Adam Mickiewicz University in Poznań
Faculty of Psychology and Cognitive Sciences
Marcin Cichosz, MA

THESIS SUMMARY

Integrated model of complex intentional actions

The dissertation addresses the issue of modeling intentional actions.

The purpose of this dissertation is to construct an integrated model of intentional actions (IMIA). It required the identification of the most important mechanisms and representations determining the course of intentional actions and the establishment of essential links between them. The following theoretical concepts and experimental results were used in the construction of the model: (1) John Searle's theory of intentionality, (2) experimental data collected within the psychology of intention, and (3) a computational model of reinforcement learning developed by neuroscientists in relation to the dopaminergic reward prediction error hypothesis. The proposed approach reveals the relationships among the various components of intentional action, as well as the mechanisms of behavioral control and the cognitive processes that shape them.

The dissertation consists of five chapters.

The first chapter shows how scattered, heterogeneous, and fragmented the theoretical and experimental studies of intentional actions are. While there have been relatively few attempts to systematize the collected material, they are limited to single-domain models. The study presents arguments justifying the need for a model that integrates knowledge about intentional actions obtained in different research fields. The subject of this chapter is complex intentional actions conceived as a system of two interacting mechanisms: (a) reinforcement learning and (ii) planning based on knowledge accumulated in the form of networks of intentional states.

Determining the influence of intentional states on behavior choice is the subject of **the second chapter** of the dissertation. It is based on John Searle's interpretation of intentionality. Drawing inspiration from the American philosopher's theory, the category of intentionality and

the key types of intentional states (prior intentions, intentions in action and beliefs) have been defined and incorporated into an integrated model of intentional action. The result of these considerations is a conceptual framework of intentional action which serves as a reference in subsequent chapters of the dissertation.

The third chapter refers to the neurobiological mechanism responsible for organizing behavior into ordered sequences. The core of this type of mechanism, according to the dopaminergic reward prediction error hypothesis (HDBPN), is the Temporal Difference Reinforcement Learning (TDRL) algorithm, implementing the reinforcement learning method. In this chapter, I discuss the principle of operation and key features of the TDRL algorithm. Analyses of the computational model considering HDBPN lead to the determination of the scope of its applicability to complex intentional actions. The starting point for these analyses is the "superpower" hypothesis of Read Montague, by means of which the American neuroscientist explains human behavior that violates the commonly accepted principle of dominance of the survival instinct.

The fourth chapter consists analyzes of the most important experimental results gathered by psychologists of intention. The data brought to light provide insight into the structure of Searle's so-called “intention in action” and its "phenomenal milieu". The most important part of this chapter is the analysis of the functional aspects of such states as (1) sense of urge, (2) reference forward to the goal object or event and (3) sense of agency.

In **the fifth chapter**, I present and discuss in detail an integrated model of complex intentional actions. The model is presented in two stages. The first stage is devoted to formulating the most important functional requirements for this model. The second stage presents the integrated model of intentional action using an iterative approach. The simplest model, using only the reinforcement learning mechanism to control behavior, is characterized first. Subsequently, increasingly sophisticated models are discussed, which make it possible to incorporate additional features of real-life intentional action. The final version of the model shows how complex the structures and mechanisms of intentional actions are, and the extent to which this picture differs from the standard mono-domain account of these actions.