Brain Inspired Cognitive Systems
August 29 – September 1, 2004
University of Stirling, Scotland, UK

DEMONSTRATING VOLITION AND AFFECTIVE DECISION MAKING IN BRAIN-INSPIRED NEURAL MODELS

Rabinder Lee, Imperial College, Exhibition Road, South Kensington, London SW7 2AZ

ABSTRACT

This paper presents an experiment concerning affective decision making in artificial neural networks (ANNs). The type of neuron used is the virtual G-RAM neuron, which was developed by Aleksander (1990) and implemented on a multi-module neuro-computational software system called Neural Representation Modeller (NRM), created by Aleksander and Dunmall (2000). I have built two neural configurations, each dealing with contrived and worldly scenarios. Each has a set of rules and facts which are pre-learned, and during execution the entity has to decide on an appropriate course of action to take, depending on the situation unfolding in the world. The entity also takes into account its current 'emotional' state. I am working on making the decisions largely non-deterministic and random, to create a true freedom of choice. However, as the entity learns about its world, it will develop preferences and habits, and these will introduce a bias in the decision making process. The neural configurations presented in this paper are novel.

INTRODUCTION

In attempting to simulate decision making resulting in voluntary action, we are posed with two important questions. The first is whether these actions will be truly free, and the assumption is that free-will exists in human beings. The second is how free action is carried out in neural structures within the brain. This knowledge is required if we are to take advantage of the efficient mechanisms already developed in humans through evolution. Several important contributions to this research question have been made by others, notably Libet (1999, 2001, 2002), Spence and Frith (1999) and Ingvar (1999). Towards this end, two separate neural configurations have been created, both which perform affective decision making in different scenarios. The accent in this paper is on the behaviour of the system, and not on its actual construction. The idea is that by investigating artificial decision making, we may get an insight into the real kind (Aleksander, 1996), and also perhaps build machines which are capable of autonomous and self-sufficient activity in their environment.

BIOLOGY AND ARTIFICIAL BEHAVIOUR

A basic functional anatomy of volition is required if we are to construct biologically plausible neural structures. Although a number of brain regions contribute to the execution of a freely willed act, one is of particular importance: the dorso-lateral prefrontal cortex (DLPFC) and its associated connections via cortico-subcortical and cortico-cortical circuits (Spence and Frith, 1999). The dorso-lateral prefrontal cortex takes part in the basal-ganglia thalamo-cortical circuits (Alexander et al, 1986), the details of which a good neuroscientist will be aware of, but not an electronic engineer like myself. This circuit is illustrated in Fig. 1. In this experiment on artificial volition, an attempt has been made to remain faithful to the biological counterparts. This has not been done robustly, since to fully reproduce complex brain structures would take a much larger project, and this project is only a preliminary experiment.

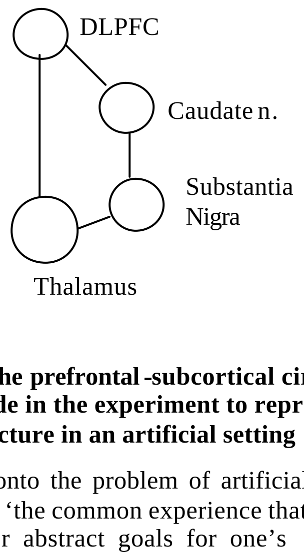

Fig. 1 One of the prefrontal-subcortical circuits. An attempt is made in the experiment to reproduce this biological structure in an artificial setting

Now lets move onto the problem of artificial free-will. Will can be defined as 'the common experience that one can produce inner concrete or abstract goals for one's future behaviour and/or cognition' (Ingvar 1999). Further, I define *free*-will as a choice, the outcome of which is non-deterministic and random. While the existence of *free*-will is still a matter of debate, one can definitely concede that there are aspects of our activity which are not free. For example, we don't always consider all our options, and our decision-making is heavily influenced by habits and preferences. However, in a simple case, such as whether we should lift our finger or not, we certainly have free choice, and this is the case which will be considered in this paper. However, I have also incorporated a feature to bias the choice, so that more complex decisions can be investigated.

METHODOLOGY

The experiment was conducted using NRM software, running on a Pentium 4 processor. A complete description of the tool can be found by downloading the software from www.sonnet.co.uk/nts. The software is the copyright of Novel Technical Solutions.

The software graphically displays the content of neural layers and sensory inputs, and also shows the connections between them. Various connection methods between two neural layers and between an input and a layer are possible. I only use 'global random' connectivity, which means that each neuron receives inputs from a randomly selected group of neurons within each layer.

