

1. Introduction

*“You don’t see something
until you have the right metaphor to let you perceive it”*

Robert Stetson Shaw

The study of higher cognitive functions includes the study of intelligent systems which exhibit a peculiar behavior: choosing courses of action that are relevant to achieving their goals. Cognitive science, defined as the study of intelligence and its computational processes, has been interested in the study of higher cognitive functions for more than a century. Since at least 1950, a branch of computer science called “artificial intelligence” has been studying the intelligence exhibited by machines, thereby implicitly bridging the gap between computations and behaviors exhibited by biological and non-biological systems. Although not everyone accepts the interchangeable usage of intelligence as applied both to animals and computers, the approach of this work follows the general framework of regarding intelligence as partially independent of the nature of the physical system in which it is implemented. This choice allows the experimenter interested in biological intelligence to simulate the system target of its study, therefore achieving a level of understanding which is beyond the capability of traditional, non-synthetic approaches. The objective of this work is to develop a computational model of higher cognitive functions using the connectionist paradigm. The psychological and neuropsychological literature is rich in models of higher cognitive functions and goal-oriented behavior (Posner, 1989), but a lack of explicit, mathematically defined, and biologically-plausible models is an obstacle both to the understanding of the underlying processes and to the maturation of cognitive science as a “hard” discipline. In this thesis, an emphasis is placed on the development of a *biologically-inspired* model of higher cognitive functions. In particular, great attention is devoted to a subset of neurophysiological and anatomical data which have pointed, in the recent past, to the involvement of specific cortical and subcortical areas, as well as neurotransmitters, in the control of higher cognitive functions and goal oriented behavior. In particular, the role of the neurotransmitter dopamine (DA) in the maintenance of short-term memory in prefrontal cortex (Pfc) neurons has been a subject of interest for the physiological, neuropsychological, neuropathological, as well as for the modeling community (Lidow et al., 1998; Goldman-Rakic et al., 2000). Such a convergence of interests is justified by the importance of Pfc and DA functioning in various high-level tasks, which are usually grouped under the broader category of “cognitive control”, as well as by the fact that these tasks are the ones more often impaired in various neurological and psychopathological conditions (schizophrenia and Parkinson disease, among others). A crucial role in the physiological functioning of Pfc is played by DA. Numerous human and animal studies have

demonstrated that the integrity of the main DAergic pathway (Ventral Tegmental Area - VTA - to Pfc) is crucial for the normal functioning of Pfc (Goldman-Rakic et al., 2000). The first part of this thesis reviews some of the main theories and models regarding the presumed role of DA in Pfc functioning and cognitive control. Tasks of different complexities are simulated in the attempt to understand how a biologically-plausible model can control the complex interactions typical of the behavior which require cognitive control. The main emphasis of the work is placed in how these properties can spontaneously arise in a system that is not designed “ad-hoc” to cope with the task. In order to fit this requirement, the model is developed step by step, with increasingly difficult tasks and with a parallel growth of model complexity. This process allows the reader to understand the need of progressive degree of complexity that the model should incorporate in order to cope with progressively increasing “cognitive functions”, and to solve otherwise unsolvable paradoxes that arises when a simple model is stretched trying to cope with complicated behavioral problems. An analogy is drawn between the rather abstract concept of cognitive control and simpler, well-known classical conditioning paradigms, which are suggested to be the prototype of these higher, articulated function. The biological grounds of the model are then presented, and a surprisingly large dataset of known anatomical and physiological data is shown to fit with the model, despite the fact that the design of the system was guided by theoretical reasons, not by the primary target of fitting neurobiological data. The model presented in the first part of this work will show how a field of Pfc neurons can store a given Short Term Memory (STM) representation and at the same time face the presentation of several distractors interleaved between the to-be-stored cue and the release of the action. The first set of simulations, however, leaves open the question of **how does the organism learn what is relevant in obtaining the reward, namely how to selectively attend to the relevant stimuli linked to the reward and to appropriately time the release of the behavior which is functional to its obtainment.** Furthermore, **how does the brain solve the paradox of being unable to store Pfc representations unless these are followed by a DAergic response, and at the same time learn contingencies between stimuli that were presented (and probably decayed) long *before* the reward was generated?** The model developed in the second part addresses the above issues. In particular, the focus is on how an organism can learn **a)** the delayed consequences of its behavior and **b)** to adaptively time the release of its behavior in order to obtain a reward. These two properties of a system are the foundations of what in cognitive science is known as of “cognitive control”, and are features of higher cognitive functions.

Summarizing, this work is an attempt to attack the problem of how the functions described in a) and b) can autonomously develop as an emergent property of a neural network model. The system developed in this work is designed to cope with a simplified task that embodies a delay between a stimulus, a reward and a response and a delay between a cue and the release of the appropriate action. It

is shown how the performance of an apparently simple task, carried out in a apparently effortless fashion even by non-primates, requires an unexpectedly complex neural machinery. The work will show how the interplay of computational, behavioral, and neurophysiological constraints can lead to the development of a coherent and ecologically plausible model of how higher cognitive functions can develop from a biologically-plausible neural model. These characteristics of the model are a truly unique feature of this work, and differentiate it with respect to the major models published so far.