughton, B. L. (1997) Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.* 17, 5900–5920.
- Schultz, W., Dayan, P., and Montague, P. R. (1997) A neural substrate of prediction and reward. *Science* 275, 1593–1599.
- Schwartz, G. (1978) Estimating the dimension of a model. *Ann. Stat.* 5, 461–464.
- Scoville, W. B. and Milner, B. (1957) Loss of recent memory after bilateral hippocampal lesions. *J. Neurochem.* 20, 11–21.

- Seth, A. K. Causal connectivity analysis of evolved neural networks during behavior. *Netw.: Comput. Neural Syst.*, in press.
- Seth, A. K., McKinsty, J. L., Edelman, G. M., and Krichmar, J. L. (2004a). Spatiotemporal processing of whisker input supports texture discrimination by a brain-based device. In: *Animals to Animats 8: Proceedings of the Eighth International Conference on the Simulation of Adaptive Behavior*. Meyer, J. A. (ed.) MIT Press, Cambridge, MA, pp. 130–139.
- Seth, A. K., McKinsty, J. L., Edelman, G. M., and Krichmar, J. L. (2004b). Visual Binding Through Reentrant Connectivity and Dynamic Synchronization in a Brain-based Device. *Cereb. Cortex* 14, 1185–1199.
- Shapiro, M. L. and Hetherington, P. A. (1993) A Simple Network Simulates Hippocampal Place Fields: Parametric Analyses and Physiological Predictions. *Behav. Neurosci.* 107, 34–50.
- Skaggs, W. E., McNaughton, B. L., Wilson, M. A., and Barnes, C. A. (1996) Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6, 149–172.
- Sporns, O. and Alexander, W. H. (2002) Neuro-modulation and plasticity in an autonomous robot. *Neural Netw.* 15, 761–774.
- Sporns, O., Almassy, N., and Edelman, G. M. (2000) Plasticity in value systems and its role in adaptive behavior. *Adapt. Behav.* 8, 129–148.
- Stewart, M. and Fox, S. E. (1990) Do septal neurons pace the hippocampal theta rhythm? *Trends Neurosci.* 13, 163–168.
- Sutton, R. S. and Barto, A. G. (1990) Time-derivative models of pavlovian reinforcement. In: *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, Moore, J. (ed.) MIT Press, Cambridge, MA, pp. 497–537.
- Taube, J. S. (1998) Head direction cells and the neurophysiological basis for a sense of direction. *Prog. Neurobiol.* 55, 225–256.
- Thierry, A. M., Gioanni, Y., Degenetais, E., and Glowinski, J. (2000) Hippocampo-prefrontal cortex pathway: anatomical and electrophysiological characteristics. *Hippocampus* 10, 411–419.
- Tononi, G. and Sporns, O. (2003) Measuring information integration. *BMC Neurosci.* 4, 31.
- Treves, A. (2004) Learning to predict through adaptation. *Neuroinformatics* 2, 361–366.
- Treves, A. and Rolls, E. T. (1994) Computational analysis of the role of the hippocampus in memory. *Hippocampus* 4, 374–391.
- Tulving, E. (1972) Episodic and semantic memory. In: *Organisation of Memory*. Donaldson, W. (ed.) Academic Press, New York, pp. 381–403.
- Uchibe, E. and Doya, K. (2004) Competitive-Cooperative-Concurrent Reinforcement Learning with Importance Sampling. In: *Animals to Animats 8: Proceedings of the Eighth International Conference on the Simulation of Adaptive Behavior*. Schaal, S., Ijspeert, A., Billard, A., Vijayakumar, S., Hallam, J., and Meyer, J. A. (eds.) MIT Press, Cambridge, MA.
- Ungerleider, L. and Mishkin, M. (1982) Two cortical visual systems. In: *Analysis of Visual Behavior*, R. Mansfield (ed.) MIT Press, Cambridge, MA, pp. 549–586.
- Ungerleider, L. G., and Haxby, J. V. (1994) ‘What’ and ‘where’ in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165.
- Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., and Mishkin, M. (1997) Differential effects of early hippocampal pathology on episodic and semantic memory. *Science* 277, 376–380.
- Weng, J. (2004) Developmental robots: Theory and experiments. *Int. J. Humanoid Robot.* 1, 199–236.
- Witter, M. P., Naber, P. A., van Haefen, T., et al. (2000a). Cortico-hippocampal communication by way of parallel parahippocampal-subicular pathways. *Hippocampus* 10, 398–410.
- Witter, M. P., Wouterlood, F. G., Naber, P. A., and Van Haefen, T. (2000b). Anatomical organization of the parahippocampal-hippocampal network. *Ann. NY Acad. Sci.* 911, 1–24.
- Wray, J. and Edelman, G. M. (1996) A model of color vision based on cortical reentry. *Cereb. Cortex* 6, 701–716.

