

# Removing Epistemological Bias From Empirical Observation of Neural Models.

Ronan Waldron\*

May 25, 1994

## Abstract

This paper addresses the application of neural network research to a theory of autonomous systems. Neural networks, while enjoying considerable success in autonomous systems applications, have failed to provide a firm *theoretical* underpinning to neural systems embedded in their natural ecological context. This paper proposes a stochastic formulation of such an embedding. A neural system derived from the cell membrane equation is shown to exhibit a stochastic dynamic which tracks an environmental process. The activity of a node is interpreted in the context of this external stochastic process, in the light of its interdependence, which is now of statistical formulation, on the nodes to which it projects.

## 1 Introduction

This paper addresses a theoretical formulation of embedded autonomous systems. The principal concern is to extend the utility of neural models to a *theoretical* science of autonomous systems. The central problem that this paper addresses is the difficulty in making impartial empirical observations of artificial systems. Without a theoretical formulation of the empirical context, any observations are subject to empirical bias, a projection of the observer's epistemology on to the network activation or system behaviour.

This research describes a model admitting of the observation of autonomous systems in an empirical context, and the assessment of the significance of the observations with respect to the underlying logic of the physical system constituting the autonomous system *and* its environment.

This paper proposes a stochastic formulation of such an embedding. Statistical methods have been applied to the study of the dynamics of Bidirectional Associative Memory networks [Figueroa 1990], but not to the modelling of network activation embedded in the environmental context. The neural system in an environment exhibits a stochastic dynamic which

tracks an environmental process. The network activity is interpreted as a random vector in this external stochastic process.

## 2 Neural Models and Systems Science

A metric of a mature theory is that it can make 'counter-intuitive' predictions that are empirically verifiable. Neural networks have failed to provide a firm *theoretical* underpinning to a neural system embedded in an ecological context, necessary for mature predictive success. Significant research (e.g. [Brooks 1990]) has been done in neural applications to autonomous systems, which provide practical engineering solutions to the problems encountered by mobile robotic systems. However, these systems are still open to intuitive bias, as the network behaviour is observed from the observer's reference frame.

System theoretic approaches (e.g [Rosen 1985]) have yet to reach the maturity for successful implementations of robotic systems. Neurobiological models have the greatest empirical success as a consequence of the strength of their mathematical formulation, and the closeness of their application area to physical systems. In Section 2.1 an example of a neurobiological model that explains an optical illusion is presented.

### 2.1 Neurobiological Models of Neural Systems

Neurobiological or electrophysiological systems provide very complex mathematical models of electrical and chemical activation levels in the nervous systems of mammals. These models are theoretically robust, as well as being empirically well founded. With the aid of complex measuring equipment, the theoretical predictions required of any scientific theory are empirically verifiable.

The 'barber pole' illusion (see Figure 1), is a classic example of a counter-intuitive empirical phe-

---

\*This research was partially funded by Hitachi Dublin Laboratory

denoted by the white arrows, and the local stimuli to the retina are all oriented laterally, denoted by the black arrows.

The physiological model of this phenomenon, the Boundary Contour System (BCS) of [Mingolla 1985], presents a theory of structure within the human visual cortex whose internal logic supports the counter-intuitive perception the natural system exhibits. The network dynamics exhibit the global behaviour as a result of local interactions, and the BCS describes mathematically the interconnections that support this dynamics.



Figure 1: The Barber Pole Illusion.

The theoretical formulation allows empirical statements to be made that are consistent with the internal logic of the nervous system, and thus the ‘illusion’ is explained. The maturity of the neurobiological model is exemplified by this level of empirical success.

The internal logic that autonomous systems obey is the necessary backdrop for any empirical theses about such systems. However practical applications of robotics control systems to the problem of autonomy have yet to reach this level of maturity. The work of [Braitenberg 1984] presents systems that have the appearance of autonomous behaviours, but no theoretical formulation that precludes *intuitive* bias. The work exemplifies precisely how easy it is to intuit too much onto an autonomous behaviour.

Subsumption architectures (e.g. [Brooks 1990]) have proved possibly the most successful autonomous systems solutions to date. They do not however provide a great degree of insight into the operation of the human nervous system, as they address systemic behaviour, rather than neurophysiological activation. The object of this research is to provide a mathematical relationship between the neural activation, and the environment within which the activation takes place, to help narrow the gap between biological and artificial autonomous systems.

