A Model of Probability Matching in a Two-Choice Task Based on Stochastic Control of Learning in Neural Cell-Assemblies

Roman V. Belavkin (R.Belavkin@mdx.ac.uk)

Christian R. Huyck (C.Huyck@mdx.ac.uk)
School of Engineering and Information Sciences
Middlesex University, London NW4 4BT, UK

Abstract

Donald Hebb proposed a hypothesis that specialised groups of neurons, called cell-assemblies (CAs), form the basis for neural encoding of symbols in the human mind. It is not clear, however, how CAs can be re-used and combined to form new representations as in classical symbolic systems. We demonstrate that Hebbian learning of synaptic weights alone is not adequate for all tasks, and that additional meta-control processes should be involved. We describe an earlier proposed architecture (Belavkin & Huyck, 2008) implementing such a process, and then evaluate it by modelling the probability matching phenomenon in a classic two-choice task. The model and its results are discussed in view of mathematical theory of learning, and existing cognitive architectures as well as some hypotheses about neural functioning in the brain.

Keywords: Artificial Intelligence, Cognitive Science, Neuroscience, Decision making, Intelligent agents, Learning, Bayesian modeling, Computational neuroscience, Human experimentation

Introduction

There exists a variety of artificial systems and algorithms for learning and adaptation. Most of them can be classified as sub-symbolic (e.g. Bayesian and connectionist networks) or symbolic systems (e.g. rule-based systems). Known natural learning systems use neural networks, and therefore can be classified as using sub-symbolic computations. A distinguishing feature of the human mind, however, is the ability to use rich symbolic representations and language.

From an information-theoretic point of view, symbols are elements of some finite set that are used to encode discrete categories of sub-symbolic information. They enable communication of information about the environment or a complex problem in a compact form. One obvious benefit is that with language, one can learn not only from one's own experience, but also from experiences of others. The benefits of reading a guidebook before going abroad are obvious.

The duality between sub-symbolic and symbolic approaches has been studied in cognitive science. There exists sub-symbolic (i.e. connectionist), symbolic (e.g. SOAR, Newell, 1990) and hybrid architectures (e.g. ACT–R, Anderson & Lebiere, 1998) for cognitive modelling. These different approaches, however, have not yet explained where the symbols are in the human mind, or how the brain implements symbolic information processing.

It was proposed by Hebb (1949) that symbols are represented in the brain not by individual neurons, but by correlated activities of groups of cells, called *cell assemblies* (CAs). The CABOT project set out to test and demonstrate this idea in an engineering task by building an artificial agent, situated in a virtual environment, capable of complex symbolic processing, and implemented entirely using CAs of simulated neurons. Some of the objectives have already been achieved and reported elsewhere (e.g. Huyck & Belavkin, 2006; Huyck, 2007; Belavkin & Huyck, 2008). The architecture and some of these works will be discussed in the next section.

The work described in this paper is concerned with a particular aspect of the project — a stochastic meta-control mechanism that modulates Hebbian learning to allow for re-use and combination of CAs into new representations, such as learning logical implications (i.e. procedural knowledge). As will be discussed in this paper, this cannot be achieved by using a Hebbian learning mechanism alone. A unique contribution of this work is evaluation of the meta-control mechanism in a cognitive model of the probability matching phenomenon in a two-choice experiment (Friedman et al., 1964). The results suggest that a proposed mechanism is a plausible model. Some neurophysiological studies and hypotheses about the brain circuitry will be discussed supporting the biological plausibility of the architecture.

Cell-Assemblies as the Basis of Symbols

In this section, we outline some of the basic features of the CABOT architecture as well as the CA hypothesis.

Neural Information Processing in CABOT

It is widely accepted that human cognition is the result of the activity of approximately 10^{11} neurons in the central nervous system (CNS) that interact with each other as well as with the outside world via the peripheral nervous system (PNS). Biological neurons are complex systems, and they have been modelled with various levels of details. In our system, we use fatiguing, leaky, integrate and fire (fLIF) neurons.

The 'integrate and fire' component is based on the classical idea that the neuron 'fires' (or spikes) if its action potential, A, exceeds a certain threshold value θ: $y = 1$ if $A \geq \theta$; $y = 0$ otherwise. The action potential, A, is a function of the inner product (integrator): $\langle x, w \rangle = \sum_{i=1}^{k} x_i w_i$, where $x \in \mathbb{R}^k$ is the stimulus vector (pre-synaptic), and $w \in \mathbb{R}^k$ is the synaptic weight vector of the neuron. Here, $\mathbb{R}^k$ is a k-dimensional real vector space, where k is the number of synapses to the neuron. We use binary signals, and therefore x is a k-dimensional binary vector.

