DEEP LEARNING

Although the ability to retain, process and project prior experience onto future situations is indispensable, the human mind also possesses the ability to override experience and adapt to changing circumstances. Cognitive scientist Stellan Ohlsson analyzes three types of deep, non-monotonic cognitive change: creative insight, adaptation of cognitive skills by learning from errors and conversion from one belief to another, incompatible belief. For each topic, Ohlsson summarizes past research, re-formulates the relevant research questions and proposes information-processing mechanisms that answer those questions. The three theories are based on the principles of *redistribution* of activation, *specialization* of practical knowledge and *resubsumption* of declarative information. Ohlsson develops the implications of those principles by scaling their consequences with respect to time, complexity and social interaction. The book ends with a unified theory of non-monotonic cognitive change that captures the abstract properties that the three types of change share.

Stellan Ohlsson is Professor of Psychology and Adjunct Professor of Computer Science at the University of Illinois at Chicago (UIC). He received his Ph.D. in psychology from the University of Stockholm in 1980. He held positions as Research Associate in the Robotics Institute at Carnegie-Mellon University and as Senior Scientist in the Learning Research and Development Center at the University of Pittsburgh before joining UIC in 1996. His work has been supported by the Office of Naval Research, the National Science Foundation and other organizations. Dr. Ohlsson has published extensively on computational models of cognition, creative insight, skill acquisition and the design of instructional software, as well as other topics in higher cognition.

Deep Learning

HOW THE MIND OVERRIDES EXPERIENCE

Stellan Ohlsson
University of Illinois at Chicago

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo, Mexico City

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521835688

First published 2011

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication data
Ohlsson, Stellan.
Deep learning : how the mind overrides experience / Stellan Ohlsson.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-521-83568-8
1. Learning, Psychology of. 2. Cognitive learning theory. 3. Mind and body. 4. Experience. I. Title.
BF318.045 2011
153.1′5–dc22 2010030593

ISBN 978-0-521-83568-8 Hardback

CONTENTS

PREFACE

The theme of this book is that human beings possess cognitive processes that enable them to override the imperatives of past experience and to act and think in novel ways, and that these processes differ from the types of cognitive processes usually envisioned in psychological theories of learning. The capability for what I call *deep learning* – or, more precisely, *non-monotonic cognitive change* – constitutes a distinct aspect of mind that follows its own laws and hence requires its own theory. The book develops this theme by summarizing and extending prior research by me and others with respect to three specific types of non-monotonic change: the creation of novelty; the adaptation of cognitive skills to changing circumstance; and the conversion from one belief to another, incompatible belief. The book offers novel theories of the mental processes operating in each of these three types of cognitive change, as well as a unified theory that captures the abstract principles that they share.

My interest in creativity, adaptation and conversion preceded my awareness that these topics are variations on a theme. As a graduate student at the University of Stockholm in the late 1970s, I tried to relate the Gestalt view of insight to the information-processing theory of problem solving proposed by A. Newell and H. A. Simon. My first attempt at such a synthesis was published in 1984, and over the years it morphed into the theory of insight in Chapter 4. I thank my Ph.D. advisor, Yvonne Waern, for her constant encouragement and strong support for this as well as other oddball activities, and for managing a weekly cognitive seminar where her students could argue about cognition. I fondly remember discussions with Yvonne herself and, among others, Ove Almkvist, Göran Hagert and Susanne Askvall. Swedish psychologists interested in cognition formed a small community at that time and I learned from my interactions with, among others, Carl Martin Allwood, Berndt Brehmer, Anders Ericsson, Henry Montgomery, Lars-Göran Nilsson, Lennart Nilsson, Rolf Sandell and Ola Svensson.

