COSC451 2010 Assignment 2 (10%)

You have a choice for this assignment: either implement a computational model, or write an essay.

Choice 1: Implementation of Miller's model of PFC

Your first option is to implement Miller's model of **task set** in prefrontal cortex (PFC), as described in Miller (2000, Box 2). (See also Lecture 4.) You should model a set of input units, a set of output units, a set of pathway units, and a set of PFC units. Your model should have the following properties:

- **Behaviourally ambiguous stimuli**: there should be a stimulus $S1$ for which two different responses $R1$ and $R2$ are required in different circumstances.
- **Habitual stimulus-response patterns**: if there are no PFC units active, $S1$ should lead to $R1$.
- **Pathway-biasing PFC units**: there should be two PFC units, $PFC1$ and $PFC2$, which bias the system towards responding to $S1$ with $R1$ and $R2$ respectively.
- **Task-selecting stimuli**: there should be 'special' stimuli which have the effect of activating PFC units. $S2$ should activate $PFC1$, and $S3$ should activate $PFC2$. Once active, a PFC unit should *stay active* until another special stimulus occurs. Thus the sequence $S2$, $S1$ should result in $R1$, while the sequence $S3$, $S1$ should result in $R2$.
- **Competition between pathway units**: there should be a fourth stimulus, $S4$, which results in $R4$ if there's no PFC activity, but which tends to be 'ignored' if the PFC is activating a competing pathway unit.

Feel free to use a neural network package if you like. (E.g. the JavaNNS simulator you used in COSC343.) However, it might be just as easy for you to implement neuron units yourself.

You should also submit a report about your code. The report should contain four sections:

- How the code works;
- How to run the code (so I can test it);
- Examples of its results;
- A brief discussion of how the code can simulate a patient whose PFC is damaged, including mention of habitual stimulus-response patterns, task-switching behaviour and stimulus-bound behaviour.

Choice 2: Essay

The other option is an essay, with the following title:

> Describe how an agent's behaviour changes following damage to the prefrontal cortex (PFC), referring both to general social behaviour and clinical tests like the Wisconsin card-sorting task and the Stroop task. Outline Miller and Cohen's computational model of the PFC. Can the model account for these behavioural effects? Finally, discuss Braver and Cohen's model of how an agent learns to update PFC state in response to incoming stimuli.

To write the essay, you should use the following reading list:

- Shallice and Burgess (1991): some example case studies of people with PFC damage.
- Miller and Cohen (2001): a more detailed introduction to PFC, and to Miller's model of the PFC.
- Braver and Cohen (2000): a computational model of when/how the PFC is updated.
- For some of the background to the Braver and Cohen model, you can consult Schultz *et al.* (1997), and if you're really keen, Sutton (1988).

The essay should be around 3000–4000 words (about 8–10 pages in a 12-point font).

Submission and marking

You should submit the code and the assignment by email to me (`alik@cs.otago.ac.nz`) by 5pm on **Friday of Week 11** (i.e. on **Friday 21 May**). 10% of available marks will be deducted for each day late.

For the code assignment, I'll give equal weight to the code and the report. (But the 'report mark' is likely to reflect the quality of the code, so in that sense the implementation is the main focus.)

For the essay, I'll give marks for an essay which is well structured, presents clear models of PFC function, and shows that you have read and thought about the papers on the reading list.

References

Braver, T. and Cohen, J. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In S. Monsell and J. Driver, editors, *Attention and Performance XVIII: Control of cognitive processes*, pages 713–737. MIT Press.

Miller, E. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience*, **1**, 59–65.

Miller, E. and Cohen, J. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, **24**, 167–202.

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

Shallice, T. and Burgess, P. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, **114**, 727–741.

Sutton, R. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, **3**, 9–44.