ARCHIVAL INSIGHTS INTO THE EVOLUTION OF ECONOMICS

MINDS, MODELS AND MILIEUX

COMMEMORATING THE CENTENNIAL OF THE BIRTH OF HERBERT SIMON

Edited by

Roger Frantz

and Leslie Marsh

Archival Insights into the Evolution of Economics

This series provides a systematic archival examination of the process by which economics is constructed and disseminated. All the major schools of economics will be subject to critical scrutiny; a concluding volume will attempt to synthesize the insights into a unifying general theory of knowledge construction and influence.

Series Editor: **Robert Leeson**

Titles include:

Robert Leeson (*editor*)
THE KEYNESIAN TRADITION

Robert Leeson (*editor*)
THE ANTI-KEYNESIAN TRADITION

Robert Leeson (*editor*)
AMERICAN POWER AND POLICY

Roger Frantz and Robert Leeson (*editors*)
HAYEK AND BEHAVIORAL ECONOMICS

Robert Leeson (*editor*)
HAYEK: A COLLABORATIVE BIOGRAPHY
Part I, Influences from Mises to Bartley

Robert Leeson (*editor*)
HAYEK: A COLLABORATIVE BIOGRAPHY
Part II, Austria, America and the Rise of Hitler, 1899–1933

Robert Leeson (*editor*)
HAYEK: A COLLABORATIVE BIOGRAPHY
Part III, Fraud, Fascism and Free Market Religion

Robert Leeson (*editor*)
HAYEK: A COLLABORATIVE BIOGRAPHY
Part IV, England, the Ordinal Revolution and the Road to Serfdom, 1931–50

Robert Leeson (*editor*)
HAYEK: A COLLABORATIVE BIOGRAPHY
Part V, Hayek's Great Society of Free Men

Robert Leeson (*editor*)
HAYEK: A COLLABORATIVE BIOGRAPHY
Part VI, Good Dictators, Sovereign Producers and Hayek's 'Ruthless Consistency'

David Hardwick and Leslie Marsh (*editors*)
PROPRIETY AND PROSPERITY
New Studies on the Philosophy of Adam Smith

Roger Frantz and Leslie Marsh (*editors*)
MINDS, MODELS AND MILIEUX
Commemorating the Centennial of the Birth of Herbert Simon

Forthcoming title:

Robert Leeson (*editor*)
HAYEK AND THE AUSTRIAN SCHOOL

Archival Insights into the Evolution of Economics
Series Standing Order ISBN: 978–1–403–99520–9 (hardback)
(*outside North America only***)**

You can receive future titles in this series as they are published by placing a standing order. Please contact your bookseller or, in case of difficulty, write to us at the address below with your name and address, the titles of the series and the ISBN quoted above.

Customer Service Department, Macmillan Distribution Ltd, Houndmills, Basingstoke, Hampshire RG21 6XS, England, UK.

Minds, Models and Milieux

Commemorating the Centennial of the Birth of Herbert Simon

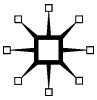
Edited by

Roger Frantz
Professor of Economics, San Diego State University, USA

and

Leslie Marsh
Senior Research Associate, The University of British Columbia, Canada

palgrave
macmillan



Softcover reprint of the hardcover 1st edition 2016 978-1-137-44249-9

First published 2016 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN 978-1-349-56680-8 ISBN 978-1-137-44250-5 (eBook)
DOI 10.1057/9781137442505

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Simon, Herbert A. (Herbert Alexander), 1916–2001, honoree. | Frantz, Roger S., editor. | Marsh, Leslie, editor.
Title: Minds, models and milieux : commemorating the centennial of the birth of Herbert Simon / edited by Roger Frantz, Professor of Economics,San Diego State University, USA, Leslie Marsh, Senior Research Associate, University of British Columbia, Canada.
Description: New York : Palgrave Macmillan, 2016. | Series: Archival insights into the evolution of economics | Includes index.
Identifiers: LCCN 2015027206
Subjects: LCSH: Economics—History—20th century. | BISAC: BUSINESS & ECONOMICS / Economic History. | BUSINESS & ECONOMICS / Economics / General. | BUSINESS & ECONOMICS / Economics / Theory.
Classification: LCC HB87 .M546 2016 | DDC 330.092—dc23
LC record available at http://lccn.loc.gov/2015027206

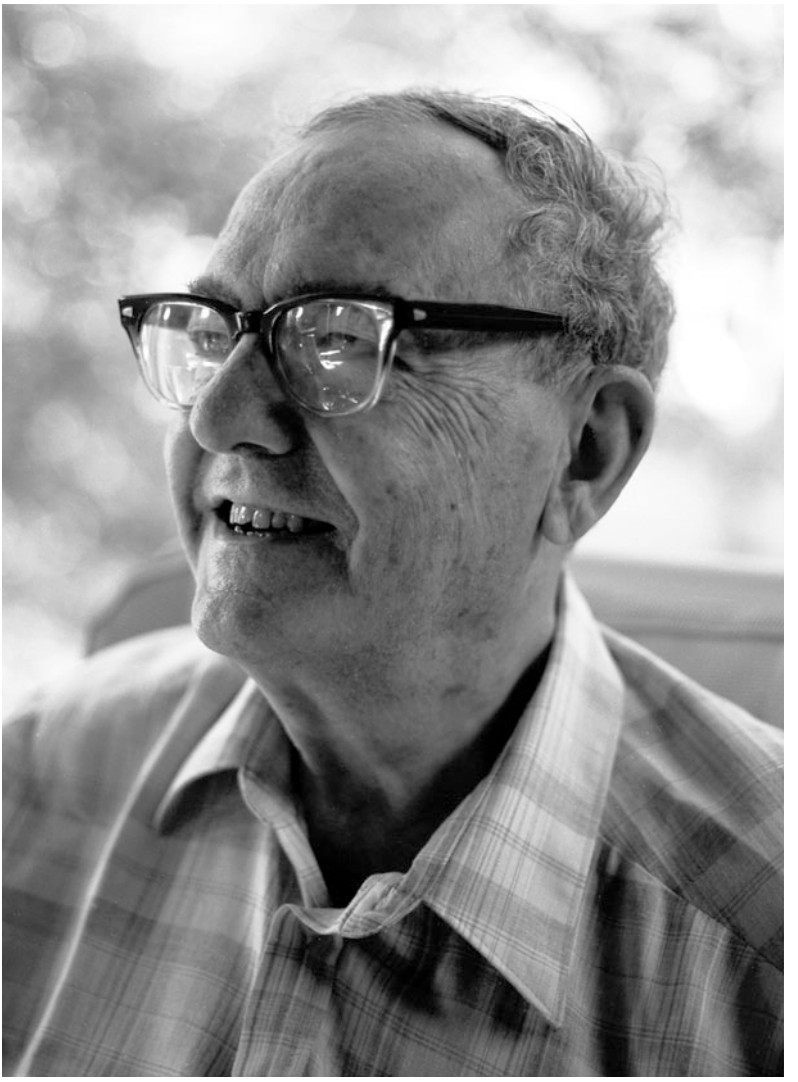
To my parents, Raymond and Shirley, and my brother Glenn,
all of whom passed away too early in life

R. F.

and for

Simon Powell

L. M.



Herbert Simon © Richard Y. Kain. Made available courtesy of Richard Y. Kain and Katherine Simon Frank

Contents

List of Figures and Tables

Figures

Tables

Foreword: Herbert Simon, the Man

How many children are conscious of their parents as different from their friends' parents? How many understand what it means to have a parent who is known to others via newspaper accounts and television appearances? Would a child sense how a 'famous' parent could be different from other parents? From my experience and observation growing up as the eldest of three of Herbert Simon's children, my answer to those questions is: zero, none, and no.

That is, those were my answers until people started asking questions and pointing out the differences. Why was your father in the newspaper? Why aren't your parents moving to another house now that your father has won the Nobel Prize? What was it like growing up with a famous father?

Outside observers (both young and old, it seems), who are strangers to or casually acquainted with a person who is publicly known, assume that somehow the relationship that others have with the 'known one' is different in some significant ways from the norm. It might be. But it doesn't have to be. My experience has been that my father's choices in parenting, lifestyle, and interactions with the world are very similar to those of the general public.

Why are these questions significant? What difference does it make? In this volume we are commemorating the 100th anniversary of Simon's birth. It's an appropriate time to consider Herbert Simon the Man as well as Herbert Simon the Scholar. I've wondered how to share some of the ways Herbert Simon integrated his personal life with his well-known work. Did one inform the other? In what ways? How consciously? The best way I've discovered to answer these questions is to look at a few instances that stand out in my recollection of my life with my father.

First and foremost, Herbert Simon thought of himself as being just like all other human beings. On several occasions he told me that he didn't consider himself special, even when he sensed that others were treating him as something other-worldly. He felt it set him at a distance from them, making for a particular loneliness. To be sure, people didn't treat him as special universally. He found comfortable relationships with family members, his personal friends, many of his colleagues, and others around the globe who, despite his reputation in academia, saw him as he saw himself.

My parents lived in Berkeley, California in the early 1940s where Herbert was director of Administrative Measurement Services under the Bureau of Public Administration at the University of California, Berkeley. During that period, he also prepared for, took, and passed his PhD prelim exams in political science at the University of Chicago. Afterwards, while still working, he wrote his dissertation. In it he laid the basis of his theory of bounded rationality that grew out of studies he'd done while employed during college in Chicago and as part of his employment in Berkeley. His dissertation, revised a few years later, became *Administrative Behavior*. During this time, I was born. Herbert Simon was 26 years old.

Shortly after my birth we moved to Chicago, where my father began to teach at the Illinois Institute of Technology. Herbert was surprised to find how easily he adapted to teaching, and found that he loved sharing his knowledge and the contact he had with students and their boundless energy and curiosity. He found that his practice early in life, of acting as if he were more social than he felt, paid off in this new work environment. The active presence of my mother in his life helped as well; she simplified his need to sustain numerous professional and personal relationships with her talent for making friends easily and organizing social occasions.

Many of the people my parents met in their early married life – colleagues, neighbors, and others – became life-long friends. With them, my father comfortably shared the simple pleasures that he pursued all his life. He made music with some, enjoyed listening to music with others. He read widely and loved to share with others what he learned through books, whether the plot of a novel, facts from non-fiction, or a new poem he'd discovered. He also loved nature. With no visible effort, he learned the Latin and common names of trees and plants, identified birds by sight and sound, and studied Geologic Survey maps for his hikes with family and friends in the mountains, on sand dunes and beaches, and, more often, in local city parks. His time spent gardening and pulling weeds (a favorite time for thinking) was more solitary, but he balanced that quiet time with frequent out-loud thinking, discussing, and debating with friends and family.

Except for his ordered chaos (he wasn't the most tidy person, and yet he was highly organized), in which his 'messes' consisted mostly of paper and books covering most flat surfaces in his office and his study at home, simplicity was a watchword for Herbert. The simple habits he developed early in his adult life freed him to think about what he considered more important matters. His aim? To avoid having to make little

decisions over and over, like what to wear or what to eat. His solution? He wore the same kinds of clothing every day – navy slacks, a pale blue or white dress shirt, dark socks, black leather shoes, and a V-neck cardigan if it was chilly. Breakfast was the same every day; a bowl of oatmeal, half a grapefruit, and a cup of black coffee sufficed! Lunch sometimes consisted of a candy bar, unless he was eating with a friend or colleague at work. Dinner was easy, as he relied on my mother to decide her home-cooked menu. This routine suited him so he could devote his active mind to interesting thoughts, new ideas, and careful observation of life around him.

As in many families, the dinner table was a place to share and catch up with the news of the day. In the kitchen before the meal, Herbert regaled my mother with daily updates about people he had seen, conversations he'd had, and analyses of minor political stresses at play, while she lent him her sympathetic ear as she prepared the meal. They were 'talking shop,' he said, whenever one of the three of us appeared to check on the cooking process. But at the dinner table he turned his attention to us, actively encouraging each to share information about what we had done that day. He never judged or criticized. He listened, responded, and sometimes even initiated a debate on a topic inspired by something someone had said. He engaged with each of us as individuals. Herbert often made funny observations, eliciting lots of laughter. Nearly every meal someone would pose a question and then pop up from the table to get the *World Almanac* or a volume of the *Encyclopedia Britannica* from the shelf in the living room. Admittedly, dinners like this might not be common in most families, but they were for us and, as pleasurable as they were, they provide the substance of many of my best childhood memories. To me, my siblings, and my mother, this wasn't an event with a 'famous man,' but just the way we were.

In the ways Herbert related to his family, he related similarly to everyone he came in contact with. He operated out of the belief that understanding other people's way of being, appreciating how they were similar and different from him, and calculating what each could offer the other, lent richness to his life. He considered the context of their culture if it was markedly different than his. Sometimes, when meeting a stranger, such as a seat-mate on an airplane, he made a point of trying to set the other at ease, gently questioning, listening, responding and questioning to know more. By the end of the conversation, the stranger would remark how good it was to share the time together, and wasn't it a wonder that they shared so much in common? He accomplished this bond merely by keenly observing and listening while he drew the other

out to talk about himself. By questioning, using the jargon the stranger introduced, and observing the stranger's style, Herbert appeared to know more about the stranger's life/work/interests than he actually did. This shy man managed to learn how to make other people feel comfortable by meeting them face-on, revealing a compatible soul.

Herbert made decisions easily, accepting the consequences and moving on. If the outcome of a decision didn't sit quite right with him, he acknowledged the lesson he learned and adjusted his next steps accordingly. How he learned to do this is not clear. But he was very aware of his talent and proclivity for decision-making. Ed Feigenbaum, one of Herbert's early graduate students, in the obituary he wrote about Simon for *Science* included this:

> in awe of his enormous knowledge and the range of his contributions, I once asked him to explain his mastery of so many fields. His unforgettable answer was, 'I am a monomaniac. What I am a monomaniac about is decision-making.' Studies and models of decision-making are the themes that unify most of Simon's contributions.
>
> How appropriate that he chose to focus his research on decision-making. (Feigenbaum, 2001)

As a child in his home, I was fortunate to have him model how seemingly effortlessly he drew conclusions and then acted on them.

How did my father teach me personally about decision-making? I can cite one incident that had a profound influence on me, but that represents Herbert's style. It was spring of my freshman year in college; I lived far from home in a dorm. I was struggling with an issue that I could not resolve. One evening, I called my father to ask him to tell me what to do. I think I probably worded it just that way! His answer, when I was finished talking, was: 'I know you'll make a good decision.' I was stunned. Clearly I had told him I wanted him to tell me what to do. But did I want him to decide? If he told me what to do, being 19, I might have done the opposite and rebelled. If he told me what to do, and if I did as he said, and then if it didn't work out, who would I blame? What would I have learned from that? After I hung up, I raged. But then I made a decision: to do nothing just then. And, parenthetically, it was a good decision. I was vehemently angry at the time, but my father's words buoyed me up later. In fact, they are always with me, giving me the courage to act on my own behalf as a responsible adult. What I've realized is that he had confidence in me, in my thought process, in my

decision-making mode, and by conveying that to me he set me free to pursue my own path.

Herbert was respectful of individuals' opinions and choices. While he didn't always agree with others, he talked with them and together they explored options. Sometimes Herbert would engage them in a fierce, challenging debate on a subject, thus forcing them to examine all sides of their point of view. He debated not to win, but rather for the mental exercise. He loved to play the role of devil's advocate.

There's more, so much more, about my father and his ways of being that I remember and that shaped me. As I write, I begin to see why others might think that I should have an answer to: 'What was it like growing up with a famous father?' Many things I took for granted in our family life are not what my friends' families experienced. But then I could argue that what they experienced, I didn't. Was that about having a famous father? Or rather about the normal range of differences in environments? Who's to say? The point of all this is more about how one man lived his life by his beliefs and values, how those beliefs and values were consistent over his work life and his personal life, and how he was so dedicated to his way of being that he was willing generously to share with the people who stopped long enough to get to know him as a human rather than as a figurehead or a hero.

As you read the chapters in this book about current research, inventions, and creations that his life's work inspired, remember that Herbert Simon was dedicated to understanding how people thought and how they made decisions; and that dedication drove him to be consistent in his relationships and behaviors.

Katherine Simon Frank
University of Minnesota

Reference

Feigenbaum, E. A. (2001). Herbert A. Simon, 1916–2001. *Science,* 16 March 2001: 2107.

Acknowledgments

Leslie would especially like to thank Doug and Dave Hardwick, Charlie Ramey, Sandy Del Vecchio and Richard Selin. Shannon Selin is, as usual, in daily practical support.

Both of us are deeply appreciative for the patience, support and efficiency of Laura Pacey, Assistant Editor of Economics at Palgrave Macmillan. Keith Povey must be commended for his judicious editing and proofing skills. Last, but by no means least, appreciation is due to the series commissioning editor, Robert Leeson.

Notes on the Contributors

Morris Altman is Professor and Head of the School of Economics and Finance at Victoria University of Wellington, New Zealand. He is the co-founding editor of the *Review of Behavioral Economics*; editor of *Behavioral Economics With Smart People* (Edward Elgar, forthcoming); and editor of *Real World Decision-Making: An Encyclopedia of Behavioral Economics* (Praeger, 2014).

Mie Augier is Associate Professor at the Naval Postgraduate School and visiting researcher at the KGC, Stanford University, USA. Her research interests include history of innovative institutions, strategy, organizations, behavioral organization theory, and the past and future of management education of business schools. Her research has been published in outlets such as *Industrial and Corporate Change*; *Organization Science*; *Management International Review*; *Journal of Economic Behavior and Organization*; among others. She co-edited with James March books in memory of Herbert Simon (2004) and Richard Cyert (2002). She is co-editor of the *Palgrave Dictionary of Strategic Management*.

Marco Castellani is Assistant Professor of Economic Sociology at the University of Brescia, Italy. His primary research interest is in individual and collective decision-making. He is a co-editor of *Proceedings of the European Conference on Modelling and Simulation*, 2014.

Shu-Heng Chen is Distinguished Professor in Economics, Director of the AI-ECON Research Center, the Experimental Economics Laboratory, and the Computational, Cognitive and Behavioral Social Science Lab at National Chengchi University, Taiwan. He also serves as editor-in-chief of the *Journal of New Mathematics and Natural Computation* and the *Journal of Economic Interaction and Coordination*. Chen holds a PhD. in Economics from University of California at Los Angeles. His research interest is mainly in computational intelligence, agent-based computational economics, behavioral and experimental economics, and computational social sciences.

Subrata Dasgupta is a scholar, teacher and author. He holds the Computer Science Trust Fund Eminent Scholar Chair in the School of

Computing and Informatics, the University of Louisiana at Lafayette, USA, where he also teaches in the Department of History. He is the author of fifteen books including, most recently, *It Began with Babbage: The Genesis of Computer Science* (2014).

Peter E. Earl is Associate Professor of Economics at the University of Queensland, Australia, and a past co-editor of the *Journal of Economic Psychology*. He has previously held academic positions at the University of Stirling, the University of Tasmania and Lincoln University, New Zealand. He was educated at the University of Cambridge, where his heavily Simon-influenced doctoral dissertation was entitled 'A Behavioural Analysis of Choice'. As well as being the author of dozens of journal articles and book chapters, he is the author or co-author of nine books. He also has edited many books, including a two-volume anthology, *The Legacy of Herbert A. Simon in Economic Analysis* (Edward Elgar Publishing, 2001).

Massimo Egidi is Professor of Economics at LUISS Guido Carli University, Rome, Italy, where he has been serving as Rector since 2007. Along with Jean-Paul Fitoussi, he is President of the International Herbert A. Simon Society. He shared a long-lasting scientific cooperation and friendship with Herbert Simon. His main research interests involve primarily the study of boundedly rational behaviors in organizations and institutions; most of his papers published in the last decade are focused on the cognitive roots of decision-making processes.

Stefano Fiori is Associate Professor of Economics in the Department of Economics and Statistics Cognetti de Martiis (University of Torino), Italy. His research fields are focused on the history of economic thought and on the connection between philosophy, economics, and other social sciences, viewed from a historical perspective. His scientific interests include pre-classical and classical economics, Austrian economics, institutional and new institutional economics, methodology of economics, theories of bounded rationality.

Katherine Simon Frank is Herbert Simon's daughter. Until adulthood, she didn't understand the impact 'famous' had on Herbert's personal life. Her life was like life in every home, she thought, until she learned that her father experienced his public life differently. Kathie, a sociologist by education and training, struggled with desiring both an academic life (learned at her father's knee), and an artistic life that she

felt passionate about. After a deeply satisfying career as Coordinator of Advising to undergraduates at the University of Minnesota, USA, she retired to follow her true passion, fiber arts. She now creates quilted wall-hangings and quilts for charity.

Roger Frantz is Professor Emeritus of Economics at San Diego State University, USA. He serves on the Board of Directors for the Society for the Advancement of Behavioral Economics. His books include the co-edited *Friedrich Hayek and Behavioral Economics* (Palgrave Macmillan, 2012); as editor, *Renaissance in Behavioral Economics. Essays in Honor of Harvey Leibenstein* (Routledge, 2007); and *Two Minds. Intuition and Analysis in the History of Economic Thought* (Springer, 2005).

