



A RECURRENT NEURAL NETWORK MODEL OF EXECUTIVE CONTROL IN THE TOWER OF LONDON TASK.

P.A. Simen,¹ T.A. Polk,² R.L. Lewis,² and E.G. Freedman³

¹Electrical Engineering and Computer Science Dept., University of Michigan, Ann Arbor

²Department of Psychology, University of Michigan, Ann Arbor

³Department of Psychology, University of Michigan, Flint

Introduction

Cognitive deficits associated with dorsolateral prefrontal cortex (DLPFC) damage are often most apparent in higher cognitive tasks that involve problem solving and managing multiple goals. Computational models of prefrontal deficits on such tasks are difficult to construct. Problem solving is most naturally modeled with symbolic systems (e.g. production systems — see below), but the effects of lesions are most naturally modeled with subsymbolic systems (neural networks). We show that when we adopt a simple and plausible model of neural computation, there is a natural and explicit mapping from symbolic, goal-driven cognition onto neural computation. We exploit this mapping to construct a neural network model that is capable of solving complex problems in the Tower of London task. The model leads to a specific hypothesis about the role of DLPFC in such tasks: namely, that DLPFC represents internally generated subgoals that modulate competition among posterior representations. When intact, the model accurately simulates the behavior of college students even on the most difficult problems. Furthermore, when the subgoal component is lesioned, it accurately simulates the behavior of prefrontal patients, including the fact that their deficits are most apparent on the most difficult tasks.

Production Systems

Production systems are automated collections of 'if-then' rules (e.g. 'if *rain = true*, then *take umbrella*').

Typical Flow of Control in a Production System:

- Determine which rules (productions) have if-conditions that match the current state of working memory and satisfy the current goal.
- Determine which of the eligible rules should 'fire', in the event that multiple eligible rules result in conflicting actions.
- Repeat.

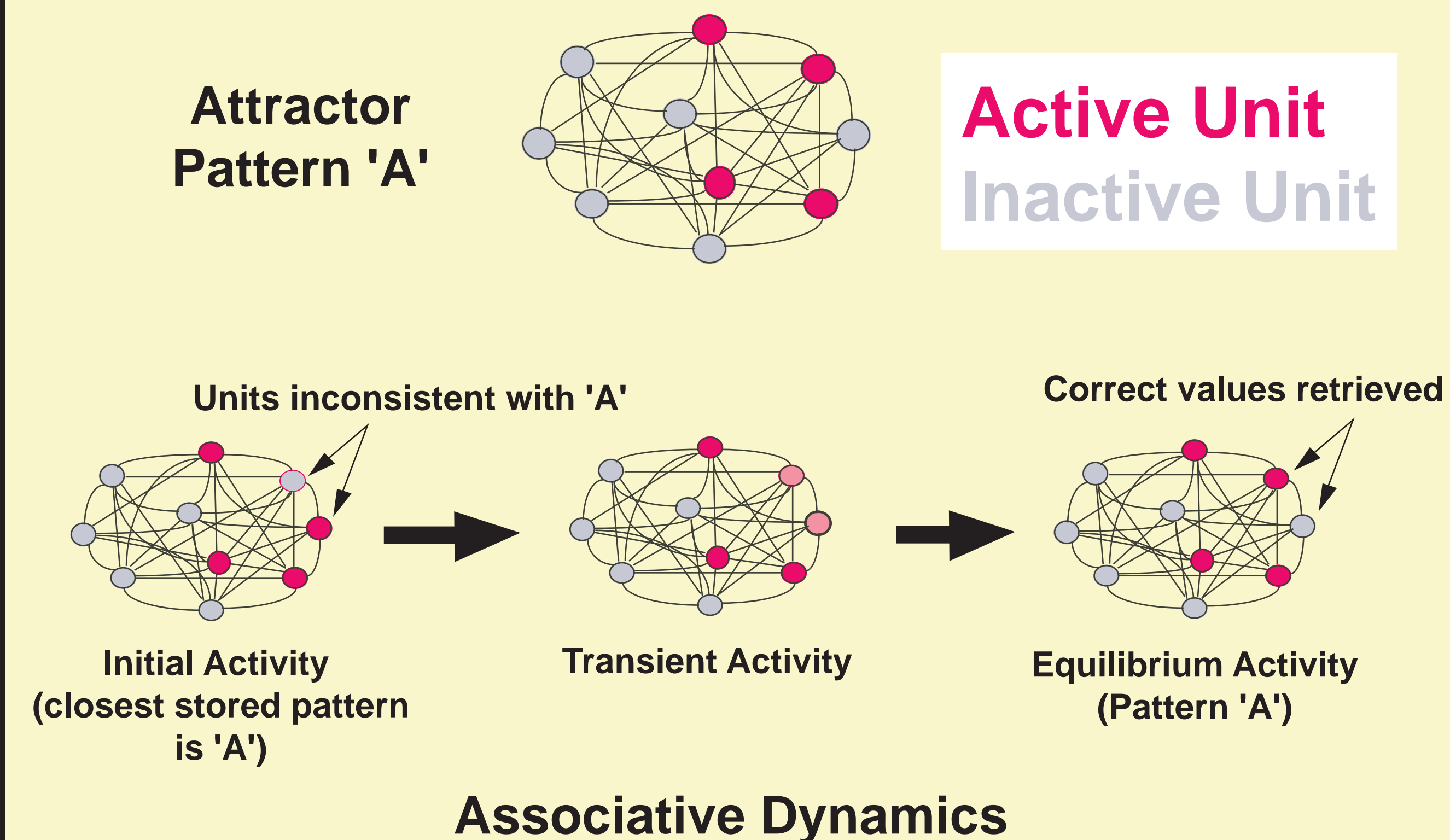
Nice properties for cognitive modeling:

- Productions provide flexible, data-driven control by constantly testing the state of the world as represented in working memory.
- Current goal provides top-down control and constrains data-driven processing toward useful ends (a natural functional explanation of goal-oriented behavior).
- Productions are associations between symbolic representations, and thus mimic associative thinking.
- It is relatively easy to construct problem-solving models with them.

Attractor Networks

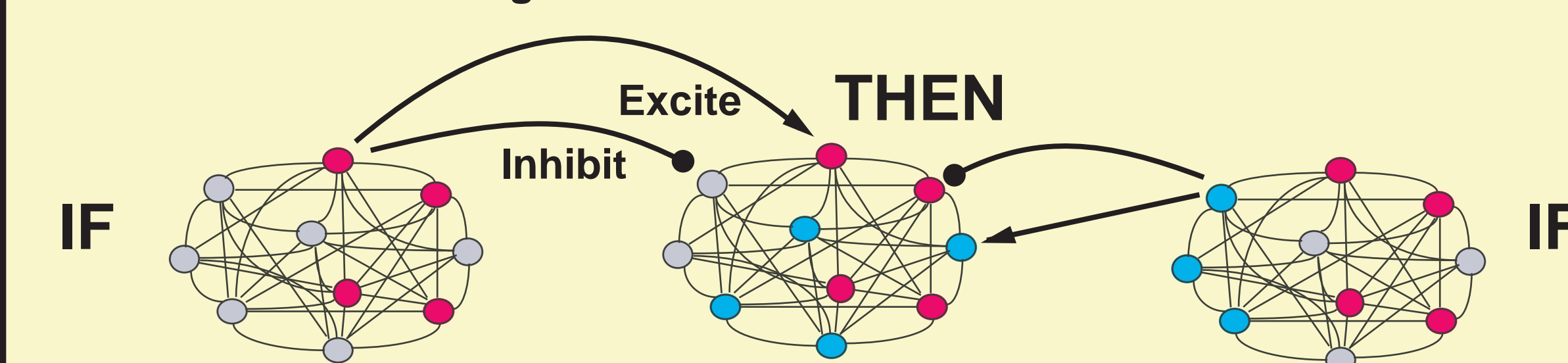
Artificial attractor networks perform useful computation by acting as 'associative memories', meaning that an item is retrieved from memory by patterns of activity that are similar to the pattern representing the item. (Hopfield, 1982)

- Recurrent connections in attractor networks are consistent with cortical connectivity.
- Attractor dynamics are consistent with primate prefrontal electrophysiological data in which some activation appears to code for working memory for items during delay tasks. (Fuster, 1997)
- Synaptic strength assignments used to store attractor patterns in the networks are consistent with strengths that Hebbian learning would produce.