The 'leaky' property refers to a more complex (non-linear) dependency of the action potential on the pre- and post-synaptic activity:

$$A_{t+1} = \frac{A_t}{d_t} + \langle x_t, w_t \rangle, \quad d_t = \begin{cases} \infty & \text{if fired } (y_t = 1) \\ d \geq 1 & \text{otherwise} \end{cases}$$

Thus, the action potential is accumulated over several time moments if the neuron does not fire. Parameter $d \geq 1$ allows for some of this activation to 'leak' away. This is the LIF model (Maas & Bishop, 2001).

The 'fatigue' property refers to a dynamic threshold that is defined as follows:

$$\theta_{t+1} = \theta_t + F_t, \quad F_t = \begin{cases} F_+ \geq 0 & \text{if fired } (y_t = 1) \\ F_- < 0 & \text{otherwise} \end{cases}$$

where values F_+ and F_- represent the *fatigue* and fatigue *recovery* rates. Thus, if a neuron fires at time t, its threshold increases, and it is less likely to fire at time $t+1$.

The fatiguing and leaky properties of the neural model allow for a non-trivial dynamics of the system. Repetitive stimulation of excitatory synapses increases the probability of a neuron to fire, even if the weights have small (positive) values. On the other hand, if the neuron fires repetitively, its threshold increases reducing the chance of it firing again. Thus, frequencies of pre- and post-synaptic activities are important factors in our system.

The weights, w, of a neuron can adapt according to the compensatory learning rule (Huyck, 2007), which is an implementation of the Hebbian principle (Hebb, 1949), where w_{t+1} depends on the correlation between the pre-synaptic, x_t, and the post-synaptic, y_t, activities.

The above described properties are known characteristics of biological neurons, and our model is a compromise between computational efficiency and biological plausibility that is important for the emerging dynamics that we discuss.

Neural Cell-Assemblies

Networks of neurons can be used as general function approximators and applied in a variety of tasks including control, pattern recognition and classification. Our system, CABOT, uses recurrent, partially connected networks (a mesh) of fLIF neurons with a largely pre-defined topology. The non-linearity of the cells and the topology of the network leads to a complex dynamics of the system similar to that in attractor and recurrent nets (e.g. Hopfield, 1982), where some of the states are more probable. These more 'stable' states can be characterised by groups of neurons that remain significantly more active than the other cells in the system. According to Hebb (1949), we refer to such reverberating groups of cells as *cell assemblies* (CAs).

In our system, the formation of CAs depends on the topology of the network, and it is facilitated by the adaptation of the weights between connected cells. Therefore, CAs can be used for pattern classification of sensory stimuli (i.e. patterns from external connections). This leads to functional *specialisation* of neurons in the network based on CAs — two cells are functionally different if they belong to different CAs, even though they are similar architecturally. Such specialisation is observed in many neural networks, such as in self-organising maps (Kohonen, 1982) and particularly in the human brain. Note that CAs are not necessarily disjoint sets of cells. A single cell may be a member of several overlapping CAs. This feature can be used to encode hierarchies of patterns (Huyck, 2007).

An important property of CAs' dynamics is their persistence. When enough neurons fire to start the reverberating circuit, the CA ignites. Once ignited, the activity within the cells in a CA may be sufficient to support itself. Many variables can contribute to this effect. In particular, the fatigue and recovery rate parameters in our system effect persistence.

A CA's activity does not only depend on the external patterns, but also on the activity of other CAs in the system as they can ignite and extinguish each other. Thus, the activity of several CAs can be characterised by different patterns of ignition order and so on. It was demonstrated earlier that such state transitions in the system of CAs are sufficiently controllable to implement a broad range of tasks simulating symbolic processing that will be discussed below.

Symbols and Human Cognition

Many models of biological neurons suggest that synaptic weights may represent the memory for statistical and sub-symbolic information of the stimulus. In particular, in many algorithms for training artificial neural networks (e.g. Oja, 1982), the weight vector $w \in \mathbb{R}^k$ corresponds to one of the principal eigenvectors of the covariance matrix $E\{xx^\dagger\}$ of input vectors $x \in \mathbb{R}^k$ that have been observed. On the other hand, human cognition, and human knowledge in particular, is encoded using symbolic representations, and the link between the symbols and neural models is less clear.

