Foreword

The general idea that brains anticipate the future, that they engage in prediction, and that one means of doing this is through some sort of inner model that can be run offline, has a long history. Some version of the idea was common to Aristotle, as well as to many medieval scholastics, to Leibniz and Hume, and in more recent times, to Kenneth Craik and Philip Johnson-Laird. One reason that this general idea recurs continually is that this is the kind of picture that introspection paints. When we are engaged in tasks it seems that we form images that are predictions, or anticipations, and that these images are isomorphic to what they represent.

But as much as the general idea recurs, opposition to it also recurs. The idea has never been widely accepted, or uncontroversial among psychologists, cognitive scientists and neuroscientists. The main reason has been that science cannot be satisfied with metaphors and introspection. In order to gain acceptance, an idea needs to be formulated clearly enough so that it can be used to construct testable hypotheses whose results will clearly support or cast doubt upon the hypothesis. Next, those ideas that are formulable in one or another sort of symbolism or notation are capable of being modeled, and modeling is a huge part of cognitive neuroscience. If an idea cannot be clearly modeled, then there are limits to how widely it can be tested and accepted by a cognitive neuroscience community. And finally, ideally, the idea will be articulated and modeled in such a way that it is not a complete mystery how it could be implemented by the brain. Though the idea that the brain models and predicts and anticipates is supported by introspection and a long history of hypotheses, it has largely failed on these latter three counts – especially compared with various theoretical competitors. And this is why the extent to which it has been embraced by cognitive science and neuroscience has been limited.

But there is good news. Mathematical tools from a number of areas, including modern control theory and signal processing, are capable of allowing for very precise mathematical formulations of the basic idea, as well as many specific versions. This allows for the ideas not only to be precisely formulated, but also to be modeled and compared to human behavioral data. And given a number of schemes for implementing these kinds of mathematical models in neural systems, it is possible to see these models as being implemented in the brain. The qualitative idea that the brain models the world is finally being clarified and quantified.

But we are still in the early stages of this process. While there are many proposals and theories that are beginning to take shape, there have been few sustained treatments of the topic that attempt to develop them in detailed and consistent ways. Rather, the applications have largely been piecemeal. In this regard the present volume represents a significant advance in the field. It offers a sustained treatment of various aspects of the general hypothesis, not only in terms of being conceptually clear and consistent, but also in terms of presenting a wide range of particular applications that illustrate the conceptual machinery in action.

It would be an overstatement to say that the idea that the brain is a modeler and predictor is revolutionary, or that the current swell in theoretical interest in the idea represents the initial stages of a revolution in cognitive neuroscience. But while talk of revolution may be overstatement, it cannot be denied that this approach to understanding brain function is beginning to take on an importance comparable to that of traditional artificial intelligence approaches and connectionist modeling approaches. The clarity, detail and quality of the ideas presented in this volume, coupled with the growing importance of this general approach, make this volume a critical contribution to our understanding of brain function, and should be read by anyone with a serious interest in understanding how the brain manages to support cognitive functions.

July 2008

Rick Grush
University of California, San Diego

Preface

Prediction is difficult – especially for the future. *Niels Bohr*

Over the last few decades, it has become increasingly clear that animals most of the time do not simply react in their world based on unconditioned or conditioned stimuli, but rather actively operate in their environment in a highly goal- and future-oriented way, and not just on the basis of current perception, but in part autonomously from environmental stimuli. Psychology now suggests that it is the goal itself that triggers behavior and attention. Learning is highly influenced by current predictive knowledge and the consequent detection of novelty. Behavioral control is most effectively controlled by the help of forward models that substitute delayed or that enhance noisy perceptual feedback. Thus, anticipations come in many forms and influence many cognitive mechanisms.

This book proposes a unifying approach for the analysis and design of artificial cognitive systems: **The anticipatory approach**. We propose a foundational view of the importance of dealing with the future, of gaining some autonomy from current environmental data, of endogenously generating sensorimotor and abstract representations. We propose a meaningful taxonomy of anticipatory cognitive mechanisms, distinguishing between the types of predictions and the different influences of these predictions on actual behavior and learning. Doing so, we sketch out a new, unifying perspective on cognitive systems. Mechanisms, that have often been analyzed in isolation or have been considered unrelated to each other, now fit into a coherent whole and can be analyzed in correlation to each other. Learning and behavior are considered increasingly intertwined and correlated with each other. Attention and action control suddenly appear as very similar processes. Goal-oriented behavior, motivation and emotion appear as related and intertwined.

While the revelation of these correlations is helpful for the analysis and comparison of different learning and behavioral mechanisms, the second benefit of the anticipatory approach is the possibility to modularly design novel cognitive system architectures. The developed taxonomy clearly characterizes which aspects are important for different anticipatory cognitive modules and how these modules may interact with each other. Thus, the second benefit of the anticipatory approach is the facilitation of cognitive system design. Building blocks of cognitive systems are proposed and exemplarily analyzed in diverse system architectures. The interaction of these building blocks then is characterized by their anticipatory nature, facilitating the design of larger, more competent autonomous artificial cognitive systems. We hope that the proposed anticipatory approach may thus not only serve for the analysis of cognitive systems but rather also as an inspiration and guideline for the progressively more advanced and competent design of large, but modular, artificial cognitive systems.

Acknowledgments

This work is supported by the EU project **MindRACES**, from Reactive to Anticipatory Cognitive Embodied Systems, funded under grant FP6-511931 under the "Cognitive Systems" initiative from the EC. Special thanks to our Project Officer, Cécile Huet, and to our Project Reviewers, Lola Cañamero and Deepak Kumar, for their valuable encouragement and advice.

July 2008

Giovanni Pezzulo
Martin V. Butz
Cristiano Castelfranchi
Rino Falcone

http://www.springer.com/978-3-540-87701-1

The Challenge of Anticipation
A Unifying Framework for the Analysis and Design of Artificial Cognitive Systems
Pezzulo, G.; Butz, M.V.; Castelfranchi, C.; Falcone, R. (Eds.)
2008, XVI, 288 p., Softcover
ISBN: 978-3-540-87701-1