Gerd Gigerenzer is Director at the Max Planck Institute for Human Development and Director of the Harding Center for Risk Literacy in Berlin, Germany. He is also Batten Fellow at the Darden Business School, University of Virginia, USA, and Fellow of the Berlin-Brandenburg Academy of Sciences and the German Academy of Sciences. His academic books include *Rationality for Mortals* (Oxford, University Press, 2010); co-author of *Simple Heuristics That Make Us Smart* (Oxford University Press, 2000) and co-editor of *Bounded Rationality: The Adaptive Toolbox* (MIT Press, 2002)). In the co-edited *Better Doctors, Better Patients, Better Decisions* (MIT Press, 2013), he shows how better informed doctors and patients can improve healthcare while reducing the costs. Gigerenzer has trained US federal judges, German physicians, and top managers in decision-making and understanding risks and uncertainties.

Fernand Gobet collaborated with Herbert Simon for six years at Carnegie Mellon and has held academic positions at the University of Nottingham and Brunel University London, UK. He is currently Professor of Decision-Making and Expertise at the University of Liverpool, UK. His main research interest is the psychology of expertise, which he studies in numerous domains, including board games, physics, language acquisition, and nursing. His research combines experimental methods with computational modeling. He has written nine books, including *Foundations of Cognitive Psychology* (2011) and *Understanding Expertise* (2015).

Bhavna Hariharan is Research Associate at the Kozmetsky Global Collaboratory and lectures at Stanford University, USA. Her research in

engineering education focuses on engineering design with underserved communities globally, identifying and studying the variables that affect the efficacy and impact of global engineering efforts with impoverished communities, understanding engineering identity formation as a global, interdisciplinary profession, and evaluating the preparedness of engineering students for global work. Her other research interests are in engineering ethics and the role of care in global engineering, and the historical and contemporary role of customers in shaping engineering education.

Ying-Fang Kao is a post-doctoral researcher at AI-ECON Research Center, National Chengchi University, Taiwan. She received her PhD in Economics and Management from the School of Social Sciences, University of Trento, Italy, and is also a research fellow at the Algorithmic Social Sciences Research Unit. Her doctoral research was on classical behavioral economics, with a particular focus on Herbert Simon's contributions. Her research interests include classical behavioral economics, computable economics, agent-based computational economics, and the history of economic thought.

Rouslan Koumakhov is Professor of Cognitive Science and Organizational Psychology at Reims Management School, France. His research expertise includes the psychology of organizations and economics, the philosophy of economics, and economic sociology.

Leslie Marsh is Senior Researcher in the Faculty of Medicine at The University of British Columbia, Canada. He is the editor of *Walker Percy, philosopher* (Louisiana State University Press, forthcoming); *Stigmergic Cognition* (Springer, forthcoming); co-editor of *Propriety and Prosperity: New Studies on the Philosophy of Adam Smith* (Palgrave Macmillan, 2014); co-editor of *A Companion to Michael Oakeshott* (Penn State University Press, 2012); and editor of *Hayek in Mind: Hayek's Philosophical Psychology* (Emerald, 2011). He is the prime mover behind the setting up of *EPISTEME: Journal of Individual and Social Epistemology*; *Cosmos + Taxis: Studies in Emergent Order and Organization*; and the Michael Oakeshott Association.

Marcin Miłkowski is Associate Professor at the Institute of Philosophy and Sociology of the Polish Academy of Sciences, Warsaw, Poland. His work focuses on the philosophy of cognitive science, in particular mechanistic and computational explanations in cognitive science. His recent

publications include *Explaining the Computational Mind* (MIT Press 2013), for which he won the annual prize of the Polish National Science Center in the Humanities and Social Sciences in 2014. He also runs a project, 'Cognitive Science in Search of Unity,' funded by the National Science Center.

Marco Novarese is Assistant Professor and Lecturer in Economics at the University of Eastern Piedmont, Italy, and is Vice General Secretary of the Herbert Simon Society.

Mark Pingle is Professor of Economics and Charles N. Mathewson Professor of Entrepreneurship at the University of Nevada, Reno, USA. He has published on behavioral economics, experimental economics, and macroeconomics. Professor Pingle is currently Book Editor for the *Journal of Behavioral and Experimental Economics*, and an Associate Editor for *Review of Behavioral Economics*. He recently served a term as President of the Society for the Advancement of Behavioral Economics (SABE), organized a recent conference for SABE, and was instrumental in helping SABE obtain recognition by the Allied Association of Social Sciences.

Robert D. Rupert is Professor of Philosophy at the University of Colorado at Boulder, USA, and Professorial Fellow at the University of Edinburgh, Scotland. He has published widely on topics in the philosophical foundations of cognitive science and the philosophy of mind, as well as on topics in metaphysics and philosophy of science. He is perhaps best known for his contribution to the debate about extended cognition, much of which can be found in his book, *Cognitive Systems and the Extended Mind* (Oxford University Press, 2009), and also for his work on mental representation.

Ron Sun is Professor of Cognitive Sciences at Rensselaer Polytechnic Institute, USA. He explores the fundamental structures of the human mind and aims for the synthesis of many ideas into a coherent model of the human mind. Specifically, the goal is to come up with a computational cognitive architecture that captures a variety of psychological processes and provides unified explanations of a wide range of data and phenomena. To do so, he advocates the use of hybrid connectionist-symbolic systems. He has been developing theories of human skill learning and human everyday reasoning as centerpieces of the cognitive architecture.

1
Herbert Simon: A Hedgehog *and* a Fox

Roger Frantz and Leslie Marsh

A Quality of Mind

If as Archilochus' famous fragment goes 'The fox knows many things, but the hedgehog knows one big thing' then Herbert Simon is, at face value, a star example of a fox. Popularized by Isaiah Berlin (1978), the fox–hedgehog distinction has been interpreted (overly simplistically as Berlin acknowledged) in terms of mutually exclusive or ideal types. Hedgehog-type intelligences are motivated by an overarching grand idea or scheme that they then apply to – or through which they filter – everything else. By contrast, fox-type intelligences are highly adaptive and come up with new ideas more suited to a specific situation or context. We are of the view that the supposed hedgehog–fox dichotomy is way too trite and one-dimensional an assessment of Simon. If there were a golden thread to Simon's work it would be the development of a more adequate theory of human problem-solving and derivatively (but no less deeply) his interest in the computer simulation of human cognition – all in the service of the former (Frantz and Marsh, 2014). The upshot is that Simon made significant contributions to economics, political science, epistemology, sociology, cognitive science, philosophy, public administration, organization theory, and complexity studies (and more besides); and while ascriptions of 'polymath' and 'Renaissance man' are not without merit, they gloss over the distinctive quality of such a mind. As Simon himself has said:

> the 'Renaissance Mind' is not broader than other intelligent minds but happens to cover a narrow swathe across the multi-dimensional space of knowledge that happens to cut across many disciplines which have divided up the space in other ways. My own narrow

> swathe happens to be the process of human problem solving and decision making, and almost everything I have done lies in that quite narrow band. (Cited in Subrata, 2003, p. 686)

This is reiterated by the Edward Feigenbaum quote that Simon's daughter, Kathie, cites in the Foreword to this collection:

> in awe of his enormous knowledge and the range of his contributions, I once asked him to explain his mastery of so many fields. His unforgettable answer was, 'I am a monomaniac. What I am a monomaniac about is decision-making.'

So without fear of paradox Herbert Simon, we contend, was both a hedgehog and a fox, a notion fully compatible with his intellectual trajectory, a career 'settled at least as much by drift as by choice' (Lindbeck, 1992). Yet despite the superlatives accorded to Simon, in an age of hyper-specialization, a tacit resentment in some academic circles can be detected, a resentment that has substantive form (ideological and/or methodological) or plain old professional sour grapes infused by misguided protectionist intent. Indeed, much of the criticism that followed his award of the Nobel Prize in 1978 was because he *wasn't* an 'economist'! This despite Shackle's recommendation that:

> To be a complete economist, a man need only be a mathematician, a philosopher, a psychologist, an anthropologist, a historian, a geographer, and a student of politics; a master of prose exposition; a man of the world with the experience of practical business and finance, an understanding of the problems of administration, and a good knowledge of four or five languages. All this in addition, of course, to familiarity with the economics literature itself. (Shackle, 2010, p. 241)

Indeed, Hayek, who like Simon had long since given up writing on technical economics by the time of his Nobel award (1974), wrote:

> exclusive concentration on a speciality has a peculiarly baneful effect: it will not merely prevent us from being attractive company or good citizens but may impair our competence in our proper field. (Hayek, 1967, pp. 123, 127)

For Simon the diverse disciplines were not conventionally discontinuous but merely different lenses through which Simon approached his

central lifelong concern – *the theorizing of human behavior, or rationality, or decision-making in complex social environments*. The situated agent necessarily has imperfect knowledge: knowledge is incomplete, distributed and error prone, subject to limited computational power and time constraints in an ever dynamic environment. All we really have are satisfactory or good enough choices if one is to function.

Marking a Centenary

Though the commemoration of Simon's centenary provides a convenient hook, the motivation behind compiling this volume has substantive intent in that over the past decade or so there has been noticeable convergence on Simon's aforementioned 'golden thread.' Perhaps Simon's most natural intellectual ally, the so-called Austrian tradition, has somewhat belatedly taken him to be an 'honorary member' of that tradition since both emphasize the lived subjectivity of experience (Doyle and Marsh, 2012). Though Simon has been associated as a founder of classical (Cartesian) artificial intelligence (Frantz, 2003), the development of his work on real life decision-making has reoriented Simon as a 'situated' theorist congenial to the non-Cartesian wing of cognitive science and philosophy of mind (Marsh, 2012). Analytical epistemology in the Plato–Descartes tradition is now supplementing its unremittingly hard-nosed individualism with acknowledgement of the ubiquitous (complex) sociality and computational constraints that orthodox economic and political rationalisms have failed so miserably to deal with adequately (Frantz, 2005). For Simon human intellectual curiosity is a virtue so long as it is tempered by epistemic modesty.

We wish to acknowledge the publication of a fine biography (Crowther-Heyck, 2005) and for all intents and purposes a most expansive *Gedenkschrift* (Augier and March, 2004). We view this present volume as continuous with the Augier and March project with several contributors generously reprising their participation along with some newer voices adding to the chorus of appreciation for Simon. A themed issue of *Mind & Society* (Novarese and Viale, 2014) grew out of the first meeting of the Herbert Simon Society (http://herbertsimonsociety.org.), and collectively these major publications act as a barometer for the sustained and growing interest in Simon. It is perhaps not too much of an overstatement to say that, with the confluence of the diverse research projects and disciplines outlined earlier, there is now plausibly something called

Simon Studies. Simon's place in the history of ideas across several dimensions is very secure, and it needs to be recognized that he (along with his near contemporary Friedrich Hayek) is very much a part of a tradition that goes back to Adam Smith.

References

Augier, M. and March, J. G. (eds). (2004). *Models of a Man: Essays in Memory of Herbert A. Simon*. Cambridge, MA: MIT Press.

Berlin, I. (1978). The Hedgehog and the Fox. In *Russian Thinkers*,ed. H. Hardy. London: Hogarth.

Crowther-Heyck, H. (2005). *Herbert A. Simon: The Bounds of Reason in America*. Baltimore: Johns Hopkins University Press.

Doyle, M. J. and Marsh, L. (2012). Stigmergy 3.0: From Ants to Economies. *Cognitive Systems Research*, 21: 1–6.

Frantz, R. (2003). Herbert Simon: Artificial Intelligence as a Framework for Understanding Intuition. *Journal of Economic Psychology*, 24, 265–77.

Frantz, R. (2005). *Two Minds: Intuition and Analysis in the History of Economic Thought*. Heidelberg: Springer.

Frantz, R. and Marsh, L. (2014). Herbert Simon. In *Real World Decision Making: An Encyclopedia of Behavioral Economics*,ed. Morris Altman. Santa Barbara: Praeger.

Hayek, F. A. (1967). *Studies on Philosophy, Politics and Economics*. Chicago: University of Chicago Press.

Lindbeck, A. (ed.). (1992). Herbert A. Simon: Biographical. In *Nobel Lectures, Economics 1969–1980*, Singapore: World Scientific Publishing Co.

Marsh, L. (2012). Mindscapes and Landscapes: Hayek and Simon on Cognitive Extension. In *Hayek and Behavioral Economics*, ed. R. Frantz and R. Leeson. Basingstoke: Palgrave Macmillan.

Novarese, M. and Viale, R. (2014). Special Issue on 'Bounded Rationality Updated'. *Mind & Society*, 13 (1).

Shackle, G. L. S. (2010). *Uncertainty in Economics and Other Reflections*. Cambridge: Cambridge University Press.

Subrata, S. (2003). Multidisciplinary Creativity: The Case of Herbert A. Simon. *Cognitive Science*, 27: 683–707.

Part I
Minds

2

Embodied Functionalism and Inner Complexity: Simon's Twenty-first Century Mind

Robert D. Rupert

Introduction

One can hardly overestimate Herbert Simon's influence on contemporary cognitive science and empirically oriented philosophy of mind. Working with collaborators at Carnegie Mellon and the Rand Corporation, he wrote Logic Theorist and General Problem Solver (GPS) and thereby helped to set the agenda for early work in Artificial Intelligence (AI) (Simon and Newell, 1971). These projects also provided AI with some of its fundamental tools: in their work in the 1950s on Logic Theorist and other programs, Simon and colleagues invented list processing (which, in John McCarthy's hands, became LISP – McCarthy, 1960), while the conceptual framework of GPS gave birth to production systems (and became SOAR – Rosenbloom *et al.*, 1991). In 1981, John Haugeland included Newell and Simon's computationalist manifesto ('Computer science as empirical enquiry: Symbols and search' – Newell and Simon 1976) in his widely read anthology *Mind Design* (Haugeland, 1981). As a result, the names of Newell and Simon became, in the philosophical world, nearly synonymous with the computational theory of the mind – at the top of the list with Fodor's (1975) and Pylyshyn's (1984). Moreover, in their early work, Simon and the Carnegie–Rand group emphasized the relative independence of an information-processing-based characterization of thought from the material components of the system so engaged (Newell *et al.*, 1958a, p. 51; 1958b, p. 163, cf. Vera and Simon, 1993, p. 9). Prominent philosophers, most notably Putnam (1960, 1963, 1967) and Fodor (1974), reified this distinctively functional level of description,

thereby formulating the late-twentieth century's dominant metaphysics of mind, functionalism, according to which mental states are, by their nature, multiply realizable functional states.

The list of Simon's influential ideas extends much further than this, however. Arguably, Simon's work on hierarchical structure and the near-decomposability of complex systems (1996, chapter 8) planted the seeds of modularity-based thinking, which blossomed in the work of Marr (1982), Fodor (1983), and evolutionary psychologists (Barkow *et al.*, 1992; Pinker, 1997). In addition, Simon's (1996) emphasis on satisficing and the bounded nature of rationality set the stage for the fascinating research programs of such varied figures as Gerd Gigerenzer (2000), Christopher Cherniak (1986), and Ron McClamrock (1995). In fact, in conjunction with Kahneman and Tversky's heuristics and biases program (Kahneman *et al.*, 1982), Simon's observations about the bounded nature of rationality yielded what is now the leading view of human cognition: it's the work of a complex but nonoptimized system that manages surprisingly well by deploying its limited resources in a context-specific way – that is, in a way that exploits a grab-bag of relatively domain-specific shortcuts that are reasonably reliable given the environments in which they're typically employed. At the same time, Simon sometimes emphasized adaptive rationality – the idea that, under a wide range of circumstances, intelligent human behavior is, given the subject's goals, a straightforward function of the structure of the task at hand – a theme appearing in the work of such luminaries as Dan Dennett (1987) and John Anderson (1990). Simon also articulated a vision of intelligent behavior as the product of simple, internal mechanisms interacting with a complex external environment and, in doing so, inspired nearly two generations of philosophers and cognitive scientists who take intelligence to be the by-product of, or to emerge from, bodily interaction with the environment (Clark, 1997; Brooks, 1999). In addition, Simon (1996, pp. 88, 94, 110) was perhaps the earliest working cognitive scientist explicitly to place aspects of internal processing on cognitive par with aspects of the environment external to the organism, a central theme in Andy Clark and David Chalmers's enormously influential essay 'The Extended Mind' (Clark and Chalmers, 1998). There could scarcely be a more humbling list of one thinker's achievements, and little has been said about Simon's contributions to our understanding of search algorithms, economics, design, or management!

As impressive as this list is, one might nevertheless wonder whether Simon's views can be integrated into a single overarching vision of

human cognition – of its structure, workings, and relation to the environment. Motivated by this kind of concern, I focus, in what follows, on an apparent inconsistency in Simon's thinking, one that stands out especially clearly against the backdrop of decades of accumulated empirical results in what is sometimes known as 'embodied cognitive science' (Varela *et al.*, 1991; Clark, 1997, 2008b; Lakoff and Johnson, 1999; Rowlands, 1999; Gallagher, 2005; Gibbs, 2006; Rupert, 2006, 2009, 2011; Shapiro, 2010; Wilson and Foglia, 2011). At first blush, embodied results seem to support certain strands of Simon's thought at the expense of others. I argue, however, that even in cases in which Simon's pronouncements about the mind were premature or his emphasis out of balance (e.g., not sufficiently oriented toward the body-based contributions to human cognition), many of his own views about the mind provide the necessary corrective and allow him to accommodate – even anticipate – embodiment-oriented insights. Simon wears an early, and perhaps slightly misshapen, version of the coat we should all gladly wear today, that of the embodied functionalist.

The Tension

Stage-setting and the Dialectic

Simon's views about the human mind fall somewhat neatly, though not perfectly, into what might appear to be two mutually antagonistic clusters. In the first cluster, one finds views associated with what is sometimes thought of as orthodox computationalism, according to which (a) the human mind operates in essentially the same way as does a human-engineered all-purpose digital computer, and (b) we best understand the functioning of the human mind by focusing on computational properties and processes defined at a level of generality that subsumes both the human mind and the full variety of human-engineered all-purpose digital computers. In the other cluster sit views more sensitive to the details of the human condition, including views that seem to register limitations on the computer metaphor of the mind and direct the attention of cognitive scientists toward interactions among fine-grained (and thereby largely distinctive) aspects of human brain and extraneural body, as well as to the interactions of the human organism with the environment beyond its boundaries.

In the remainder, I further articulate these clusters of ideas and attempt to neutralize the apparent tension between them, but first a few words about motivation. After all, why should we care whether Simon's corpus presents a coherent picture of the human mind? What's it to us,

today? Readers who care deeply about the history of ideas have, by dint of that commitment alone, sufficient reason to reconstruct and evaluate Simon's vision of the mind. In addition, there is a related issue of fairness: to the extent that Simon's work set the stage for the development of widely held views in contemporary cognitive science – perhaps even embodiment-oriented views – Simon should be duly credited; more generally, we should want to acknowledge fully Simon's contribution to the course of cognitive science.

Perhaps of greatest interest to many readers, however, is the possibility that, by revisiting Simon's views on mind and cognition, we clarify and help to resolve current debates in the philosophy of cognitive science. The present chapter aspires to this goal by examining relations between contemporary embodiment-oriented cognitive science and the theoretical commitments of historical cognitivism. Many contemporary embodiment theorists (Varela *et al.*, 1991; van Gelder, 1995; Glenberg, 1997; Lakoff and Johnson, 1999; Gibbs, 2006; and various essays in Stewart *et al.*, 2010) claim that their view, *qua* embodied view, stands in stark contrast to the (purportedly waning) computational functionalist orthodoxy[1] in cognitive science. I maintain that this way of casting the contemporary debate about embodiment rests on a fundamental misunderstanding of functionalism and of the historical contributions of cognitivists. A careful examination of Simon's views bolsters this charge. It is true that, in some of his definitive-sounding, synoptic pronouncements about the mind, Simon seems to claim that it can be understood in complete ignorance of its bodily basis (e.g., 1996, pp. 58–9, 73–4); in fact, he sometimes uses (although not entirely approvingly) the language of disembodiment (1996, pp. 22, 83) to describe cognition. Furthermore, given Simon's status as a founding contributor to cognitive science, these pronouncements take on the glow of canon, appearing to sit at the heart of the computationalist orthodoxy in cognitive science and providing fair foil to those primarily concerned with bodily contributions to cognition. But, to narrate Simon's story thusly would be a travesty; one need not read far from his remarks about disembodiment and the negligibility of the neural in order to find mitigating observations and commitments, some of which explicitly acknowledge the contribution of the contingent details of our bodily materials to the determination of the contingencies of our cognitive processing. The subtlety of Simon's view of the mind should thus give us pause whenever we read that orthodox cognitive science must be overthrown if we are to accommodate the embodied perspective. For, if the views of Simon, an ur-computationalist at the center of the cognitive revolution, can accommodate, and in fact lay the theoretical groundwork

for, embodied cognitive science, one should wonder whether the stated target of today's embodiment revolutionaries, as well as the ways in which they situate their own position in the broader cognitive scientific scheme, mustn't somehow be premised on a mistake.

To be clear, various aspects of contemporary cognitive science, including some embodiment-related results, effectively challenge some ideas associated with historical cognitivism: anyone inclined to think that intelligence is what all and only universal Turing machines engage in should have flown a white flag long ago. All the same, if such untenable claims don't fairly represent the historical views of the ur-practitioners of cognitivism (e.g., Simon), then to object to such views is not to take issue with the cognitivist tradition or to show that the tools and perspectives central to the cognitive revolution must be supplanted by the fruits of a new, embodied movement.

