Preface

Embodiment, the idea that the body is required for intelligence, has been around for two decades, but an awful lot has changed since then. Research labs and leading technology companies around the world have produced or are developing a host of sometimes science fiction-like creations: almost frighteningly realistic humanoid robots, robot musicians, wearable technology, robots controlled by biological brains, robots that can walk without a brain, real-life cyborgs, robots in homes for the elderly, robots that literally put themselves together, artificial cells grown automatically, and simulated genetic regulatory networks for growing virtual creatures. This new breed of technology, along with many significant theoretical advances, is the direct result of the embodied approach to intelligence. Along the way, many of the initially vague ideas have been elaborated and the arguments sharpened, and the diverse outcomes are beginning to form into a coherent structure. Thus, it seemed like a good opportunity to work out the first steps toward a theory of intelligence and write a book about it.

From a personal perspective, I (Rolf) have given many seminars and lectures to nonspecialized audiences, and many of them were able to relate in very direct and natural ways to the ideas presented: the ideas I talked about seemed to hold relevance for their own interests and specialties. What most people found intriguing was that this research demonstrates how things can always be seen differently. We all have our strong prejudices and often think, "It's got to be like that, there is no other way!" For example, if you want to build a fast-running robot you must have fast electronics; an object-collecting robot must have a means for recognizing the objects it is supposed to gather; or an insect with six legs needs a centralized control program somewhere in its brain to coordinate all its legs while walking. Surprisingly, it turns out that none of these are true, as we will see later.

So, I felt that rather than writing another specialized textbook like *Understanding Intelligence*, co-authored with Christian Scheier in 1999, a popular science–style publication accessible to a wide readership might be a more suitable undertaking. Science and technology are no longer isolated fields. They closely interact with the corporate, political, and social aspects of our society; and that interaction, among its other effects, increases the need to justify basic research. Convinced that we might be able to provide a novel perspective not only on artificial intelligence but more generally on how we view ourselves and the world around us, we took up the challenge of trying to translate the scientific results and insights we have gained into everyday language. The result is this book.

Aims and Scope

The goal of this book is twofold: on the one hand it is to explore the implications of embodiment (how having a body affects intelligence), to work out the first steps toward a theory of intelligence, and finally to demonstrate the wide applicability of these ideas. On the other, we will try to show that things can always be seen differently. So, the book is conceptual, and is geared toward a broad audience in education, business, information technology, engineering, entertainment, the media, as well as academics from virtually all disciplines and levels, but especially those involved in psychology, neuroscience, philosophy, linguistics, and biology. And last but not least, this book is also intended for anyone interested in technology, its future, and its implications for society. No special training or education is required for understanding the ideas presented: we have tried to provide background information, examples, and pointers to further reading for the more difficult-to-grasp concepts.

The core of the theory consists of a set of "design principles for intelligent systems." The reason for choosing the form of design principles for our theory is that they are a compact way of describing insights about intelligent systems in general and they provide convenient heuristics for actually building artificial systems, like robots in particular. And actually building systems is crucial because we want to design and construct intelligent artificial systems so that we can understand intelligent systems in general: this is the synthetic methodology—the basic methodology of artificial intelligence—which can be characterized as understanding by building. As we will show with many examples, by building artificial

systems we can learn about biology, but also about intelligence in general. An exciting prospect is that this enables us not only to study natural forms of intelligence, but to create new forms of intelligence that do not yet exist; "intelligence as it could be," to adapt a quote by the founder of the field of artificial life, Chris Langton. Thus, by building robots, our intention is to learn something about intelligence, and not so much to build technologically sophisticated robots. So, the book is not so much about the intricacies of the engineering process or the details of how to build robots, but rather about the basic insights that arise as a result of building robots.

We have tried to support our claim that the ideas developed in this book have broad applicability beyond the field of artificial intelligence proper by providing illustrations from the fields of ubiquitous computing, strategic management, human memory, and robotic technology in everyday life. We hope that the reader will enjoy these case studies and will feel encouraged to apply the ideas to areas of his or her own interest.

We should perhaps briefly comment on the term *artificial intelligence* here before continuing. With the introduction of the notion of embodiment about 20 years ago, the field has undergone fundamental changes, so that sometimes the term *embodied artificial intelligence* is employed, and we even published a book with precisely that title (Iida et al., 2004). In this book we will avoid that usage, because it somehow suggests that there is the "real" field of artificial intelligence—the overarching, encompassing discipline—and then there is this small subarea called "embodied" artificial intelligence. We feel that this perspective is somewhat inappropriate. As we will elaborate later, there are essentially two directions in artificial intelligence: one concerned with developing useful algorithms or robots; and another direction that focuses on understanding intelligence, biological or otherwise. In order to make progress on the latter, an embodied perspective is mandatory. In this research branch, artificial intelligence *is* embodied.