It was proposed by Hebb (1949) that CAs may be considered as the neural basis of symbols. Indeed, as discussed in the previous section, CAs can be easily mapped to some discrete categories of the stimuli, and their activity patterns can model serial processing typical for symbolic algorithms. Testing this hypothesis experimentally is one of the main objectives of the CABOT project. However, many challenges had to be overcome to make a purely CA-based system performing some non-trivial symbol processing task.

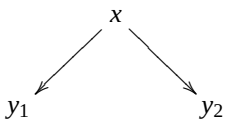
Previously, we reported a system performing a counting task that consisted of 7 modules and 40 CAs (Huyck & Belavkin, 2006). A more recent system, CABOT 2, is an artificial agent functioning in a virtual 3D environment that has a model of visual information processing, and is capable of natural language processing and action selection (Belavkin & Huyck, 2008). One of the advantages of such a CA-based architecture is that neural CAs, that we associate with symbolic representations, integrate also all the sensory (i.e. sub-symbolic) information, which can be a natural solution to the *symbol grounding* problem. An associated phenomenon of symbolic processing is *grounding transfer* — combination and re-use of existing symbols to form new representations.

The re-use of symbols is also important for learning procedural knowledge. Indeed, a logical implication (i.e. a production rule) may use combinations of symbols both in the antecedent and the consequent, and generally there are many more possible combinations than the number of rules that are actually used. Hybrid architectures, such as ACT–R, rely on statistical (sub-symbolic) computations to 'filter' out the unwanted rules in the process called *conflict resolution*. In CABOT, associations between CAs are learnt due to the Hebbian learning mechanism. However, as will be pointed out below, this mechanism alone is not sufficient to implement learning of particular associations between CAs representing existing symbols. To resolve this problem, an additional stochastic meta-control mechanism, moderating the Hebbian learning, has been introduced (Belavkin & Huyck, 2008). Here, we use this mechanism to model the probability matching phenomenon in a classical two-choice experiment, and this way evaluate its plausibility.

Stochastic Meta-Control of Learning

Two-Choice Task

Let x, y_1 and y_2 be three symbols, where x represents a stimulus (antecedent), and y_1, y_2 represent two alternative responses (consequents). Thus, we have a conflict between two implications $x \to y_1$ and $x \to y_2$ shown on the diagram below



This is a simplest two-choice task (a more complex two-choice task may involve a set of different stimuli). The choice of y_1 or y_2 is followed by some reinforcement event E that may have different utility values (e.g. a success after choosing y_1 or a failure after choosing y_2). Learning the associations between the choices and the utility values, such as $u(x \to y_2) \leq u(x \to y_1)$, leads to a preference $y_2 \lesssim y_1$, and therefore learning rule $x \to y_1$. If the reinforcement event is not deterministic, but occurs with some probability $P(E) = \pi \in [0,1]$, then the preference of y_1 to y_2 may also be stochastic. As demonstrated in many experiments with animals and human participants, the frequency of choosing y_1 adapts to probability π of reinforcement with high utility — a phenomenon referred to as the *probability matching*. This phenomenon can be explained based on the theories of optimal statistical decisions (Wald, 1950) and information value (Stratonovich, 1965).

Principles of Statistical Learning

Let us consider an abstract system with input $x \in X$ and output $y \in Y$. Any learning system can be characterised by some optimisation criteria and information constraints (Belavkin, 2009). Optimisation corresponds to some preference relation on the input-output pairs $(x,y) \in X \times Y$. In a deterministic setting, this preference relation can be represented by a utility function $u : X \times Y \to \mathbb{R}$, while in stochastic setting one considers conditional probability distributions $P(u \mid x,y)$ on values of utility $u \in \mathbb{R}$. If the utility function $u = u(x,y)$ or the joint distribution $P(u,x,y)$ is known (and hence $P(u \mid x,y)$), then given input x, the optimal output $\hat{y} \in Y$ maximises the expected utility:

$$\hat{y}(x) = \arg\max_y E_P\{u \mid x,y\}$$

where $E_P\{\cdot\}$ denotes the expected value with respect to distribution P (in the deterministic case, $E_P\{u \mid x,y\}$ coincides with $u = u(x,y)$). The *greedy* strategy of always choosing the optimal output can be expressed as follows:

$$P(y \mid x) = \begin{cases} 1 & \text{if } y = \hat{y}(x) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Information constraints mean that either the utility function $u = u(x,y)$ or the distribution $P(u,x,y)$ is not known. Instead, one has some data from past occurrences of $(u,x,y) \in \mathbb{R} \times X \times Y$ which can be used to estimate $\tilde{u}(x,y) \approx E_P\{u \mid x,y\}$. In this case, the greedy strategy for choosing the system's output is not optimal. The optimal policy is the following exponential ('soft-max') distribution (e.g. Belavkin, 2009):

$$\hat{P}(y \mid x) = Q(y \mid x) \exp\{\beta \tilde{u}(x,y) - \Psi(\beta,x)\} \tag{2}$$

where $Q(y \mid x)$ is the distribution corresponding to the minimum of information (e.g. no data), parameter β is related to the amount of information available in the data, and $\Psi(\beta,x)$ is defined from the normalisation condition (i.e. $\Psi(\beta,x) = \ln \sum_Y Q(y \mid x) \exp\{\beta \tilde{u}(x,y)\}$). Distribution (2) is obtained by solving the following variational problem

$$U(I) = \sup_P \{E_P\{u\} : I(P,Q) \leq I\}$$

where $I(P,Q)$ is the Kullback-Leibler divergence of distribution $P(u,x,y)$ from $Q(u,x,y)$ representing information amount I contained in the data. Parameter β^{-1} appears in the solution as the Lagrange multiplier related to information constraint I by the derivative of $U(I)$:

$$\beta^{-1} = U'(I) \tag{3}$$

The function above is decreasing so that $\beta^{-1} \to 0$ (or $\beta \to \infty$) as information increases. Note that the exponential distribution (2) converges to the greedy strategy (1) as $\beta \to \infty$.

Exponential distributions are often used for selecting the output of a system in machine learning and stochastic optimisation algorithms. It is also used in the ACT–R cognitive architecture to model some stochastic properties of behaviour. In particular, it was used in the ACT–R model of the two-choice experiment, discussed below. However, the 'temperature' parameter β^{-1} is usually set to some constant value or determined from some arbitrary 'annealing' schedule. The relation of β^{-1} to entropy of success in ACT–R was proposed in (Belavkin, 2002/2003), and it was shown that it improves the match between the models and data. The derivation of optimal function $\beta^{-1} = U'(I)$ can be found in (Stratonovich, 1965) and more generally in (Belavkin, 2009).

Meta-Control of Hebbian Learning

The output of a neuron depends on its weight vector $w \in \mathbb{R}^k$, which, according to Hebb's hypothesis, adapts to the correlation between the pre- and post-synaptic activities x and y in the past. It is attractive to conclude, therefore, that Hebbian learning is a particular implementation of the statistical learning. However, the utility is clearly missing in this description of neural plasticity. What criteria does such a process of changing the weights optimise? If in a two-choice task the system accidentally chooses the 'incorrect' cell-assembly y_2, then the weights associating x with neurons in y_2 increase due to the correlation-based Hebbian learning. This can only increase the chance of $x \to y_2$ igniting in the future, even though the reinforcing event E following the choice of $x \to y_2$ has a low utility (i.e. a failure). Thus, some additional process should be involved to increase the chance of the 'correct' combination $x \to y_1$ after the reinforcing event E. Such a process appears to be especially useful if the CA-based symbolic representations, formed earlier, are to be re-used. Below we describe a neural implementation of such a meta-control of Hebbian learning based on the utility feedback (Belavkin & Huyck, 2008) following principles of statistical learning.

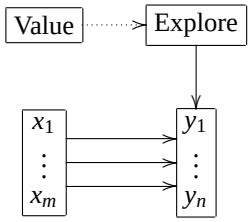


Figure 1: Components and connections of the Value and Explore modules controlling Hebbian learning of connections between CAs in modules X and Y. Solid and dashed arrows show excitatory and inhibitory connections respectively.