Simon and Computational Functionalism

Orthodox cognitive science is associated with (although perhaps not fully exhausted by, see note 1) the views that (a) intelligence has its basis in computational processing and (b) there is a distinctively functional level of description, explanation, or reality apposite to the scientific study of cognition or the philosophical understanding of its nature. Many of Simon's programmatic statements about mind, cognition, and cognitive science seem to reflect these core elements of computational functionalism.

Physical Symbol Systems

Simon endorses the Physical Symbol Systems Hypothesis (PSS) about intelligence (Newell and Simon, 1976, p. 116), which appears to manifest an uncompromising computationalism. According to the PSS, only physical symbol systems are intelligent, and whatever they do (that's sufficiently complex or organized) is the exercise of intelligence. What is a physical symbol system?

> A physical symbol system holds a set of entities, called symbols . . . [It] possesses a number of simple processes that operate upon symbol structures – processes that create, modify, copy, and destroy symbols . . . Symbols may also designate processes that the symbol system can interpret and execute. Hence the programs that govern the behavior of a symbol system can be stored . . . in the system's own memory, and executed when activated. (1996, p. 22, and see ibid., p. 19)

In a nutshell, intelligence is the activity of all-purpose, stored-program computers.

Computational Models of Problem-Solving

There is no doubt that Simon intends such claims to cover human intelligence; as much as Simon's description of physical symbol systems might put the reader in mind of artificial, computer processing, these operations 'appear to be shared by the human central nervous system' (ibid, p. 19).

Moreover, it was clear from early on (Newell *et al.*, 1958b) that Simon saw the computational modeling of human intelligence to be an explanation of how humans solve problems. Simon and colleagues were out to demystify human thought; by appealing to information processing models, they meant to 'explain how human problem-solving takes place: what processes are used, and what mechanisms perform these processes' (ibid., p. 151). In describing their work on programs that, for example, prove logical theorems, play chess, and compose music, Simon and colleagues referred to their 'success already achieved in synthesizing mechanisms that solve difficult problems in the same manner as humans' (1958a, p. 3).

Functionalism

Functionalism in philosophy of mind is, in a nutshell, the view that cognition is as cognition does, the idea that the nature of mental states and cognitive processes is determined by, and exhausted by, the role they play in causal networks, not a matter of their being constituted by some particular kind of materials. Simon unequivocally endorses a functionalist understanding of computing: 'A computer is an organization of elementary functional components in which, to a high approximation, only the function performed by those components is relevant to the behavior of the whole system' (1996, pp. 17–18). This theme continues over the pages that follow, with repeated comparisons to the human case: 'For if it is the organization of components, and not physical properties, that largely determines behavior, and if computers are organized somewhat in the image of man, then the computer becomes an obvious device for exploring the consequences of alternative organizational assumptions for human behavior' (ibid., p. 21).

In Simon's hands, the description of the functional states at issue – at least at the levels of programming relevant to cognitive scientific enquiry – lines up with our common-sense way of describing the domains in question (they are semantically transparent, in the sense

of Clark, 1989, 2001). This is reflected partly in Simon's use of subject protocols to inform his computational modeling of human cognition (Simon and Newell, 1971, pp. 150, 152; Newell *et al.*, 1958b, p. 156), not merely as a stream of verbal data that must be accounted for (by, for example, our best models of speech production), but as reports the contents of which provide a reasonably reliable guide to the processes operative in problem-solving (cf. Newell *et al.*, 1958a, p. 40 n1). This is worthy of note because, in the philosophical imagination, functionalism has often been understood as a claim about the kind of mental states one can specify in everyday language and the appearance and operation of which, in oneself, can be tracked by introspection. This has tended to obscure the live possibility that one can construct computational models of at least fragments of human cognition (a) that deal in representations of features or properties that have no everyday expression in natural language and to which we have no conscious access, and (b) that may well tell the entire story about the fragment of cognition in question. In other words, Simon's work with protocols and his account of, for instance, the processes of means-end reasoning and theorem-proving – accounts that seem to involve representations that match naturally with the fruits of introspection – reinforce the image of Simon as a coarse-grained computational functionalist who makes no room for subtle, subconscious body-based processing. (Cf. Simon's discussion of memory for chess positions [1996, pp. 72–3]; Simon generates his description of the relevant cognitive process by auto-report and couches his account of the process in everyday chess-playing terms.) In contrast, many embodiment-oriented models deal in features and processes difficult to express in natural language and better captured only mathematically or in a computational formalism that has no natural, everyday expression (cf. Clark's [1989] characterization of connectionist networks as detecting and processing microfeatures). A computational functionalism of the latter, fine-grained sort is computational functionalism nevertheless, for it invokes multiply realizable, functionally characterized, representational states to account scientifically for the data distinctively associated with the mind. What more could one want from a legitimate psychofunctionalism (Block 1978)?

Multiple Realization and Metaphysical Autonomy

As a philosophical theory of the nature of mental states, functionalism entails multiple realizability (MR), that the very same mental or cognitive properties or states can appear in, be implemented by, or take the form of significantly different physical structures (different with

respect to how the physical sciences would characterize those structures and their properties). Although Simon's comments often seem oriented toward methods of investigation or kinds of explanation, he also appears to be committed to metaphysical functionalism and the associated MR thesis.

Newell, Shaw, and Simon express the view in this way:

> We do not believe that this functional equivalence between brains and computers implies any structural equivalence at a more minute anatomic level (for example, equivalence of neurons with circuits). Discovering what neural mechanisms realize these information processing functions in the human brain is a task for another level of theory construction. Our theory is a theory of the information processes involved in problem-solving and not a theory of neuronal or electronic mechanisms for information processing. (1964, p. 352 – quoted in Gardner, 1987, p. 148; cf. Newell *et al.*, 1958a, p. 51)

And, from a more recent paper by Vera and Simon (a paper about situated cognition, no less!): 'And, in any event, their physical nature is irrelevant to their role in behavior. The way in which symbols are represented in the brain is not known; presumably, they are patterns of neuronal arrangements of some kind' (1993, p. 9).

Moreover, when Simon discusses the functional equivalence of computer and human, he says that 'both computer and brain, when engaged in thought, are adaptive systems, seeking to mold themselves to the shape of the task environment' (1996, p. 83). Taking such remarks at face value, Simon seems to be talking about the very things and properties in the world, as they are, not merely, for example, how they are best described for some practical purpose.

MR and Methodological Autonomy

At the same time, we should recognize that Simon is often interested in matters methodological or epistemological, in how one should go about investigating intelligent systems or explaining the output of intelligent systems. He frequently talks about our interests, and about discovery, theory construction, and description. Even the equivalence referred to above could just as well be equivalence vis-à-vis our epistemic interests rather than equivalence in what exists independently of those interests.

Frequently these metaphysical and epistemic messages run parallel to each other (perhaps because Simon has both in mind, and the

metaphysical nature of the properties and patterns in question determines the important epistemic and methodological facts). For instance, he says that

> many of the phenomena of visualization do not depend in any detailed way [note the use of 'detailed'] upon underlying neurology but can be explained and predicted on the basis of quite general and abstract features of the organization of memory. (1996, pp. 73–4)

Although one might wonder about the nature of the dependence being discussed, the passage's emphasis on the predictive and explanatory value of postulated features lends itself naturally to a primarily epistemological reading.

This is not an exercise in hair splitting. A failure to distinguish between the metaphysical and epistemic or methodological dimensions of computational functionalism can easily obscure the relation between computational functionalism as a theory of mind and computational functionalism as it's sometimes been practiced, the former of which makes plenty of room for embodied cognitive science, even if some computational functionalists have not, in practice, focused on bodily contributions to cognition.

Adaptive Rationality

The idea of adaptive rationality plays a central role in *Sciences of the Artificial*. After describing what they take to be the small number of parameters constraining human thought – including, for example, that human short-term memory can hold approximately seven chunks – Newell and Simon say:

> [T]he system might be almost anything as long as it meets these few structural and parametral specifications. The detail is elusive because the system is adaptive. For a system to be adaptive means that it is capable of grappling with whatever task environment confronts it . . . its behavior is determined by the demands of that task environment rather than by its own internal characteristics. (1971, p. 149; and see Simon, 1996, pp. 11–12)

On this view, we can, for many purposes, think of internal workings of cognitive systems as black boxes; but for a small number of parameters determined by their material constitution, human cognitive systems – in fact, any intelligent systems – are organized in some way or another

so as to adapt to task environments. As intelligent systems they thus exhibit similar behavior across a wide range of circumstances (where goals are shared).

Simplicity of the Inner

Simon opens a central chapter of *Sciences of the Artificial* with the tale of an ant picking its way across a windblown beach toward its nest, encountering many small obstacles along the way. To an onlooker, the ant's path might seem to reflect a complex internal process. According to Simon, though, the path's 'complexity is really a complexity in the surface of the beach, not a complexity in the ant' (1996, p. 51). He extends the moral of the story to human cognition and behavior: 'Human beings, viewed as behaving systems, are quite simple. The apparent complexity of our behavior over time is largely a reflection of the complexity of the environment in which we find ourselves' (1996, p. 53).

This is a striking claim, not entailed by adaptivity; for various highly organized systems might all, nevertheless, be organized so as to behave adaptively, and one might even think it necessary that a system have a complex structure in order to respond effectively in a wide range of task environments. I emphasize the claim of inner simplicity at this point in the discussion because the claim seems to stand so deeply at odds with the embodied program, which takes human cognition, in all of its nuance and complexity, to be largely a function of fine-grained facts about the distinctive structural and causal organization of the inner workings of the human system.

Throughout this section I have portrayed Simon as a representative of a brand of computational functionalism often derided by embodied theorists for its utter disregard for the material basis of cognition and for its tendency instead to fetishize such projects as a formal analysis of task domains and the exploration of algorithms of search through them, algorithms that have no connection to distinctively human, body-based strategies.

Embodied Functionalism and Contingency-making

The preceding portrayal significantly misrepresents Simon's approach to the mind, and here's a way to begin to see why. One of the most important aspects of the embodied view is a commitment to what might be reasonably called 'contingency-making'. On the embodied view, contingent facts about our bodily existence color and shape human cognition (and presumably, something similar holds for other creatures as well).

Humans have a particular body shape and orientation and particular networks of muscles that interact in a finely-tuned orchestra of counterbalancing forces to keep the organism alive and functioning in the face of ongoing perturbation by environmental forces. Moreover, patterns of neural firings involved in such orchestration provide resources to be co-opted for other cognitive purposes, from the encoding of items in working memory (Wilson, 2001) to the metaphorical understanding of abstract concepts (Lakoff and Johnson, 1999). In this picture, the resources used in real time by human organisms – in all of their *ad hoc*, makeshift, exapted, kludgey (i.e., made up from poorly matched components), but stunningly effective, glory – bear the marks of this distinctive body-based orchestra. Thus, according to embodiment theorists, there is simply no way to understand what's going on with human cognition absent the thorough investigation of these contingency-making, bodily contributions.

Simon's picture, too, is rife with contingency-making, arguably of the bodily sort (although he does not often use such language), which places him – or at least the second cluster of his views – squarely in a camp with the embodiment theorists. In fact, one might reasonably contend that Simon's uncompromising interest in the contingencies of the human case helped to set the stage for the embodied movement.

It is the burden of the next section to make a detailed case for this claim, but first a door must be pried open. Don't embodiment theorists routinely criticize, reject, and even demonize computationalism, functionalism, and multiple realizability? How could I possibly, in all seriousness, present Simon as a proto-embodiment theorist, given his clear commitments to computational functionalism and related views, as documented above? The answer requires that we revisit these much-derided orthodox positions and clarify their relation to embodied cognitive science. In doing so, we clear the way for a proper understanding of Simon's view and of embodied functionalism, generally speaking.

Fine-grained Functionalism: Respecting the Bodily Source of Functionally Specified Contingencies

Many embodiment theorists criticize functionalism and computationalism as *disembodied* and reject them outright. If these critics are correct, then, Simon – *qua* ur-computationalist committed to functionalism, computationalism, multiple realizability, and the irrelevance of the physical – can't possibly be cast as an embodiment theorist, not without radical revision or wholesale denial of significant aspects of his thought.

This way of framing the dialectic gains no traction against the conciliatory picture of Simon's thought being constructed here, for it rests on a misunderstanding of computationalism and functionalism (cf. Rupert, 2006, 2009, ch. 11; Clark, 2008a, Shapiro, 2010). What, after all, does it mean to claim that functionalist and computationalist approaches are disembodied? One version of the complaint seems to be that the physical material of the body or brain doesn't play a significant metaphysical role in the functionalist's or computationalist's theory of mind or cognition; the body has been left out. If that, however, correctly represents the operative complaint, embodied theorists have misunderstood the metaphysical relation of their views to classical cognitive research. According to the classic versions of functionalism, the physical body (the brain, in particular, but this applies to whatever matter realizes the computational system) determines, in the strictest metaphysical sense, which functions the human cognitive system computes. That is the very nature of the realization-relation. Thus, according to classical computational functionalism, the body is the root and determinant of our cognitive being. Unless we construe embodied theories as type-type identity theories – which raises a host of its own problems (Clark, 2008a) – embodied views are in precisely the same boat, metaphysically, as functionalism. Embodied views in cognitive science are, in fact, functionalist views. They are, however, fine-grained functionalist views, in that they tend to build fine details of the workings of the human brain, body, and world into their functionalist models, by individuating the functional properties in those cognitive models in a way that reflects the details of the causal profiles of the relevant physical materials (although stopping well short of any attempt to capture the full causal profiles of those materials). So long as more than one collection of physical matter could exhibit the same, relevant causal profile as described in an embodied model of cognition, the processes in question fit functionalist metaphysics perfectly. They may not be recognized by common-sense or analytic functionalists, but they are functionalist in the sense relevant to cognitive science; they are the fruits of a psychofunctionalist approach (Block, 1978, p. 269), which looks to science – embodied cognitive science, as it turns out – to characterize the functional properties in question.

The embodied approach recommends a different methodology than was pursued by many classical computationally oriented cognitive scientists. The embodied functionalist lets such things as bodily activity be her guide, epistemically, when attempting to figure out, for instance, which algorithms (or other abstract processes – a nod to those who lean toward dynamical-systems-based modeling [e.g., Chemero,

2009]) govern human cognition. This has no bearing on the truth of functionalism, however (except to confirm functionalism by the success of the embodied strategy!), but rather it stands in opposition to a certain empirical bet that many computationalist-functionalists made in the early days of cognitive science; that the relevant algorithms and the location of the machinery that executes them could be identified from the armchair or by reflection on the way everyday people typically talk and think about human thought processes. (Dennett makes this kind of point about neuromodulators and neuroanatomy in the context of a discussion of functionalism and consciousness [2005, pp. 17–20].)

A now relatively common way to capture the differences among functionalist views is to recognize a continuum running from coarse-grained to fine-grained versions of functionalism. A given functionally individuated state can be individuated by anything from a simple pattern of relations among very few states, inputs, and outputs (consider the stock example of a Coke machine – Block, 1978, p. 267) to a massively populated space of states, including millions of possible input and output states (as is the case with the typical human cognitive state, the presence of which is determined by the incredibly complex network of functional and computational states realized, and thus determined, by the brain). In his official pronouncements, Simon seems to commit himself to a coarse-grained functionalist view, and it's clear enough why. If human internal cognitive operations are so few, and their workings straightforward, constrained only by a small number of biologically set parameters easily measured and modeled using behavioral data alone, then human cognitive operations are broadly multiply realizable and largely independent of the fine details of neural structure and processing. But, Simon could be wrong about this even if functionalism and computationalism are true. Moreover, the turn toward a fine-grained computational functionalism – even one that focuses on body-related contributions – may be inspired by some of Simon's own work.

The Body of Simon's Work

In this section, I explore Simon's views that fall into the second of the two clusters identified at the outset. Simon's explicit methodological pronouncements often jibe with what one would expect from an embodiment theorist, and many of the theoretical constructs central to his work can be most fruitfully understood as ways of discovering the body's contingent effects on human cognitive processing. Or, so I argue in the remainder of this section.

Simon, Environmental Interaction, and Bodily Determined Conceptual Variation

What methodology does an embodied perspective in cognitive science call for? One common embodied approach tracks the cognitive effects of interaction with the environment, either on an evolutionary, learning, or developmental scale (Rupert, 1998); this can be done via simulation of embodied agents (Beer, 2003; Husbands *et al.*, 1995), construction of robots (Brooks, 1999; Steels and Spranger, 2008), or observation of human infants (Thelen and Smith, 1994). Much of the embodied work focuses not only on the diachronic issue of changes in systems over time, but also on the real-time interaction with the environment, seeking to uncover the contribution that bodily motion and interaction with the environment make to intelligent behavior (Kirsh and Maglio, 1994). Simon's remarks about the ant would appear to provide clear inspiration for this kind of interactive work, which is central to the history of embodied cognitive science.

Of significant interest, too, is the modeling of fine-grained details of the neural contributions to bodily control (Grush, 2004), including the contribution of neural representations (or other mechanisms or processes) that track bodily interaction with the environment (Iriki *et al.*, 1996). Such a project proceeds by combining at least two forms of reasoning: (a) the use of behavioral data together with our best guesses, given the task, about which functions are likely being performed by areas of the brain active during task performance; and (b) sensitivity to activity at the neural level during task performance to help us generate and select theories of the functions being carried out, which might be quite fine-grained, species-specific, and based on bodily contingencies (while still being multiply realizable, at least in principle).

Simon and Newell state this strategy explicitly as part of their 11-point strategy for explaining human problem-solving:

> Begin to search for the neurophysiological counterparts of the elementary information processes that are postulated in the theories. Use neurophysiological evidence to improve the problem-solving theories, and inferences from the problem-solving theories as clues for the neurophysiological investigations. (1971, p. 146)

Thus, there is nothing contradictory in Simon's (and Newell's) being a proto-embodied-functionalist – with regard to method, not only metaphysics – while at the same time pressing ahead with a largely information-processing-based research program (ibid., pp. 157–58), one

inspired by behavioral data and *a priori* theorizing about the functional organization of the system producing that behavior.

Note, too, their endorsement of contingency:

> These properties – serial processing, small short-term memory, infinite long-term memory with fast retrieval but slow storage – impose strong constraints on the ways in which the system can seek solutions to problems in larger problem spaces. A system not sharing these properties – a parallel system, say, or one capable of storing symbols in long-term memory in milliseconds instead of seconds – might seek problem solutions in quite different ways from the system we are considering. (Ibid., p. 149)

'Different bodies, different concepts' is one of the most widely known themes from the embodiment literature (Shapiro, 2010), and according to leading embodiment theorists (Lakoff and Johnson, 1999), processing profiles individuates concepts (that is, determine whether two concepts are the same or different). Thus, Simon and Newell seem to be marching in step with the embodiment theorists: a system with a significantly different material basis might well solve problems in different ways, which entails difference in processing; and, since the concepts at issue are individuated by their processing profiles, solving problems in different ways would seem to entail the use of different concepts.

Although Simon sometimes seems to advocate for the 'disembodiment of mind' (Simon 1996, p. 83), he explicitly qualifies such remarks: 'It would be unfortunate if this conclusion were altered to read that neurophysiology has nothing to contribute to the explanation of human behavior. That would of course be a ridiculous doctrine . . . It is to physiology that we must turn for an explanation of the limits of adaptation' (ibid.). The human system doesn't adapt perfectly to the problems it faces, and when its behavior is not driven optimally by the structure of the problem at hand, it is because of the constraints placed on it by neurophysiology.

Simon's two recurring examples of such constraints are that working memory can hold only about seven chunks of information and that it takes about eight seconds for an item to be transferred into long-term memory. Simon should have known better, however, given his various commitments and insights, than to emphasize the smallness of the number of such constraints and to push them to the theoretical margins, as merely the limits of adaptation. He acknowledges, for example, that 'Neurophysiology is the study of the inner environment' (ibid.), and,

although he seems to identify explicitly only two or three interesting parameters determined by that inner environment – in his official pronouncements about said number, at least – he seems to invoke many such parameters and identify many effects of such parameters across the range of his work. Moreover, his general theoretical outlook gives him permission to identify many more such parameters – as many as the data demand – as is reflected by his actual practice.

The Body Ascendant

Beyond the general themes discussed in the previous section, there are various ways in which Simon's own theorizing seems to encourage an embodied perspective, as this subsection demonstrates.

Stress, Duress, and Taxing Environments

Notice the way in which Simon and Newell talk about the discovery of parameters: 'Only when the environment stresses its capacities along some dimension . . . do we discover what those capabilities and limits are, and are we able to measure some of their parameters (Simon, 1969, Ch. 1 and 2)' (1971, p. 149). And, unsurprisingly, their flagging of Simon 1969 (which I've been citing in its most recent edition, 1996) is spot on. When a system is taxed, Simon claims, or makes a mistake, the properties of the system's material substrate, which would not otherwise show themselves, become behaviorally manifest (1996, pp. 12, 58, 83).