One last comment is necessary before we turn to the contents of the book. In spite of the fact that the materials presented are often a bit theoretical and require concentration on the part of the reader, we have tried to make the book fun to read by providing many examples. Also, the web site for the book (www.ifi.unizh.ch/groups/ailab/HowTheBody) contains many links to videos and other supporting material, as well as a discussion forum. To make the book even more appealing, we have engaged an artist and computer scientist, Shun Iwasawa of the

University of Tokyo, who, with great talent, technical skill, and understanding of the subject matter of the book, created Japanese Manga–style illustrations that, we hope, will stimulate the reader's interest and communicate the fun, forward-thinking style of this field of study.

Road Map to the Book

There are three parts to this book. Part I is introductory, familiarizing the reader with the contents of the book and the basic concepts. Part II, the core section, summarizes our attempts to develop a theory of intelligence. Part III applies the theory, in particular the design principles developed in part II, to a number of case studies beyond the field of artificial intelligence proper. Finally, part IV provides a summary of the major points made in the book.

In the first chapter we will introduce what the terms *thinking*, *cognition*, and *intelligence* mean, discuss why intelligence has fascinated people from all walks of life throughout history, and introduce the field of artificial intelligence and the embodied view of intelligence. Chapter 2 presents an overview of the intellectual landscape of artificial intelligence. This should give the reader a flavor of the kinds of research questions that are out there, as well as the fascination, but also the difficulties, of navigating and actually doing research in this highly rugged interdisciplinary field.

Part II is an attempt to formulate the first steps toward a theory of intelligence. It is the central part of the book, and so it is a bit on the heavy side, conceptually speaking. But we have tried to include many examples to illustrate the abstract ideas and to support our arguments. Chapter 3 outlines what type of theory we are looking for and introduces a general framework provided by a number of important notions such as diversity-compliance, frame of reference, the synthetic methodology, time perspectives, and emergence. This chapter contains a bit of philosophy of science, which we use to outline the nature of the theory, and describes what it means to make progress and to do work in the field. Just to take one example, there are three time frames at which we can study behavior: "here and now"; learning and development; and evolution. Chapters 4, 5, and 6 are organized around these time frames.

Chapter 4 first describes properties of real-world agents and then sketches a set of design heuristics—what we call the design principles for intelligent systems—that can be used to guide us in engineering such

agents but also to help us understand biological ones. These design principles mostly concern the "here-and-now" time frame. Chapter 5 explores design and analysis issues from a developmental perspective, and asks how high-level cognition can emerge during a process of ontogenetic development; how cognition emerges as the agent matures into an "adult." For example, how is it possible that something discrete, such as abstract symbol processing, can arise in a completely continuous system—and we are all continuous systems—over time? A specific, somewhat provocative instance of this question is what walking—or, more generally, locomotion—has to do with thinking, which we will explore in detail in that chapter. The chapter concludes with a set of design principles at the developmental time perspective. Chapter 6 looks at how we can harness ideas from biological evolution in order to design agents—complete with bodies, sensors, motors, and brains—from scratch. Here, we as designers step back and let simulated evolution do the work for us. The point is to let evolution design virtual agents that perform increasingly complex tasks, so that at some point we might be inclined to use the term *cognitive* to characterize their behavior. One of the goals of this chapter is to demonstrate the power of artificial evolution. Specifically, we will give some impressive examples of where it has outperformed humans. While chapter 5 focuses on the lifetime of an individual, in chapter 6 we extend the time frame to encompass many generations of agents, and widen our view to consider not just single agents but populations of them. Again, we summarize the main results as a set of design principles, this time for evolutionary systems.

The implications of considering populations rather than individuals are discussed in chapter 7. There, we look at emergent phenomena that arise in populations of agents; that is, phenomena, or global behavioral patterns, in the group that come about as the agents interact with each other without knowing about the global pattern. These kinds of emergent behaviors are often referred to as collective intelligence. We will also introduce another kind of collective intelligence, namely modular robots: i.e., robots that are composed of many modules, which, as they interact with each other, can achieve interesting collective behaviors. In modular robotics, the modules can be viewed as agents, in addition to the robot itself. The main points in this chapter are also captured in a set of design principles for collective intelligence.