The meta-control process involves two specialised modules: Value and Explore. Their connections in the system are shown on Figure 1. Here, $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_n\}$ are sets of CAs representing m stimuli and n responses respectively. Initially, there are excitatory connections from every CA in X to all CAs in Y, which means that all pairs (x,y) (i.e. all rules $x \to y$) are equally preferred. Thus, given input $x \in X$, any response $y \in Y$ can be selected. However, due to Hebbian learning, the connection $x \to y$ is reinforced if a particular pair of CAs ignite together, giving the pair a higher chance to ignite together in the future. Thus, simply by virtue of Hebbian learning, the system can learn eventually to prefer some random pairs. The purpose of the Value and Explore modules is to make this process selective according to the utility value of the feedback.

The output activity of the Value module represents the utility values u associated with the pair (x,y) selected on the previous step. The input of the module can be configured according to the application (e.g. using sensory information).

The purpose of the Explore module is to randomise the activity of the response CAs (i.e. CAs in set Y). The Explore module contains cells that can be active without any external stimulation due to spontaneous activation. The cells in the Explore module send excitatory signals to all CAs in Y, and the weights of these connections do not change. Thus, the activity in the Explore module can trigger randomly any response CA, and this process does not have a memory. The Explore module implements the effect of parameter β^{-1} in the exponential distribution.

The Value module sends inhibitory connections to the Explore module, so that high activity of the Value cells may shut down the activity in the Explore module. As a result, any response CA that has been ignited in set Y will persist longer because it is less likely to be shut down by another CA. Such a connectivity implements the following learning scheme: If a particular pair (x,y) results in a high utility value, then high activity of the Value module inhibits the Explore module, and the responsible (x,y) pair is allowed to persist longer, and the $x \to y$ connection increases relative to others due to Hebbian learning.

Learning the 'correct' rules (subset $R \subset X \times Y$) contributes to a better performance of the system (i.e. higher expected utility). As a consequence, the average activity of the Value module increases with time, while the activity of the Explore module decreases. This dynamic also corresponds to a decrease of parameter β^{-1} as information increases making the system less random and more deterministic.

Modelling Probability Matching

To test how adequately the above mechanism can represent properties of human cognition, we evaluate its performance against data from a classic two-choice experiment due to Friedman et al. (1964). The choice of this dataset was motivated not only by its quality and detailed description of the procedures, but also because it was used to 'calibrate' stochastic properties of other cognitive architectures, such as ACT–R (Anderson & Lebiere, 1998). The complete description of the experiment and data can be found in the original paper (Friedman et al., 1964). Here we give a basic outline.

Experiment Description and Previous Work

In this experiment, participants were asked to select one of two responses on presentation of a stimulus. After the response was selected, a reinforcement event E occurred with probability $P(E) = \pi$ that did not depend on the response. Each participant had to perform this task in three sessions, each session consisting of 8 blocks, each block consisted of 48 trials. The probability $P(E) = \pi$ changed between each 48–trial block. This paper will report only simulations of results in Sessions 1 and 2. In these two sessions, blocks 1, 3, 5 and 7 had $P(E) = .5$, and blocks 2, 4, 6, and 8 were with $P(E) \in \{.1, .2, .3, .4, .6, .7, .8, .9\}$ that was assigned according to a random pattern. Thus, probability $P(E) = \pi$ was alternating between .5 and some value above or below .5 between

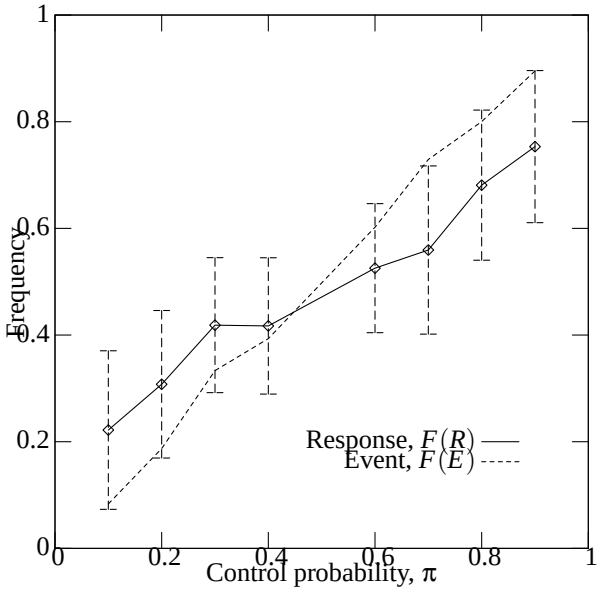


Figure 2: Frequency of response (ordinates) as a function of the probability of reinforcing this response (abscissae). Points and error bars represent average response and standard deviations in 48–trials of two-choice task from 80 participants, reported in (Friedman et al., 1964). Dashed line shows frequency of the reinforcing event itself.