Interestingly, contemporary embodiment theorists often emphasize that cognition is time pressured and kludgey and that this affects cognition through and through. What embodiment theorists add to Simon's perspective, then – which addition Simon can naturally accommodate – is that the limiting properties of the inner system virtually always show through. Because thought is commonly time pressured – or in some other way pressured by context, including internal context – the human cognitive system is almost always being taxed, and thus significant and contingent aspects of the material substrate of cognition are continually on display. A constant stream of external stimuli, together with the ongoing fluctuations in internal context (neuromodulators, hormone levels, neural correlates of emotional states and moods), create a shifting cognitive context, a series of 'perturbing influences' on cognitive processing; and under these conditions, the fine-grained physical properties of our brains – and bodies and external environments, perhaps – reveal themselves in all sorts of ways as relevant to our understanding of human performance and intelligent behavior.

Beyond Duress

I've tried to extend the scope of Simon's category of a system under duress, so that it covers a wide range of cases in which embodied contributions to cognition reveal themselves. Circumstances involving what is more clearly conceived of as breakdown also reveal fascinating aspects of the human cognition, in ways that Simon would have no reason to resist. Over the decades, cognitive scientists have learned no small amount about the human cognitive system – its architecture and component mechanisms – from the study of subjects suffering from autism spectrum disorders, Capgras syndrome, hemispatial neglect, prosopagnosia, Balint's syndrome, and many more disorders.

Experimental manipulations, too, have led to the discovery of a variety of parameters (and their values), as well as systemic properties that emerge from interaction among sometimes quirkily functioning component parts, all of which Simon can gladly take on board. Consider, for instance, Pylyshyn's proposal that there are four FINST (fingers of instantiation) pointers (Pylyshyn, 2000), or studies showing that tactile discrimination becomes more sensitive when experimenters activate a visual representation of the area being touched (Serino and Haggard 2010), or that one's report of the timing of one's conscious intention to press a button shifts when pre-SMC (sensorimotor contingency) is subject to TMS (transcranial magnetic stimulation) *after* one has pressed the button (Lau *et al.*, 2007). Presumably, the material basis of human cognition, including materials that have more or less to do with specifically sensory or motor contributions to cognition, determine the presence in humans of such cognitively relevant architectural elements and their specific causal profiles.

In his attempt to understand the complexities of human problem-solving, Simon himself identifies numerous contingent architectural elements of the human cognitive system and phenomena that would seem to arise contingently from the interaction of these elements, and many of these seem amenable to an embodied treatment. Among Simon's most enduring contributions to cognitive scientific theorizing are his notions of bounded cognition and the associated idea of satisficing. Humans have limited cognitive resources, limited long- and short-term memory capacities, for example. But, by satisficing (looking for a satisfactory, or 'good enough,' solution), we compensate for the limited (that is, bounded) nature of our cognitive systems and thereby solve (well enough) problems we couldn't easily (or even possibly!) solve optimally, by, for example, exhaustive search through, and evaluation of, all possible options. Most of us don't undertake the project of finding

the very best wine to go with our dinner; instead, we find one that's good enough. We achieve this partly by employing search heuristics (we buy a wine we've heard of, that seems priced about right, and that has a high Wine Spectator score posted beneath it).

At this point, one might reasonably wonder whether Simon wasn't onto one of the most substantive insights of the embodiment-based literature. Think of how deeply nonoptimizing is the human use of heuristics and biases. Cognitive scientists with time on their hands and lots of computing power can explain to us why the shortcuts we use represent reasonable approaches to time- and resource-pressured problem-solving. But, *we*, in situ, can't do such calculations; we would lose the advantage of employing heuristics to satisfice if we were to attempt to calculate the costs and benefits of using those heuristics in a given case (McClamrock, 1995). Instead, their use must come naturally, in some sense, and the only plausible ways for them to come naturally are the ones emphasized by embodiment theorists: we solve problems by relying *automatically* on shortcuts built into the neural system, into our bodily structure, and into our environmental tools; and by learning to exploit, relatively automatically, ways in which the fine-grained details of our bodily and neural structures interact with fine-grained aspects of our environments.

More of Simon's Own

Many more of Simon's theoretical constructs or posits dovetail with, or even lend support to, an embodied perspective:

a. Heuristics and search. In a discussion of search strategies in chess, Simon and Newell say, 'The progressive deepening strategy is not imposed on the player by the structure of the chess task environment. Indeed, one can show that a different organization would permit more efficient search' (1971, p. 153). And note that, as their following paragraphs make clear, Simon and Newell are interested, not only in the way humans happen to play chess, but also the way in which the memory resources available – themselves bodily determined – help to determine the search strategy humans choose.
b. Along related lines, Newell and Simon identify in subjects a tendency, as part of a planning phase, to abstract from the details of a particular problem, without (somewhat surprisingly) maintaining a clear distinction between the abstracted space used in planning (or strategizing) and the problem space ultimately appropriate to the

problem at hand (1971, p. 156). An embodied theorist might wonder whether the human moves easily between such spaces because, in some sense (geometrically, perhaps?), the bodily resources for one way of thinking are nested within the resources for the other. And, other things being equal, there's no reason Simon shouldn't pursue this kind of answer to his questions about the selection of problem spaces to work in – perhaps seeing ease of migration from one space to a more abstract version of it as another body-determined parameter.

c. The human's retrieval of information from long-term memory is itself governed by further processes of interest (Simon, 1996, p. 82). This process includes the use of associative memory, which, according to Simon, is best understood as the manipulation of list structures. This central aspect of the organization of the human mind, which shows its distinctive character in experimental settings, is plausibly a function of the fine-grained properties of the material basis of the human cognitive system. Moreover, it's particularly plausible that the resources for such associations include images and motor patterns – the sorts of resources emphasized by embodiment theorists.
d. Simon observes and attempts to explain deviations from economic rationality: 'Affected by their organizational identifications, members frequently pursue organizational goals at the expense of their own interests' (1996, p. 44). What drives such identification? How do analytically capable individuals become company men and company women, even though the individual irrationality of doing so is demonstrable? Presumably, this results from ways in which the material basis of human cognition contingently determines our affective states and processing, the contributions of which to cognition have come to be understood better and better in recent decades (see the literature on the Iowa Gambling task, for example – Bechara *et al.*, 2005) and is often associated with the embodied perspective.
e. In Simon's discussion of design (1996, pp. 128–30), he emphasizes the importance of style, of the choices architects and composers face at various points in what are, essentially, processes of creative problem-solving. The results of the choices made – in effect, the expressions of different styles – can be seen as contributions to generate-and-test cycles, according to Simon. The choices contribute to the generation of a variety of possibilities, the availability of which will, among other benefits, increase the probability of access to a satisfactory option. And in the paragraph that follows, Simon extends this vision to the design of cities and economies.

What, however, determines such differences in style? Learning history? Genetic differences? Perhaps, but how would such variations in history and genes create variation in design choices? Particularly when it comes to such intensely visual and spatial domains as design, one would think they do so by affecting the sorts of cognitive resources – sensorimotor routines, for example – of interest to embodied theorists.

f. Simon's work suggests many more possibilities for the application of an embodied perspective. Why does the human production system employ the difference-reduction operators it does (1996, p. 94) rather than others? Why does human attention work in the way it does (ibid., pp. 143–44)? What is the bodily basis of our tendency to discount the future, which Simon himself recognizes as another 'of the constraints on adaptation belonging to the inner environment' (ibid., p. 157)?

As in the cases discussed above – of associative memory, use of search heuristics, the contribution of varying styles to design processes, and the contribution of 'irrational' factors to economic cognition – the profile of human processing seems unaccounted for by adaptive rationality alone, partly because it is shot through with contingency. And, while these contingencies can be functionally and computationally (or at least mathematically) characterized, that hardly makes them elements of a universal disembodied rationality. The contingencies themselves are determined by the fine-grained contingencies of the bodily materials that realize the functions in question.

The Co-opting of Problem Spaces

According to Simon, 'Every problem-solving effort must begin with creating a representation for the problem – a problem space in which the search for the solution can take place' (1996, p. 108). A subject's choice of a problem space is the way the 'particular subject represents the task in order to work on it' (Simon and Newell, 1971, p. 151). About these matters, Simon is characteristically honest, 'The process of discovering new representations is a major missing link in our theories of thinking' (Simon, 1996, p. 109; cf. Simon and Newell, 1971, p. 154).

To my mind, questions about the acquisition and selection of problem spaces are some of the most fascinating in cognitive science. One such question is under what conditions the subject co-opts an already acquired representation of a problem space for use in the solution of a new problem. Simon discusses this issue in connection with a simple

card game. After describing its rules, Simon notes that the game has essentially the same structure as the familiar game of tic-tac-toe (1996, pp. 131–32), which, I imagine, comes as a surprise to most readers. But, once one has seen the structural equivalence, one can immediately transfer strategies for playing tic-tac-toe to Simon's card game.

Three comments are in order, two of which concern the connection to embodied theorizing and the ways in which embodied theorizing might help to answer some of Simon's open questions about problem-solving. First, many of Simon's examples of problem-solving involve mathematical, or otherwise symbolic, reasoning. It's worth noting, then, relatively recent work on the role of gesture in the development of problem-solving skills in arithmetic (Goldin-Meadow, 2003). Second, questions to do with encoding – about representing the problem within the problem space – have also been given an embodied treatment. This research demonstrates the great extent to which symbolic reasoning depends on spatial arrangements in the external world and the way they interact with our perceptuo-motor resources (Landy *et al.*, 2014).

My third observation distances the discussion somewhat from specific claims about mathematical or symbolic cognition. Rather, it is to note the depth and scope of Simon's observations regarding the co-opting of problem spaces. On Simon's view, cognition is bounded and, as a result, we take cognitive shortcuts. We do so partly by mapping new problems into a stock of existing problem spaces that are themselves amenable to the use of effective heuristics. This mapping of problems – typically more complex problems – into less complex or more familiar problem spaces is one of the primary themes of Daniel Kahneman's recent overview of his life's work (Kahneman, 2011). This is also one of the central themes of the work of one of the champions of embodiment, George Lakoff. Much of Lakoff's work has been an effort to show that a relatively small number of body-based source domains provide the materials for a metaphorical understanding of what we think of as abstract concepts and domains (Lakoff and Johnson, 1999). It's plausible enough, then, that important aspects of the availability of Simon's heuristics, as well as their structural natures, are determined by the details of our bodily experience. And given the centrality of these issues to Simon's picture of the mind, embodiment would emerge as central to a theory of cognition, wedding, in this way, the work of Simon, Kahneman, and Lakoff.

In the end, Simon's vision need have very little added to it to become a deeply embodied position. And what must be added conflicts in no way with his big-picture theoretical commitments, *qua* computational functionalist, and is at least suggested by many of his other most important theoretical commitments.

Conclusion

In this chapter, I identified two clusters of Simon's views, some of which clearly align him with computational functionalism and others of which either implicitly or explicitly presuppose or support significant aspects of the embodied perspective. I also argued that the former views do not conflict with the embodied perspective. On balance, Simon's picture of the mind emerges as a coherent precursor to contemporary embodied functionalism.

There are broader lessons to be learned here as well. Virtually everyone in the philosophy of cognitive science, from orthodox computationalist to embodiment theorist, is in the same boat: they are all metaphysically disembodied[2] and epistemically embodied, and always have been. Unless embodiment theorists choose to embrace type-type identity theory as a metaphysics of mind, their view is every bit as functionalist and disembodied as was early cognitivism. According to both groups, the physical materials fully determine what happens at the cognitive level but cannot be literally identified with those cognitive processes.[3] Even those who reject computationalism as a modeling methodology (in favor of, say, dynamical-systems-based modeling – Chemero, 2009) do not identify their theoretical types with physical types: there is no brain state that is, literally, the one and only kind of state that is identical to, for example, a periodic attractor or an attractor well; all parties are up to their necks in multiply realizable kinds, at the very least in the sense that they make explanatory use of abstract kinds or properties that take a variety of forms (see Weiskopf, 2004 for a related problem). At the same time, from the epistemic or methodological standpoint, cognitive science has always been concerned with embodiment. From Hebb (1949) to Lettvin, Maturana, McCulloch, and Pitts (1959) to Hubel and Wiesel (1962) to Sperry, Gazzaniga, and Bogen (1969) to Bliss and Lømo (1973) to Warrington (1975), the early decades of cognitive science saw a steady stream of highly influential work in neuroscience that was of broad interest specifically because it was taken to have cognitive implications. The real history of cognitive science – not the 'disembodied' caricature of it – was of a piece with embodied functionalism, even though some major figures, including Simon in some moods, may have hoped they could generate computational models of cognition without attending to the bodily materials that they quite well knew determined those computational functions in humans.

Notes

1. A few words about terminology are in order. Functionalism holds that the nature of a mental state is its causal role – its characteristic causal relations to inputs, outputs, and other mental states – not what kind of physical stuff it is made of; functionalism thus allows that two creatures with different kinds of material bodies could be in the same mental state, so long as both creatures are in some state or other, of whatever composition, that plays the causal role individuative of the kind of mental state in question. Computational states are a kind of functional state, to be sure, individuated in terms of what they contribute to computational processing regardless of their physical composition. But, computational states are not necessarily a kind of mental state; so one could endorse a computationalism of sorts, without thinking computation has anything to do with cognition or the mind. In contrast, computationalism in cognitive science asserts that at some level of description (perhaps the neural level, perhaps the subpersonal level), a computational formalism provides the best way to model cognition-related processing or, ontologically speaking, that computational processing contributes significantly to the production of intelligent behavior (or other data of cognitive science). Computational functionalism results when a functionalist adopts an explicitly computational perspective on mental states, holding that mental states are, metaphysically speaking, functional states the nature of which is to play causal roles characteristic of computational states. For further discussion, see Piccinini (2010). Of course, to the extent that the category of the mental remains contested and blurred at the boundaries, it may remain unclear the degree to which one's computational models of the processes that produce intelligent behavior must 'seem genuinely mental' in order for one to count as a computational functionalist, as opposed to a mere computationalist.
2. In the sense that at least some of the properties that play a causal-explanatory role are not identical to properties of independent interest in the physical sciences. They are, instead, multiply realizable, even if some of the properties – the embodiment-oriented ones – place stringent constraints on the range of physical structures that can realize them.
3. There are ways for embodiment theorists to resist. For instance, there's an enormous literature on the so-called grounding problem (Harnad, 1990), which suggests that for any mental representation to have content, it must be associated with sensorimotor states. An embodiment theorist who endorses this sensorimotor constraint on content *and* holds that the nature of sensorimotor states can't be captured functionally will have genuinely set herself against functionalism. But, to make this anti-functionalist position credible, the embodiment theorist must give an alternative scientific account of said sensorimotor states, and that has been missing, thus far, from the embodiment literature. Similar remarks apply to embodiment theorists who assign a privileged role to conscious experiences connected to the body or to certain biological processes. Consciousness and the maintenance of organismic integrity might play privileged roles in our understanding of cognition, but that does not itself speak against functionalism, unless accompanied by a scientifically respectable nonfunctionalist account of consciousness and of life.

References

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Lawrence Erlbaum.

Barkow, J., Cosmides, L. and Tooby, J. (eds) (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Bechara, A., Damasio, H., Tranel, D. and Damasio, A. (2005). The Iowa Gambling Task and the Somatic Marker Hypothesis: Some Questions and Answers. *Trends in Cognitive Sciences* 9, 4, 159–62.

Beer, R. (2003). The Dynamics of Active Categorical Perception in an Evolved Model Agent. *Adaptive Behavior* 11, 209–43.

Bliss, T. V. P. and Lømo, T. (1973). Long-lasting Potentiation of Synaptic Transmission in the Dentate Area of the Anaesthetized Rabbit following Stimulation of the Perforant Path. *Journal of Physiology*, 232, 331–56.

Block, N. (1978). Troubles with Functionalism. In *Perception and Cognition. Issues in the Foundations of Psychology, Minnesota Studies in the Philosophy of Science*, vol. 9, ed. C. Savage. Minneapolis: University of Minnesota Press, 261–325.

Brooks, R. (1999). *Cambrian Intelligence: The Early History of the New AI*. Cambridge, MA: MIT Press.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Cherniak, C. (1986). *Minimal Rationality*. Cambridge, MA: MIT Press.

Clark, A. (1989). *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. Cambridge, MA: MIT Press.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. (2001). *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford: Oxford University Press.

Clark, A. (2008a). Pressing the Flesh: A Tension in the Study of the Embodied, Embedded Mind. *Philosophy and Phenomenological Research* 76, 1, 37–59.

Clark, A. (2008b). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

Clark, A. and Chalmers, D. J. (1998). The Extended Mind. *Analysis*, 58, 7–19.

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

Dennett, D. C. (2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press.

Fodor, J. A. (1974). Special Sciences. *Synthese*, 28, 77–115.

Fodor, J. A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Gallagher, S. (2005). *How the Body Shapes the Mind*. New York: Oxford University Press.

Gardner, H. (1987). *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books

Gibbs, R. (2006). *Embodiment and Cognitive Science*. Cambridge: Cambridge University Press.

Gigerenzer, G. (2000). *Adaptive Thinking*. Oxford: Oxford University Press.

Glenberg, A. (1997). What Memory is For. *Behavioral and Brain Sciences* 20, 1–19.
Goldin-Meadow, S. (2003). *Hearing Gesture: How Our Hands Help Us Think*. Cambridge: Belknap Press/Harvard University Press.
Grush, R. (2004). The Emulation Theory of Representation: Motor Control, Imagery, and Perception. *Behavioral and Brain Sciences,* 27, 377–96.
Harnad, S. (1990). The Symbol Grounding Problem. *Physica D*, 42, 335–46.
Haugeland, J. (ed.) (1981). *Mind Design*. Cambridge, MA: Bradford/MIT Press.
Hebb, D. (1949). *The Organization of Behavior*. New York: Wiley & Sons.
Hubel, D. H. and Wiesel, T. N. (1962). Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *Journal of Physiology*, 160, 106–54.
Husbands, P., Harvey, I. and Cliff, D. (1995). Circle in the Round: State Space Attractors for Evolved Sighted Robots. *Robotics and Autonomous Systems* 15, 83–106.
Iriki, A., Tanaka, M. and Iwamura, Y. (1996). Coding of Modified Body Schema during Tool Use by Macaque Postcentral Neurones. *Neuroreport*, 7, 2325–30.
Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus & Giroux.
Kahneman, D., Slovic, P. and Tversky, A. (eds). (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
Kirsh, D. and Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, 18, 513–49.
Lakoff, G. and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.
Landy, D., Allen, C. and Zednik, C. (2014). A Perceptual Account of Symbolic Reasoning. *Frontiers in Psychology*, 5, 1–10 (Article 275).
Lau, H. C., Rogers, R. D. and Passingham, R. E. (2007). Manipulating the Experienced Onset of Intention after Action Execution. *Journal of Cognitive Neuroscience*, 19, 1, 1–10.
Lettvin, J. Y., Maturana, H. R., McCulloch, W. S. and Pitts, W. A. (1959). What the Frog's Eye Tells the Frog's Brain. *Proceedings of the Institute of Radio Engineers*, 47, 11, 1940–51.
Marr, D. (1982). *Vision*. New York: W. H. Freeman.
McCarthy, J. (1960). Recursive Functions of Symbolic Expressions and their Computation by Machine, part I. *CACM,* 3, 4, 184–95.
McClamrock, R. (1995). *Existential Cognition: Computational Minds in the World*. Chicago: University of Chicago Press.
Newell, A., Shaw, J. C. and Simon, H. A. (1958a). The Process of Creative Thinking. Rand Corporation technical report P-1320.
Newell, A., Shaw, J. C. and Simon, H. A. (1958b). Elements of a Theory of Human Problem Solving. *Psychological Review*, 65, 3, 151–66.
Newell, A., Shaw, J. C. and Simon, H. A. (1964). The Process of Creative Thinking. In *Contemporary approaches to creative thinking*, ed. H. Gruber, G. Terell and M. Wertheimer. New York: Atherton.
Newell, A. and Simon, H. A. (1976). Computer Science as Empirical Enquiry: Symbols and Search. *Communications of the ACM*, 19, 113–26.
Piccinini, G. (2010). The Mind as Neural Software? Understanding Functionalism, Computationalism, and Computational Functionalism. *Philosophy and Phenomenological Research*, 86, 2, 269–311.
Pinker, S. (1997). *How the Mind Works*. New York: W. W. Norton.