The overall activity of each simulation is to explore various possible courses of action (or trajectories), and then to choose one which is most appealing to its current affective state. Each stage of each trajectory is associated with an emotion. The entity starts by depicting its current situation, then constructs (imagines) a possible sequence of future states, then comes back to its current situation again, and then starts the process once more. In humans or animals, each stage may also be characterised by an image, ephemeral logic or language. In the NRM configurations I have built, each stage simply consists of an image and an associated emotion.

LABORATORY EXPERIMENTS 1

The first simulation involved 11 neural layers, with a total of approximately 12,600 neurons (see **Fig. 2**). Only the trajectory exploration stage is shown in the the figure. Neural layer 3 can be thought of as depicting the 'emotional' content of the entity. It reflects positive and negative emotions, like fear or joy, and hence this neural layer can be compared to the paleocortex of the amygdala or the layers of nerve cells within the forebrain's outer covering of neocortex. This layer forms the hinge of the decision making process, as it provides feedback to the trajectory exploration layers, thus influencing the final choice made by the entity.

The connections between the world and neural layer 2 can be seen as the cortico-geniculate pathway within the brain. These constantly explore possible courses of action using the visual imagination. Initially, the entity is programmed to go through all possible trajectories, although this is not biologically realistic. (The second simulation performs the trajectory search in a better way – see later). Emotional feedback from neural layer 3 eventually causes a bias to set in, after which the entity makes a decision as to which act to perform.

The trajectories which the entity goes through were contrived for this particular application, but one could conceive of a system which actually learns the actions it could take, rather than an external programmer entering these in. This would entail another set of neural modules, the implementation of which is outside the scope of this experiment. The trajectories which the entity goes through are shown in **Fig. 3**.

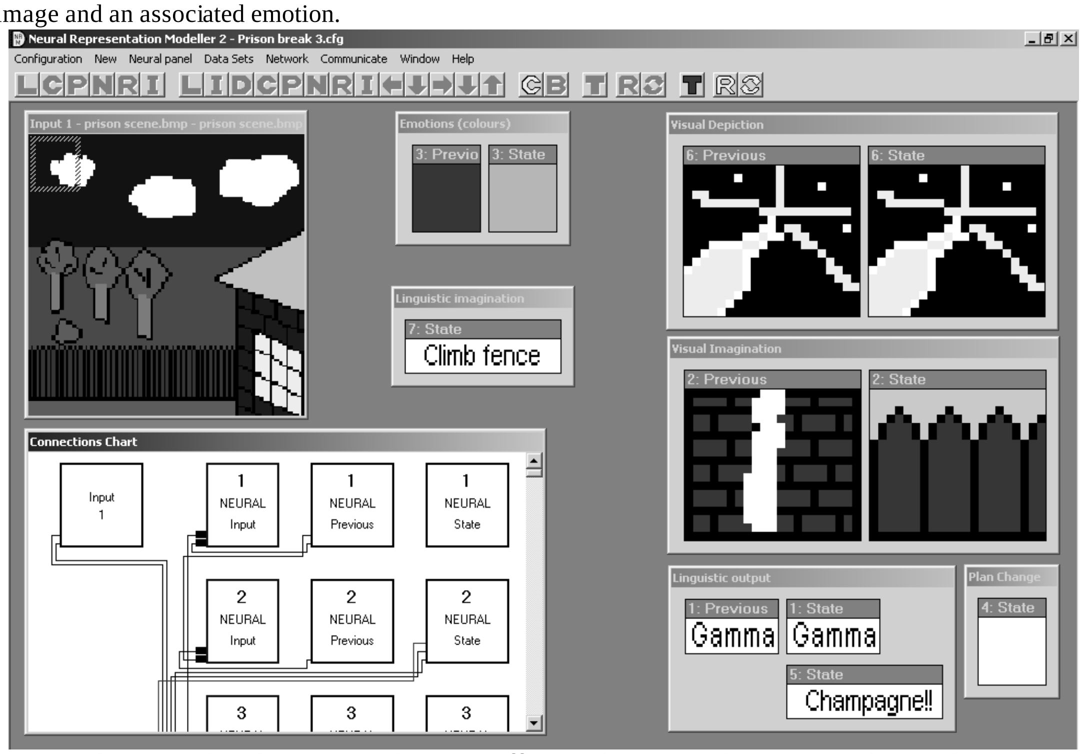

Fig. 2 The prison break scenario: the NRM affective decision making system during the trajectory exploration stage. Each pixel in the 'State' windows is the output of a single neuron.