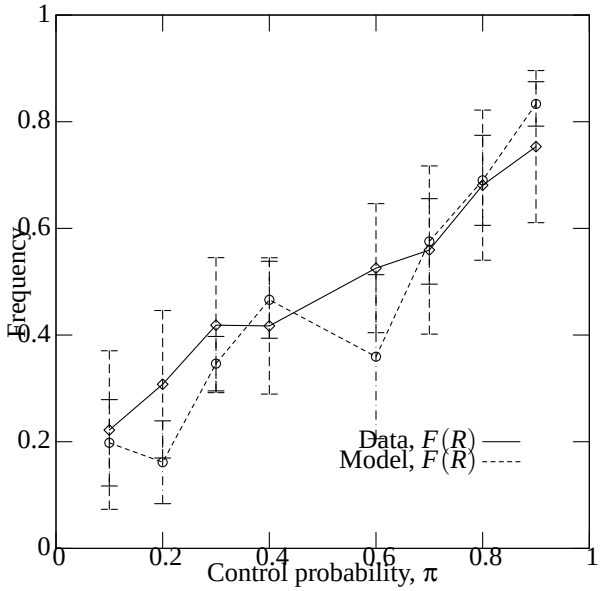


Figure 3: Comparison of response frequency produced by the CABOT model with response frequency by participants in (Friedman et al., 1964). RMSE=8.937%.

48-trial blocks. The data recorded the number of times Response 1 was chosen in each 48-trial block.

Figure 2 shows the results of these experiments, reported by Friedman et al. (1964). The charts show frequencies of Response 1, $F(R)$, and reinforcement events, $F(E)$, as functions of the control probability $P(E) = \pi$. One can see that the frequency of the reinforcement event $F(E)$ approximates the the control probability $F(E) \approx P(E)$. The response frequency $F(R)$ also matches the probability $P(E)$, but it differs significantly at the lower and higher ends of the range: When $P(E)$ is low ($\pi = .1$), the participants overestimate the probability ($F(R) \geq P(E)$); when $P(E)$ is high ($\pi = .9$), the participants underestimate it ($F(R) \leq P(E)$). Thus, the response appears to be less certain than the reinforcing event.

As suggested by Anderson and Lebiere (1998), this experimental evidence indicates against using the greedy strategy (1) for choosing the response. The data was modelled in ACT–R by sampling responses from exponential distribution with some $\beta^{-1} > 0$. This agrees with equations (2) and (3), where $\beta^{-1} \to 0$ only when information $I \to \sup I$. We now describe a model of this experiment implemented in CABOT.

Model Description

The model used the architecture shown on Figure 1, where module X consisted of CAs representing one or more stimuli, and module Y contained two CAs representing two alternative responses. There were excitatory connections with low weights from module X to all CAs in module Y. The weights on these connections, however, could adapt according to a Hebbian rule increasing associations $x \to y$ between active CAs. The fatigue and leak parameters of the Y network were set in such a way that CAs ignite only when an external stimuli are present. The CAs in Y inhibited each other so that only one of the CAs in Y was active at any moment. The Explore module had excitatory connections with a small proportion of cells in module Y. These connections were distributed uniformly, and the weights did not adapt. Spontaneous activation in the Explore module could randomly trigger any of the two response CAs in module Y. The activity of the Explore module could be inhibited by the output activity from the Value module that was triggered in each trial according to probability $P(E) = \pi$ of the reinforcement event, controlled by the experimental sequence.

When the Explore module is inhibited by the reinforcing activity of the Value module, the active pair (x,y) is allowed to persist longer, strengthening the connections $x \to y$ relative to other connections. We found that the robustness of this effect depends on the time (i.e. number cycles) these CAs are allowed to persist. In this model, it takes approximately between 10–20 cycles for a response CA in Y to ignite, and if the Explore module is active, then the response CA may change during another 10–20 cycles. In this experiment, the system ran for 100 cycles per trial which was sufficient for the control of learning to have a robust effect. The complete code of the simulation is available online from the CABOT project website.