Putnam, H. (1960). Minds and Machines. S. Hook (ed.), *Dimensions of Mind*. New York: New York University Press, 57–80.
Putnam, H. (1963). Brains and Behavior. In *Analytical Philosophy Second Series*, ed. R. Butler. Oxford: Basil Blackwell.
Putnam, H. (1967). Psychological Predicates. In *Art, Mind, and Religion,* ed. W. H. Capitan and D. D. Merrill. Pittsburgh: University of Pittsburgh Press, 37–48.
Pylyshyn, Z. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.
Pylyshyn, Z. (2000). Situating Vision in the World. *Trends in Cognitive Sciences*, 4, 5, 197–207.
Rosenbloom, P., Laird, J., Newell, A. and McCarl, R. (1991). A Preliminary Analysis of the Soar Architecture as a Basis for General Intelligence. *Artificial Intelligence*, 47, 289–325.
Rowlands, M. (1999). *The Body in Mind: Understanding Cognitive Processes*. Cambridge: Cambridge University Press.
Rupert, R. (1998). On the Relationship between Naturalistic Semantics and Individuation Criteria for Terms in a Language of Thought. *Synthese*, 117, 95–131.
Rupert, R. D. (2006). Review of *Embodiment and Cognitive Science* by R. Gibbs. *Notre Dame Philosophical Reviews*, 2006.08.20.
Rupert, R. D. (2009). *Cognitive Systems and the Extended Mind*. Oxford: Oxford University Press.
Rupert, R. D. (2011). Embodiment, Consciousness, and the Massively Representational Mind. *Philosophical Topics*, 39, 99–120.
Serino, A. and Haggard, P. (2010). Touch and the Body. *Neuroscience and Biobehavioral Reviews*, 34, 224–36.
Shapiro, L. (2010). *Embodied Cognition*. New York: Routledge.
Simon, H. A. (1996). *Sciences of the Artificial,* 3rd edn (first edition published in 1969). Cambridge, MA: MIT Press.
Simon, H. A. and Newell, A. (1971). Human Problem Solving: The State of the Theory in 1970. *American Psychologist*, 26, 2, 145–59.
Sperry, R. W., Gazzaniga, M. S. and Bogen, J. E. (1969). Interhemispheric Relationships: The Neocortical Commissures; Syndromes of Hemispheric Disconnection. In *Handbook of Clinical Neurology,* vol. 4, ed. P. J. Vinken and G. W. Bruyn. Amsterdam: North-Holland, 273–90.
Steels, L. and Spranger, M. (2008). The Robot in the Mirror. *Connection Science*, 20, 4, 337–58
Stewart, J., Gapenne, O. and Di Paolo, E. (eds) (2010). *Enaction*. Cambridge, MA: MIT Press.
Thelen, E. and Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition andAaction*. Cambridge, MA: MIT Press.
van Gelder, T. (1995). What Might Cognition Be, If Not Computation? *Journal of Philosophy* 92, 345–81.
Varela, F., Thompson, E. and Rosch, E. (1991). *The Embodied Mind*. Cambridge, MA: MIT Press.
Vera, A. H. and Simon, H. A. (1993). Situated Action: A Symbolic Interpretation. *Cognitive Science* 17, 7–48.
Warrington, E. K. (1975). The Selective Impairment of Semantic Memory. *Quarterly Journal of Experimental Psychology,* 27, 635–57.

Weiskopf, D. (2004). The Place of Time in Cognition. *British Journal for the Philosophy of Science*, 55, 87–105.

Wilson, M. (2001). The Case for Sensorimotor Coding in Working Memory. *Psychonomic Bulletin & Review*, 8, 1, 44–57.

Wilson, R. and Foglia, L. (2011). Embodied Cognition. *The Stanford Encyclopedia of Philosophy* (Fall 2011 edition), ed. Edward N. Zalta (http://plato.stanford.edu/archives/fall2011/entries/embodied-cognition/).

3

Towards a Rational Theory of Heuristics

Gerd Gigerenzer

Herbert Simon left us with an unfinished task, a theory of bounded rationality. Such a theory should make two contributions. First, it should describe how individuals and institutions *actually make* decisions. Understanding this process would advance beyond as-if theories of maximizing expected utility. Second, the theory should be able to deal with situations of uncertainty where 'the conditions for rationality postulated by the model of neoclassical economics are not met' (Simon, 1989, p. 377). That is, it should extend to situations where one cannot calculate the optimal action but instead has to satisfice, that is, find either a better option than existing ones or one that meets a set aspiration level. This extension would make decision theory particularly relevant to the uncertain worlds of business, investment, and personal affairs.

Simon proposed satisficing as a general alternative to optimizing and also used the term to refer to a specific decision-making heuristic. Consider his account of why he studied political science and economics: 'I simply picked the first profession that sounded fascinating' (Simon, 1978, p. 1). This process is the essence of the *satisficing heuristic*, to set an aspiration level that defines what a satisfactory option would be and then choose the first alternative that meets that level. Satisficing can deal with uncertainty, that is, with situations where not all alternatives and consequences can be foreseen. The same rule is used for business decisions. Developers of high-rise office buildings and malls report that they decide in favor of an investment if they can get at least x return in y years (Berg, 2014a), and BMW dealers price used cars by setting an aspiration level and lowering it when the car is not sold after about 30 days (Artinger and Gigerenzer, 2015). Satisficing is a heuristic in the *adaptive toolbox* of individuals and organizations, but not the only one. When

making his own retirement investments, economist Harry Markowitz did not use his Nobel Prize winning 'optimal' mean-variance model, but a simple heuristic. 'I thought, "You know, if the stock market goes way up and I'm not in it, I'll feel stupid. And if it goes way down and I'm in it, I'll feel stupid,"' Markowitz recalls, 'so I went 50-50' (interview with Bower, 2011, p. 26). An equal split between bonds and equities is an instance of the 1/*N* heuristic. When wealth is allocated across a menu of *N* stocks (instead of stocks and bonds, as Markowitz did), studies indicate that 1/*N* typically outperforms the mean-variance model in the real world of finance, where the assumptions of the mean-variance model are not met (DeMiguel *et al.*, 2009).

Simon himself never systematically studied the heuristics in the adaptive toolbox, nor did he analyze the conditions under which these heuristics are successful – their 'ecological rationality' (Gigerenzer and Selten, 2001). Simon was well aware that he had provided a direction, but not a theory.[1] As he wrote to me shortly before his death, 'I did not want to give the impression that I thought I had "solved" the problem of creating an empirically grounded theory of economic phenomena. What I was trying to do is to call attention for the need for such a theory' (see Gigerenzer, 2004, p. 406). Earlier, he had wondered why his call for realism was received with 'something less than unbounded enthusiasm' and 'largely ignored as irrelevant for economics' (Simon, 1997, p. 269). I believe the answer is that he challenged two profound methodological commitments of neoclassical economists, the twin allegiance to optimization and as-if theories (Berg, 2014b). Going beyond these, Simon called for a shift towards:

1. *Uncertainty:* Analyze decision-making under uncertainty, where the optimal action cannot be determined.
2. *Process:* Design formal theories of the process of decision-making rather than as-if theories.

Let me explain. In this chapter, I use the term *risk* to describe situations in which all alternatives, outcomes, and probabilities are known for sure. The prototype is the choice between monetary gambles where all payoffs are well-defined. Correspondingly, I use the term *uncertainty* for situations where not all is known or can be foreseen. Similar distinctions have been made before. Knight (1921) distinguished between measurable probabilities, that is, frequencies and propensities, and those that cannot be measured empirically: 'a *measurable* uncertainty, or "risk" proper . . . is so far different from an *unmeasurable*

one that it is not in effect an uncertainty at all' (p. 20). L. J. Savage drew a comparable line between small worlds where Bayesian decision theory applies and situations where it does not. For instance, Savage (1972, p. 16) believed it would be 'utterly ridiculous' to apply Bayesian decision theory to problems such as planning a picnic or playing chess, and for different reasons. Planning a picnic, like choosing a profession, is an ill-defined situation, where it is impossible to foresee every possible outcome and where surprises may happen. Thus, the best course of action cannot be calculated in advance. Chess, in contrast, is a well-defined game with an optimal sequence of moves, which, however, no machine or mind can find. In technical terms, the game is computationally intractable – as are most problems that computer scientists work on (Tsotsos, 1991). This differs from tic-tac-toe, where players can easily determine the best sequence of moves, which makes it monotonous for all but small children.

The prototype of an as-if theory is the Ptolemaic model, with the Earth in the center and the planets and the sun orbiting around it in circles and epicycles. Few astronomers believed that planets actually move in such odd-looking epicycles; rather, the theory was designed as a guide for making predictions about planetary positions. Based on Copernicus's heliocentric revolution, Kepler's laws of planetary motion model the actual process of movement, with the planets moving around the sun in ellipses. After Ptolemy's as-if theory was overthrown in favor of a theory of the process, theoretical realism eventually led to better predictions. In the natural sciences, moving from as-if to process is considered progress. Not so in neoclassical economics. In his defense of as-if models, Friedman (1953) argued that the goal is not realism, but prediction. Yet, as I will argue, and as seen in the above example, increased realism is likely to improve prediction.

These two methodological commitments, optimization and as-if, are closely related. The ideal of optimization requires full knowledge of the relevant conditions and thus promotes as-if theories of economic agents who inhabit a world of known risks, not uncertainty. Yet in the real world of business, these risks (such as the space of all possible outcomes and the probability distributions over outcomes and payoffs) are often not known. The standard procedure of neoclassical economics is to transform situations of uncertainty into those of risk in order to be able to determine the best course of action. Whether this optimal solution is actually optimal in the real situation (i.e., under uncertainty) remains up in the air. Consider chess

again, where it is impossible to calculate the optimal sequence of moves although it does exist. In order to optimize, an as-if modeler could reduce the 8x8 board to a 4x4 board with a smaller set of figures. Yet this method would contribute little to mastering the real game. The alternative program is to accept that optimization has its limits and instead focus on analyzing the heuristics that chess masters and computer programs actually use. For some economists, however, a model without optimization does not belong to economics and is therefore inadmissible. Methodological commitments can unite a discipline but may also prove to be a mental straightjacket that inhibits progress.

The Neoclassical Counterrevolution

If we think of Simon's vision as a revolutionary program, the reason why it has been largely ignored can be called a 'neoclassical counterrevolution' supported, surprisingly, by the majority of behavioral economists.

First, neoclassical economists have declared bounded rationality to be nothing but full rationality in disguise. It is nothing new, so the argument goes, and we can therefore ignore it. For instance, in his essay in memory of Herbert Simon, Arrow (2004) insisted that 'boundedly rational procedures are in fact fully optimal procedures when one takes account of the cost of computation in addition to the benefits and costs inherent in the problem as originally posed' (p. 48). In many economists' view, bounded rationality is simply as-if optimization under constraints; Simon's bounds can be modeled by merely adding new constraints, such as those of memory and problem-solving ability, to the standard budget constraints. Simon once told me that he had considered suing colleagues who misused his term *bounded rationality* to refer to optimization.

Second, consider behavioral economics. Simon was one of its creators, but soon dropped out when Kahneman, Tversky, Thaler, and their followers took over and changed its course. Contrary to Simon, these researchers argued that there is nothing wrong with the theory of expected utility maximization but that the fault lies with people who do not follow it. 'Our research attempted to obtain a map of bounded rationality, by exploring the systematic biases that separate the beliefs that people have and the choices they make from the optimal beliefs and choices assumed in rational-agent models' (Kahneman, 2003, p. 1449).[2] Yet, suboptimal beliefs were not what Simon had in mind; as he points

out, 'bounded rationality is not irrationality' (Simon, 1985, p. 297). Nevertheless, many psychologists have come to believe that bounded rationality is the study of deviations from rationality.

Although behavioral economists started out with the promise of greater psychological realism, most have surrendered to the as-if methodology. Cumulative prospect theory, inequity-aversion theory, and hyperbolic discounting are all as-if theories. They retain the expected utility framework and merely add free parameters with psychological labels (Berg and Gigerenzer, 2010), which is like adding more Ptolemaic epicycles in astronomy. The resulting theories tend to be more unrealistic than the expected utility theories they are intended to improve on. Behavioral economics has largely become a repair program for expected utility maximization.

In sum, both neoclassical and behavioral economists altered and fitted Simon's program of bounded rationality to their programs, emphasizing rationality and irrationality, respectively. Despite this apparent contradiction, both groups regard the classical utility maximization framework as the sole way to model rational decisions. The behavioral economics 'revolution,' as it was once called, has boiled down to defending the neoclassical commitments to optimization and as-if theories.

Adding Parameters to the Utility Function Helps Predict the Past, Not Necessarily the Future

But what is wrong, one might ask, with these commitments, given that they provide a general framework of rationality? The price paid for the lack of realism is predictive power, which is exactly what Milton Friedman held to be the benchmark of a successful theory. Adding more parameters to the utility function leads to a better fit, but may mean losing predictive power. For instance, a study showed that cumulative prospect theory excelled in predicting the past, that is, when its parameters were fitted to known data, but when predicting the future, it did systematically worse than a simple rule called the *priority heuristic* (for hard choices, that is, gambles with similar expected values) and than expected value theory (for easy choices; see Brandstätter *et al.*, 2006). This result is not accidental. Neither the priority heuristic nor expected value theory has free parameters, and thus both avoid error due to parameter estimation, a source of prediction error I consider below. This simplicity can be a strength; the priority heuristic implies the fourfold pattern of risk attitude and other violations of expected utility theory without needing different sets of parameters for different classes

of violations (Katsikopoulos and Gigerenzer, 2008). Note that predictive performance is not the same as R^2 in data fitting, which amounts to hindsight; prediction is about foresight, as in out-of-sample prediction when an inference is made from a sample to another sample or a population.

In their review of half a century of research, D. Friedman, Isaac, James, and Sunder (2014) analyzed the empirical evidence for how well Bernoulli functions – such as utility of income functions, utility of wealth functions, and the value function in prospect theory – perform in terms of predictive accuracy. They concluded: 'Their power to predict out-of-sample is in the poor-to-nonexistent range, and we have seen no convincing victories over naïve alternatives' (p. 3). Similarly, Stewart, Reimers, and Harris (2014) experimentally showed that no stable mapping exists between attribute values and subjective equivalents, as assumed in expected utility theories and their modifications, such as prospect theory and hyperbolic discounting theory. This instability was documented long ago in psychophysical research (Brunswik, 1934; Parducci, 1965). If D. Friedman *et al.* (2014) are correct, then expected utility theories and their modifications fail both at describing the process of decision-making and at accurately predicting the actual outcomes.

The Argument: Better Realism, Better Prediction

In the following, I start with Milton Friedman's statement that the measure of a good theory is its predictive power and derive Simon's realism – rather than Friedman's as-if – as a logical consequence. Friedman (1953, p. 41) wrote that a theory should be evaluated 'only by seeing whether it yields predictions that are good enough for the purpose in hand or that are better than predictions from alternative theories.' I will argue: (1) that simple heuristics can predict better than highly parameterized models under a wide range of conditions; and also (2) model the process of how individuals and organizations actually make decisions; and therefore, (3) that Friedman's goal of prediction implies studying simple heuristics, not only as-if theories. In this way, I derive Simon's call for realism from Friedman's call for good theories.

Specifically, I show that the error in prediction (unlike in data fitting) has two components that we can influence: bias and variance. Prediction by simple heuristics tends to decrease the variance component, while adding free parameters increases it (Gigerenzer and Brighton, 2009). Next, I distinguish three ways of reducing error from variance

and show that these correspond to three classes of heuristics that humans rely on. Finally, I show that in natural environments, the bias component of error generated by heuristics appears to be surprisingly low. Together, the analysis of bias and variance specifies the conditions for when simple heuristics predict better than complex 'rational' models and provides an explanation of why simple heuristics can be rational.

The Ecological Rationality of Heuristics

In my own work, I have tried to lay the foundations for a theory of bounded rationality. Such a theory addresses not only Simon's descriptive question (how *do* people make decisions?), but also a normative one (how *should* people make decisions *under uncertainty*?). The study of the adaptive toolbox asks the descriptive question: What is the repertoire of heuristics available to an individual or organization? Its methods are experimentation and observation, and the results are algorithmic models of heuristics, such as satisficing and 1/*N*. The study of the 'ecological rationality' of heuristics asks the normative question: What are the conditions under which a heuristic leads to a better outcome than a competing strategy? Its methods are analysis and computer simulation, and the results are the conditions under which a class of heuristics is successful according to a metric such as predictive accuracy.

The study of ecological rationality reaches beyond Simon's call for descriptive process models. However, it was inspired by an analogy of Simon's: 'Human rational behavior (and the rational behavior of all physical symbol systems) is shaped by a scissors whose two blades are the structure of the task environment and the computational capabilities of the actor' (Simon, 1990, p. 7). We have fleshed out his analogy into a systematic theory of ecological rationality (Gigerenzer *et al.*, 2011; Gigerenzer and Selten, 2001). The results explain when and why one should rely on simple heuristics in order to make predictions superior to those of highly parameterized models.

Such a theory of bounded rationality is not about human failure. Rather, it explains how and when people can make good decisions by using less information. In what follows, I will deal exclusively with the ecological rationality of heuristics, focusing on the conditions of their predictive power. For general reviews, see Gigerenzer, Todd, and the ABC Research Group (1999); Hertwig, Hoffrage, and the ABC Research Group (2013); Todd, Gigerenzer, and the ABC Research Group (2012).

The Bias–Variance Dilemma

The cause of error is sometimes conceived as

$$\text{Error} = \text{bias} + \varepsilon \tag{1}$$

where ε is unsystematic noise (mean zero and uncorrelated with bias), and bias is the systematic difference between the (average) prediction a model makes and the true state. For instance, if the true temporal trajectory of a variable is a polynomial of third degree and a linear regression is used to predict the variable, the model has a systematic bias. Equation (1) is implicit in the argumentation of the heuristics-and-biases program (Kahneman, 2011), where a cognitive error is defined in terms of a systematic bias that arises from ignoring information such as base rates. In this view, if the bias is eliminated, rational judgments are obtained. Yet equation (1) is appropriate only in a world of known risks or data fitting, not for making predictions.

Enter prediction. Consider the problem of estimating the true value μ in a population on the basis of random samples. Each of S samples ($s = 1, \ldots, S$) generates an estimate x_s. The variability of these estimates x_s around their mean $\bar{x}$, which is called *variance* in machine learning, is another source of prediction error (Brighton and Gigerenzer, 2012; Geman *et al.*, 1992). The variance component reflects the sensitivity of the predictions to different samples drawn from the same population. Thus the prediction error (the sum of squared error) can be captured in the equation:

$$\textit{Prediction error} = \text{bias}^2 + \text{variance} + \varepsilon \tag{2}$$

where

- $\text{bias} = \bar{x} - \mu$, that is, the average deviation of the mean of the sample estimates from the true value, and
- $\text{variance} = \frac{1}{s}\sum(x_s - \bar{x})^2$, that is, the mean squared deviation of the sample estimates from their mean $\bar{x}$.

Figure 3.1 provides a visual depiction of bias and variance. The bull's eye represents the true value, and each dart the estimate from a sample. Mr. Bias, whose darts landed on the left dartboard, has a systematic bias but low variance. Mr. Variance, whose darts landed on the right dartboard, has no bias because the darts line up exactly around the bull's eye. However, his dart throws show considerable variance. Thus, in prediction, two sources of error (ignoring noise) need to be considered, not one.

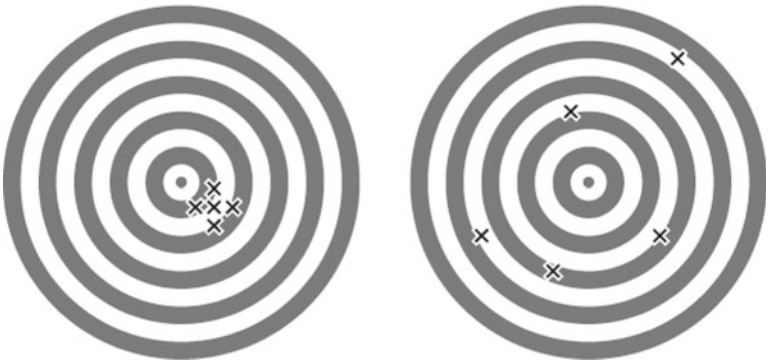


Figure 3.1 A visual depiction of bias and variance
A visual analogy of the two components of prediction error: bias and variance. The bull's eye is the unknown true value μ (here: 0,0) to be predicted. Each dart represents a predicted value x_s based on a random sample from the population with the true value μ. Bias is zero if the mean prediction 'hits' the target. Left: Mr. Bias's strategy results in a systematic bias, whose size is the distance between the mean of the darts thrown and the bull's eye ($\bar{x} - \mu,$), and a low variance, that is, the darts are close together. Right: Mr. Variance's strategy results in zero bias ($\bar{x} = \mu,$), that is, the darts are lined up exactly around the bull's eye, but with considerable variance.

A moderate bias with low variance (left) may lead to better results than would a zero bias with high variance.

The dart analogy in Figure 3.1 does not capture the trade-off between bias and variance. Adding free parameters to a model, which happens when replacing expected utility theory with prospect theory, is likely to reduce the bias component of error, but at the cost of increased variance. By taking away free parameters, such as when replacing expected utility with expected value theory, the opposite happens, a likely reduction in variance at the cost of higher bias. Variance is also influenced by how the parameters are combined, which is determined by the functional form of the model (e.g., multiplicative or exponential). A strategy without any free parameter likely has some bias but no variance. The reason is that the strategy is not sensitive to specific samples and always produces the same prediction. Consider again the $1/N$ heuristic and the mean-variance model for allocating one's wealth across N stocks. Mean-variance needs to estimate its numerous parameters from stock data and will generate error from both variance and bias. In contrast, $1/N$ does not estimate any parameters but in fact ignores past data and thus does not generate error from variance but likely from higher bias.