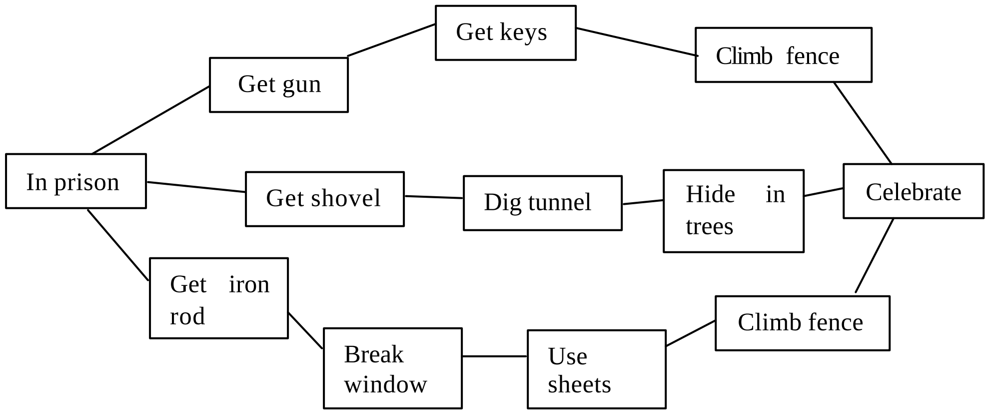

Fig. 3 The various courses of action available to the entity, during the trajectory exploration stage. The entity was taught these during training, so it had to do very little hands-on learning.

LABORATORY EXPERIMENT 2

The second simulation involved 7 neural layers, with a total of approximately 11,000 neurons (see **Fig. 4**). Neural layer B can be seen as equivalent to the dorso-lateral prefrontal cortex within the brain, since it contributes to the performance of chosen, or 'willed' actions.

The basic functioning of this system is as follows: initially, neural layer B causes the visual imagination (or the cortico-geniculate pathway in the brain) to visualise a particular course of action. When the instruction is given to act (input 3), the system actually performs the action, and through proprioceptive feedback, acquires an emotional response to the activity. This emotional response provides feedback to layer B, and either causes the entity to look for a new course of action, or to continue doing the same thing. The system diagram is shown on the next page (Fig. 5), in terms of connections between neural layers and inputs.

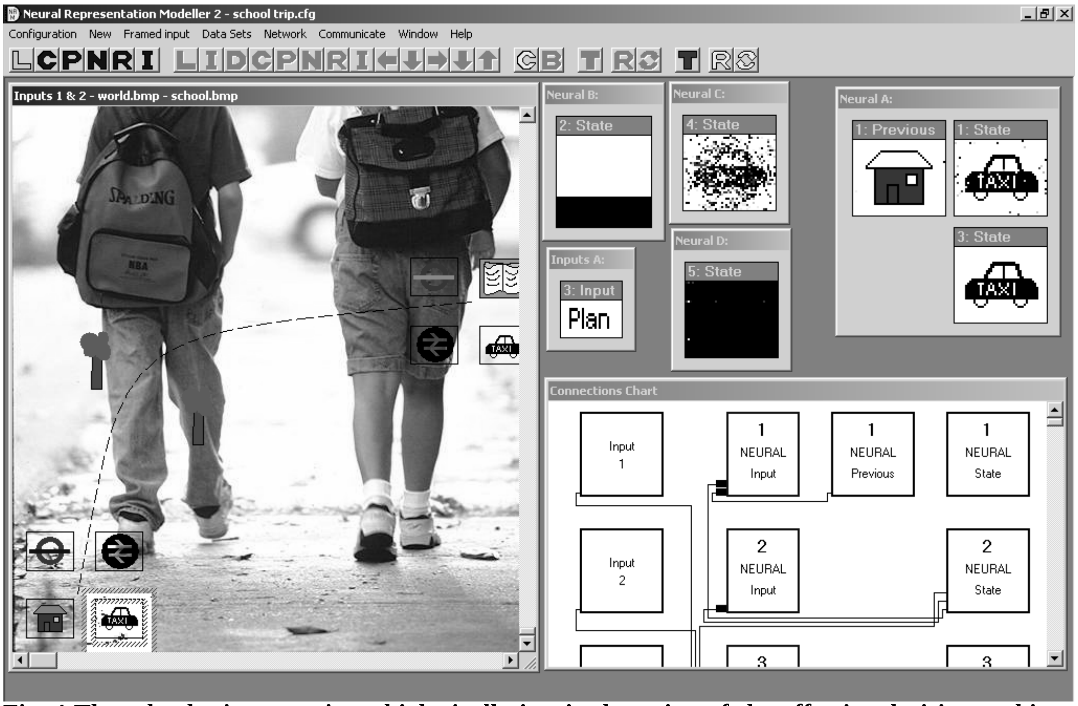

Fig. 4 The school trip scenario: a biologically inspired version of the affective decision making system. Neural layer B is equivalent to the dorso-lateral prefrontal cortex, which is responsible for voluntary selection of action in humans. Through proprioceprive feedback from the world, this layer controls the trajectory planning layers, thereby attaining the desired affective response.

