Extending Mechanics to Minds

This book deploys the mathematical axioms of modern rational mechanics to help the reader understand minds as mechanical systems that exhibit actual, not metaphorical, forces, inertia, and motion. Using precise mental models developed in artificial intelligence, the author analyzes motivation, attention, reasoning, learning, and communication in mechanical terms.

These analyses provide psychology and economics with new characterizations of bounded rationality, provide mechanics with new types of materials exhibiting the constitutive kinematic and dynamic properties characteristic of different kinds of minds, and provide philosophy with a rigorous theory of hybrid systems combining discrete and continuous mechanical quantities. The resulting mechanical reintegration of the physical sciences that characterize human bodies and the mental sciences that characterize human minds opens traditional philosophical and modern computational questions to new paths of technical analysis.

JON DOYLE is SAS Institute Distinguished Professor of Computer Science at North Carolina State University in Raleigh, North Carolina. His formative contributions helped generate widespread studies of reason (or truth) maintenance and nonmonotonic logic; self-governing reflective agents; rational and economic foundations for reasoning, representation, and learning; and qualitative decision theory. Dr. Doyle is a Fellow of the American Association for Artificial Intelligence (AAAI). He has served as a member of the AAAI Executive Council, as chairman of ACM SIGART, and as a member of several editorial boards, including those of the *Journal of Artificial Intelligence Research*, *Computational Intelligence*, *ACM Computing Surveys*, the *Journal of Logic, Language and Information*, and *AI Communications*.

Extending Mechanics to Minds

The Mechanical Foundations of Psychology and Economics

JON DOYLE
North Carolina State University

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/9780521861977

First published 2006

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

Library of Congress Cataloging-in-Publication data
Doyle, Jon.
Extending mechanics to minds : the mechanical foundations of psychology
and economics / Jon Doyle.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-86197-7 (hardcover)
1. Mechanics, Applied – Mathematics. 2. Artificial intelligence. I. Title.
TA350.D69 2006
006.3 – dc22 2006002823

ISBN 978-0-521-86197-7 Hardback

For Carol

Wife of noble character,
Clothed with strength and dignity,
Worth far more than rubies,
Gift of God.

O unaussprechlich süßes Glück! Wer ein holdes Weib errungen, Stimm in unsern Jubel ein! Nie wird es zu hoch besungen, Retterin des Gatten sein.	O unspeakably sweet fortune! Let him who has won a fair wife join in our rejoicing! Our song can never praise too much the savior of her husband.

Beethoven, Bouilly, & Sonnleithner, *Fidelio*, Act II Finale

Contents

Preface

This book uses concepts from mechanics to help the reader understand and formalize theories of mind, with special concentration on understanding and formalizing notions of rationality and bounded rationality that underlie many parts of psychology and economics. The book provides evidence that mechanical notions including force and inertia play roles as important in understanding psychology and economics as they play in physics. Using this evidence, it attempts to clarify the nature of the concepts of motivation, effort, and habit in psychology and the ideas of rigidity, adaptation, and bounded rationality in economics. The investigation takes a mathematical approach. The mechanical interpretations developed to characterize mechanical reasoning and rationality also speak to other questions about mind, notably questions of dualism and materialism.

More generally, the exposition sketches the development of psychology and economics as subfields of mechanics by showing how one might formalize representative psychological and economic systems in such a way that these formalized systems satisfy modern axiomatic treatments of mechanics. This formalization explicates psychological and economic concepts under study by identifying corresponding properties of certain mechanical systems. Not all concepts of psychology and economics correspond to mechanical notions, and among those that do, not all concepts currently popular in psychology and economics correspond to natural mechanical ones. The concepts studied nevertheless permit natural identifications of familiar psychological or economical notions as natural mechanical notions, formalizing the memory of an agent as its mass and internal configuration, and the agent's motivations as forces that produce changes in the agent's momentum by means of changes in its mass and velocity.

Psychology and economics have endured many attempts at physical analogy, most perhaps deserving the apparent failure they reaped as a result of inadequate formal basis, lack of ties to concrete problems, or sheer wrong-headedness about the phenomena and concepts of physics, psychology, or economics. The present effort avoids a similar fate by combining modern mathematical axiomatizations of mechanics developed by Walter Noll (1958, 1963, 1972, 1973) with concrete examples of proven interest from computation and artificial intelligence. The result shows that at least some minds constitute actual mechanical systems, not merely participants in occasional mechanical analogies.