Modern work on skill acquisition began with a 1979 article by Y. Anzai and H. A. Simon at Carnegie-Mellon University (CMU). They reported a computer simulation model of a single subject learning a new problem-solving strategy. As a graduate student, I had the opportunity to visit CMU in the fall of 1978, at the very moment when this line of work began. Anders Ericsson, a fellow graduate student from Stockholm, was already at CMU as a post-doctoral Fellow, and I thank him for his generosity in letting me stay at his house for several months. I appreciate the willingness of CMU faculty members John R. Anderson, David Klahr, Allen Newell, Lynn Reder, Robert Siegler, Herbert A. Simon and their students and associates – including Patrick Langley, David Neves, John Laird and Paul Rosenbloom – to engage intellectually with a student visitor. Pat in particular took me under his wing. We spent many hours debating computational models of skill acquisition, and our collaboration continues to this day. The multiple-mechanism theory of adaptation presented in Chapter 6 is a descendant of those discussions.

My work acquired an educational aspect during my years as Senior Scientist at the Learning Research and Development Center (LRDC) at the University of Pittsburgh. I continued work on insight in collaboration with Jonathan Schooler, which resulted in a widely cited paper on the relation between insight and language. During those years my work on skill acquisition led to the theory of learning from error that is the centerpiece of Chapter 7. I also branched out into research on intelligent tutoring systems. Pat Langley and I had previously investigated the application of machine learning techniques to the problem of online diagnosis of student errors, but my understanding of tutoring systems was much improved at LRDC by discussions and collaborations with Jeffrey Bonar, Bruce Buchanan, Alan Lesgold, Johanna Moore and Kurt VanLehn. My collaboration with Bruce and Johanna on the automatic generation of explanations for medical patients strengthened my long-standing interest in the philosophy of explanation. The reader will encounter this topic in Chapter 2.

The focus on explanation led in turn to an interest in the nature of declarative knowledge generally. My understanding of this topic owes much to interactions with Michelene ("Micki") Chi, James Greeno, Lauren Resnick, James Voss and others. The years at LRDC touched other aspects of my professional development as well. From Glynda Hull I learned that the prose of scholarly texts does not have to be dull and boring, and I hope the reader can see the effects of this lesson in the present book. From Gaia Leinhardt I learned to respect the skills of classroom teachers. Robert Glaser and Lauren Resnick taught me the elements of grantsmanship. There was a steady stream of visitors

passing through LRDC. Andreas Ernst, a student from Germany, now professor of environmental systems analysis at the University of Kassel, spent a year with me teaching cognitive skills to the HS simulation model that stars in Chapters 7 and 8. My interactions with Erno Lehtinen provided an opportunity to think through the function of abstraction in declarative knowledge. Similarly, I benefited from my conversations with David Perkins, then and in later visits with his group at Harvard University. During the LRDC years, I was privileged to have Nancy Bee, Ernest Rees and James J. Jewett working with me in their various capacities. I thank John Anderson, Micki Chi, Susan Chipman and Lauren Resnick for their assistance at a crucial moment in my career.

When I moved to the University of Illinois at Chicago (UIC) in 1996 I continued all three lines of research. Guenther Knoblich, then a graduate student at the Max Planck Institute in Munich, Germany, spent the better part of a year with me in Chicago. We pushed the theory of insight beyond what I had been able to do in previous publications, and we conducted experiments to support it. The theory in Chapter 4 is a revised version of the cognitive mechanisms we identified. Our experimental work benefited from our collaboration with my UIC colleague Gary Raney, who contributed his expertise in eye-tracking methodology. I thank Guenther for arranging an opportunity to continue this work during a six-week visit to the Max Planck Institute in the spring of 1998, and Institute Director Professor Wolfgang Prinz for his support and hospitality.

My work on the design of intelligent tutoring systems for cognitive skills has advanced in two important ways at UIC. The first advance occurred when I was contacted in 1996 by Antonija ("Tanja") Mitrovic, a computer scientist who was in the process of escaping strife in her former homeland and re-settling herself and her family in New Zealand. Tanja wanted to use the theory of constraint-based learning from error that the reader finds in Chapter 7 to guide the design of intelligent tutoring systems. Tanja is now a leading researcher in that field, and I thank her for the thrill of seeing the ideas we talked about become real in the series of intelligent tutoring systems that she and her co-workers and students have produced at Canterbury University in New Zealand. The second important advance was the arrival at UIC of Barbara Di Eugenio, a computational linguist with expertise in tutoring whom I already knew from LRDC. We have studied tutorial dialogues in order to base the design of tutoring systems on a solid empirical basis. The all-too-brief statement about the application of the constraint-based approach to tutoring in Chapter 7 summarizes a few of the insights gained through my collaborations with Tanja and Barbara and their students.