### 3 The Stochastic Network

A generic network architecture based on the ‘cell membrane’ equation is considered in this paper. The

of Plonsey and Fleming [Plonsey 1969], amongst others, is a simple model of the electrical activation of a neuron under stimulation. It has been used as the fundamental building block in many network architectures notably those in the series of Adaptive Resonance Theory (ART) architectures [Grossberg 1976, Carpenter 1990].

#### 3.1 The Cell Membrane Equation

A mathematical formulation of the cell membrane equation is given in Equation 1. Nodal activation as a function of time is represented by  $x_i(t)$ . Letting  $\mathcal{N}_{E_i}$  denote the net excitatory inputs, and  $\mathcal{N}_{I_i}$  the net inhibitory inputs, we have

$$\frac{dx_i(t)}{dt} = -Ax_i(t) + (B - x_i(t))\mathcal{N}_{E_i} - (x_i(t) + C)\mathcal{N}_{I_i} \quad (1)$$

If  $x_i(t)$  is greater than  $B$ , all three terms in  $\frac{dx_i(t)}{dt}$  are negative, and so  $x_i$  decreases. Thus it is bounded above by  $B$ . Similarly, if  $x_i(t)$  is less than  $-C$ , then all three terms are positive, and  $x_i$  increases. This implies that the nodal activity is bounded in a discrete range  $[-C, B]$ .

Typical excitatory and inhibitory inputs are given in Equations 2 and 3.  $\mathcal{J}_i$  denotes the set of nodes to which  $x_i$  is connected, with  $z_{ji}(t)$ ,  $j \in \mathcal{J}_i$ , denoting the excitatory synaptic inter-connectivity.  $\mathcal{K}_i$  denotes inhibitory projections, and  $z_{ki}$ ,  $k \in \mathcal{K}_i$ , the inhibitory synaptic connections.

$$\mathcal{N}_{E_i}(t) = \sum_{j \in \mathcal{J}} B_j z_{ji}(t) + I_i, \quad (2)$$

$$\mathcal{N}_{I_i}(t) = - \sum_{k \in \mathcal{K}} C_k z_{ki}(t) \quad (3)$$

Numerous network architectures can be constructed from these equations. The different choices of the net excitatory and inhibitory inputs, together with the overall connectivity, govern the network behaviour.

#### 3.2 Network Architecture

The construction of a simply network, with exponential decay, zero inhibition and simple excitation, following from this equation is trivial. We consider the following architecture, where  $J$  denotes the set of nodes in the network, and  $x_i$ ,  $i \in J$ , denote the nodal activations.

$$\frac{dx_i(t)}{dt} = -Ax_i(t) + (B - x_i(t))\mathcal{N}_{E_i}, \quad (4)$$

Let the synaptic connections be governed by the following equation, where the time index  $t$  is not necessarily the same as in Equation 4.

$$\frac{dz_{ji}(t)}{dt} = (-z_{ji} + \delta_{ji} \sum_{j \in \mathcal{J}} z_{ji}) f(x_i(t)) \quad (5)$$

$$\text{Where } \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

This describes a simple dynamics for the synaptic connections, with exponential decay for each node, and gated excitation on  $z_{ii}(t)$ . The function  $f(x_i(t))$  that describes the gating is arbitrary, providing it is constant with respect to the time scale of the synaptic dynamics.

### 3.3 Probability Measure

Synaptic activity will be interpreted as a probability measure on the measurable space composed of the network nodes  $J$ , and its power set,  $\mathcal{J} = 2^J$ , representing all possible projections. The measurable space  $(J, \mathcal{J})$ , will be presented as the state space of the stochastic process in the environment, allowing unbiased interpretations of network activation.

The definition of a probability measure is given in Definition 3.1.

**Definition 3.1** *Given a measure  $z : \mathcal{J} \rightarrow [0, 1]$  on a measurable space  $(J, \mathcal{J})$ , and a disjoint partition  $\mathcal{A}_{\mathcal{J}} \equiv \{\mathcal{J}_i \mid i \in \mathcal{I}\}$  of  $J$ , then  $z$  is a probability measure if and only if*

$$\int_{\mathcal{I}} z(\mathcal{J}_i) d\mathcal{A}_{\mathcal{J}} = 1, \quad (6)$$

for all partitions  $\mathcal{A}$ .