Program

There is no philosophy that is not founded upon knowledge of the phenomena, but to get any profit from this knowledge it is absolutely necessary to be a mathematician.

Daniel Bernoulli, 1763 letter to John Bernoulli III (Truesdell 1984b, pp. 19–20)

I approach the foundations of psychology and economics dissatisfied with the analytical and mathematical concepts typically employed to describe these fields, concepts that, especially in psychology, frequently vex the student, fail the scholar, and leave the subject needlessly disconnected from the rest of science. At the same time, I rejoice in the panoply of concepts and methods current mathematics offers for ordering the world, to borrow Jaffe's (1984) apt phrase, concepts that have been applied relentlessly to understanding physics and more recently computation and economics but less so to psychology. The past century of mathematics has provided astounding progress in understanding logic, computation, meaning, and ideal rationality, but one need not assume that in coming to a mathematical understanding of mind these contributions will continue to play the roles current sensibilities might assign them.

The search for better ways of formulating problems and possible solutions forms the most fundamental activity of any field, and indeed of much of thinking and computing. Hamming (1962) wrote that "The purpose of computing is insight, not numbers," and Minsky (1974, pp. 78, 56) wrote that "thinking begins first with suggestive but defective plans and images, that are slowly (if ever) refined and replaced by better ones," and that "[t]he primary purpose in problem solving should be better to understand the problem space, to find representations within which the problems are easier to solve."

In searching for better and more appropriate formulations, each field of thought seizes upon successive sets of fundamental concepts and uses these conceptual "atoms" as the basis for its views of the problems and theories of the field. Each set of atoms has its day, with intensive exploration revealing how well it illuminates the problems and how well it eases their solution. As its limits become clearer, theorists introduce some new atom or set of atoms to provide an even better point of view. The aim of the search is to find the most appropriate conceptual atoms, those that cleanly divide the phenomena and problems in powerful ways. As Miller (1986) points out, most conceptualizations of psychology in the past have focused on sets of atoms inadequate to the task, "dismembering cognition," in his vivid phrase, with a set of concepts that "leaves its object shattered in lumps" rather than a set that "carves a topic at its joints." Mathematicians use the terms *beauty* and *depth* as terms of approbation for good theories, expressing something of the same theoretical esthetic as Miller, reviling the logger and applauding the wright. One can observe the search process especially clearly in mathematics. In the large, one finds alternative foundations for all of mathematics, foundations based on logic, set theory, category theory, and intuitionism. In the small, specific mathematical theories are decomposed into parts (matrix theory into noncommutative algebra and representations over specific rings) and recast by exchanging conclusions and axioms (exchanging natural numbers and arithmetic for zero, a successor function, and an exclusion or comprehension principle).

I use the term *rational psychology* to name the branch of mathematics aiming to investigate psychology by means of the most fit mathematical concepts, that is, the mathematical concepts that yield the simplest, most elegant, most powerful, and most insightful formulation of psychological theories (Doyle 1983f). *Rational* here refers to conceptual, mathematical analysis, not to any putative rationality or irrationality of the systems under study. Historically the term referred to the philosophical study of psychology. My use of the term follows the model of rational mechanics, which has named the study of the foundations of mechanics since the time of Newton.

Rational psychology studies mathematically possible organizations for agents. I call these possible organizations "psychologies" (Doyle 1982a, 1990a). One can thus view rational psychology as seeking to classify all possible psychologies in the same way that group theory seeks to classify all possible groups. The special laws of mechanics represent such classificatory devices. Identifying special psychological materials or structures characteristic of interesting classes of agents represents a central method of the field.

The present effort draws directly on the model of rational mechanics by eschewing theories formulated in terms of representations in favor of theories cast in terms of the behaviors themselves, and by treating the problem of understanding different types of mental organization and behavior as formalization of distinct types of psychological materials, rather than as a search for the "true" theory of psychology or economics. Each type of psychological material obeys not only the general mechanical laws applicable to every material but also special laws characteristic of the specific material.

