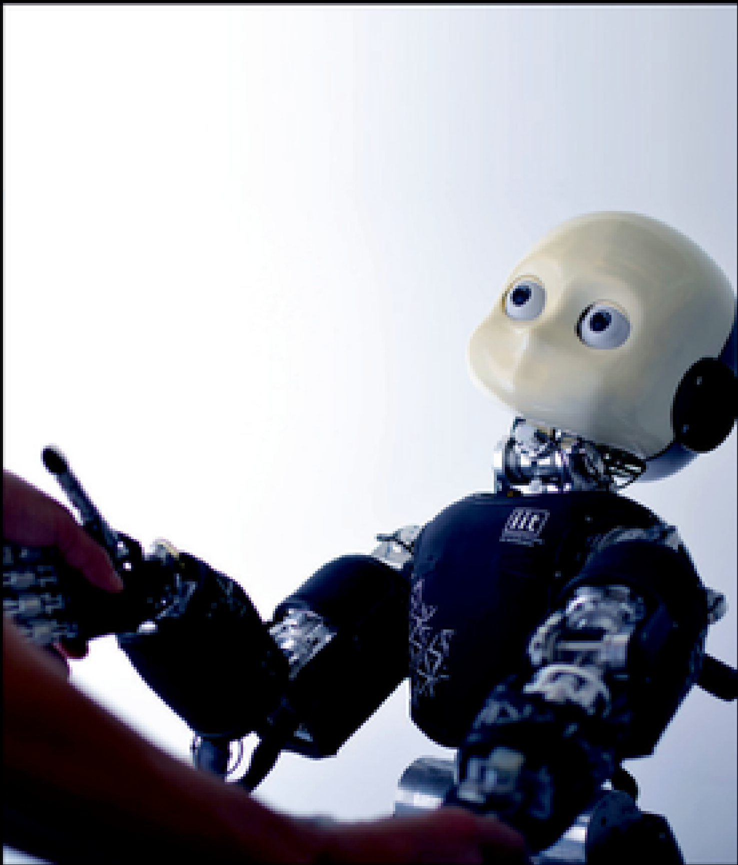

OXFORD SERIES ON COGNITIVE MODELS AND ARCHITECTURES

EXPLORING ROBOTIC MINDS

Actions, Symbols, and Consciousness
as Self-Organizing Dynamic Phenomena

JUN TANI

OXFORD

Exploring Robotic Minds

OXFORD SERIES ON COGNITIVE MODELS AND ARCHITECTURES

Integrated Models of Cognitive Systems
Edited by Wayne D. Gray

In Order to Learn: How the Sequence of Topics Influences Learning
Edited by Frank E. Ritter, Joseph Nerb, Erno Lehtinen, and Timothy O'Shea

How Can the Human Mind Occur in the Physical Universe?
By John R. Anderson

Principles of Synthetic Intelligence PSI: An Architecture of Motivated Cognition
By Joscha Bach

The Multitasking Mind
By David D. Salvucci and Niels A. Taatgen

How to Build a Brain: A Neural Architecture for Biological Cognition
By Chris Eliasmith

Minding Norms: Mechanisms and Dynamics of Social Order in Agent Societies
Edited by Rosaria Conte, Giulia Andrighetto, and Marco Campennì

Social Emotions in Nature and Artifact
Edited by Jonathan Gratch and Stacy Marsella

Anatomy of the Mind: Exploring Psychological Mechanisms and Processes with the Clarion Cognitive Architecture
By Ron Sun

Exploring Robotic Minds: Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena
By Jun Tani

Exploring Robotic Minds

Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena

Jun Tani

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

Library of Congress Cataloging-in-Publication Data
Names: Tani, Jun, 1958– author.
Title: Exploring robotic minds : actions, symbols, and consciousness as self-organizing dynamic phenomena / Jun Tani.
Description: Oxford; New York: Oxford University Press, [2017] |
Series: Cognitive models and architectures | Includes bibliographical references and index.
Identifiers: LCCN 2016014889 (print) | LCCN 2016023997 (ebook) |
ISBN 9780190281069 (hardcover : alk. paper) | ISBN 9780190281076 (UPDF)
Subjects: LCSH: Artificial intelligence. | Robotics. | Cognitive neuroscience.
Classification: LCC Q335 .T3645 2017 (print) | LCC Q335 (ebook) |
DDC 629.8/9263—dc23
LC record available at https://lccn.loc.gov/2016014889

9 8 7 6 5 4 3 2 1

Printed by Sheridan Books, Inc., United States of America

Contents

Foreword

Frank E. Ritter

This book describes the background and results from Jun Tani's work of over a decade of building robots that think and learn through interaction with the world. It has numerous useful and deep lessons for modelers developing and using symbolic, subsymbolic, and hybrid architectures, so I am pleased to see it in the *Oxford Series on Cognitive Models and Architectures*. It is work that is in the spirit of Newell and Simon's (1975) theory of empirical exploration of computer science topics and their work on generation of behavior, and also takes Newell and Simon's and Feynman's motto of understanding through generation of behavior seriously. At the same time, this work extends the physical symbol hypothesis in a very useful way by suggesting by example that the symbols of human cognition need not be discrete symbols manually fed into computers (which we have often done in symbolic cognitive architectures), but can instead be composable neuro-dynamic structures arising through iterative learning of perceptual experience with the physical world.

