This book offers a critical reconstruction of the fundamental ideas and methods of artificial intelligence research. Through close attention to the metaphors of AI and their consequences for the field's patterns of success and failure, it argues for a reorientation of the field away from thought in the head and toward activity in the world.

By considering computational ideas in a philosophical framework, the author eases critical dialogue between technology and the humanities and social sciences. AI can benefit from new understandings of human nature, and in return, it offers a powerful mode of investigation into the practicalities and consequences of physical realization.

Computation and human experience

Learning in doing: Social, cognitive, and computational perspectives

GENERAL EDITORS: ROY PEA
JOHN SEELY BROWN

Sociocultural Psychology: Theory and Practice of Doing and Knowing, edited by Laura M. W. Martin, Katherine Nelson, and Ethel Tobach

Sociocultural Studies of Mind, edited by James V. Wertsch, Pablo del Rio, and Amelia Alvarez

The Computer as Medium, edited by Peter Bogh Andersen, Berit Holmqvist, and Jens F. Jensen

Distributed Cognitions: Psychological and Educational Considerations, edited by Gavriel Salomon

Understanding Practice: Perspectives on Activity and Context, edited by Seth Chaiklin and Jean Lave

Street Mathematics and School Mathematics, by Terezinha Nunes, David William Carraher, and Analucia Dias Schliemann

Situated Learning: Legitimate Peripheral Participation, by Jean Lave and Etienne Wenger

The Construction Zone: Working for Cognitive Change in School, by Denis Newman, Peg Griffin, and Michael Cole

Plans and Situated Actions: The Problem of Human Machine Communication, by Lucy A. Suchman

Mind and Social Activity, edited by Ethel Tobach, Rachel Joffe Falmagne, Mary B. Parlee, Laura Martin, and Aggie Schribner Kapelman

Computation and human experience

PHILIP E. AGRE
University of California, San Diego

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521384322

First published 1997

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication data
Agre, Philip E.
Computation and human experience / Philip E. Agre.
p. cm. - (Learning in doing)
ISBN 0-521-38432-X(hardback). ISBN 0-521-38603-9(pbk).
1. Artificial Intelligence - Data processing. 2. Human behavior.
I. Title II. Series
Q336.A37 1997
006.3'01 - dc 20 96-14823
CIP

ISBN 978-0-521-38432-2 Hardback
ISBN 978-0-521-38603-6 Paperback

Cambridge University Press & Assessment has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate. Information regarding prices, travel timetables, and other factual information given in this work is correct at the time of first printing but Cambridge University Press & Assessment does not guarantee the accuracy of such information thereafter.

Joshu asked Nansen: "What is the path?"

Nansen said: "Everyday life is the path."

Joshu asked: "Can it be studied?"

Nansen said: "If you try to study, you will be far away from it."

Joshu asked: "If I do not study, how can I know it is the path?"

Nansen said: "The path does not belong to the perception world, neither does it belong to the nonperception world. Cognition is a delusion and noncognition is senseless. If you want to reach the true path beyond doubt, place yourself in the same freedom as sky. You name it neither good nor not-good."

At these words Joshu was enlightened.

Mumon's comment: Nansen could melt Joshu's frozen doubts at once when Joshu asked his questions. I doubt though if Joshu reached the point that Nansen did. He needed thirty more years of study.

In spring, hundreds of flowers; in autumn, a harvest moon;
In summer, a refreshing breeze; in winter, snow will accompany you.
If useless things do not hang in your mind,
Any season is a good season for you.

Ekai, *The Gateless Gate,* 1228

Contents

Preface

Artificial intelligence has aroused debate ever since Hubert Dreyfus wrote his controversial report, *Alchemy and Artificial Intelligence* (1965). Philosophers and social scientists who have been influenced by European critical thought have often viewed AI models through philosophical lenses and found them scandalously bad. AI people, for their part, often do not recognize their methods in the interpretations of the critics, and as a result they have sometimes regarded their critics as practically insane.

When I first became an AI person myself, I paid little attention to the critics. As I tried to construct AI models that seemed true to my own experience of everyday life, however, I gradually concluded that the critics were right. I now believe that the substantive analysis of human experience in the main traditions of AI research is profoundly mistaken. My reasons for believing this, however, differ somewhat from those of Dreyfus and other critics, such as Winograd and Flores (1986). Whereas their concerns focus on the analysis of language and rules, my own concerns focus on the analysis of action and representation, and on the larger question of human beings' relationships to the physical environment in which they conduct their daily lives. I believe that people are intimately involved in the world around them and that the epistemological isolation that Descartes took for granted is untenable. This position has been argued at great length by philosophers such as Heidegger and Merleau-Ponty; I wish to argue it technologically.