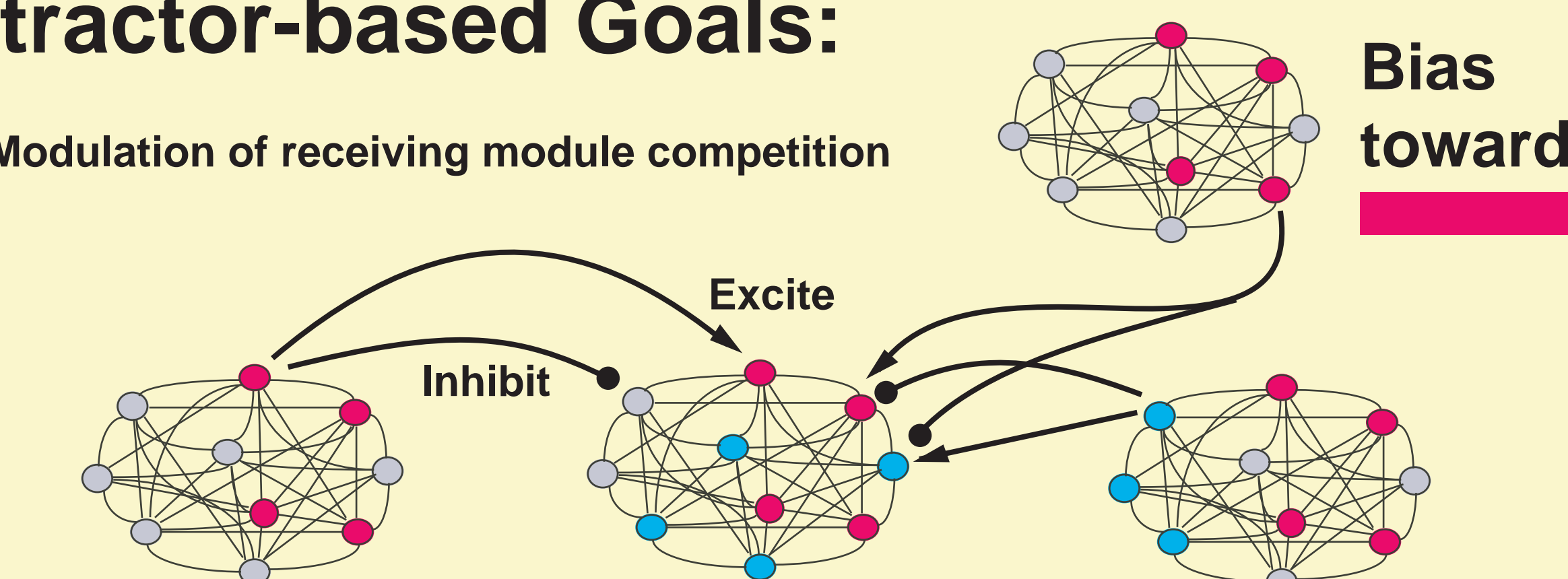
Attractor-based Productions:

- Feedforward excitation/inhibition by modular attractor networks
- Low-level conflict resolution through winner-take-all dynamics within receiving module



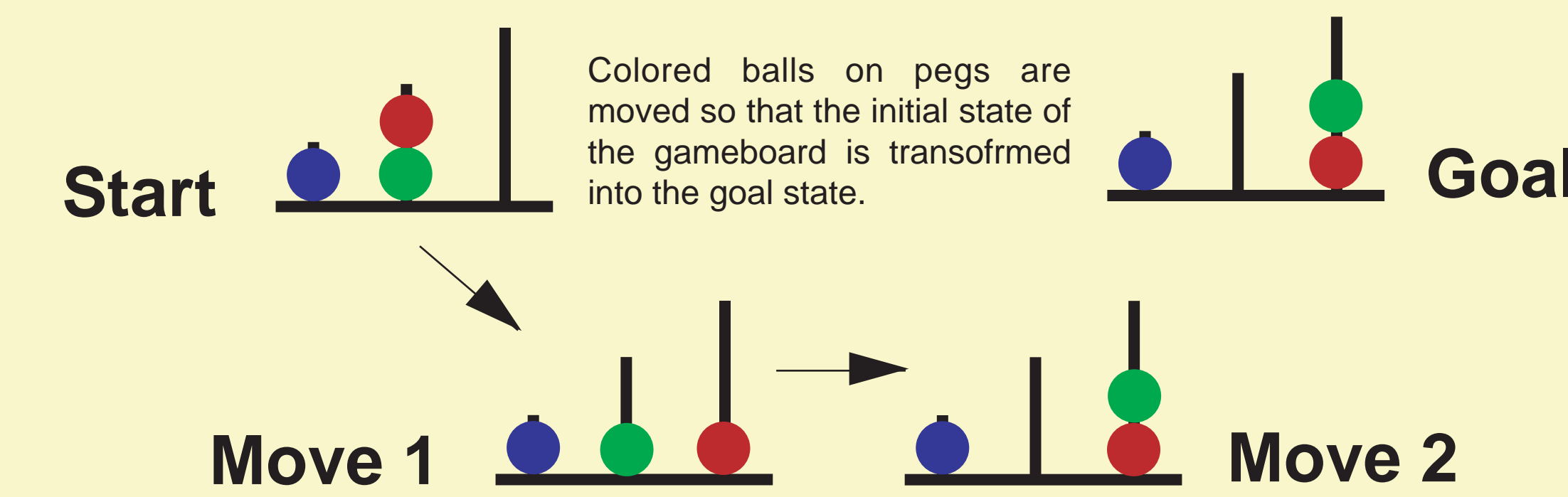
Attractor-based Goals:

- Modulation of receiving module competition



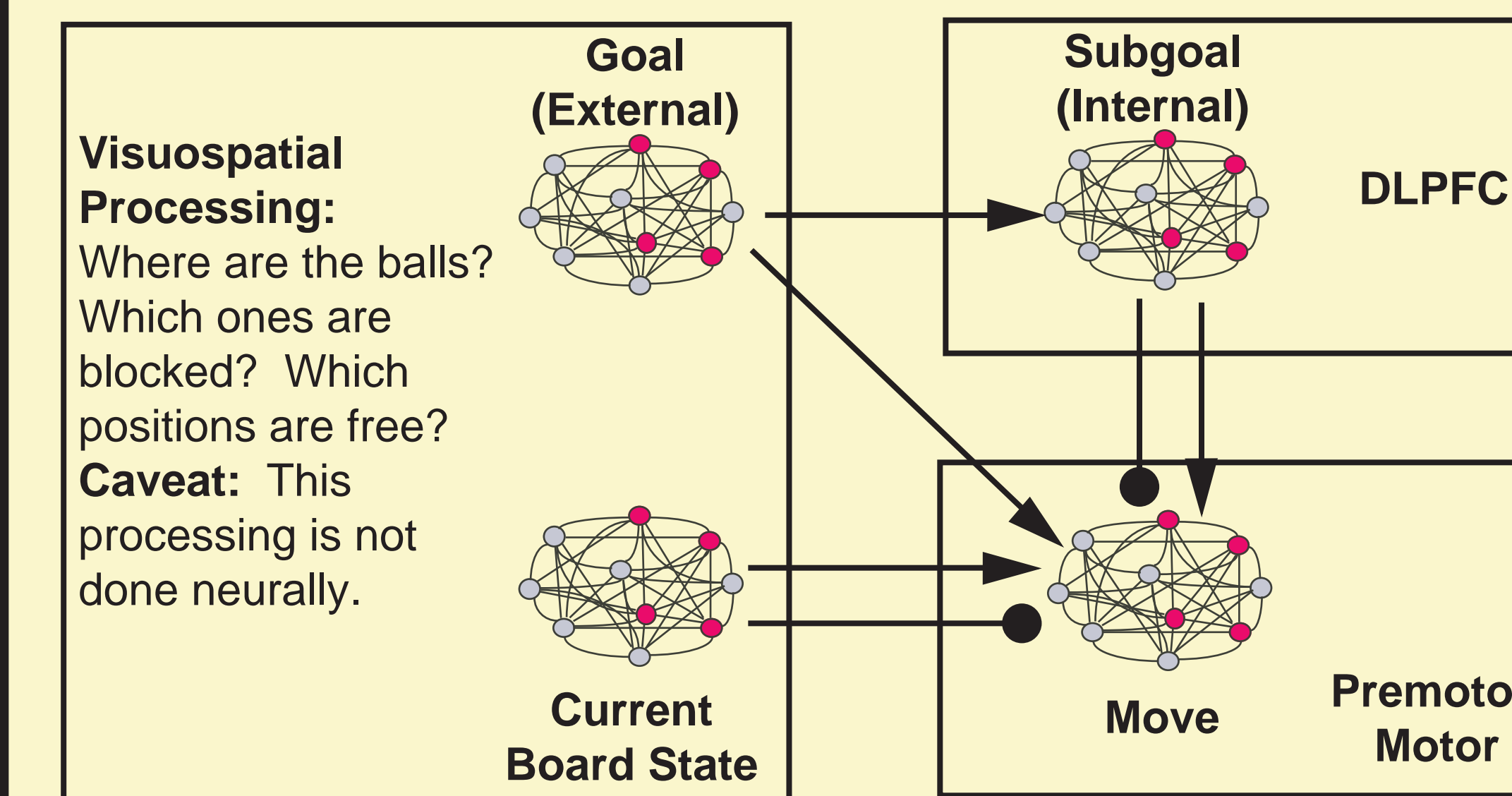
Tower of London

Task: Achieve solution