Tani's work has explored some of the deep issues in embodied cognition, about how interaction with the environment happens, what this means for representation and learning, and how more complex behavior can be created or how it arises through more simple aspects. These lessons include insights about the role of interaction with the environment, consciousness and free will, and lessons about how to build neural net architectures to drive behavior in robots.

The book starts with a review of the foundations of this work, including some of the philosophical foundations in this area (including the symbol grounding problem, phenomenology, and the role of time in thinking). It argues for a role of hierarchy in modeling cognition, and for modeling and understanding interaction with an external world. The book also notes that state space attractors can be a useful concept in understanding cognition, and, I would add, this could be a useful additional way to measure fit of a model to behavior. This review also reminds us of areas that current symbolic models have been uninformed by—I don't think that these topics have been so much ignored as much as put on a list for later work. These aspects are becoming more timely, as Tani's work shows they can be. The review chapters make this book particularly useful as an advanced textbook, which Tani already uses it for.

Perhaps more importantly, in the second half of the book (Chapters 6 to 11) Tani describes lessons from his own work. This work argues that behavior is not always programmed or extant in a system, but that it can or often should arise in systems attempting to achieve homeostasis—that there are positions of stability in a mental representation (including modeling others, imitation), and that differences in knowledge between the levels can give rise to effects that might be seen to be a type of consciousness, a mental trace of what lower levels should do or are doing, or explanations of what they have done based on predictions of the agent's own behavior, a type of self-reflexive mental model. These results suggest that more models should model homeostasis and include more goals and knowledge about how to achieve it.

His work provides another way of representing and generating behavior. This way emphasizes the dynamic behavior of systems rather than the data structures used in more traditional approaches. The simple ideas of evolution of knowledge, feedback, attractors, and further concepts provide food for thought for all systems that generate behavior. These components are reviewed in the first part of the book. The second part of the book also presents several systems used to explore these ideas.

Lessons from this book could and should change how we see all kinds of cognitive architectures. Many of these concepts have not yet been noticed in symbolic architectures, but they probably exist in them. This new way to examine behavior in architectures has provided insights already about learning and interaction and consciousness. Using these concepts in existing architectures and models will provide new insights into how compositional thoughts and actions can be generated without facing the notorious problems of the symbol grounding problem or, ultimately, the mind–body problem.

In his work about layers of representation, he has seen that higher levels might not just lead the lower levels, but also follow them, adjusting their own settings based on the lower levels' behavior. An interpretation of the higher levels trying to follow or predict the lower levels provides a potential computational description and explanation of some forms of consciousness and free will. I found these concepts particularly intriguing. Not only that higher levels could follow and not lead lower levels, but that the mismatch could lead to a kind of postdiction in which intention becomes consciously aware after action. We might see this elsewhere as other architectures, their environments, and their interaction with the environment become more complex, and indeed should look for it.

I hope you find the book as useful and suggestive of new areas of work and new aspects of behavior to consider for including in architectures as I have.

Preface

The mind is ever elusive, and imagining its underlying mechanisms remains a constant challenge. This book attempts to show a clear picture of how the mind might work, based on tangible experimental data I have obtained over the last two decades during my work to construct the minds of robots. The essential proposal of the book is that the mind is comprised of *emergent phenomena*, which appear via intricate and often conflictive interactions between the top-down subjective view for proactively acting on the external world and the bottom-up recognition of the resultant perceptual reality. This core idea can provide a scaffold to account for the various fundamental aspects of the mind and cognition. Allowing entangling interactions between the top-down and bottom-up processes means that the skills we need to generate complex actions, knowledge, and concepts for representing the world and the linguistic competency we need to express our experiences can naturally develop—and the *cogito*[1] that allows this "compositional" yet fluid thinking and action appears to be embedded in dynamic neural structures.