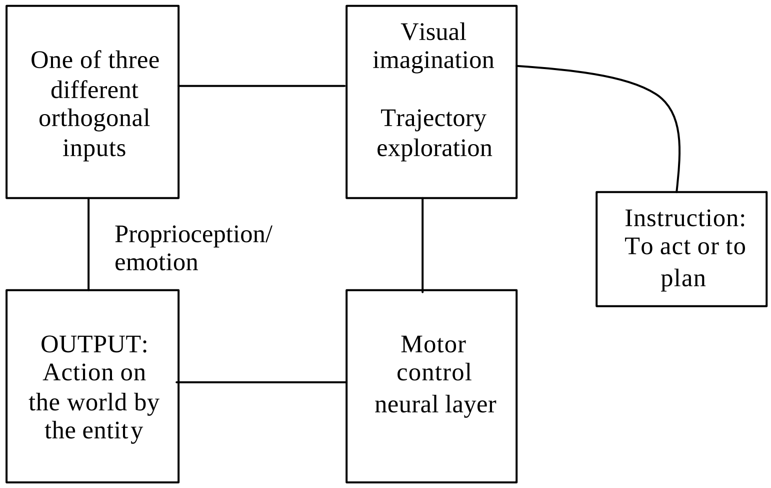

Fig. 5 System diagram of an affective decision making configuration. The feedback is crucial: it enables a bias to be formed in the making of a choice, incorporating preference and emotion into the decision.

The trajectories available to the entity are simple: it can choose to go from home to school in three different ways: by taxi, by bus or by train. There are many possible ways to make the system more complicated. For example, each journey can be associated with three variables: quality, cost and length. To make the decision as to which mode of transport to use, the entity will have to take into account its own emotional state and its fatigue level, as well as the properties of each journey mode. For example, it's quite expensive to take a taxi, but it's more convenient. If the entity is tired, it will have to make a difficult choice: should it save money and travel by bus, or should it travel in comfort? The decision will be non-deterministic, and will not be pre-programmed, and hence unpredictable – which is the way it happens in human beings. Constructing such an entity will form the goal of future work.

DISCUSSION AND CONCLUSION

Volition is an enigmatic field of research, and over the last few decades, there have been many quantitative studies made on the topic. Techniques like functional neuro-imaging have shown that a characteristic pattern of activity is associated with the experience of inducing a freely willed act. In this way, the main brain regions responsible for voluntary action have been identified and analysed. The results from these experiments can now be used to construct biologically inspired models of free action and affective decision-making – that is, decisions which are made on the basis of emotional evaluation. This paper has proposed two such models, both of which are preliminary and tentative in their scope.

The potential usefulness of research into affective autonomous behaviour is great. It can be used in robotics, or in virtual entities which require this ability. In considering the usefulness of voluntary action in machines, let us for a moment consider its usefulness in human beings. In humans of course, we actually debate the existence of free-will – but the crucial thing is, we know for sure that we at least have the illusion of possessing it. Therefore the evolution of this complex behaviour must have resulted from external environmental pressures, and it must offer some survival advantage. One conceivable benefit is that the possessing the ability to experience free-will goes hand in hand with the ability to make a distinction between the us and the external world (Spence & Frith, 1999). This is related to a second benefit: that of perceiving other entities in the world as separate to us, and attributing independent thoughts and motives to them, so that we may be able to act accordingly to survive. And thirdly, any obstacle in the world which requires us to perform a non-routine action would require us to have the capability of freely willed, voluntary movement. All-in-all, free-will is crucial to our functioning in a complex environment, and therefore research into its artificial equivalent is of paramount importance.

ACKNOWLEDGMENTS

I wish to thank Igor Aleksander for his supervision of my research work. I also thank Barry Dunmall for many hours of discussion on many wide-ranging topics, including free-will and neural architectures in NRM.

REFERENCES

Aleksander, I. (1990). "*Neural Systems Engineering: towards a unified design discipline?*" Computing and Control Journal **I**: 259-265

Aleksander, I. (1996). "*Impossible minds: my neurons, my consciousness*" London: Imperial College

Libet, B. (1999). "*Do we have free-will?*" Journal of Consciousness Studies, **6**, No. 8-9, pp. 47-57

Libet, B. (2001). "*Consciousness, free action and the brain*" Journal of Consciousness Studies, **8**, No. 8, pp. 59-65

Libet, B. (2002). "*The timing of mental events: Libet's experimental findings and their implications*" Consciousness and Cognition, **11**, pp. 291-299

Spence, S., & Frith, C. (1999) "*Towards a functional anatomy of volition*" Journal of Consciousness Studies, **6**, No. 8-9, pp. 11-29

Ingvar, D. (1999) "*On volition*" Journal of Consciousness Studies, **6**, No. 8-9, 1999, pp. 1-10