in minimum number of moves.



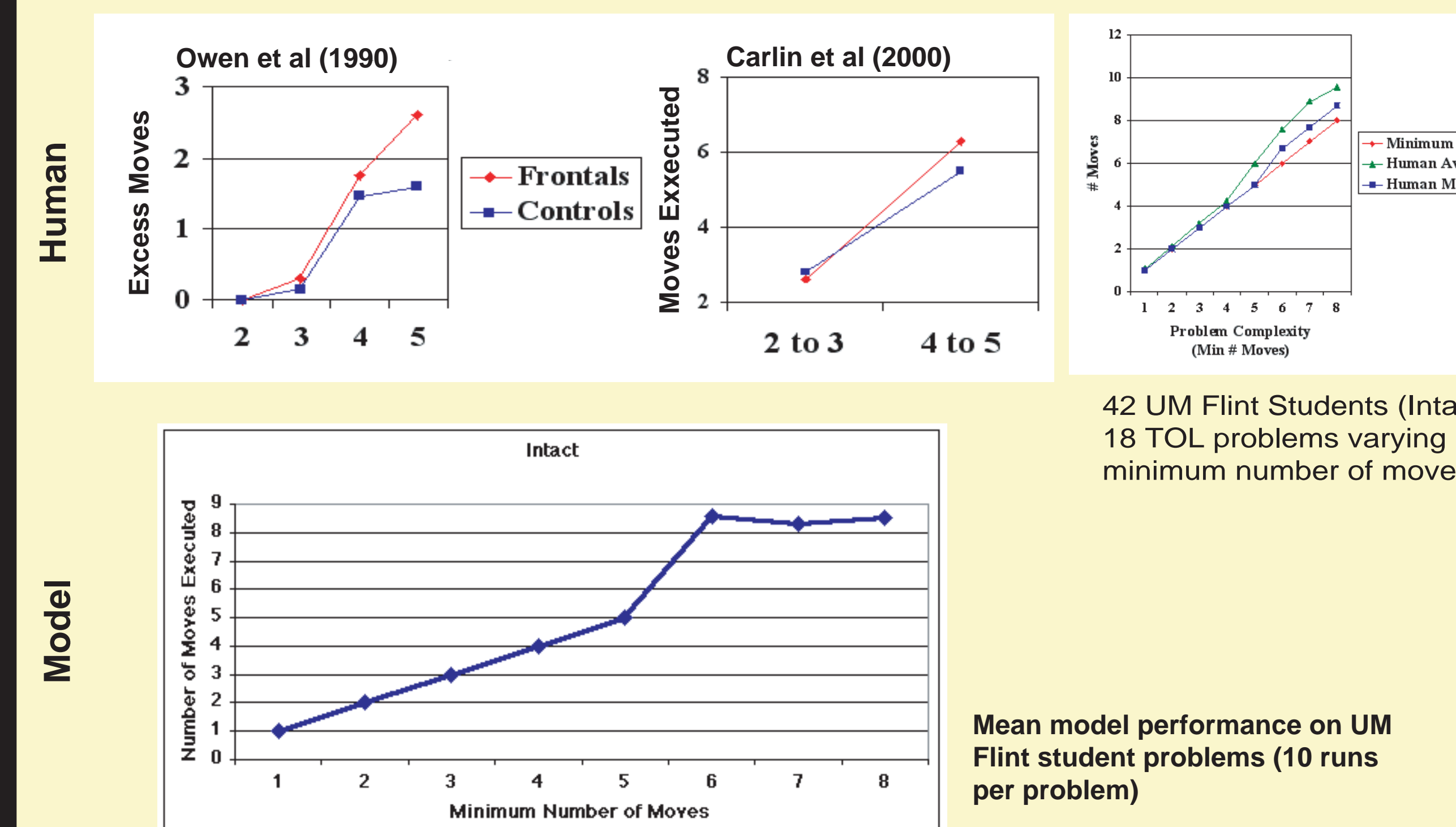
Tower of London Model

Neural model implemented in Matlab.