The crucial argument here is that this cogito is free from the problems inherent in Cartesian dualism, such as that of interaction and how a nonmaterial mind can cause anything in a material body and world, and vice versa. We avoid such problems because the cogito embedded

1. Cogito is from a Latin philosophical proposition by Rene Descartes "Cogito ergo sum," which has been translated as "I think, therefore I am." Here, cogito denotes a subject of cognizing or thinking.

in the continuous state space of dynamic neural systems is also *matter*, rather than nonmatter composed of a discrete symbol system or logic. Therefore, the cogito can interact *physically* with the external world: As one side pushes forward a little, the other side pulls back elastically so that a point of compromise can be found in conflictive situations through iterative dynamics. It is further proposed that even the nontrivial problem of consciousness (what David Chalmers has called the *hard problem of consciousness*) and free will can become accessible by considering that consciousness is also an emergent phenomenon of *matter* arising inevitably from such conflictive interactions. The matter here is alive and vivid in never-ending trials by the cogito to comprehend an ever-changing reality in an open-ended world. Each of these statements—my proposals on the workings of the mind—will be examined systematically by reviewing multidisciplinary discussions, largely from the fields of neuroscience, phenomenology, nonlinear dynamics, psychology, cognitive science and cognitive robotics. Actually, the book aims for a unique way of understanding the mind from rather an unordinary but inspiring combination of ingredients such as humanoid robots, Heidegger's philosophy, deep learning neural nets, strange attractor from chaos theory, mirror neurons, Gibsonian psychology, and more.

The book has been written with a multidisciplinary audience in mind. Each of the chapters start by presenting general concepts or tutorials on each discipline—cognitive science, phenomenology, neuroscience and brain science, nonlinear dynamics, and neural network modeling—before exploring the subjects specifically in relation to the emergent phenomena which I believe constitute the mind. By providing a brief introduction to each topic, I hope that a general audience and undergraduate students with a specific interest in this subject will enjoy reading on to the more technical aspects of the book that describe the neurorobotics experiments.

I have debts of gratitude to many people. First of all, I thank Jeffrey White for plenty of insightful advice on this manuscript in regard to its contents, as well as for editing in English and examining every page. I would like to commend and thank all members of my former laboratory at RIKEN as well as of the current one in the Korean Advanced Institute of Science and Technology (KAIST) who, over the years, have contributed to the research described in this book. I am lucky to have many research friends with whom I can have in-depth discussions about shared interests. Takashi Ikegami has been one of the most inspiring. His stroke of genius and creative insights on the topics of life and the mind are irreplaceable. I admit that many of my research projects described in this book have been inspired by thoughtful discussions with him. Ichiro Tsuda provided me deep thoughts about possible roles of chaos in the brain. The late Joseph Goguen and late Francisco Varela generously offered me much advice about the links between neurodynamics and phenomenology. Karl Friston has provided me thoughtful advice in the research of our shared interests on many occasions. Michael Arbib offered insight into the concept of action primitives and mirror neuron modeling. He kindly read my early draft and sent it to Oxford University Press. I have been inspired by frequent discussions about developmental robotics with Minoru Asada and Yasuo Kuniyoshi. I would like to express my gratitude and appreciation to Masahiro Fujita, Toshitada Doi, and Mario Tokoro of Sony Corporation who kindly provided me with the chance to start my neurorobotics studies more than two decades ago in an elevator hall in a Sony building. I must thank Masao Ito and Shun-ichi Amari at RIKEN Brain Science Institute for their thoughtful advice to my research in general. And, I express my gratitude for Miki Sagara who prepared many figures. I am grateful to Frank Ritter as the Oxford series editor on cognitive models and architectures who kindly provided me advice and suggestions from micro details to macro levels of this manuscript during its development. The book could not have been completed in the present form without his input. I wish to thank my Oxford University Press editor Joan Bossert for her cordial support and encouragement from the beginning. Finally, my biggest thanks go to my wife, Tomoko, who professionally photographed the book's cover image; my son, Kentaro; and my mother, Harumi. I could not have completed this book without their patient and loving support.

This book is dedicated to the memory of my father, Yougo Tani, who ignited my interest in science and engineering before he passed away in my childhood. Some additional resources such as robot videos can be found at https://sites.google.com/site/tanioupbook/home. Finally, this work was partially supported by RIKEN BSI Research Fund (2010-2011) and the 2012 KAIST Settlement and Research of New Instructors Fund, titled "Neuro-Robotics Experiments with Large Scale Brain Networks."

Part I

On the Mind

1

Where Do We Begin with Mind?