This is a formidable task, given that many AI people deny that such arguments have any relevance to their research. A computer model, in their view, either works or does not work, and the question is a purely technical one. Technical practice is indeed a valuable way of knowing, and my goal is not to replace it but to deepen it. At the same time, AI has had great trouble understanding certain technical impasses that philosophical methods both predict and explain. The problem is one of consciousness: the AI community has lacked the intellectual tools that it needs to comprehend its own difficulties. What is needed, I will argue, is a critical technical practice – a technical practice for which critical reflection upon the practice is part of the practice itself. Mistaken ideas about human nature lead to recurring patterns of technical difficulty; critical analysis of the mistakes contributes to a recognition of the difficulties. This is obvious enough for AI projects that seek to model human life, but it is also true for AI projects that use ideas about human beings as a heuristic resource for purely technical ends.

Given that no formal community of critical technical practitioners exists yet, this book necessarily addresses two very different audiences, technical and critical. The technical audience consists of technical practitioners who, while committed to their work, suspect that it might be improved using intellectual tools from nontechnical fields. The critical audience consists of philosophers and social scientists who, while perhaps unhappy with technology as it is, suspect that they can make a positive contribution to its reform.

In my experience, the first obstacle to communication between these audiences is the word "critical," which for technical people connotes negativity and destruction. It is true that critical theorists are essentially suspicious; they dig below the surface of things, and they do not expect to like what they find there. But critical analysis quickly becomes lost unless it is organized and guided by an affirmative moral purpose. My own moral purpose is to confront certain prestigious technical methodologies that falsify and distort human experience. The purpose of critical work, simply put, is to explain how this sort of problem arises. Technical people frequently resist such inquiries because they seem to involve accusations of malfeasance. In one sense this is true: we all bear some responsibility for the unintended consequences of our actions. But in another sense it is false: critical research draws attention to structural and cultural levels of explanation – to the things that happen through our actions but that exist beneath our conscious awareness.

In the case of AI, I will argue that certain conceptions of human life are reproduced through the discourses and practices of technical work. I will also argue that the falsehood of those conceptions can be discerned in their impracticability and that new critical tools can bring the problem into the consciousness of the research community. The point, therefore, is not to invoke Heideggerian philosophy, for example, as an exogenous authority that supplants technical methods. (This was not Dreyfus's intention either.) The point, instead, is to expand technical practice in such a way that the relevance of philosophical critique becomes evident *as a technical matter.* The technical and critical modes of research should come together in this newly expanded form of critical technical consciousness.

I am assuming, then, that both the technical and critical audiences for this book are sympathetic to the idea of a critical technical practice. Writing for these two audiences simultaneously, however, has meant reckoning with the very different genre expectations that technical and critical writing have historically entailed. Technical texts are generally understood to report work that their authors have done; they are focused on machinery in a broad sense, be it hardware, software, or mathematics. They open by making claims – "Our machinery can do such and such and others' cannot" – and they confine themselves to demonstrating these claims in a way that others can replicate. They close by sketching future work – more problems, more solutions. Critical texts, by contrast, *are* the work that their authors have done. Their textuality is in the foreground, and they are focused on theoretical categories. They open by situating a problematic in an intellectual tradition, and they proceed by narrating their materials in a way that exhibits the adequacy of certain categories and the inadequacy of others. They close with a statement of moral purpose.

When a technical audience brings its accustomed genre expectations to a critical text or vice versa, wild misinterpretations often result, and I have spent too many years revising this text in an attempt to avoid seeming either scandalous or insane. The result of this effort is a hybrid of the technical and critical genres of writing. By intertwining these two strands of intellectual work, I hope to produce something new – the discursive forms of a critical technical practice. This being a first attempt, I will surely satisfy nobody. Everyone will encounter whole chapters that seem impossibly tedious and other chapters that seem unreasonably compressed. I can only ask forbearance and hope that each reader will imagine the plight of other readers who are approaching the book from the opposite direction.

This amalgamation of genres has produced several curious effects, and rather than suppress these effects I have sought to draw them out and make them explicit. One such effect is a frequent shifting between the levels of analysis that I will call reflexive, substantive, and technical. In one section I will explicate the ground rules for a contest among substantive theories, and on the next I will advocate one of these theories over the others. In one chapter I will criticize technical uses of language, and in the next I will start using language in precisely those ways in order to show where it leads.

Another such effect is a frequent overburdening of terms. Sometimes, as with "logic," a term has evolved in different directions within the critical and technical traditions, so that it is jarring to bring the divergent senses together in the same text. In other cases, as with "problem" and "model," technical discourse itself employs a term in at least two wholly distinct senses, one methodological and one substantive. And in yet other cases, technical and critical vocabulary together have drawn so many words out of circulation that I cannot help occasionally using some of them in their vernacular senses as well. I have tried to make the senses of words clear from context, and in most cases I have provided notes that explain the distinctions.