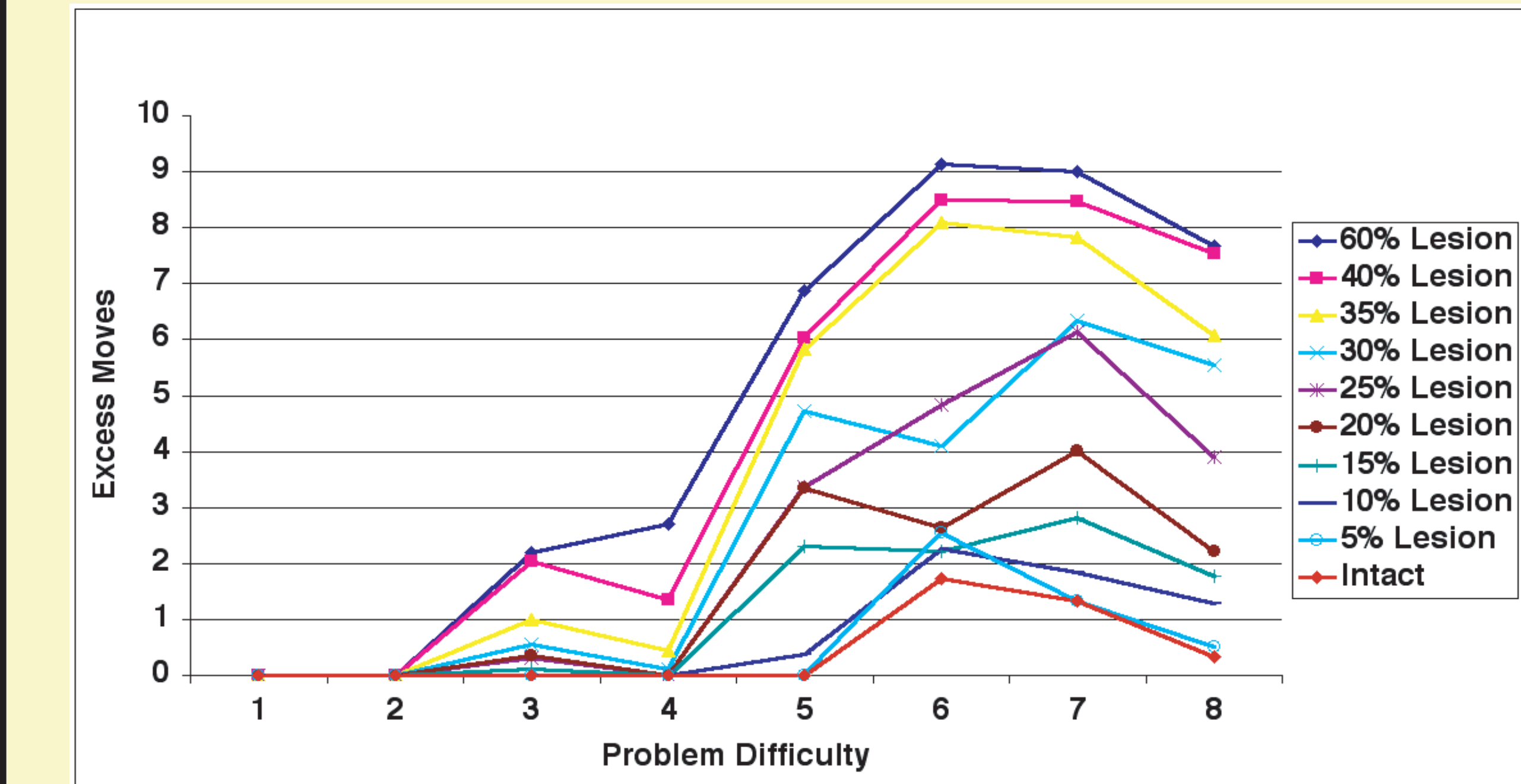


The prefrontal component of the model produces working memory for a current goal or subgoal. Each goal representation produces a corresponding control signal that biases the competition among legal move representations so that goal-achieving moves tend to be preferred.