How do our minds work? Sometimes I notice that I act without much consciousness, for example, when reaching for my mug of coffee on the table, putting on a jacket, or walking to the station for my daily commute. However, if something unexpected happens, like I fail to grasp the mug properly or the road to the station is closed due to roadwork, I suddenly become conscious of my actions. How does this consciousness arise at such moments? In everyday conversation, my utterances are generated smoothly. I automatically combine words in the correct order and seldom consciously manipulate grammar when speaking. How is this possible? Although it seems that many of our thoughts and actions are generated either consciously or unconsciously by utilizing knowledge or concepts in terms of images, rules, and symbols, I wonder how they are actually stored in our memories and how they can be manipulated in our minds. When I'm doing something like making a cup of coffee, my actions as well as thoughts tend to shift freely from getting out the milk to looking out the window to thinking about whether to stay in for lunch today. Is this spontaneous switching generated by my *will*? If so, how is such will initiated in my mind in the first place? Mostly, my everyday thinking or action follows routines, habituation, or social conventions. Nevertheless, sometimes some novel images, thoughts, or acts can be created. How are they generated? Finally, a somewhat philosophical question arises: How can I believe that this world really exists without my subjectively thinking about it? Does my subjective mind subsume the reality of the world or is it the other way around?

The mind is one of the most curious and miraculous things. We know that the phenomena of the mind, like those just described, originate in the brain: We often hear scientists saying that our minds are the products of "entangled" activities of neurons firing, synapse modulations, neuronal chemical reactions, and more. Although the scientific literature contains an abundance of detailed information about such biological phenomena in the brain, it is still difficult to find satisfactory explanations about how the mind actually works. This is because each piece of detailed knowledge about the biological brain cannot as yet be connected together well enough to produce a comprehensive picture of the whole. But understanding the mind is not only the remit of scientists; it is and has always been the job of philosophers, too. One of the greatest of philosophers, Aristotle, asserted that "The mind is the part of the soul by which it knows and understands" (Aristotle, Trans. 1907). It is hard, however, to link such metaphysical arguments to the actual biological reality of the brain.

Twenty-five years ago, I was a chemical plant engineer with no such thoughts about the brain, consciousness, and existence until something wonderful happened by chance to start me thinking about these things seriously. One day I traveled to a chemical plant site in an isolated area in northern Japan to examine a hydraulic system consisting of piping networks. The pipeline I saw there was huge, with a diameter of more than 1.5 m and a total length of around 20 km. It originated in a ship yard about 10 km away from the plant and inside the plant yard it was connected to a complex of looping networks equipped with various functional components such as automatic control valves, pumps, surge accumulators, and tanks.

I was conducting an emergency shutdown test of one of the huge main valves downstream in the pipeline when, immediately after valve shutdown, I was terrified by the thundering noise of the "water hammer" phenomenon, the loud knocking heard in a pipe caused by an abrupt pressure surge upstream of the valve. Several seconds later I heard the same sound arising from various locations around the plant yard, presumably because the pressure surge had propagated and was being reflected at various terminal ends in the piping network. After some minutes, although the initial thunderous noise had faded, I noticed a strange coherence of sounds occurring across the yard. I heard "a pair" of water hammers at different places, seeming to respond to each other periodically. This coherence appeared and disappeared almost capriciously, arising again in other locations. I went back to the plant control room to examine the operation records, plotting the time history of the internal pressure at various points in the piping network. As I thought, the plots showed some oscillatory patterns of pressure hikes appearing at certain points and tending to transform to other oscillatory patterns within several minutes. Sometimes these patterns seemed to form in a combinatory way, with a set of patterns appearing in different combinations with other sets. At that point I jumped on a bicycle to search for more water hammers around the plant yard even though it was already dusk. Hearing this mysterious ensemble of roaring pipes in the darkness, I felt as if I was exploring inside a huge brain, where its consciousness arose. In the next moment, however, I stopped and reflected to myself that this was not actually a mystery at all but complex transient phenomena involving physical systems, and I thought then that this might explain the spontaneous nature of the mind.

I had another epiphany several months later when, together with my fellow engineers, I had the chance to visit a robotics research laboratory, one of the most advanced of its kind in Japan. The researchers there showed us a sophisticated mobile robot that could navigate around a room guided by a map preprogrammed into the robot's computer. During the demonstration the robot maneuvered around the room, stopped in front of some objects, and said in a synthesized voice, "This is a refrigerator," "This is a blackboard," and "This is a couch." While we all stood amazed at seeing the robot correctly naming the objects around us, I asked myself how the robot could know what a refrigerator meant. To me, a refrigerator means the smell of refreshing cool air when I open the door to get a beer on a long hot summer day. Surely the robot didn't understand the meaning of a refrigerator or a chair in such a way, as these items were nothing more to it than landmarks on a registered computational map. The meanings of these items to me, however, would materialize as the result of my own experiences with them, such as the smell of cool air from the refrigerator or the feeling of my body sinking back into a soft chair as I sit down to drink my beer. Surely the meanings of various things in the world around us would be formed in our brains through the accumulation of our everyday experiences interacting with them. In the next moment I started to think about building my own robot, one that could have a subjective mind, experience feelings, imagine things, and think about the world by interacting in it. I also had some vague notion that a subjective mind should involve dynamic phenomena fluttering between the conscious and unconscious, just as with the water hammers that had captured my imagination a few months earlier.