The partition  $\mathcal{A}$  represents the set of subsets  $\mathcal{J}_i$  of  $J$  that describes the projections of the network nodes  $x_i$ , and so the possible network architectures.

### 3.4 Theorem Statement

The following theorem proves that this network architecture supports the interpretation of synaptic interconnectivity as a probability measure on the space defined by the connectivity, and hence of network activity as a stochastic kernel.

bility measure on  $\mathcal{J}$  for all  $t \leq t_0$ ,  $\kappa > 0$ , and

$$\left. \frac{d}{dt} z_t(\mathcal{J}_i) \right|_{t_0} = \kappa z_{t_0}((\mathcal{J}_i)'), \quad (7)$$

$$\left. \frac{d}{dt} z_t(\mathcal{J}_j) \right|_{t_0} = -\kappa z_{t_0}(\mathcal{J}_j), \quad (8)$$

where  $\mathcal{J}_i$  is an event in  $(J, \mathcal{J})$  and  $\{\mathcal{J}_j \mid j \in \mathcal{I}^i\}$  form a partition  $\mathcal{W}_{(\mathcal{J}_i)'}$  of the complement of  $\mathcal{J}_i$ , then  $z_t$  remains a probability measure for all  $t \geq t_0$ .

We require to show that

$$\int_{\mathcal{I}} z_{t_1}(\mathcal{J}_i) d\mathcal{A}_{\mathcal{J}} = 1$$

given

$$\int_{\mathcal{I}} z_{t_0}(\mathcal{J}_i) d\mathcal{A}_{\mathcal{J}} = 1.$$

**Proof:** We begin by solving Equation 7 and Equation 8 for  $z_t$  as a function of time. By simple integration, noting that  $z_t((\mathcal{J}_i)') = 1 - z_t(\mathcal{J}_i)$ , we get

$$z_{t_1}(\mathcal{J}_i) = 1 - (1 - z_{t_0}(\mathcal{J}_i)) e^{-\kappa(t_1 - t_0)}, \quad (9)$$

$$z_{t_1}(\mathcal{J}_j) = z_{t_0}(\mathcal{J}_j) e^{-\kappa(t_1 - t_0)}. \quad (10)$$

Now,

$$\begin{aligned} & \int_{\mathcal{I}} z_{t_1}(\mathcal{J}_i) d\mathcal{A}_{\mathcal{J}} \\ &= z_{t_1}(\mathcal{J}_i) + \int_{\mathcal{I}^i} z_{t_1}(\mathcal{J}_j) d\mathcal{A}_{(\mathcal{J}_i)',} \quad (11) \end{aligned}$$

$$\begin{aligned} &= 1 - (1 - z_{t_0}(\mathcal{J}_i)) e^{-\kappa(t_1 - t_0)} \\ &+ \int_{\mathcal{I}^i} z_{t_0}(\mathcal{J}_j) e^{-\kappa(t_1 - t_0)} d\mathcal{A}_{(\mathcal{J}_i)',} \quad (12) \end{aligned}$$

$$\begin{aligned} &= 1 - ((1 - z_{t_0}(\mathcal{J}_i)) - \\ & \int_{\mathcal{I}^i} z_{t_0}(\mathcal{J}_j) d\mathcal{A}_{(\mathcal{J}_i)',}) e^{-\kappa(t_1 - t_0)} \\ &= 1 - (1 - \\ & \int_{\mathcal{I}} z_{t_0}(\mathcal{J}_j) d\mathcal{A}_{\mathcal{I}}) e^{-\kappa(t_1 - t_0)} \quad (13) \end{aligned}$$