Results

The model was used to simulate Sessions 1 and 2 of eight 48-trial blocks each with variable control probabilities π (Friedman et al., 1964). The results comparing response fre-

quency of the model with the data are shown on Figure 3. The model approximates the data fairly well (RMSE=8.937%) showing the probability matching effect that also overestimates and underestimates the low and high value of the control probability π respectively. Note that unlike the ACT–R model, where the estimated parameter β^{-1} in the exponential distribution was constant (Anderson & Lebiere, 1998), the activity of the Explore module randomising the response is dynamic.

Conclusions

In this paper, we discussed the CABOT architecture and some challenges associated with implementing the CA hypothesis of symbolic processing in the brain. The problem of re-use and combination of symbols, particularly in learning procedural knowledge, pointed at one significant shortcoming of the standard Hebbian learning mechanism — adaptation of weights based purely on correlations does not take into account the optimisation criteria that a system may have to satisfy. To resolve this problem, stochastic meta-control based on utility feedback was introduced into the system.

It is attractive to speculate about the existence of the Value and Explore modules in the brain. Some researchers have proposed that tonically active cholinergic neurons in the basal ganglia and striatal complex play an important role in conflict resolution and learning procedural knowledge (Granger, 2006). These neurons account for a small proportion of the connections that are quite uniform and non-topographic, and the activity of these neurons was suggested to play the role of stochastic noise, similar to the activity of cells in the Explore module (see Fig. 1). Interestingly, the activation of the tonically active cholinergic neurons is inhibited by the activation from the reward path, similar to the function of the Value module in our system. Other studies of mechanisms for exploratory behaviour in the brain are also in favour of the exponential distribution model (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006).

Setting these speculations aside, this work has demonstrated that the proposed mechanism can be used for controlling Hebbian learning in networks of relatively biologically faithful models of neurons. The mechanism allows for selective learning of connections between specialised groups of cells (CAs), and following Hebb's hypothesis it shows not only that CAs can indeed be associated with symbols, but also shows how such representations can be re-used and combined to learn new knowledge. Simulation of the probability matching effect has demonstrated that the mechanism is also a plausible cognitive model. We anticipate that the proposed architecture can also be used to model other psychological phenomena, such as the effect of reinforcement values on speed of learning, and this is one possible direction of our future research.

Acknowledgements

This work was supported by EPSRC grant EP/DO59720.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.

Belavkin, R. V. (2003). *On emotion, learning and uncertainty: A cognitive modelling approach*. PhD thesis, The University of Nottingham, Nottingham, UK.

Belavkin, R. V. (2009). Bounds of optimal learning. In *2009 IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning* (pp. 199–204). Nashville, TN, USA: IEEE.

Belavkin, R. V., & Huyck, C. (2008). Emergence of rules in cell assemblies of fLIF neurons. In *The 18th European Conference on Artificial Intelligence.*

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879.

Friedman, M. P., Burke, C. J., Cole, M., Keller, L., Millward, R. B., & Estes, W. K. (1964). Two–choice behaviour under extended training with shifting probabilities of reinforcement. In R. C. Atkinson (Ed.), *Studies in mathematical psychology* (pp. 250–316). Stanford, CA: Stanford University Press.

Granger, R. (2006, July). Engines of the brain: The computational instruction set of human cognition. *AI Magazine*, *27*(2), 15–32.

Hebb, D. O. (1949). *The organization of behavior*. New York: John Wiley & Sons.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the USA*, *79*, 2554–8.

Huyck, C. (2007). Hierarchical cell assemblies. *Connection Science*.

Huyck, C., & Belavkin, R. V. (2006, April). Counting with neurons, rule application with nets of fatiguing leaky integrate and fire neurons. In D. Fum, F. D. Missier, & A. Stocco (Eds.), *Proceedings of the Seventh International Conference on Cognitive Modeling.* Trieste, Italy: Edizioni Goliardiche.

Kohonen, T. (1982). Self–organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59–69.

Maas, W., & Bishop, C. (2001). *Pulsed neural networks*. MIT Press.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, Massachusetts: Harvard University Press.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, *15*, 267–273.

Stratonovich, R. L. (1965). On value of information. *Izvestiya of USSR Academy of Sciences, Technical Cybernetics*, *5*, 3–12. (In Russian)

Wald, A. (1950). *Statistical decision functions*. New York: John Wiley & Sons.