Part III discusses a number of case studies demonstrating the application of the concepts and design principles developed in part II to

problems that lie outside the area of artificial intelligence proper: we will look at ubiquitous computing, management, the psychology of human memory, and robotic and artificial intelligence technologies in our everyday lives. We will show that the perspective of embodiment can shed new light on these topics. The case studies are self-contained and can be read in any sequence after the reader has finished chapters 1 through 7. In chapter 8, we discuss ubiquitous computing, a rapidly expanding discipline in informatics which in fact shares many ideas with artificial intelligence. In this new field, the goal is to explore the potential of "putting computing everywhere": into cars, clothes, cups, shoes, buildings, appliances, mobile phones, consumer goods in general; and embedding them into communication networks of ever-increasing size and complexity. Chapter 9, written by Simon Grand and Rolf Pfeifer, is an initial attempt to apply the perspective of embodied intelligence to the business world, and in particular to the design and construction of new products, businesses, and companies in an intrinsically uncertain, complex, and unpredictable world. That chapter is meant to demonstrate that the design principles indeed have wide potential applicability. Chapter 10 presents a case study on human memory that illustrates on the one hand how embodiment provides a new perspective on old problems, and on the other how it can be employed to better understand recent trends in memory research. Chapter 11 tries to assess the feasibility, desirability, and economic reality of developing all kinds of robots, and in particular humanoid robots, that could enter and participate in the everyday lives of humans and our society.

Part IV, the last part of the book, will summarize the main points of our theory and provide a review of the design principles in a single chapter, chapter 12. There we also present a list of selected highlights that sum up what we feel are the key insights gained throughout the book. In conclusion, we will return to one of the central goals of the book: we will present a collection of examples illustrating how things can always be seen differently.

Acknowledgments

We would like to thank all the members of the Artificial Intelligence Laboratory of the University of Zurich for continued discussions, excellent research, enthusiasm, and the many ideas that finally found their way into this book. Big thanks go to our friends Yasuo Kuniyoshi, Olaf

Sporns, Akio Ishiguro, Hiroshi Yokoi, Koh Hosoda, Fumio Hara, and Hiroshi Kobayashi, who kept encouraging us to move ahead with the project as quickly as possible. Josh would like to express his warmest thanks to Hod Lipson for selflessly providing time and space within which to work on this project. We would also like to express our thanks to the many funding agencies that have made the research described in this book possible, in particular the Swiss National Science Foundation and the IST Program of the European Union. Moreover, I (Rolf) would like to express my very personal thanks to Yasuo Kuniyoshi, Tomomasa Sato, Hirochika Inoue, Yoshi Nakamura, and all the other members of the Department of Mechano-Informatics, for inviting me to the University of Tokyo to be a twenty-first-century COE (Center of Excellence) professor of Information Science and Technology, where many draft chapters of this manuscript were written. Their perspective on intelligent agents as complex dynamical systems has strongly influenced the contents of the book.

Our great thanks go also to Gabriel Gomez, who has researched many issues concerning the project. Also, we highly appreciate the contributions of Max Lungarella, who, in particular with his PhD thesis but also with many personal discussions, is largely responsible for the quality of chapter 5. Also, the ideas of Fumiya Iida, to whom we owe the title of chapter 5, "From Locomotion to Cognition," have been instrumental and have strongly helped shape the ideas of chapter 5. We are extremely grateful to Shun Iwasawa for his outstanding, inspiring, and instructive illustrations: they further help bridge the gap between science, engineering, and entertainment.

Thanks go also to all the researchers around the world from many different disciplines for their ideas that have provided inspiration for our arguments so that we were able to put together a coherent story. And, of course, to Rodney Brooks, for having started this exciting research field in the first place, and for providing the foreword.

We would also like to express our sincere thanks to the MIT Press, in particular to Bob Prior for supporting this project, to the copy editor, Suzanne Schafer, and to a number of anonymous reviewers who, with their constructive criticisms, have enormously helped improve the quality of the manuscript. Our thanks go also to Britta Glatzeder, who was involved in the initial stages of the project and to whom we owe the title of the book. Rolf would also like to thank Claudia Wirth, who kept everybody off his back when he was trying to finalize the manuscript and who kept the lab running during his long absence.

There are many, many others—faculty, staff, students, friends, and family—who have provided support one way or another to both authors, and without whom this book would never have been finished, let alone begun: we are deeply in debt to all of you. I (Josh) would like to thank my family—Toby, Carol, and Ralph—for understanding and helping me in my long journey to get to this point. I (Rolf) would like to thank in particular my two sons, Serge and Mischa, who have always encouraged me to continue when times were hard.