At UIC, I have had multiple opportunities to develop my interest in the nature of declarative knowledge. Andrew Johnson, Jason Leigh and Thomas Moher are three UIC computer scientists who specialize in virtual reality and related technologies. Together we built and field tested a learning environment for teaching children that the Earth is spherical rather than flat. The instructional intervention was not as powerful as we had hoped, but the design and data collection stimulated our thinking about the nature of declarative knowledge and belief. My interest in the philosophy of explanation has also benefited from discussions with Nicholas Huggett, Jon Jarrett and Colin Klein, colleagues in the philosophy department at UIC. Micki Chi invited me in 2004 to co-author a review paper that summarized the cognitive mechanisms behind the acquisition of complex declarative knowledge. That effort stimulated me to develop a new theory of belief revision. I thank Gale Sinatra for encouraging me to put that theory in writing, and for making room for it in the pages of the *Educational Psychologist*. The reader will find the current version of that theory in Chapter 10.

Like many other cognitive scientists, I often find it difficult to explain to people in other professions what I do for a living. One defense against such social embarrassment is to talk about the implications of cognitive science for everyday life. The question arises as to what those implications are. How do the consequences of cognitive processes scale up to long periods of time and across levels of complexity? Do the details of individual cognition matter for the groups, teams and organization in which human beings normally operate? These questions have stimulated my interest in computer simulation of the connection between individual and social cognition. Two UIC colleagues stand out as sources of inspiration in this regard. Siddartha Bhattacharyya and I have collaborated on a computer model of social creativity using a technique called agent-based modeling. My understanding of this enterprise has been greatly advanced by interactions with my colleague James Larson, a social psychologist whose experiments are as elegant as his simulation models of group decision making and problem solving. What I have learned from these colleagues has informed my treatment of the relations between the individual and the collective in Chapters 5 and 8.

Throughout my years at UIC, I have had the privilege of working with a large group of graduate students: Bettina Chow, Andrew Corrigan-Halpern, David Cosejo, Thomas Griffin, Joshua Hemmerich, Trina Kershaw, Timothy Nokes, Justin Oesterreich, Mark Orr, Shamus Regan and Robert Youmans. The reader will see glimpses of their work here and there throughout the book. I thank each and every one of them for our many stimulating discussions.

Pursuing the three topics of insight, skill acquisition and belief revision in parallel over multiple years inevitably led to the question of how these three types of cognitive change are related. In the 1990s, I became fascinated by the complex systems revolution that swept through both the natural and the social sciences, and it dawned on me that this new view of reality directly impacts my own work: If both nature and society are chaotic, complex and turbulent, then how must the mind be designed to enable people to function in that kind of world? The question led to a different synthesis of my three interests from any that I had envisioned previously.

A 2004 sabbatical year at the Computer Science Department at Canterbury University in New Zealand provided the opportunity to attempt a synthesis. I thank the Erskine Foundation for the fellowship that made this visit possible. I thank my friend and colleague Tanja Mitovic, department head Timothy Bell and the staff of the Erskine Foundation for bearing the burden of the paperwork and the other practical arrangements associated with my visit. As befitting the laptop lifestyle of the contemporary era, this book was written in coffee shops rather than in offices. The first draft was hammered out in a charming café in the Cashmere Hills, just southwest of the Canterbury plain, called, appropriately enough, The Cup, while the inevitable rewriting was done in Starbucks and Barnes & Noble coffee shops in the Gold Coast neighborhood of Chicago. I thank the staff at these places for their friendliness, their patience with a customer who never leaves, and their diligence in keeping those cappuccinos coming. In the course of my writing, colleagues at UIC and elsewhere who have helped by responding to various questions and requests for comments and materials include John Anderson, Tibor Bosse, Daniel Cervone, William Clancey, Stephanie Doane, Renee Elio, Susan Goldman, David Hilbert, Ben Jee, Jim Larson, Michael Levine, Matthew Lund, James MacGregor, Clark Lee Merriam of the Cousteau Society, Thomas Ormerod, David Perkins, Michael Ranney, Steven Smith, Terri Thorkildsen, Jan Treur, Endel Tulving, Jos Uffink, David Wirtshafter and Beverly Woolf.