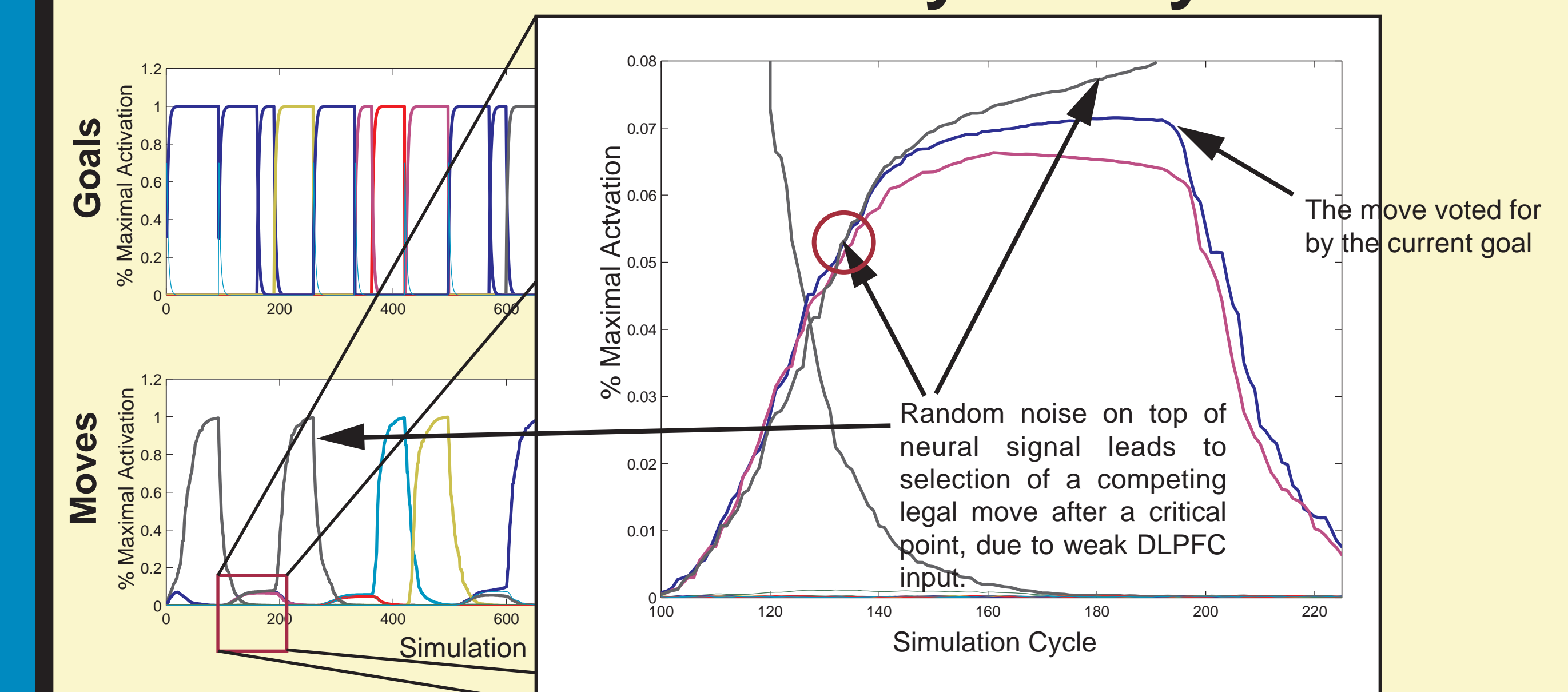
Human/Model Comparison



Graded Lesion Effects



The Source of Variability: Noisy Neurons



Discussion

The Tower of London task is a classic test of impairment in problem solving in prefrontal patients, and appears to highlight the role of DLPFC in flexible, symbolic cognitive processing. (Shallice, 1982; Owen et al, 1990) The model presented here captures both normal and patient behavior in this task. It supports a hierarchical theory of cognitive organization, in which lower-level processes are unaffected by PFC damage, but in which control implemented by PFC leads to greater flexibility. More specifically, it predicts that DLPFC is particularly critical for representing internally generated subgoals. As one of the few instantiations of a neural problem solver, the model illustrates the potential of a particular mapping of symbolic processing onto neural processing.