Prerequisites

The ideas presented in this book draw on a broad range of fields, including mathematics, physics, philosophy, psychology, economics, politics, logic, and computation. The ideal reader of this book would bring to the reading a solid grasp of the foundations and development of mathematics, classical and relativistic rational mechanics, quantum theory, logic, probability, the theory of computation, psychology, artificial intelligence, and mathematical, psychological, and political economics. Undoubtedly a keen appreciation of Aeschylus and Aristophanes would also help, but even the former list probably rules out everyone, including the author. Cognizant of the demands of the material, the book has been written with hopes that readers possessing common college acquaintance with mathematical physics, computation, and artificial intelligence will find something intelligible and interesting in the book. Familiarity with (not necessarily mastery of) the mathematical way of thinking, the broad spectrum of mathematical ideas, and their basic mechanical applications would be very valuable; Mac Lane's (1986) useful survey represents a good model of this knowledge. The interior survey chapters of Penrose's (1989) book and Gurtin's (1981) introduction to continuum mechanics might offer useful supplements. Truesdell's (1968b; 1984c) essays and Benvenuto's (1991) history of mechanics provide valuable perspective on the recency and impermanence of current interpretations of mechanical concepts. Russell and Norvig's (2002) textbook of artificial intelligence employs economic rationality to convey a unified perspective of that field in a manner reasonably consonant with the approach taken here.

Plea

The player who stakes his whole fortune on a single play is a fool, [but the science of mathematics] merely shows that other players are greater fools.
(Julian Lowell Coolidge 1909, p. 189)

I am fortunate that virtually every element of the development presented here is known, even well known, in at least one field of scholarly thought. While I know much about some of these fields, *la vida breve* frustrates gaining thorough expertise in each. Accordingly, much as I love books that present what for a generation or two seem the most elegant, general, or final forms of ideas, completing this book required continuous battle against the urge to perfect. My aim in writing instead was to develop each component enough to show its place in the theory and to make the basic ideas formal enough to enable interested parties to find their proper mathematical form. To reapply another sentiment earlier expressed by Julian Lowell Coolidge (1940, p. xii), while my own inadequacy for such a task has been abundantly evident to me, it did not seem sufficient reason for not making the attempt.

Although many research monographs provide a chapter outlining future work, such a chapter would seem a joke here, for one may view the entire book as outlining a program of future work. I am told that Paul Halmos once ended a talk by saying words to the effect that "If I stopped now, I'd be stopping before the most important point, but I don't have anything more to say" and then walking away. I am sure this book stops before the most important point, and I hope the reader will press on regardless.

Because of the diversity of the subject matter and background, a careful reading is likely to require substantial effort, even for a reader familiar with the technical developments of a relevant field. I know of no short-term remedy for this difficulty, but I try to provide some repetition of explanation and motivation appropriate to points at which readers of different interests might take up the text.

Completing this exposition proved difficult because numerous possible alternative treatments presented themselves at every point of the development, and at times I despaired of pulling together sensible selections from the great and bewildering variety of alternative views. At such times, the best hope for completing any exposition seemed to be to write a collection of interrelated and potentially competing vignettes of each alternative view of each topic, in the style of Minsky's (1986) *Society of Mind*. Although that approach has something to recommend it, not least that it may offer the only feasible possibility for describing systems as complicated as the mind with instruments as simple as the mind, I rebelled against the approach because mechanics, unlike psychology, possesses an extended and ordered conceptual development. Surely, I thought, a mechanical understanding of mind must admit at least some of this logical development.

Even if one finds the development presented here compelling or convincing, one should not assume the main interpretations presented here to be the best ones, no matter how many interesting and intuitive properties they exhibit. Although I believe many interpretations presented here lie on the right path, there might be more than one reasonable mechanical interpretation of some psychological or economic systems. Sufficiently abstract systems might admit alternative possible mechanical interpretations. Common computational models, such as Turing machines (Turing 1936), finite automata (Rabin & Scott 1959), and the random-access stored program machine, or RASP (Elgot & Robinson 1964), might fall in this class. One can blame some of the difficulty I experienced in completing this exposition on the interpretational ambiguities I encountered in analyzing the more structured equilibrium-transition model of the Reason Maintenance System, or RMS (Doyle 1983e, 1994), which served as the primary focus of my formalization efforts for many years.

Dana Scott (1989, p. 5) once wrote that "[y]ou cannot have a clear conscience in Mathematics if you do not follow up the possibilities," and in this respect my conscience is not clear, for I regard the theory presented here merely as a down payment on a true mechanics of mind. In some ways, the theory that follows hews much too closely to traditional mechanical formalism, and thus raises the suspicion that it simplifies too much to capture the true complexities of the mind. I would love to learn that one loses nothing by these simplifications and that this traditional mechanical form suffices to characterize even the most subtle mental phenomena and properties, but I do not expect to live long enough to see such good news or its refutation. The present volume gives some indications of developments in the direction of these mental subtleties. I hope to write another examining even more in further detail.