Equation 11 follows from a separation of  $\mathcal{J}$  into the two sections  $i$  and  $j$ . Equation 12 follows from direct substitution of Equation 9 and 10 into Equation 11. Equation 13 follows as Equation 6 is independent of the index  $t$ , Equation 6 states that the integral in Equation 13 is unity, and

$$\therefore \int_{\mathcal{I}} z_{t_1}(\mathcal{J}_i) d\mathcal{A}_{\mathcal{J}} = 1 \quad (14)$$

which simply states Definition 3.1.  $\diamond$

Theorem 3.1 is expressed in a stochastic notation, while the network equations are couched in more conventional algebraic notation. Ignoring the functional and indexed expressions of temporal dependence in Equations 5, 8 and 7, the notational equivalence is between

$$z(\mathcal{J}_i) : \mathcal{J}_i \rightarrow [0, 1] \equiv z_{ji} \quad j \in \mathcal{J}_i,$$

where  $\mathcal{J}_i = \{x_j\}$ . Thus we can interpret  $z(\mathcal{J})$  as the synaptic connectivity to the nodes  $x_i \in \mathcal{J} \subset J$ .

Thus we have concluded that Equation 5 is compatible with a probabilistic interpretation of synaptic activity. We have that

$$z_j \equiv z(\{x_i\}) : \mathcal{J}_i \rightarrow [0, 1]$$

is a measure on the set  $\mathcal{J}_i$  of nodes for each  $x_j$  to which  $x_i$  projects. This permits interpretation of  $z_{ji}$  as a stochastic kernel, such that

$$z_{ji} : J \times \mathcal{J} : (i, j) \rightarrow x_i.$$

Due to the one to one correspondence between stochastic processes and kernels, this kernel can be interpreted as a model of a stochastic process. Letting the environmental source of the inputs  $I_i$  to the network be denoted by a probability space  $(E, \mathcal{E})$ , the process may be represented

$$\mathcal{Z} : E \rightarrow \mathcal{J}_i : I_i \rightarrow x_i.$$

The synaptic activity can thus be interpreted as a model for the process  $\mathcal{Z}$ , where  $z_{ji}$  is the conditional probability of  $I_i$  given  $x_i$ .

## 4 Conclusion

Theorem 3.1 proves that the network dynamics arising in a network based on the cell membrane equation is compatible with an interpretation of network activity that views the synaptic connectivity as probability measures. The nodal activation  $x_i$  is interpreted as a random vector on the probability space  $(E, \mathcal{E})$ . The random variables  $\mathcal{Z}_i : I_i \rightarrow x_i$  define the semantics of the network activation.

### 4.1 Conversations about Neural States

Any empirical observations of  $x_i$  or  $z_{ji}$  may be related to the environmental model  $(E, \mathcal{E})$ . The benefit of having a theoretical model of the environment is that empirical observations of an autonomous system take place in the environment, and so verification can

of reference of the empirical phenomenon. Thus it is possible to talk about neural activation, without introducing observational bias that cannot be analysed within the theoretical framework.

## References

- [Braitenberg 1984] V. Braitenberg. *Vehicles: Experiments in Synthetic Psychology*. MIT Press, London 1984.
- [Brooks 1990] R. A. Brooks. *Artificial Life and Real Robots*. in Towards a Practice of Autonomous Systems. Proceedings of the First European Conference on Artificial Life. P. 3, MIT Press, Boston, USA 1991.
- [Carpenter 1990] G. Carpenter and S. Grossberg. *ART 3 Hierarchical Search: Chemical Transmitters in Self-Organising Pattern Recognition*. IJCNN-90-Wash-dc II-30
- [Figueroa 1990] J. G. Figueroa, M. Romero, E. Vargas and C. Flores. *Langevin Equations and the Formal Foundations of Neural Networks*. IJCNN-90-Wash-dc I-385
- [Grossberg 1976] S. Grossberg. Adaptive pattern classification and universal recoding: Ii. feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 23:187–202, 1976b.
- [Mingolla 1985] *Neural Dynamics of Perceptual Grouping: Textures, Boundaries and Emergent Segmentations*. S. Grossberg and E. Mingolla. Perception and Psychophysics, Vol. 38, 141-171, (1985)
- [Plonsey 1969] R. Plonsey and D. Fleming. *Bioelectrical Phenomena*. McGrath-Hill, New York, 1969.
- [Rosen 1985] R. Rosen. *Anticipatory Systems: Philosophical, Mathematical and Methodological Foundations*. Pergamon Press, Canada, 1985.