

## Abstract

*Approved for Public Release*

The goal of this project is to integrate state-of-the-art advances in the brain, cognitive and computer sciences, to build a coherent, computationally explicit model of how the human brain gives rise to its remarkable — and still unique — flexibility and autonomy of function. Artificial systems have now achieved levels of performance that match or exceed humans in domains previously considered to be the exclusive province of the human brain, such as face recognition, game playing, and some natural language processing tasks. However, no existing system approaches the competence of the human brain in as *many* domains, its *flexibility* to rapidly acquire competence in new ones, nor its *autonomy* in deciding on its own when and how to do so. Advances in brain and cognitive science have generated computationally explicit models for many of the subsystems thought to contribute to these remarkable abilities, including mechanisms underlying object recognition, associative and supervised learning, decision making, attention, working memory, and simple forms of cognitive control. However, these subsystems have been modeled largely independently of one another, and none alone can account for the flexibility and autonomy of human behavior. This project aims to take on the challenge of explaining these phenomena in terms of interactions among known subsystems, by integrating existing models into a single, coherent system; bringing them under the regulation of adaptive mechanisms of control; and testing the ability of this system to function and adapt in a wide range of laboratory tasks and more complex, naturalistic environments.

The proposed work is organized around three closely related objectives. The first will address the ability to form and use abstract, general purpose representations required for the flexibility of behavior. The second will address the ability to strategically control behavior required for autonomy of function, including the ability of a system to acquire “meta-knowledge” of its own capabilities, and use this to direct future learning. This work will integrate subsets of existing models that address simple, well-controlled laboratory tasks to address more complex ones, and then integrate these into a more comprehensive model to address performance in more naturalistic settings. Finally, the third objective, pursued in interaction with the first two, will address the methodological needs for model integration and testing in more complex and naturalistic environments, such as online, realtime games, that approximate the complexity of real world settings, while providing sufficient structure and access to detailed quantitative data required to support scientific research.

This work promises to produce transformative advances in our understanding of the unique flexibility and autonomy of the human brain, which is unlikely to be explained by models of isolated functions or capabilities. Rather, as with other complex devices made up of closely interacting components parts, such as airplanes or computers, full understanding requires building and analyzing complete prototypes, that allow components to interact with one another, and testing these in realistically complex environments.

Building a model that explains the mechanisms in the brain that give rise to these capabilities would have fundamental value, with consequences directly relevant to DoD (such as more sophisticated approaches to human factors design, and applications to human computer interaction [HCI], brain computer interfaces [BCI], and skill acquisition) as well as the design of artificial systems that better approximate the flexibility and autonomy of human behavior. It would also have obvious and profound consequences for other fields concerned with human behavior, such as psychiatry and neurology, economics, and sociology.