I thus expect that further examination of these topics will find ways to justify and perfect the interpretation given here, to identify and justify other interpretations, or both. My own limitations—in knowledge, in competence, in vision, and in time—require me to leave most such investigations to others.

Past

This book represents part of a larger effort aiming to come to a mathematical understanding of rationality and, more generally, all of thinking. My work on mathematical and computational analysis of thinking began in about 1973, yielding the early works *A Truth Maintenance System* (Doyle 1979), *Explicit Control of Reasoning* (de Kleer *et al.* 1977), *Non-Monotonic Logic I* (McDermott & Doyle 1980), and *A Model for Deliberation, Action, and Introspection* (Doyle 1980). Those works bear few traces of the mechanical perspective explored here, apart from a concern with the kinematical, mechanistic approach that underlies most machine computation and modern artificial intelligence, but they provided the central concrete example of the nonmonotonic RMS and the insight used to construct the mechanical interpretation of thinking presented here.

The present investigation of mechanical treatments of mind began in about 1981 as I sought to understand RMS conservatism in terms of least action principles and Lagrangian formalism. The first outlines and partial drafts of this book date from 1982–1983, during which time the importance of inertia for psychology and economics became clear. In 1984 I completed the first extended draft, numerous portions of which survive in the present version. A variety of interruptions delayed completion, so that the present exposition took form in 1998–1999 and underwent gradual expansion before coming into nearly final form in 2003, with brief synopses circulated along the way (Doyle 2001, 2002a).

The prolonged period of completion is misleading, however. Mechanical investigations provided the direct inspiration for and some technical elements of numerous ideas presented in earlier monographs, including *The Foundations of Psychology* (Doyle 1982a, 1990a), *What Is Church's Thesis?* (Doyle 1982b, 2002b), *Some Theories of Reasoned Assumptions* (Doyle 1983e), *Artificial Intelligence and Rational Self-Government* (Doyle 1988a), *Rationality and Its Roles in Reasoning* (Doyle 1992a), and *Reasoned Assumptions and Rational Psychology* (Doyle 1994). The present book provides numerous references to these writings, wherein the interested reader might find additional details.

Acknowledgments

The first debt of the present work accrues to the writings of Clifford Truesdell and Walter Noll. As a first approximation, one might regard the contributions of the present work as constituting a footnote to Noll's achievements.

I owe great debts to the Laboratory for Computer Science of the Massachusetts Institute of Technology (1988–2001), to the Computer Science Department of Carnegie Mellon University (1981–1988), and to the Computer Science Department of Stanford University (1980–1981) for supporting me materially and intellectually during the development of these ideas, as well as to the Fannie and John Hertz Foundation for supporting my graduate work at the Massachusetts Institute of Technology Artificial Intelligence Laboratory (1975–1980). Since 2001, a SAS Institute Professorship in the Computer Science Department of North Carolina State University has provided me the freedom to complete this work.

I owe much to stimulation afforded by conversations over the years about mechanics, physics, and related ideas with Harold Abelson, John Baez, Jaime Carbonell, Randall Davis, Johan de Kleer, Merrick Furst, Clark Glymour, Robert Hartwig, Robert Hermann, David Israel, Norman Margolus, Matthew Mason, John McCarthy, Marvin Minsky, Joel Moses, Allen Newell, Walter Noll, Ramesh Patil, Michael Rabin, Larry Rudolph, Elisha Sacks, Dana Scott, Herbert Simon, Richard Statman, Guy Steele, Ernest Stitzinger, Richmond Thomason, Tommaso Toffoli, Robert Hugh Walker, Michael Wellman, and Jack Wisdom. I owe much to the instruction of my late teachers David Bourgin, John Mac Nerney, and Jürgen Schmidt. I also thank Matthias Stallman for help with translations from the German.

I especially thank Joseph A. Schatz, Gerald Sussman, and Peter Szolovits for boundless insight and encouragement. Joe started me thinking about many of these problems and taught me the perspective needed to persevere. Many

were the times the words of the Teacher rang in my ears as I wondered if all this were foolishness and vanity. Joe helped remind me that even if all this proved foolishness, which he doubted, I would find it greater foolishness not to, as Rilke says, try the last. In earlier years Gerry provided encouragement sufficient for seven and taught me how to think about physics, computation, and intelligence. In later years he organized informal seminars on mechanics and quantum theory with Jack Wisdom that helped keep truth, beauty, and key questions in view. Pete provided me a home constantly filled with life, ideas, and friendship.