The peculiarity of my project might also be illustrated by a comparison with Paul Edwards's (1996) outstanding recent history of AI. In the language of social studies of technology (Staudenmaier 1985), Edwards opposes himself to "internalist" studies that explain the progress of a technical field purely through the logic of its ideas or the economics of its industry. He observes that internalist studies have acquired a bad name from their association with the sort of superficial, self-justifying history that Kuhn (1962) lamented in his analysis of "normal science." In response to this tendency, Edwards (1996: xiv) positions his work as a "counterhistory," drawing out the interactions among cultural themes, institutional forces, and technical practices that previous studies have inadvertently suppressed.

While I applaud this kind of work, I have headed in an entirely different direction. This book is not only an internalist account of research in AI; it is actually a work *of* AI – an intervention within the field that contests many of its basic ideas while remaining fundamentally sympathetic to computational modeling as a way of knowing. I have written a counterhistory of my own. By momentarily returning the institutional dimensions of AI to the periphery, I hope to permit the esoteric practices of the field to emerge in their inner logic. Only then will it be possible to appreciate their real power and their great recalcitrance.

Not only is the daily work of AI firmly rooted in practices of computer system design, but AI researchers have also drawn upon deeper currents in Western thought, both reproducing and transcending older ideas despite their conscious intentions. Would-be AI revolutionaries are continually reinventing the wheel, regardless of their sources of funding, and I hope to convey some idea of how this happens. AI is not, however, intellectually consistent or static; to the contrary, its development can be understood largely as successive attempts to resolve internal tensions in the workings of the field. I want to recover an awareness of those tensions through a critical exhumation of their sources.

My expository method is hermeneutic. I want to exhibit AI as a coherent totality, and then I want to turn it inside out. To do this, I have painted a big picture, examining the most basic concepts of computing (bits, gates, wires, clocks, variables, seriality, abstraction, etc.) and demonstrating their connection to seemingly more contentious ideas about such matters as perception, reasoning, and action. I will pass through some of these topics several times in different ways. My purpose is not to produce an exhaustive linear history but to assemble a complex argument. Technical ways of knowing are irreducibly intuitive, and each pass will open up a new horizon of intuition based on technical experience. AI has told variations on a single story about human beings and their lives; I believe that this story is wrong, and by forcing its tensions to the surface I hope to win a hearing for a completely different story. My goal, however, is not to convince everyone to start telling that same story. Computer modeling functions as a way of knowing only if the modelers are able to hear what the materials of technical work are trying to tell them, and if they respond by following those messages wherever they lead. The only way out of a technical impasse is through it.

I owe many debts. John Brace taught me mathematics, and Chuck Rieger and Hanan Samet introduced me to the field of AI. At MIT, many of the ideas in this book arose through conversations with David Chapman. He is also the principal author of the computer program that I discuss in Chapter 13. John Batali, Gary Drescher, Ken Haase, and Ian Horswill were early members of the debating society in which I first practiced the arguments. I appreciate the supportively skeptical comments of Randy Davis, Ken Forbus, Pat Hayes, and Dan Weld, and the philosophical assistance of Jon Doyle.

As the internal problems in AI became clear, I went looking for people who could explain them to me. Hubert Dreyfus, Harold Garfinkel, and Lucy Suchman introduced me to the phenomenological tradition; Jean Comaroff, John Comaroff, Bill Hanks, and Jean Lave introduced me to the dialectical tradition; and George Goethals introduced me to the psychoanalytic tradition.

Parts of this book began life in dissertation research that I conducted at the MIT Artificial Intelligence Laboratory, and I am indebted to Mike Brady, Rod Brooks, Gerald Jay Sussman, and Patrick Winston for their guidance. Aside from the people I have already mentioned, Jonathan Amsterdam, Mike Dixon, Carl Feynman, and Eric Saund wrote helpful comments on drafts of that work. I was supported for most of that time by a graduate fellowship from the Fannie and John Hertz Foundation, and support for the AI Laboratory's artificial intelligence research was provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-85-K-0124.

Work on this manuscript began at the University of Chicago, and I appreciate the support and comments of Tim Converse, Kris Hammond, Charles Martin, Ron McClamrock, and the other members of the University of Chicago AI group. As the manuscript evolved and became a book, it benefited greatly from extensive comments by Steve Bagley, David Chapman, Julia Hough, Frederic Laville, Lucy Suchman, and Jozsef Toth, and from the assistance of John Batali, Margaret Boden, Bill Clancey, David Cliff, Mike Cole, Bernard Conein, Johan de Kleer, Bruce Donald, Yrjö Engeström, Jim Greeno, Judith Gregory, David Kirsh, Rob Kling, Jim Mahoney, Ron McClamrock, Donald Norman, Beth Preston, Stan Rosenschein, Penni Sibun, Rich Sutton, Michael Travers, and Daniel Weise. Paul Edwards and Brian Smith were kind enough to provide me with drafts of their own books prior to publication. Mario Bourgoin directed me to the epigraph (which is taken from Paul Reps, ed., *Zen Flesh, Zen Bones* [Anchor Press, n.d.], by permission of Charles E. Tuttle Publishing Company of Tokyo, Japan). My apologies to anybody I might have omitted.