Sometime later I went back to school, where I studied many subjects related to the mind and cognition, including cognitive science, robotics, neuroscience, neural network modeling, and philosophy. Each discipline seemed to have its own specific way of understanding the mind, and the way the problems were approached by each discipline seemed too narrow to exchange ideas and views with other disciplines. No single discipline could fully explain what the mind is or how it works. I simply didn't believe that one day a super genius like Einstein would come along and show us a complete picture of the mind, but rather I suspected that a good understanding, if attainable, would come from a mutual, relational understanding between multiple disciplines, enabling new findings and concepts in one domain to be explainable using different expressions in other disciplines.

It was then it came to me that building robots while taking a multidisciplinary approach could well produce a picture of the mind. The current book presents the outcome of two decades of research under this motivation.

* * *

This book asks how natural or artificial systems can host cognitive minds that are characterized by higher order cognitive capabilities such as compositionality on the one hand and also by autonomy in generating spontaneous interactions with the outer world either consciously or unconsciously. The book draws answers from examination of synthetic neurorobotics experiments conducted by the author. The underlying motivation of this study differs from that of conventional intelligent robotics studies that aim to design or program functions to generate intelligent actions. The aim of synthetic neurorobotics studies is to examine experimentally the emergence of nontrivial mindlike phenomena through dynamic interactions, under specific conditions and for various "cognitive" tasks. It is like examining the emergence of nontrivial patterns of water hammer phenomena under the specific operational conditions applied in complex pipeline networks.

The synthetic neurorobotics studies described in this book have two foci. One is to make use of dynamical systems perspectives to understand various intricate mechanisms characterizing cognitive minds. The dynamical systems approach has been known to be effective in articulating mechanisms underlying the development of various functional structures by applying the principles of self-organization from physics (Nicolis & Prigogine, 1977; Haken, 1983). Structures and functions to mechanize higher order cognition, such as for compositional manipulations of "symbols," concepts, or linguistic thoughts, may develop by means of self-organization in internal neurodynamic systems via the consolidative learning of experience. The other focus of these neurorobotics studies is on the embodiment of cognitive processes crucial to understanding the circular causality arising between body and environment as aspects of mind extend beyond the brain.

This naturally brings us to the distinction between the subjective mind and the objective world. Our studies emphasize *top-down intentionality* on the one hand, by which our own subjective images, views, and thoughts consolidated into structures through past experience are proactively projected onto the objective world, guiding and accompanying our actions. Our studies also emphasize *bottom-up recognition* of the perceptual reality on the other hand, which results in the modification of top-down intention in order to minimize gaps or errors between our prior expectations and actual outcomes. The crucial focus here is on the *circular causality* that emerges as the result of iterative interactions between the two processes of the top-down subjective intention of acting on the objective world and the bottom-up recognition of the objective world with modification of the intention. My intuition is that the key to unlocking all of the mysteries of the mind, including our experiences of consciousness as well as free will, is hidden in this as yet unexplored phenomenon of circular causality and the structure within which it occurs. Moreover, close examination of this structure might help us address the fundamental philosophical problem brought to the fore in mind/body dualism: how the subjective mind and the objective world are related. The synthetic robotics approach described in this book seeks to answer this fundamental question through the examination of actual experimental results from the viewpoints of various disciplines.

This book is organized into two parts, namely "*Part I On the Mind*" from chapter 1 to chapter 5 and "*Part II Emergent Minds: Findings from Robotics Experiments*" from chapter 6 to chapter 11. In Part I, the book reviews how problems with cognitive minds have been explored in different research fields, including cognitive science, phenomenology, brain science, neural network modeling, psychology, and robotics. These in-depth reviews will provide general readers with a good introduction to relevant disciplines and should help them to appreciate the many conflicting arguments about the mind and brain active therein. Part II starts with new proposals for tackling these problems through neurorobotics experiments, and through analysis of their results comes out with some answers to fundamental questions about the nature of the mind. In the end, this book traces my own journey in exploration of the fundamental nature of the mind, and in retracing this journey I hope to deliver an intuitively accessible account of how the mind works.

2

Cognitivism

One of the main forces having advanced the study of the mind over the last 50 years is cognitivism. Cognitivism regards the mind as an externally observable object that can be best articulated with symbol systems in computational metaphors, and this approach has become successful as the speed and memory capacity of computers has grown exponentially. Let us begin our discussion of cognitivism by looking at the core ideas of cognitive science.