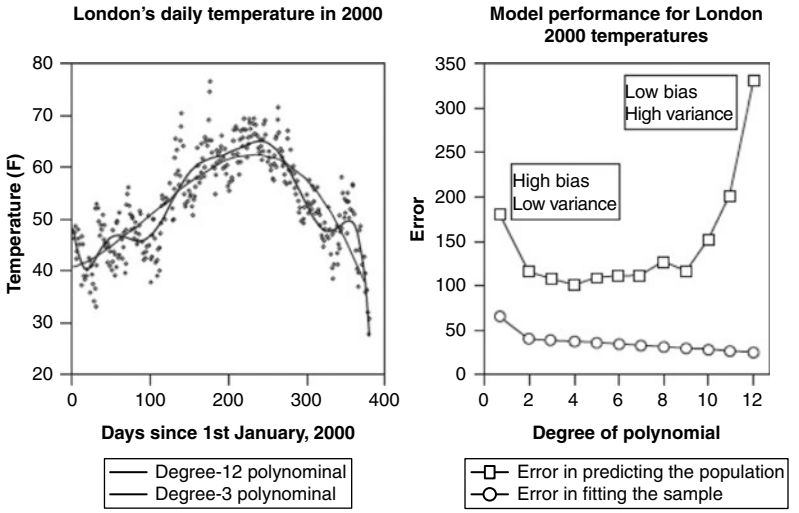


Figure 3.2 Empirical illustration of the bias–variance trade-off when predicting the average daily temperature in London

The bias–variance dilemma in prediction. Left: Data fitting. Each point is the average temperature in London for one of 365 days in the year 2000. The figure shows the best-fitting degree-3 polynomial (thin line) and degree-12 polynomial (thick line), using the least squares method. Clearly, the 12-degree model fits the data best. The error is the sum of squared error. Right: Prediction. The task is to predict the average temperature in London on every day, based on random samples of 30 days. Although the fit increases with higher-degree polynomials (lower curve), the prediction error does not follow this pattern. Rather, there is a u-shaped function between prediction error and complexity of polynomial. For instance, a degree-1 polynomial (i.e., a straight line) generates less error than the degree-12 polynomial, which has less bias but more variance. Adapted from Gigerenzer and Brighton (2009).

Figure 3.2 provides an empirical illustration of the bias–variance trade-off when predicting the average daily temperature in London. The left panel shows the temperature for each day, and a 3-degree and a 12-degree polynomial fitted to the data. The right panel shows that the fit improves (i.e., the error decreases) when the polynomial grows in complexity. A polynomial of degree 364 would guarantee a perfect fit, so that a line can be drawn through each point. But that is not true for prediction. The u-shaped curve in the right panel reveals the trade-off between bias and variance in prediction. Bias is highest for the 1-degree polynomial and lowest for the 12-degree polynomial, while variance shows the opposite pattern. The 4-degree polynomial has the best trade-off between bias and variance, that is, the lowest total error. Note that the 12-degree polynomial, which has the best fit and thus the least bias,

predicts less accurately than the 1-degree polynomial that has a strong bias but less variance.

Let me summarize. The bias–variance dilemma decomposes the total prediction error into bias, variance, and noise. The variance component can be reduced by decreasing the number of parameters and by increasing the sample size. To arrive at good predictions, simpler models predict more accurately to a point, which represents the bias–variance trade-off, while further simplification may lead to an increase in error. Thus, to minimize total error, a certain amount of bias is needed to counteract variance, which is the error due to oversensitivity to characteristics of specific samples. Bias *per se* is not the problem, as assumed in the heuristics-and-biases program. Rather, it can be part of the solution.

Simple Heuristics

How to Reduce Prediction Error

Consider predicting which of two alternatives will have a higher value on a variable of interest. Assume that the true state of nature can be represented by a linear equation with n attributes (predictors). We do not know what the weights are and want to reduce prediction error due to variance. There are several ways to proceed, each corresponding to a class of simple heuristics that people rely on (Gigerenzer and Gaissmaier, 2011):

1. *One-reason heuristics*. The prediction is based solely on a single predictor among $n > 1$ observable predictors; the other $n - 1$ are ignored. This class of heuristics can be seen as a special case of sequential search heuristics with only one predictor (next class).
2. *Sequential search heuristics*. The prediction is based on a lexicographic rule, that is, if the first predictor does not allow for a decision, then the second is used, and so on. The predictors are ordered by simple correlations between each and the variable of interest, ignoring dependencies (i.e., the covariance structure) among predictors.
3. *Tallying heuristics*. The prediction is based on all n predictors by assigning equal weights to each one and then summing their values.

In each of these classes of heuristics, error due to variance is reduced (relative to a full linear model) because the number of parameters to be estimated is reduced. For instance, the common rationale for all three classes is to avoid the prediction error that results from estimating the $n(n+1)/2$ covariances. Tallying does not estimate the order of the

predictors either, but only their signs (positive or negative). The price for reducing variance is that bias is likely increased (but not necessarily; see below). Consider a few cases.

One-Good-Reason Heuristics

Companies such as catalog retailers, airlines, and hotel chains target their previous customers with product information and special offerings. Not all customers are active, that is, will buy in the future, and predicting which are active is important when reducing marketing costs. The state-of-the-art approach is the Pareto/NBD model and its variants (Schmittlein and Petersen, 1994), where NBD stands for negative binomial distribution. For each previous customer, it yields the probability that he or she is still active, based on the following assumptions (Wübben and von Wangenheim, 2008):

> *Pareto/NBD model:* While the customer is active, purchases follow a Poisson process with purchase rate λ. Customer lifetime is exponentially distributed with dropout rate μ. The purchase rates λ and the dropout rates μ are distributed according to a gamma distribution across the population of customers. The rates λ and μ are distributed independently of each other.

Although this model estimates the probability that a customer is active, it has found little acceptance among experienced managers. Instead, business executives rely on a toolbox of simple heuristics (Verhoef *et al.*, 2002). Wübben and von Wangenheim (2008) observed that managers in an airline and an apparel retailer relied on a recency-of-last-purchase (hiatus) rule:

> *Hiatus heuristic:* If a customer has not made a purchase for nine months or longer, classify him/her as inactive, otherwise as active.

The hiatus heuristic is an instance of the class of one-reason heuristics. It considers the hiatus only and ignores other information used by the Pareto/NBD model, such as the number of purchases made. Given that the hiatus heuristic uses only a subset of the relevant information used by the Pareto/NBD model and does not estimate any parameters (if the hiatus is fixed), it might appear to be second best. Equation (2) shows the mistake behind that assumption. The real question is whether the total error that the Pareto/NBD model generates is higher or lower than that of the hiatus heuristic. Wübben and von Wangenheim (2008) put

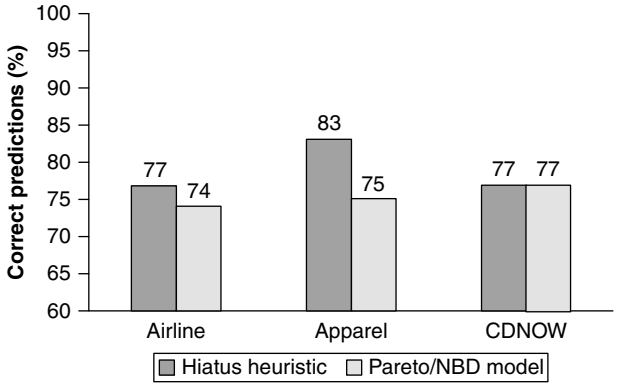


Figure 3.3 Hiatus heuristic made more correct predictions than the Pareto/NBD model

How to predict which customers will buy in the future? Shown is a competitive test of the Pareto/NBD model from marketing science against the hiatus heuristic managers rely on. The heuristic better predicts customer behavior for the airline and the apparel business, and equally well for the CDNOW retailer. With a fixed hiatus (such as 9 months), the heuristic has no free parameter and thus does not make errors due to variance. Note that the heuristic uses only a subset of the data the Pareto/NBD model uses, that is, makes better predictions with less effort. Adapted from Wübben and von Wangenheim (2008).

the issue to an empirical test. They calibrated the Pareto/NBD model to the customer databases of three companies, using 40 weeks of data, and used this calibrated model to predict the next 40 weeks of activities. The third company was the online CD retailer CDNOW, using a six-month hiatus. Figure 3.3 shows that the hiatus heuristic made more correct predictions than did the Pareto/NBD model for the airline customers, with 77% versus 74%, and for the apparel customers, with 83% versus 75%, and matched the number of correct predictions for the CDNOW customers. Less information can be more.

With a fixed hiatus, the hiatus heuristic has bias, but no variance. The Pareto/NBD model likely has less bias, but additional error due to variance because it needs to estimate four parameters from the sample data. The success of the heuristic suggests that its error due to bias is less than the total error made by the Pareto/NDP model.

Sequential Search Heuristics

The *take-the-best heuristic* was the first heuristic that the ABC Research Group systematically studied (Gigerenzer and Goldstein, 1996; Luan *et al*., 2014). It helps decision makers choose between two alternatives

based on n attributes, not only one, as with the hiatus heuristic. It orders attributes ($i = 1, \ldots, n$) by their simple validities v_i defined as the proportion of correct decisions c_i:

$$v_i = c_i / t_i \tag{3}$$

The denominator t_i gives the number of cases where the values of the two alternatives on the attribute i differ. If the values on the first attribute do not differ, the next is looked up, and so on in a lexicographic way. The decision is based on the first attribute that differentiates; all other attributes are ignored. Note that their order is determined by simple validities v_i – unlike beta weights in multiple regression, which require estimating the covariance matrix – and that these validities are estimated from samples. For simplicity, I assume here that there is a positive correlation between each attribute and the outcome (dependent) variable. Take-the-best entails three steps:

Search rule: Look through predictors in the order of their validity.
Stopping rule: Stop search when the first predictor is found where the values of the two alternatives differ.
Decision rule: Predict that the alternative with the higher predictor value has the higher value on the outcome variable.

How well does take-the-best predict compared to multiple regression? Figure 3.4 shows the results for 20 prediction tests on economic, demographic, and societal questions, such as which of two houses will have a higher selling price, or which school will have a higher drop-out rate (Czerlinski *et al.*, 1999). In every test, half of the data points were used to fit the parameters and the other half was predicted, a procedure known as cross validation. On average, multiple regression had the better fit, but take-the-best predicted better. To excel in fitting and fail in prediction is known as overfitting.

The take-the-best heuristic was also more frugal than regression, requiring, on average, only 2.4 predictors compared to 7.7 for regression. Like the hiatus heuristic and the Pareto/NBD model, take-the-best used only a subset of the information used by multiple regression, which protected against estimation error from variance.

Tallying Heuristics

Unlike take-the-best, tallying relies on all predictors but uses equal weights. Figure 3.4 shows that, on average, tallying predicted better

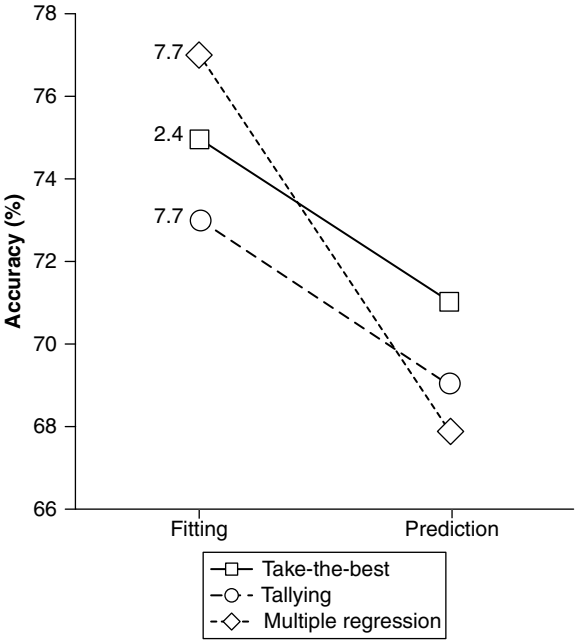


Figure 3.4 Results for 20 prediction tests on economic, demographic, and societal questions

Less-is-more effects across 20 prediction tasks in economics, business, biology, and other fields. Two heuristics, take-the-best and tallying, are tested competitively against multiple regression (Czerlinski *et al.*, 1999). Note that many of the tasks are taken from textbooks on regression. The number of attributes ranged between 3 and 18, and the number of alternatives between 11 and 395. Take-the-best orders attributes (predictors) in a simple way (without analyzing dependencies between cues) and uses only the first cue that differentiates between the alternatives. Tallying introduces a different bias; it uses all attributes that regression uses but does not estimate their weights, instead using the same weight for each. Prediction is tested by letting the three strategies estimate their parameters from half of the data set and then testing performance on the other half (cross-validation). Multiple regression estimates beta weights, take-the-best estimates only the order of cues, and tallying only the sign of the cues. By introducing bias, both heuristics make more accurate predictions than regression. For comparison, the performance in fitting data is shown. Regression excels in data fitting because it has more free parameters, but makes fewer correct predictions. Adapted from Gigerenzer and Brighton (2009).

than multiple regression, although not as well as take-the-best. Similarly, Åstebro and Elhedhli (2006) used a version of tallying to forecast the commercial success for early-stage ventures and reported that the heuristic predicted 86% of successes and failures correctly, compared to a log-linear regression model that predicted 79% correctly.

Two Misconceptions

These results demonstrate the bias–variance trade-off in the real world of prediction. They also help to correct two widespread misconceptions about heuristics. First, a common explanation for why people rely on heuristics is the accuracy–effort trade-off: Heuristics reduce effort but pay for this with less accuracy (Conlisk, 1996). The effort is often called deliberation costs. Although such a general trade-off sounds plausible, it is incorrect. Heuristics indeed reduce effort, but that does not necessarily reduce accuracy, as the empirical results in this section demonstrate. More generally, the bias–variance dilemma implies that there is no general accuracy–effort trade-off. It also explains when and why higher accuracy results from less effort. These situations are known as less-is-more effects.

A second misunderstanding is the claim that the study of heuristics is unnecessary because a heuristic can always be rewritten as a special case of the general linear model. Indeed, take-the-best and tallying can (Martignon and Hoffrage, 2002), but rewriting does not help to make better predictions. By generalizing a heuristic to a linear model, one can actually lose predictive power by creating more error due to variance, as Figure 3.4 illustrates. After all, rewriting the law of falling bodies as a general polynomial does not add to understanding physics.

Empirical Evidence

There is a large body of empirical studies showing that the classes of heuristics described in the previous sections are good models for how people make decisions, and that people tend to use them in an adaptive way, that is, in situations where they are ecologically rational. For instance, sequential search heuristics, such as in take-the-best, have been studied extensively in both laboratory experiments (e.g., Bröder, 2012; Bergert and Nosofsky, 2007) and 'in the wild' (Gigerenzer *et al.*, 2011). This research showed that decisions made by experts – from airport customs officers to police officers to professional burglars – are often best predicted by take-the-best or similar lexicographic rules, while novices and undergraduates try to consider all n attributes (Pachur and Marinello, 2013). Hence, experts appear to know intuitively how to reduce error due to variance, which makes their decisions more efficient.

Simon started with the concept of satisficing. Today, we have a large body of empirical evidence for other classes of heuristics, together with formal models of them (Gigerenzer and Gaissmaier, 2011). These formal

models are a scientific leap from the prematheatical period of vague labels such as 'availability' and 'System 1' (Kahneman, 2011). Formal models can make testable predictions and enable the normative study of their ecological rationality.

Environmental Structures and Bias

So far, we have focused on variance. But, according to the bias–variance dilemma, the crucial question is how much bias a heuristic produces by reducing variance. To assess bias, one needs to compare predicted outcomes to actual outcomes. Assume again that the actual outcome can be represented by a linear equation with n predictors. Consider again a choice between objects A and B, based on n predictors, where the value of the ith predictor is represented by x_i and weighted in the linear payoff function by w_i. To simplify, assume that the predictors are binary and the weights are nonnegative. The class of strategies we consider are sequential search (lexicographic) heuristics, with one-reason decision-making as a special case with only one predictor.

Environmental Structures

The term *environment* refers to the alternatives, outcomes, payoffs, and all other factors in the model exogenous to the decision maker. Simple environments may be described as a joint distribution of predictors and outcome variables, which induce payoff distributions that depend on actions in the decision maker's choice set. With this broad interpretation of *environment*, we can then investigate which environmental structures 'help' lexicographic heuristics perform well so that they have a small or even zero bias (in addition to small variance). We know of three structural features: noncompensatoriness, dominance, and cumulative dominance.

Noncompensatoriness. The weights $w_1, w_2, w_3, \ldots, w_n$ are noncompensatory if they satisfy the $n-1$ inequality constraints:

$$w_i > \sum_{n=i+1}^{n} w_j, \quad i = 1, 2, \ldots, n-1 \tag{4}$$

An example is the set of weights {1, ½, ¼, ⅛} (see Figure 3.5). If the weights are noncompensatory, then a linear rule (with the same order of predictors) will always lead to the same choice as a lexicographic rule (Martignon and Hoffrage, 2002). Take the example of weights above.

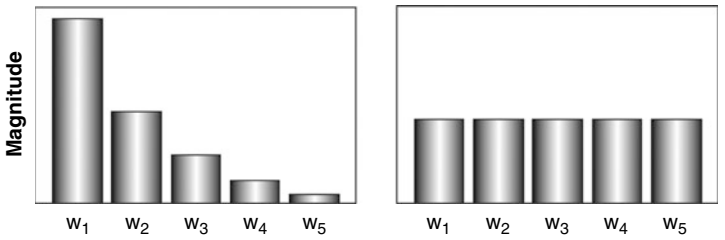


Figure 3.5 Environmental structures and bias
Left: A graphical illustration of noncompensatoriness. If the weights of a linear rule are noncompensatory, such as 1, ½, ¼, ⅛, and 1⁄16, then a lexicographic heuristic with the same order of attributes will always make the same prediction as the linear rule. Therefore, if the true state of nature can be represented by a linear rule and noncompensatoriness holds, a lexicographic heuristic has no bias. *Right*: For comparison, a set of weights where tallying has no bias. Adapted from Martignon and Hoffrage (2002).

If the lexicographic rule yields decisions on the basis of the first predictor (with weight 1), every linear rule will match this choice, because the sum of all other weights (½+¼+...) will always be smaller than the weight of the first predictor. Thus, if the true state of nature is linear and noncompensatoriness holds, then a lexicographic heuristic with the same order of predictors has no bias.

Dominance. If alternative *A* has a value higher than or equal to alternative *B* in all *n* predictors, and a higher value on at least one predictor, then alternative *A* dominates alternative *B*. Figure 3.6 (top) illustrates dominance. If *A* dominates *B*, a lexicographic heuristic (and tallying) will arrive at the same prediction as a linear rule. In terms of a linear rule, dominance means that all differences $w_i \Delta x_i = w_i(x_{iA} - x_{iB})$ are nonzero and at least one is positive; thus, the linear rule chooses *A* over *B*. This result holds for any (nonnegative) weights and does not depend on noncompensatory ones. Thus, if the true state of nature is linear and dominance holds, a lexicographic heuristic has no bias.

Cumulative dominance. The cumulative profile of an alternative consists of *n* values, where the *i*th value is the sum of the first *i* values. Alternative *A* cumulatively dominates *B* if its cumulative profile exceeds or equals the cumulative profile of *B* in every term and exceeds it in at least one term (Baucells *et al.*, 2006). If cumulative dominance holds, then a linear rule (with the same order of predictors) predicts the cumulative dominant object, just as a lexicographic rule does. Consider the example in Figure 3.6 (bottom). Unlike in the top panel, dominance does not hold. To check for cumulative dominance, one first compares

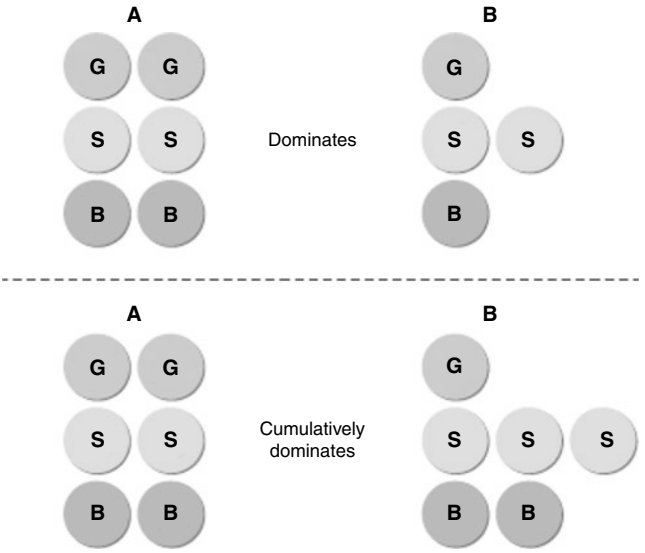


Figure 3.6 Cumulative dominance
A pictorial illustration of dominance and cumulative dominance. Which of two alternatives, *A* and *B*, is more valuable? The alternatives vary on three attributes, gold, silver, and bronze coins. In the top panel, *A* dominates *B* because it has more gold and bronze coins, and as many silver coins. In the bottom panel, dominance does not hold, but cumulative dominance does. To check for cumulative dominance, one first compares the number of gold coins, then the number of gold and silver coins, and finally the number of all coins. In every comparison, alternative *A* has at least as many coins as *B*, and more coins in one. It has more gold coins, an equal number of gold and silver coins, and an equal number of gold, silver, and bronze coins. Thus, *A* cumulatively dominates *B*. If dominance or cumulative dominance holds, a linear model makes the same choice (alternative A) as a lexicographic heuristic with the same order of cues. Adapted from Şimşek (2014).