My greatest debts accrue to my parents, Leo and Marilyn Doyle, for sustenance and shaping in ways I continue to learn to appreciate; to my bride Carol, son Alan, and daughter Emily for their patient love, encouragement, and help in retaining what vestiges of humanity I bear; and to God for giving me life, time, joy, and inspiration. If there be goodness here, *soli Deo gloria*.

Raleigh, North Carolina — Jon Doyle
July 2005

Outline of the book

This book motivates the mechanical study of intelligence and rationality, reviews modern mechanics and its historical relations to psychology, adapts mechanical axioms to cover hybrid and discrete systems, presents illustrative formalizations of representative rational systems in psychology and economics, and reflects on the character of mechanical laws and theories. My exposition of these ideas divides the development into several parts.

Part I: Reconciling Natural and Mental Philosophy

Part I introduces the problem, the aims of the project, and some of its background.

- Chapter 1 introduces the subject and ideas of the book in the context of understanding the mind and constructing mechanical persons.
- Chapter 2 discusses the benefits of the mechanical approach, especially in shedding new light on questions of materialism and new methods for characterizing limits to rationality.
- Chapter 3 explores in greater detail the mechanical viewpoint and its history, briefly relating the project to past efforts on mechanical interpretations of psychological and economic phenomena as an aid to understanding better the subsequent development.

Part II: Reconstructing Rational Mechanics

Part II explains the structure of modern rational mechanics and reformulates the axiomatic development in a manner appropriate to hybrids of continuous, discrete, physical, and mental mechanical subsystems.

- Chapter 4 summarizes the theoretical structure of modern rational mechanics, including the modern conception of physical law and the division of mechanical laws into general and special laws.
- Chapter 5 develops the kinematic axioms of mechanics, generalizing the usual axioms to accommodate discrete aspects of space, hybrid mechanical systems, and indeterministic worlds.
- Chapter 6 develops the dynamical axioms, which parallel the usual developments in most respects, evidencing the modesty of the reformation of mechanics needed to cover psychology and economics.
- Chapter 7 reconsiders several characteristics of mechanical systems in light of the reconstruction of kinematical and dynamical axioms, including determinism, continuity, conservation principles, least action principles, reversibility, and locality.

Part III: Mechanical Minds

Part III presents mechanical formalizations of key psychological and economical notions.

- Chapter 8 notes the wide variety of mental organizations presumed or postulated by theorists in many fields, identifies one special class involving plural, discrete, affective cognition for special examination, and summarizes the structure of the Reason Maintenance System, or RMS, that illustrates this class of psychologies.
- Chapter 9 uses the mechanical formalism to examine mind–body duality and the plurality inherent in mental organization and faculties.
- Chapter 10 sets out the basic framework of discrete mechanical motion, mass, and force underlying the analyses of the following chapters.
- Chapter 11 takes a detailed look at simple reasoning patterns as exemplifying mechanical forces and then offers speculative relations between these simple forces and more complex mathematical concepts and mechanical phenomena.
- Chapter 12 analyzes the mechanical nature of rationality and limits on rationality, including effort, volition, inherent intelligence, and the forces generated by desire, intention, habit, refraction, and other rational motives.
- Chapter 13 characterizes learning in terms of mechanical concepts of mass, persistent configuration, plastic deformation, and relaxation responses to applied forces.

- Chapter 14 studies mental uncertainty as a mechanical phenomenon, constructing a straightforward theory of measurement that yields structures akin to subjective probabilities and weakness of will, and then presenting a speculative subjective measurement structure with connections to concepts of quantum theory.

Part IV: The Metaphysics of Mechanics

Part IV discusses a number of mainly philosophical characteristics of physical theories in relation to mechanics.

- Chapter 15 discusses implications of the mechanical axioms for traditional philosophical questions of materialism, especially similarities between the numerous past broadenings of concepts of materialistic theories and the present further broadening to the materials of psychology and economics.
- Chapter 16 addresses the reducibility of physical law to behaviors of elementary particles and considers related topics including the possibility of discovering additional physical laws, the uniformity of physical laws, and other topics connected with the completeness of physics.
- Chapter 17 relates mental mechanics to notions of effective computation, both to understand computation in mental terms and to understand the relation of effectiveness to mechanical theory.
- Chapter 18 examines issues pertaining to the finiteness or infiniteness of the universe, especially as these relate to discrete models of mechanics as developed here.

Part V: Conclusion of the Matter

- Chapter 19 summarizes and assesses the work, identifies some additional issues for future exploration, and reflects on the history of some of the ideas.