2.1. Composition and Recursion in Symbol Systems

The essence of cognitivism is represented well by the principle of compositionality (i.e., the meaning of the whole is a function of the meaning of the parts), but specifically that as expounded by Gareth Evans (1982) in regard to language. According to Evans, the principle asserts that the meaning of a complex expression is determined by the meanings of its constituent expressions and the rules used to combine them (sentences are composed from sequences of words). However, its central notion that the whole can be decomposed into reusable parts (or primitives) is applicable to other faculties, such as action generation. Indeed, Michael Arbib (1981) in his motor schemata theory, which was published not long before Evans' work on language, proposed that complex, goal-directed actions can be decomposed into sequences of behavior primitives. Here, behavior primitives are sets of commonly used behavior pattern segments or motor programs that are put together to form streams of continuous sensory-motor flow. Cognitive scientists have found a good analogy between the compositionality of mental processes, like combining the meanings of words into those of sentences or combining the images of behavior primitives into those of goal-directed actions "at the back of our mind," and the computational mechanics of the combinatorial operations of operands. In both cases we have concrete objects—symbols—and distinct procedures for manipulating them in our brains. Because these objects to be manipulated—either by computers or in mental processes—are symbols without any physical dimensions such as weight, length, speed, or force, their manipulation processes are considered to be cost free in terms of time and energy consumption. When such a symbol system, comprising arbitrary shapes of tokens (Harnad, 1992), is provided with *recursive* functionality for the tokens' operations, it achieves compositionality with an infinite range of expressions.

Noam Chomsky, famous for his revolutionary ideas on generative grammar in linguistics, has advocated that recursion is a uniquely human cognitive competency. Chomsky and colleagues (Hauser, Chomsky, & Fitch, 2002) proposed that the human brain might host two distinct cognitive competencies: the so-called *faculty of language in a narrow sense* (FLN) and the *faculty of language in a broad sense* (FLB). FLB comprises a sensory-motor system, a conceptual-intentional system, and the computational mechanisms for recursion that allow for an infinite range of expressions from a finite set of elements. FLN, on the other hand, involves only recursion and is regarded as a uniquely human aspect of language. FLN is thought to generate internal representations by utilizing syntactic rules and mapping them to a sensory–motor interface via the phonological system as well as to the conceptual–intentional interface via the semantic system.

Chomsky and colleagues admit that some animals other than humans can exhibit certain recursion-like behaviors with training. Chimps have become able to count the number of objects on a table by indicating a corresponding panel representing the correct number of objects on the table by association. The chimps became able to count up to around five objects correctly, but one or two errors creep in for more than five objects: The more objects to count, the more inaccurate at counting the chimps become. Another example of recursion-like behavior in animals is cup nesting, a task in which each cup varies in size so that the smallest cup fits into the second smallest, which in turn can be "nested" or "seriated" into larger cups. When observing chimps and bonobos cup nesting, Johnson-Pynn and colleagues (1999) found that performance differed by species as well as among individuals; some individuals could nest only two different sizes of cups whereas others could pair three by employing a subassembly strategy, that is, nesting a small cup in a medium size cup as a subassembly and then nesting them in a large cup. However, the number of nestings never reliably went beyond three. Similar limitations in cup nesting performance have been observed in parrots (Pepperberg & Shive, 2001) and the degu, a small rat-size rodent (Tokimoto & Okanoya, 2004).

These observations of animals' object counting and nesting cup behaviors suggest that, although some animals can learn to perform recursion-like behaviors, the depth of recursion is quite limited particularly when contrasted with humans in whom almost an infinite depth of recursion is possible as long as time and physical conditions allow. Chomsky and colleagues thus speculated that the human brain might be uniquely endowed with the FLN component that enables infinite recursion in the generation of various cognitive behaviors including language. What then is the core mechanism of FLN? It seems to be a recursive call of logical rules. In counting numbers, the logical rule of "add one to the currently memorized number" is recursively called: Starting with the currently memorized number set to *0*, it is increased to *1*, *2*, *3*, . . . , *infinity* as the "add one" rule is called at each recursion. Cup nesting can be performed infinitely when the logical rule of "put the next smallest cup in the current nesting cup" is recursively called. Similarly, in the recursive structure of sentences, clauses nest inside of other clauses, and in sentence generation the recursive substitution of one of the context-free grammar rules for each variable could generate sentences of infinite length after starting with the symbol "S" (see Figure 2.1 for an illustrative example).