A and *B* on the top attribute (gold): *A* has two gold coins, and *B* only one. Then *A* and *B* are compared on the sum of the top two attributes; here the number of coins is the same. Finally, the comparison is made on all three attributes, and again there is no difference. Because there is one difference in favor of *A*, and otherwise zero difference, *A* cumulatively dominates *B*.

Thus, unlike simple dominance, cumulative dominance requires an order of predictors or attributes. The cumulative difference can be defined as:

$$\Delta x_i' = \sum_{j=1}^{i} \Delta x_j \quad i = 1, 2, \ldots, n \tag{5}$$

Alternative A cumulatively dominates B if all terms $w_i'\Delta x_i'$ are nonnegative and at least one of them is positive, where $w_i' = w_i - w_{i+1}, i = 1, 2, \ldots, n-1$, and $w_n' = w_n$. Note that dominance implies cumulative dominance, but not vice versa.

These three environmental conditions influence the bias component of the error. Noncompensatoriness refers to the relative strength of the predictors in the environment, while the two dominance conditions refer to the relative quality of alternatives (Katsikopoulos, 2011). In sum, if either noncompensatoriness or dominance or cumulative dominance holds, then lexicographic heuristics will have no bias, relative to a linear rule.

How Often Do These Conditions Hold in the Real World?

It is not easy to answer this question because there is no way to define the set of all prediction problems and draw a random sample. But it is possible to investigate a large and diverse number of natural data sets. Şimşek (2013) analyzed 51 data sets from online repositories, textbooks, research publications, packages for R statistical software, and individual scientists collecting field data. The data sets spanned content areas as diverse as biology, business, computer science, ecology, economics, education, engineering, and medicine, among others. The number of attributes ranged from 3 to 21, which were numeric or binary; the number of objects (alternatives) from 12 to 601, corresponding to numbers of possible pairwise comparisons ranging from 66 to 180,300. Each of these comparisons amounts to a prediction being made.

How often was one or more of the three conditions – noncompensatoriness, dominance, and cumulative dominance – satisfied? The result was surprising. The median for the 51 data sets was 90% (Şimşek, 2013). That is, in half of the data sets, more than 90% of the decisions encountered were such that a lexicographic rule yielded the same prediction as a linear model. When the predictors were dichotomized at the median, this number increased to 97%. In other words, in the majority of the cases, the lexicographic heuristics had the same bias as a linear model. Together with the potential for reducing variance, this result explains why simple heuristics often outperform linear models in prediction, as shown in Figure 3.4.

In summary, the prediction error has two primary components that can be influenced, bias and variance. Variance can be decreased by decreasing the number (and kind) of free parameters and by increasing the sample size. Simple heuristics decrease variance because they have

few or zero free parameters. To analyze the bias component, one needs to know the true process that generates actual outcomes, which was assumed to be a linear function of *n* predictors or attributes whose order is known. Under those assumptions, three environmental conditions were described that guarantee that lexicographic heuristics have no bias when choosing between two alternatives: noncompensatoriness, dominance, and cumulative dominance. Alternatively, if the true process is unknown, the bias is equivalent to that of a linear model. An analysis of a diverse collection of data showed that one or more of these conditions were in place in 90% (97% for binary attributes) of the cases. Together, these results provide an explanation of less-is-more effects, that is, situations where using a subset of the available information (e.g., ignoring all data except the hiatus) leads to better predictions than using all available information.

Methodological Principles

Finally, the study of the adaptive toolbox and that of ecological rationality entail adherence to three methodological principles:

1. *Algorithmic models of heuristics*, not verbal labels (such as availability, System 1, and near-tautologies; see Gigerenzer, 1996; Gigerenzer *et al.*, 2012).
2. *Tests of prediction*, not fitting data.
3. *Competitive testing* (such as testing the predictions of two fully specified models), not testing of a single model.

For instance, the research on the hiatus heuristic uses an algorithmic model of the heuristic, tests its performance in prediction, and compares it to the Pareto/NBD model. Although these methodological principles should be obvious they are not widely followed in those parts of the behavioral economics literature that propose verbal labels rather than algorithmic models of heuristics or fit models, such as prospect theory, to data sets without out-of-sample prediction and without testing them competitively against models of simple heuristics.

Conclusion

In this chapter, I started with Milton Friedman's dictum that the measure of a good theory is its predictive power and derived Simon's realism, rather than as-if, as the logical consequence. Specifically, I showed that

the error in prediction (unlike in data fitting) has two components, bias and variance. The use of simple heuristics likely decreases variance, while the use of models with many free parameters tends to increase it. Moreover, an analysis of natural environments indicates that simple heuristics generate a surprisingly small bias, relative to linear models. Investigating the balance between bias and variance allows for deriving the conditions under which simple heuristics can predict more accurately than complex 'rational' models, and vice versa. These results can unite Friedman's dictum with Simon's call for realism. If rationality means making better predictions, we should seriously investigate the adaptive toolbox of humans and the ecological rationality of heuristics rather than adding more parameters to as-if utility functions.

As a consequence, the use of simple heuristics by economic agents should not be routinely attributed to mere deliberation costs or even irrationality. Instead, it should be recognized that some degree of bias actually enables better performance in situations of uncertainty. Risk and uncertainty can require different sets of tools, of statistics, and of heuristics. In Simon's words (1981, p. 36), uncertainty 'places a premium on robust adaptive procedures instead of strategies that work well only when finely tuned to precisely known environment.'

Many years ago, a well-known behavioral economist told me with utmost conviction: 'Look, either reasoning is *rational* or it's *psychological*.' This false opposition between what is regarded as rational versus psychological has haunted me since. It is time to rethink the nature of rationality. A theory of bounded rationality based on the twin foundations of the adaptive toolbox and ecological rationality can be a start. Pursuing this goal is a step towards making progress on the unfinished task Simon left behind and may even contribute to unifying economics and psychology.[3]

Notes

1. Simon developed his thinking over decades, a complex process that I cannot give due justice to in this article. For instance, he introduced a satisficing heuristic in 1955, but in the Appendix of that article he presented an optimization model that maximizes expected value of the sales price, similar to the optimization under constraints model that Stigler (1961) later proposed. In the introduction to this article in his collected papers, Simon (1979, p. 3) made it clear that he thinks of satisficing as nonoptimizing: 'Satisficing, aiming at the good when the best is incalculable, is the key device.' Similarly, in early writings he sometimes linked bounded rationality to cognitive limitations, while later he linked it to cognition and environment – the scissors analogy – and argued that it is impossible to understand behavior by

looking at the one blade of cognitive limitations only. Thus, part of the misinterpretations of Simon's concept of bounded rationality I point out later may be due to his own development in thinking.

2. Although Simon's bounded rationality is commonly presented as a forerunner of Kahneman's heuristics and biases program, the latter's relation to bounded rationality appears to be an afterthought. In fact, Simon is not cited at all in Kahneman and Tversky's major early papers (all reprinted in Kahneman, Slovic, & Tversky, 1982). Simon is briefly mentioned in the preface to this anthology, apparently more as a nod to a distinguished figure than an acknowledgment of a significant intellectual debt (Lopes, 1992).
3. For helpful comments I would like to thank Florian Artinger, Nathan Berg, Henry Brighton, Ralph Hertwig, Perke Jacobs, Konstantinos Katsikopoulos, Shenghua Luan, Thorsten Pachur, and Özgür Şimşek.

References

Arrow, K. J. (2004). Is Bounded Rationality Unboundedly Rational? Some Ruminations. In *Models of a Man: Essays in Memory of Herbert A. Simon*, ed. M. Augier and J. G. March. Cambridge, MA: MIT Press, 47–55.

Artinger, F., and Gigerenzer, G. (2015). Aspiration-Adaptation, Price Setting, and the Used-Car Market. Unpublished manuscript.

Åstebro, T. and Elhedhli, S. (2006). The Effectiveness of Simple Decision Heuristics: Forecasting Commercial Success for Early-Stage Ventures. *Management Science*, 52, 395–409.

Baucells, M., Carrasco, J. A., and Hogarth, R. M. (2006). *Cumulative Dominance and Heuristic Performance in Binary Multi-Attribute Choice*. Available at SSRN: http://www.researchgate.net/publication/23695656_Cumulative_Dominance_and_Heuristic_Performance_in_Binary_Multi-Attribute_Choice.

Berg, N. (2014a). Success from Satisficing and Imitation: Entrepreneurs' Location Choice and Implications of Heuristics for Local Economic Development. *Journal of Business Research,* 67, 1700–09. doi: 10.1016/j.jbusres.2014.02.016

Berg, N. (2014b). The Consistency and Ecological Rationality Schools of Normative Economics: Singular versus Plural Metrics for Assessing Bounded Rationality. *Journal of Economic Methodology, 21*, 375–95.

Berg, N. and Gigerenzer, G. (2010). As-if Behavioral Economics: Neoclassical Economics in Disguise? *History of Economic Ideas, 18*, 133–65. doi: 10.1400/140334

Bergert F. B. and Nosofsky, R. M. (2007). A Response-Time Approach to Comparing Generalized Rational and Take-the-Best Models of Decision Making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 331, 107–29.

Bower, B. (2011). Simple Heresy. Rules of Thumb Challenge Complex Financial Analyses. *Science News*, 179, 26.

Brandstätter, E., Gigerenzer, G., and Hertwig, R. (2006). The Priority Heuristic: Making Choices without Trade-offs. *Psychological Review*, 113, 409–32. doi: 10.1037/0033-295X.113.2.409

Brighton, H. and Gigerenzer, G. (2012). Are Rational Actor Models 'Rational' Outside Small Worlds? In *Evolution and Rationality: Decisions, Co-operation and Strategic Behavior*, ed. S. Okasha and K. Binmore. Cambridge: Cambridge University Press, 84–109.

Bröder, A. (2012). The Quest for Take-the-Best. In *Ecological Rationality: Intelligence in the World*, ed. P. M. Todd, G. Gigerenzer and the ABC Research Group. New York: Oxford University Press, 216–40.

Brunswik, E. (1934). *Wahrnehmung und Gegenstandswelt: Grundlegung einer Psychologie vom Gegenstand her.* Leipzig: Deuticke.

Conlisk, J. (1996). Why Bounded Rationality? *Journal of Economic Literature*, 34, 669–700.

Czerlinski, J., Gigerenzer, G. and Goldstein, D. G. (1999). How Good are Simple Heuristics? In *Simple Heuristics that Make Us Smart*, ed. G. Gigerenzer, P. M. Todd, and the ABC Research Group. New York: Oxford University Press, 97–118.

DeMiguel V, Garlappi L, and Uppal R. (2009). Optimal versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies*, 22, 1915–53.

Friedman, D., Isaac, R. M., James, D. and Sunder, S. (2014). *Risky Curves. On the Empirical Failure of Expected Utility.* New York: Routledge.

Friedman, M. (1953). *Essays in Positive Economics*. Chicago: University of Chicago Press.

Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, 4, 1–58.

Gigerenzer, G. (1996). On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky. *Psychological Review*, 103, 592–6.

Gigerenzer, G. (2004). Striking a Blow for Sanity in Theories of Rationality. In *Models of a Man: Essays in Memory of Herbert A. Simon*, ed. M. Augier and J. G. March. Cambridge, MA: MIT Press, 389–409.

Gigerenzer, G., and Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science*, 1, 107–43.

Gigerenzer, G., Fiedler, K. and Olsson, H. (2012). Rethinking Cognitive Biases as Environmental Consequences. In *Ecological Rationality: Intelligence in the World*, ed. P. M. Todd, G. Gigerenzer and the ABC Research Group. New York: Oxford University Press, 80–110.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, 62, 451–82.

Gigerenzer, G. and Goldstein, D. G. (1996). Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review, 103*, 650–69. doi: 10.1037/0033-295X.103.4.650

Gigerenzer, G., Hertwig, R. and Pachur, T. (eds) (2011). *Heuristics: The Foundations of Adaptive Behavior.* New York: Oxford University Press.

Gigerenzer, G., and Selten, R. (eds). (2001). *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press.

Gigerenzer, G., Todd, P. M. and the ABC Research Group. (1999). *Simple Heuristics that Make Us Smart*. New York: Oxford University Press.

Hertwig, R., Hoffrage, U. and the ABC Research Group. (2013). *Simple Heuristics in a Social World*. New York: Oxford University Press.

Kahneman, D. (2003). Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice. In *Les Prix Nobel: The Nobel Prizes 2002*, ed. T. Frangsmyr. Stockholm: Nobel Foundation, 1449–89.

Kahneman, D. (2011). *Thinking Fast and Slow*. London: Allen Lane.

Kahneman, D., Slovic, P. and Tversky, A. (eds). (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Katsikopoulos, K. V. (2011). Psychological Heuristics for Making Inferences: Definition, Performance, and the Emerging Theory and Practice. *Decision Analysis*, 8, 10–29.

Katsikopoulos, K. V. and Gigerenzer, G. (2008). One-Reason Decision-making: Modeling Violations of Expected Utility Theory. *Journal of Risk and Uncertainty*, 37, 35–56.

Knight, F. (1921). *Risk, Uncertainty and Profit* (Vol. XXXI). Boston: Houghton Mifflin.

Lopes, L. L. (1992). Three Misleading Assumptions in the Customary Rhetoric of the Bias Literature. *Theory & Psychology*, 2, 231–6. doi: 10.1177/0959354392022010

Luan, S., Schooler, L. J. and Gigerenzer, G. (2014). From Perception to Preference and on to Inference: An Approach-Avoidance Analysis of Thresholds. *Psychological Review*, 121(3), 501–25. doi: 10.1037/a0037025.

Martignon, L. and Hoffrage, U. (2002). Fast, Frugal, and Fit: Lexicographic Heuristics for Paired Comparison. *Theory and Decision*, 52, 29–71. doi: 10.1023/A:1015516217425

Pachur, T. and Marinello, G. (2013). Expert Intuitions: How to Model the Decision Strategies of Airport Customs Officers? *Acta Psychologica*, 144, 97–103.

Parducci, A. (1965). Category Judgment: A Range-Frequency Model. *Psychological Review*, 72, 407–18.

Savage, L. J. (1972). *The Foundations of Statistics* (2nd edn). New York: Wiley.

Schmittlein, D. C., and Peterson, R. A. (1994). Customer Base Analysis: An Industrial Purchase Process Application. *Marketing Science*, 13, 41–67. doi: 10.1287/mksc.13.1.41

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics,* 69, 99–118. doi: 10.2307/1884852

Simon, H. A. (1978). Rationality as Process and as Product of Thought. *American Economic Review*, 68, 1–16.

Simon, H. A. (1979). *Models of Thought.* New Haven: Yale University Press.

Simon, H. A. (1981). *The Sciences of the Artificial* (2nd edn). Cambridge, MA: MIT Press.

Simon, H. A. (1985). Human Nature in Politics: The Dialogue of Psychology with Political Science. *American Political Science Review*, 79, 293–304.

Simon, H. A. (1989). The Scientist as Problem Solver. In *Complex Information Processing: The Impact of Herbert A. Simon*, ed. D. Klahr and K. Kotovsky. Hillsdale, NJ: Elbaum, 375–98.

Simon, H. A. (1990). Invariants of Human Behavior. *Annual Review of Psychology*, 41, 1–19. doi: 10.1146/annurev.ps.41.020190.000245

Simon, H. A. (1997). *Models of Bounded Rationality, Volume 3: Empirically Grounded Economic Reason*. Cambridge, MA: MIT Press.

Şimşek, Ö. (2013). Linear Decision Rule as Aspiration for Simple Decision Heuristics. In *Advances in Neural Information Processing Systems: Vol. 26: 27th Annual Conference on Neural Information Processing Systems 2013 [online version]*, ed. C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K.Q. Weinberger. Red Hook, NY: Curran Associates, 2904–12.

Şimşek, Ö. (2014). How Natural Environments Support Simple Decision Heuristics. Talk presented at the ABC Workshop, Max Planck Institute for Human Development, Berlin, October.

Stewart, N., Reimers, S. and Harris, A. J. L. (2014). On the Origin of Utility, Weighting, and Discounting Functions: How They Get their Shapes and How to Change their Shapes. *Management Science, 61*, 687–705. doi: 10.1287/mnsc.2013.1853

Stigler, G. J. (1961). The Economics of Information. *Journal of Political Economy*, 69, 213–25. doi: 10.1086/258464

Todd, P. M., Gigerenzer, G. and the ABC Research Group. (2012). *Ecological Rationality: Intelligence in the World*. New York: Oxford University Press.

Tsotsos, J. (1991). Computational Resources do Constrain Behavior. *Behavioral Brain Sciences,* 14, 506–7.

Verhoef, P. C, Spring, P. N., Hoekstra, J. C. and Leeflang, P. S. H. (2002). The Commercial Use of Segmentation and Predictive Modeling Techniques for Database Marketing in the Netherlands. *Decision Support Systems*, 34, 471–81.

Wübben, M. and von Wangenheim, F. (2008). Instant Customer Base Analysis: Managerial Heuristics Often 'Get It Right.' *Journal of Marketing*, 72, 82–93.

4

From *The Sciences of the Artificial* to Cognitive History

Subrata Dasgupta

Introduction

It is well known that Herbert Simon was a twentieth-century scientific polymath who made seminal contributions to the social sciences, behavioral sciences, design theory, computer science, and the philosophy of science (Dasgupta, 2003a, 2003b). My interest in Simon in this essay, however, lies in his remarkable and highly original book, *The Sciences of the Artificial* (Simon, 1996). In this work (henceforth referred to as *Sciences*), which in a sense unifies his multidisciplinary contributions, Simon dwells on the concept and nature of the human-made or *artificial* world, the things that populate it – *artifacts* – and in what sense and how the making of the artificial yields to scientific investigation.

In this chapter I wish to explore a particular *consequence* of the ideas put forth in *Sciences*. I wish to show how some of the key concepts advanced in it affords a conceptual framework for a (relatively) new historical discipline for the study of human creativity, the creative tradition, and the intellectual tradition. This discipline is called *cognitive history.*

The Making of the Artificial

To understand the connection between the concept of the artificial *à la* Simon and human creativity, we must first recognize that creativity refers to the intellectual–cognitive–sociocultural processes whereby *new* and *consequential* things are brought into existence (Dasgupta, 2011). The products of the creative process are thus human-made, thus *artifacts,* which are in some sense original and consequential and sometimes even world view changing. Artifacts as I use the term can be not only material in form (the conventional interpretation of the word), but also

abstract or symbolic (such as ideas, concepts, mathematical theorems, algorithms, plans, designs, and so on). The creative being is thus fundamentally immersed in the artificial world. He or she is in the business of making the artificial: an *artificer*.

Simon's main concerns in *Sciences* were with the characteristics of the artificial and how it is distinct from the natural. The problem that particularly intrigued him was to 'show how empirical propositions can be made at all' about artificial phenomena whose fundamental features are so distinct from natural phenomena (Simon 1996, p. xi). Herein lay the *sciences* of the artificial for, to Simon, making empirical – thus testable and potentially falsifiable – propositions about things in the world and subjecting them to experimental or observational tests is the cornerstone of empirical science (Newell & Simon, 1976; Simon, 1995). As he put it in 2000, 'I consistently maintain that science is concerned with describing and explaining how the world is'[1] – in other words, how the artificial could be subjected to empirical inquiry was Simon's concern in *Sciences*.

For creativity researchers and historians of the creative tradition (in science, art, literature, engineering, music, philosophy, etc.) the domain of interest is also the making of the artificial – broadly speaking, the structure of means and ends whereby, and the contexts wherein, original, valuable (in some sense), and consequential artifacts are brought into being by their human artificers. The concept of the artificial is, thus, ubiquitous in creativity studies.

The 'Oughtness' of the Artificial

One of Simon's major insights into artificial phenomena and how they differ from the natural is his distinction between *is* and *ought* – a distinction that goes back to his doctoral dissertation on administrative decision making in the early 1940s (Simon, 1976, pp. 45–60), and which he attributed to his reading of A. J. Ayer's *Language, Truth and Logic* (1936).[2]

The natural world just *is*. Natural phenomena are not endowed with purpose, goals, or needs. Natural scientists *qua* natural scientists belong to *is-culture*. They do not pose questions of what a natural phenomenon is *for*? Contrast this with the world of artifacts. As Simon pointed out, the artificial makes no sense without purpose (Simon, 1996, p. 5). Bridges, alloys, algorithms, machines, blast furnaces, a country's constitution, administrative organizations – all artifacts, some material, some abstract – are what they are because their creators *wanted* them to be

as they are. Artifacts are endowed with their makers' – their artificers' – purpose because they believed that the artifacts *ought to be* of a certain kind. There is *agency* at work in the artificial world.

Thus, artificers and their artifacts belong to *ought-culture*. Because of this there is an 'air of contingency' about the artificial (Simon, 1996, p. xi). An artifact *could* have been something other than what it is if the artificer's choices and decisions had been different. There is nothing *necessary* about the Tower Bridge in London, or the Notre Dame Cathedral in Paris, or Picasso's *Les Demoiselles d'Avignon* hanging in New York, or a nation's constitution. Each could have been otherwise. This oughtness, then, must be part of the issue for those whose business is not making the artificial but making *sense* of the making of the artificial.