The specific investigations that underpin the theoretical formulations in this book were made possible primarily by grants from the Office of Naval Research (ONR). Very special thanks to Susan Chipman, who as program officer dealt with 20 years' worth of grant proposals with analytical acumen, broad knowledge of the field, much good advice and some mercy. In addition, I have been the grateful recipient of grants from the National Science Foundation (NSF) and the Office for Educational Research and Improvement (OERI). Seed grants from UIC helped get some of these investigations under way.

Throughout the six years of writing this book, the staff at Cambridge University Press has been patient, to say the least, with my repeatedly postponed deadlines. I thank the editors, the copy editor and the members of the production department for their work in bringing the manuscript through the production process. I have Deborah Roach to thank for the jacket photo.

Although an author of this kind of book has many people to thank, writing is a solitary endeavor for which all the rewards arrive well after the work is done. As the work stretches over multiple years, moments arrive when it is difficult to sustain belief in the enterprise. I thank my wife, Elaine C. Ohlsson, for her upbeat encouragement and her unwavering belief that the book would one day be done, and that it would be worth the sacrifice of the many hours we could have spent together if I had not been glued to my keyboard.

I have benefited from my interactions with all the individuals mentioned here and with many others. I am solely responsible for the use I have made of what I have learned, and any errors and mistakes, conceptual or technical, are entirely my own.

This book can be read in different modes. It can be read as a review of research in the three areas of creativity, skill acquisition and belief revision. The reader in this mode should be forewarned that Chapters 3, 6 and 9 are not neutral summaries. They are designed to lead the reader to the conclusion that existing theories are insufficient to answer the relevant questions, and thereby prepare the ground for my own theoretical proposals. That said, I have tried to mention every good idea that I have encountered in 35 years of reading the cognitive research literature, and I believe the book could serve as the text for a graduate seminar on cognitive change. Readers in this mode are encouraged to pay attention to the notes; I put most of the history and background material there. Regarding issues in human cognition, I cite original research articles. Regarding matters outside my specialty, I allow myself to cite secondary sources. I believe that a newcomer to the study of cognitive change has no need to repeat my extensive idea mining of the cognitive literature, but can take the present book as his* starting point and move forward, but perhaps that is an author's conceit.

A second reading mode is to focus on the technical contributions, that is, the three specific theories proposed in Chapters 4, 7 and 10, and to evaluate each on its own terms as a contribution to the relevant research area. This

* For brevity and elegance of expression, I use "he," "his" throughout as synonyms for "he or she," "his or her." This is a stylistic choice and not a statement about gender.

mode will be natural for cognitive psychologists. Readers in this mode will find that I emphasize the theoretical ideas themselves and use the broader canvas of a book to discuss them in more detail than can be fitted into the standard research article. My goal throughout has been conceptual clarity and deep explanation, not coverage of laboratory findings.

In a third mode, the reader would focus on the goal of understanding non-monotonic change as a category of cognitive change that poses unique theoretical puzzles and therefore requires its own principles and explanatory schemas. In this reading mode, the synthesis of the three theories into a unified theory in Chapter 11 is the most important contribution of the book.

The core contributions are necessarily technical in nature, but I have tried to write in such a way that an educated layperson can read this book as an extended reflection on the turbulence of the human condition. My ambition has been to write the kind of book I enjoy reading: A serious contribution to science that spells out the broader implications of the research for everyday life. Readers in this mode might want to skim the central sections of Chapters 4, 7 and 10, but I hope they enjoy the rest of the book.

For very busy people, there is yet another way to approach this book: Read the first sentence of Chapter 1 and the last sentence of Chapter 12, and postpone the stuff in between until after retirement.

The Gold Coast, Chicago
May 2010

DEEP LEARNING