Chomsky and colleagues' crucial argument is that the core aspect of recursion is not a matter of what has been learned or developed over a lifetime but what has been implemented as an *innate* function in the faculty of language in a narrow sense (FLN). In their view, what is to be learned or developed are the interfaces from this core aspect of recursion

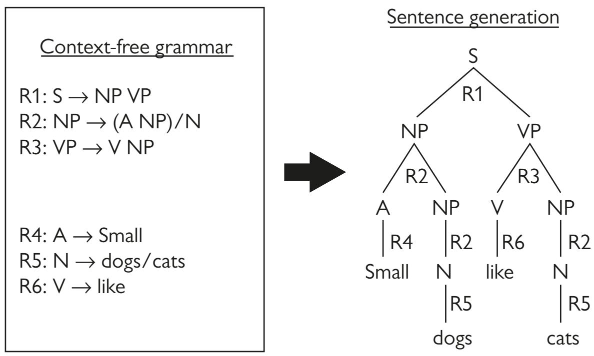

Figure 2.1. On the left is a context-free grammar (CFG) consisting of a set of rules and on the right is an example sentence that can be generated by recursive substitutions of the rules with the starting symbol "S" allocated to the top of the parsing tree. Note that the same CFG can generate different sentences, even those with infinite length, depending on the nature of the substituting rules (e.g., repeated substitutions of R2: NP→A NP).

ability to the sensory–motor systems or semantic systems in the faculty of language in a broad sense (FLB). They assert that the unique existence of this core recursive aspect of FLN is an innate component that positions human cognitive capability at the top of the hierarchy of living systems.

Such a view is contentious though. First, it is not realistic to assume that we humans perform infinite recursions in everyday life. We can neither count infinitely nor generate/recognize infinite-length sentences. Chomsky and colleagues, however, see this not as a problem of FLN itself but as a problem of external constraints (e.g., a limitation in working memory size in FLB in remembering currently generated word sequences) or of physical time constraints that hamper performing infinite recursions in FLN. Second, are symbols actually manipulated recursively somewhere in our heads when counting numbers or generating/recognizing sentences? If there are fewer than six objects on a table, the number would be grasped analogically from visual patterns; if there are more than six objects, we may start to count them one by one on our fingers. In our everyday conversations we generally talk without much concern for spoken grammar: Our colloquialisms seem to be generated not by consciously combining individual words following grammatical rules, but by automatically and subconsciously combining phrases. However, when needing to write complex embedded sentences such as those often seen in formal documents, we sometimes find ourselves consciously dealing with grammar in our search for appropriate word sequences. Thus, the notion of there being infinite levels of recursion in FLN might apply only rarely to human cognition. In everyday life, it seems unlikely that an infinite range of expressions would be used.

Many cognitive behaviors in everyday life do still of course require some level of manipulation that involves composition or recursion of information. For example, generating goal-directed action plans by combining behavior primitives into sequences cannot be accounted for by the simple involuntary action of mapping sensory inputs to motor outputs. It requires some level of manipulation of internal knowledge about the world, yet does not involve infinite complexity. How is such processing done? One possibility might be to use the core recursive component of calling logical rules in FLN under the limitation of finite levels of recursions. Another possibility might be to assume *subrecursive* functions embedded in analogical processes rather than logical operations in FLB that can mimic recursive operations for finite levels. Cognitivism embraces the former possibility, with its strong conviction that the core aspect of cognition should reside in symbol representation and a manipulation framework. But, if we are to assume that symbols play a central role in cognition, how would symbols comprising arbitrary shapes of tokens convey the richness of meaning and context we see in the real world? For example, a typical artificial intelligence system may represent an "apple" with its features "color-is-RED" and "shape-is-SPHERE." However, this is merely to describe the meaning of a symbol by way of other symbols, and I'm not sure how my everyday experience with apples could be represented in this form.

2.2. Some Cognitive Models

This section looks at some cognitive models that have been developed to solve general cognitive tasks by utilizing the aforementioned symbolist framework. The General Problem Solver (GPS) (Newell & Simon, 1972; Newell, 1990) that was developed by Allen Newell and Herbert A. Simon is such a typical cognitive model, which has made a significant impact on the subsequent direction of artificial intelligence research.

Numerous systems such as Act-R (Anderson, 1983) and Soar (Laird et al., 1987) use this rule-based approach, although it has a crucial problem, as is shown later.

The GPS provides a core set of operations that can be used to solve cognitive problems in various task domains. In solving a problem, the problem space in terms of the goal to be achieved, the initial state, and the transition rules are defined. By following a means-end-analysis approach, the goal to be achieved is divided into subgoals and GPS attempts to solve each of those. Each transition rule is specified by an action operator associated with a list of precondition states, a list of "add" states and a list of "delete" states. After an action is applied, the corresponding "add" states and "delete" states are added to and deleted from the precondition states. A rule actually specifies a possible state transition from the precondition state to the consequent state after applying the action.

Let us consider the so-called *monkey–banana problem* in which the goal of the monkey is to become not hungry by eating a banana. The rules defined for GPS can be as shown in Table 2.1.

By considering that the goal is ["not hungry"] and the start state is ["at door," "on floor," "has ball," "hungry," "chair at door"], it can be seen that the goal state ["not hungry"] can be achieved by applying an action of "eat bananas" in Rule 5 if the precondition state of ["has bananas"] is satisfied. Therefore, this precondition state of ["has bananas"] becomes the subgoal to be achieved in the next step. In the

Table 2.1. Example Rules in GPS

Rule #	Action	Precondition	Add	Delete
Rule 1	"climb on chair"	"chair at middle room," "at middle room," "on floor"	"at bananas," "on chair"	"at middle room," "on floor"
Rule 2	"push chair from door to middle room"	"chair at door," "at door"	"chair at middle room," "middle room"	"chair at door," "at door"
Rule 3	"walk from door to middle room"	"at door," "on floor"	"at middle room"	"at door"
Rule 4	"grasp bananas"	"at bananas," "empty handed"	"has bananas"	"empty handed"
Rule 5	"eat bananas"	"has bananas"	"empty handed," "not hungry"	"has bananas," "hungry"

same manner, the subgoal ["has bananas"] can be achieved by applying an action of ["grasp bananas"] with the precondition of ["at bananas"], which can be achieved again by applying another action of ["climb on chair"]. Repetitions of backward transition from a particular subgoal to its sub-subgoal by searching for an adequate action enabling the transition can result in generation of a chain of actions, and the goal state can be achieved from the start state by applying the resulting action sequence.

The architecture of GPS is quite general in the sense that it has been applied to a variety of different task domains including proving theorems in logic or geometry, word puzzles, and chess. Allen Newell and his colleagues (Laird et al., 1987) developed a new cognitive model, Soar, by further extending GPS. Of particular interest is its primary learning mechanism, *chunking*. Chunking is involved in the conversion of an experience of an action sequence into long-term memory. When a particular action sequence is found to be effective to achieve a particular subgoal, this action sequence is memorized as a chunk (a learned rule) in long-term memory. When the same subgoal appears again, this chunked action sequence is recalled rather than deliberating over and synthesizing it again. For example, in the case of the monkey–banana problem, the monkey may learn an action sequence of "grasp bananas" and "eat bananas" as an effective chunk for solving a current "hungry" problem, and may retain this chunk because "hungry" may appear as a problem again in the future.

The idea of chunking has attracted significant attention in cognitive psychology. Actually, I myself had been largely influenced by this idea after I learned about it in an artificial intelligence course given by John Laird, who has led the development of Soar for more than two decades. At the same time, however, I could not arrive at full agreement with the treatment of chunking in Soar because the basic elements to be chunked are symbols rather than continuous patterns even at the lowest perceptual level. I speculated that the mechanism of chunking should be considered at the level of continuous perceptual flow rather than symbol sequences in which each symbol already stands as an isolated segment within the flow. Later sections of this book explore how chunks can be structured out of continuous sensory–motor flow experiences. First, however, the next section introduces the so-called *symbol grounding problem*, which cognitive models built on symbolist frameworks inevitably encounter.

2.3. The Symbol Grounding Problem

The symbol grounding problem as conceptualized by Steven Harnad (1990) is based on his assertion that the meanings of symbols should originate from a nonsymbolic substrate like sensory-–motor patterns and as such, symbols are grounded bottom up. To give shape to this thought, he proposed, as an abstract model of cognitive systems, a hybrid system consisting of a symbol system in the upper level and a nonsymbolic pattern processing system in the lower level. The nonsymbolic pattern processing system functions as the interface between sensory–motor reality and abstract symbolic representation by categorizing continuous sensory–motor patterns into sets of discrete symbols. Harnad argued that meaning, or semantics, in the hybrid system would no longer be parasitic on its symbol representation but would become intrinsic to the whole system operation, as such representation is now grounded in the world. This concept of a hybrid system has similarities to that of FLN and FLB advocated by Chomsky and colleagues in the sense that it assumes a core aspect of human cognition in terms of logical symbol systems, which can support up to an infinite range of expressions, and peripheries as the interface to a sensory–motor or semantic system that may not be involved in composition or recursion in depth.

This idea of a hybrid system reminds me also of Cartesian dualism. According to Descartes the mind is a thinking thing that is nonmaterial whereas the body is nonthinking matter, and the two are distinct. The nonmaterial mind may correspond to FLN or symbol systems that are defined in a nonphysical discrete space, and the body to sensory–motor processes that are defined in physical space. The crucial question here is how these two completely distinct existences that do not share the same metric space can interact with each other. Obviously, our minds depend on our physical condition and the freshness of the mind affects the swiftness of our every move. Descartes showed some concern about this "problem of interactionism," asking how a nonmaterial mind can cause anything in a material body, and vice versa. Cognitive scientists in modern times, however, seem to consider—rather optimistically I think—that some "nice" interfaces would enable interactions between the two opposite poles of nonmatter and matter.

Let's consider the problem by examining a problem in robot navigation as an example, reviewing my own work on the subject (Tani, 1998). A typical mobile robot, which is equipped with simple range sensors, may travel around an office environment while taking the range reading that