Importing Is-Culture into Ought-Culture

Simon's *Sciences* actually concerned itself with the question: How can we *import* is-culture into ought-culture? For example, it is true that a machine is the outcome of a designer's goals and purposes and, thus, choices. It belongs to ought-culture. But *given* the known purposes or goals, and *given* knowledge about the structure and behavior of the machine, what can we say about the *relationship* between goals/purposes and the machine? In this situation, the artifact-purpose complex can be treated *as if* it is a natural object; as something that is just *is*, to be investigated as one would a natural phenomenon. The sciences of the artificial then concern the question of how the scientist of the artificial (who is not necessarily the artificer) can make sense of this relationship between goals and artifacts, between ends and means. The problem of ought is transposed into a problem of is. Thus, if the artificer's oughts implied subjective values, these values are then treated in the sciences of the artificial as objective entities. The scientist of the artificial does not pass judgment on the values – on the oughts. She must not allow her own values to intrude into the inquiry. The relevant part of the artificial world becomes, for the purpose of the science, analogous to the natural world – the world as if it just is. As Simon insisted:

> 'How the world is' – how it really is – is independent of the . . . observer, although s/he must hold certain values in order to do effective science . . . Among the values s/he must have is the value of not permitting his or her own values to influence the interpretation of the empirical evidence of how the world is.[3]

The Creative Phenomenon and Its Historicity

Consider now creativity. As noted, a creative being – artist, scientist, designer, engineer, inventor, composer, writer, etc. – lives and works in the artificial world. She is an artificer. Creativity entails making artifacts that are original and consequential. In particular, consider the following scenarios:

1. There is a specific, created artifact (material, as in the case of machines, buildings, paintings, sculpture; or symbolic and abstract, as in the case of mathematical theorems and proofs, algorithms, plans, designs, methods and procedures, concepts and ideas). Usually the creator is an individual but could be a cluster of individuals (e.g., a modern piece of biological research may entail two dozen, or even more, biologists working collaboratively). Assume that the artifact is significantly original in some important sense.[4] We are interested in the nature of the singular *act of creation* that produced the artifact.
2. There is a specific *individual*: a painter, sculptor, scientist, writer, poet, playwright, engineer, composer, inventor, philosopher, or social thinker. We wish to determine the nature of this person's creative life, as characterized, for example, by his *cognitive style* (Dasgupta, 2005).
3. There is a specific creative (artistic, literary, scientific, technological, social, philosophical) *movement* across a space-time frame.[5]

The individual act of creation, the individual's cognitive style over the course of a creative life, and the individual creative movement signify three different 'levels' of creativity. Collectively let me call them *creative phenomena*.

All creative phenomena share an important trait: *historicity*. To understand and explain a phenomenon, the scenario is of one or more of the following sort:

A. There is a creative phenomenon.
B. The 'facts' pertaining to it are available, to a greater or lesser extent, to the creativity researcher in some archival fashion.
C. The researcher then attempts to construct a *theory* that explains the phenomenon.

This explanatory theory will be a theory of something that belongs to the past – immediate or distant. The phenomenon *already* belongs to history. It must therefore be examined as an historical event.[6] This is the first aspect of the historicity of creativity.

The second aspect of historicity lies in that the explanatory theory itself will be an historical explanation. (As R. G. Collingwood stated, the past survives in the present [Collingwood, 1946].) That is, in order for the researcher to offer an explanation of the creative phenomenon (no matter how 'current' it is) she must reach back to the past: in part to the creative being's personal history (the biographical past); and in part to the social, cultural, and epistemic states of the world in the space-time frame in which the phenomenon resided, these states of the world being themselves historically shaped. The creativity researcher must enter into the archives; or she must enter into a biographical engagement with the creator in the context of individual acts of creation or of a creative life; or she must enter into a socially, culturally, epistemically informed historical engagement with the creative movement.

The third aspect of the historicity of creativity is a consequence of the first two. Because an explanation will be an historical explanation, this will necessarily take the form of a *narrative* – a story (of course supported by evidence) which connects elements of the phenomenon in a causally ordered form (Bruner, 1990). In offering a theory the researcher will be saying: 'This is how it began, then this happened because of such and such, then this' and so on. The narrative, an unfolding over time, *is* the theory.

Thus, Cognitive History

The historicity of creativity thus makes the creativity researcher into a historian. But a historian of a particular sort. If the aim of history (including historical biography) is to uncover aspects of the past that are hidden, unknown, or unnoticed (Collingwood, 1946; Evans, 2000), then the task of the creativity researcher *qua* historian (or vice versa) is to render visible, known, and noticeable the *cognitive* characteristics of creative phenomena. The creativity researcher becomes a *cognitive historian*.

To the best of my knowledge the term cognitive history is due to Nancy Nersessian (Nersessian, 1995), which, she proposed, 'joins historical inquiry with those carried out in the sciences of cognition in order to explain the "thinking practices" whereby "scientists create, change and communicate their representations of nature"' (ibid., p. 194). Her vision of cognitive history was thus basically confined to scientific creativity.

But the scope of cognitive history can be expanded to embrace the whole of the *creative tradition* – which actually reaches back to the age of *Homo habilis* some two and a half million years ago, with their invention of primitive stone tools. It should embrace not just science but art, technology, engineering, design, literary production, music, scholarship, and so on. More generally then, we may propose that:

> *Cognitive history is a symbiosis of the methods and tools of historical and biographical investigation and the theories, models and methods of cognitive science. Its aim is to understand and explain actual creative phenomena taken from the history of the creative tradition.*

The cognitive historian asks certain types of questions pertaining to creative phenomena as they were encountered in history: What kinds of *knowledge* and *beliefs* did an individual bring to bear on her creative work? What *goals* were established in the course of that work? How did these goals *originate*? What kinds of knowledge were *generated* and *how*? What forms of *reasoning* are evident in a person's creative life and work? Can we elicit a sense of his/her *cognitive style*? The cognitive historian attempts to provide a theory or a narrative as an explanation that is shaped and influenced by cognitive theories and models. Cognitive history *does not ignore* the social and the cultural. But it does stand apart from cultural, social, and intellectual histories in its emphasis on the cognitive.[7]

Nersessian introduced the term cognitive history in 1995 but before her the marriage of cognitive science and history as a pathway to 'explaining science' was briefly touched upon by philosopher of science Ronald Giere (Giere, 1988, pp. 18–19). To my knowledge the first *practitioner* of cognitive history (without using the term) was cognitive psychologist Howard Gruber in his study (published in 1974) of Darwin's ideation of the principle of natural selection (Gruber 1981). Since then cognitive history has been practiced implicitly or explicitly by a range of creativity researchers, not only in the natural sciences (Langley *et al.,* 1987; Kulkarni and Simon, 1988; Miller, 1986; Gooding, 1992; Tweney, 1992; Thagard and Croft 1999; Dasgupta, 1999) but also in the artificial sciences, technology, and engineering design (Carlson and Gorman, 1992; Dasgupta, 1994a, 1994b), art (Dasgupta, 2005), literary composition (Wallace, 1989; Jeffrey, 1989; Ippolito and Tweney, 2000), and art history scholarship (Kozbelt, 2008).

Simon's Model of the Artificial

Cognitive science is a congeries of a half-dozen (if not more) 'root disciplines',[8] so we should not be surprised that the cognitive–historical approach is as many-chambered as is cognitive science itself. Perhaps it is more accurate to speak of cognitive *histories* just as some speak of cognitive *sciences*.

Thus, Simon's *Sciences* is but one of several theoretical influences on cognitive history. (For a somewhat different framework, see Gruber, 1989.) However, its power and immediacy as a shaper of cognitive history lie in two key aspects: (a) its domain is, quite explicitly, the artificial and the making of the artificial. In this explicit recognition of the artificial Simon and his *Sciences* stand apart from other influences; (b) the *sciences* of the artificial *à la* Simon are, fundamentally, cognitive in nature. Grounded in these foundational concepts Simon constructed a *model of the artificial* that became the object of his sciences. In *very* brief (for lack of space) this model can be described as follows:

One. Artifacts represent their artificer's beliefs about how the world *ought to be*. They embody the goals and desires of their artificers.

Two. An artifact serves as an *interface* between an *outer environment* (wherein the artifact must reside and with which it must interact) and an *inner environment* (meaning, the 'substance and organization of the artifact itself' [Simon, 1996, p. 6]).

Three. The complexity of an artifact is a reflection of the complexity of the outer environment.

Four. If there is a match between outer and inner environments the artifact meets its artificer's intended goal and is said to be *adapted* to its environment.

Five. Because of the artificer's imperfect or incomplete knowledge about the outer environment, or limits to his cognitive processing capacity, the artificer is constrained by limits to his ability to make optimal decisions. That is, the artificer suffers from *bounded rationality*.[9]

Six. Consequently, the artificer makes decisions that are *satisficing* (good enough) rather than optimizing (the best) (Simon, 1987b).

Seven. Consequently, the artificer seeks solutions to their problems by *searching* through the 'problem space,' employing *heuristic* strategies to do so.

A Simonian Pentimento for Cognitive History

This model of the artificial is by no means sufficient for cognitive–historical explanations of creative phenomena. In particular, Simon does not deal with the historicity issue; nor with the conditions that determine why and how the artificer is deemed creative; nor with the crucial role, in creative phenomena, of the consumer(s) of the artifact; nor with the vexing problem that much (perhaps most) of cognition occurs in the unconscious; nor with the social and cultural aspect of the outer environment; nor with the role of emotion in creativity; or the place of aesthetics in creative acts and creative lives.

But Simon's model of the artificial offers a powerful foundation for an approach to cognitive history. Simon himself, along with his collaborators, has conducted cognitive–historical studies in the sciences that draw upon his model of the artificial (Langley *et al*., 1987; Kulkarni and Simon, 1988). Elsewhere, I have offered a narrative framework for creativity that begins with Simon's model of the artificial, but enriches it so as to accommodate the above-mentioned issues. Very briefly, this *metanarrative*[10] envisions all creative phenomena as comprising of one or more *creative encounters* between a producer (artificer) and one or more consumers *via* the produced artifact. The consumers and their shared culture constitute the producer's 'outer environment'. The producer's (and, in fact, the consumer's) 'inner environment' comprises a complex of *beliefs/knowledge, need/goal,* and *emotion spaces* that interact with one another by way of *cognitive processes*, giving rise to the artifact. However, because so much of a cognitive process occurs in the unconscious – not the Freudian variety but what is called the cognitive unconscious (Kihlstrom, 1987) – what is more discernible is the producer's *cognitive style,* that is, one or more identifiable patterns or regularities underpinning the goals, knowledge, perception, and/or reasoning she brings to bear in the course of cognitive processes (Dasgupta, 2003a, p. 686). (This concept of cognitive style finds a trace in Simon's brief discussion of 'style in design' in *Sciences*, and in much greater detail in a later paper [Simon, 1975].) Likewise, the consumer's cognitive process gives rise to a *response* to the artifact. A creative encounter is effected between producer/artificer and consumer, to the extent that the consumer *identifies* with the producer.[11]

Like a pentimento, Herbert Simon's model of the artificial (even his view of style) is clearly visible beneath this cognitive–historical metanarrative.[12]

Notes

1. Simon, H. A. to S. Dasgupta, personal communication, email May 26, 2000.
2. Simon, H. A. to S. Dasgupta, personal communication, email, Dec. 1, 1999.
3. Simon, H. A. to S. Dasgupta, personal communication, email, May 26, 2000.
4. The issue of *originality* is complicated. Unfortunately space does not allow me to dwell on it here. For discussions see Boden 1991, Dasgupta 1996, Dasgupta 2011.
5. Examples of creative movements I happen to have studied are: (a) a nineteenth-century intellectual, literary, and social awakening which occurred in British India in the eastern region of Bengal. This movement came to be called The Bengal Renaissance. See Dasgupta 2007, Dasgupta 2011; (b) the techno-scientific movement spanning the middle third of the twentieth century that gave birth to computer science. See Dasgupta 2014.
6. This is why *real* creativity, as it has occurred in the course of history, cannot be adequately studied in a laboratory setting.
7. For a contemporary interpretation of what cultural history, social history, and intellectual history are 'about', see the relevant essays in Cannadine 2002.
8. See, e.g., the papers commemorating the thirtieth anniversary of the Cognitive Science Society Conference in *TopiCS, Topics in Cognitive Science*, 2, 3, July 2010.
9. For a compact description of this concept (the invention of which led to his Nobel Prize) see Simon 1987a.
10. I am, of course, aware that by proposing a 'metanarrative' I am condemned forever by postmodernists who avow 'incredulity towards metanarrative'! See Lyotard, 1984.
11. This metanarrative was originally described in Dasgupta, 2007: 7–20. A much more comprehensive account is presented in Dasgupta, forthcoming.
12. I thank Wenceslao Gonzales for his comments on an earlier version of this essay.

References

Boden, M. (1991). *The Creative Mind*. New York: Basic Books.

Bruner, J. L. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.

Cannadine, D. (ed.) (2002). *What is History Now?* Basingstoke: Palgrave Macmillan.

Carlson, W. B. and Gorman, M. E. (1992). A Cognitive Framework to Understanding Technological Creativity. In *Inventive Minds: Creativity in Technology*, ed. R. W. Weber and D. N. Perkins. New York: Oxford University Press, 48–79.

Collingwood, R. G. (1946). *The Idea of History*. Oxford: Clarendon Press.

Dasgupta, S. (1994a). *Creativity in Invention and Design*. New York: Cambridge University Press.

Dasgupta, S. (1994b). Testing the Hypothesis Law of Design: The Case of the Britannia Bridge. *Research in Engineering Design*, 6, 1, 38–57.

Dasgupta, S. (1996). *Technology and Creativity*. New York: Oxford University Press.
Dasgupta, S. (1999). *Jagadis Chandra Bose and the Indian Response to Western Science*. New Delhi: Oxford University Press.
Dasgupta, S. (2003a). Multidisciplinary Creativity: The Case of Herbert A. Simon. *Cognitive Science*, 27, 683–707.
Dasgupta, S. (2003b). Innovation in the Social Sciences: Herbert A. Simon and the Birth of a Research Tradition. In *The International Handbook on Innovation*, ed. L. V. Shavinina. Amsterdam: Elsevier Science, 458–70.
Dasgupta, S. (2005). Cognitive Style in Creative Work: The Case of the Painter George Rodrigue. *PsyArt*. http://www.clas.ufl.edu/ipsa/journal/2005_dasgupta01.shtml
Dasgupta, S. (2007). *The Bengal Renaissance: Identity and Creativity from Rammohun Roy to Rabindranath Tagore*. New Delhi: Permanent Black.
Dasgupta, S. (2011). Contesting (Simonton's) Blind Variation, Selective Retention Theory of Creativity. *Creativity Research Journal*, 32, 166–82.
Dasgupta, S. (2014). *It Began with Babbage: The Genesis of Computer Science*. New York: Oxford University Press.
Eatwell, J., Milgate, M. and Newman, P. (ed.) (1987). *The New Palgrave: A Dictionary of Economics*. London: Macmillan.
Evans, R. J. (2000). *In Defence of History*. London: Granta.
Giere, R. (1988). *Explaining Science*. Chicago: University of Chicago Press.
Gooding, D. C. (1992). Putting Agency Back into Experiment. In Pickering, A. (ed.), *Science as Practice and Culture*. Chicago: University of Chicago Press, 65–112.
Gruber, H. E. (1981). *Darwin on Man*, 2nd edn. Chicago: University of Chicago Press.
Gruber, H. E. (1989). The Evolving Systems Approach to Creative Work. In *Creative People at Work* (1989), ed. D. B. Wallace and H. E. Gruber. New York: Oxford University Press, 3–29.
Ippolito, M. F. and Tweney, R. D. (2000). The Journey to Jacob's Room: The Network of Enterprise of Virginia Woolf's First Experimental Novel. *Creativity Research Journal*, 15, 1, 25–44.
Jeffrey, L. R. (1989). Writing and Rewriting Poetry: William Wordsworth. In *Creative People at Work* (1989), ed. D. B. Wallace and H. E. Gruber. New York: Oxford University Press, 69–90.
Kihlstrom, J. F. (1987). The Cognitive Unconscious. *Science*, 237, Sept 17, 1445–52.
Kozbelt, A. (2008). E. H. Gombrich on Creativity: A Cognitive Historical Case Study. *Creativity Research Journal*, 20, 1, 93–104.
Kulkarni, D. and Simon, H. A. (1988). The Processes of Scientific Discovery: The Strategy of Experimentation. *Cognitive Science*, 12, 139–76.
Langley, P., Simon, H. A., Bradshaw, G. L. and Zytkow, J. (1987). *Scientific Discovery*. Cambridge, MA: MIT Press.
Lyotard, J-F. (1984). *The Postmodern Condition: A Report on Knowledge*. (G. Bennington and B. Massuni, tr.). Minneapolis: University of Minnesota Press, xxiv.
Miller, A. I. (1986). *Imagery in Scientific Thought*. Cambridge, MA: MIT Press.
Nersessian, N. J. (1995). *Opening the Black Box: Cognitive Science and History of Science*. Osiris (2nd series), 10, 194–211.

Newell, A. and Simon, H. A. (1976). Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM*, 19, 3, 113–26.
Simon, H. A. (1975). Style in Design. In *Spatial Synthesis in Computer-Aided Building Design*, ed. C.M. Eastman. New York: John Wiley, 287–309.
Simon, H.A. (1976). *Administrative Behavior* (3rd edn). New York: The Free Press.
Simon, H. A. (1987a). Bounded Rationality. In *The New Palgrave: A Dictionary of Economics,*, ed. J. Eatwell, M. Milgate and P. Newman. London: Macmillan, 266–8.
Simon, H. A. (1987b). Satisficing. In *The New Palgrave: A Dictionary of Economics*, ed. J. Eatwell, M. Milgate and P. Newman. London: Macmillan, 243–5.
Simon, H. A. (1995). Artificial Intelligence: An Empirical Science. *Artificial Intelligence*, 77, 95–127.
Simon, H. A. (1996). *The Sciences of the Artificial* (3rd edn). Cambridge, MA: MIT Press.
Thagard, P. R. and Croft, D. (1999). Scientific Discovery and Technological Innovation: Ulcers, Dinosaur Extinction and the Programming Language Java. In *Model Based Reasoning in Scientific Discover*y, ed. L. Magnani, N. J. Nersessian and P. R. Thagard. New York: Kluwer, 125–37.
Tweney, R. D. (1989). Fields of Enterprise: On Michael Faraday's Thoughts. In *Creative People at Work*, ed. D. B. Wallace and H. E. Gruber. New York: Oxford University Press, 91–106.
Tweney, R. D. (1992). Inventing the Field: Michael Faraday and the Creative 'Engineering' of Electromagnetic Field Theory. In *Inventive Minds: Creativity in Technology*, ed. R. W. Weber and D. N. Perkins. New York: Oxford University Press, 31–47.
Wallace, D. B. (1989). Streams of Consciousness and Reconstruction of Self in Dorothy Richardson's Pilgrimage. In *Creative People at Work*, ed. D. B. Wallace and H. E. Gruber. New York: Oxford University Press, 147–70.
Wallace, D. B. and Gruber, H. E. (ed.) (1989). *Creative People at Work*. New York: Oxford University Press.
Weber, R. W. and Perkins, D. N. (ed.) (1992). *Inventive Minds: Creativity in Technology*. New York: Oxford University Press.

5
Rationality and the True Human Condition

Ron Sun

Introduction

The notion of rationality is important to many fields in social and behavioral sciences. Herbert Simon's seminal work on bounded rationality and satisficing led to broadened conceptions of rationality, which significantly impacted on the social and behavioral sciences. In this chapter, I would like to further explore the notion of rationality, on the basis of Simon's work.

First, in this regard, it may be necessary, I believe, to go beyond Simon's notions of bounded rationality and satisficing, for example, in dealing with limitations and variations of actual human rationality. Furthermore, I believe that it may be necessary to go beyond the notion of rationality as optimization of utility functions. I argue that we need to take into serious consideration the true human condition in this regard, that is, actual human nature (especially actual human psychology), in defining or understanding the notion of rationality.

In the 1950s, Simon proposed his theory of bounded rationality, which tried to reflect real human abilities to reason and make decisions, in relation to his work in economics and organization theory (Simon, 1957, 1991). Simon's notion of a limited kind of rationality may, in some way, have enabled the social sciences to move beyond the then prevailing theories (in economics and in other branches of social sciences). But the questions now are: does it go far enough in respecting human reality? What is the true human condition in this regard? Is the true human condition sufficiently captured by Simon's notion of rationality?

It is my belief that the true human condition (especially human psychology) has, unfortunately, not yet been sufficiently addressed in this line of work. This point applies to Simon's treatment of rationality, and

also to Simon's approach to studying cognition within the realm of cognitive science and artificial intelligence (which paralleled his work in economics and organization theory). In contrast, in this chapter, what I want to emphasize is exactly the true human condition that I believe has not been sufficiently examined in Simon's approach. I will do so based on the framework of a comprehensive computational theory of the human mind, that is, a computational cognitive architecture, taking into account some very human facets of human nature.

Below I first discuss what rationality means in various contexts (while questioning this very notion). Then I discuss some findings regarding the unconscious mind, which shows various tendencies that appear to contradict the notion of rationality. I then present a theoretical framework that addresses all of these facets in mechanistic and process-based ways (but not necessarily mathematically). The application of this framework in exploring issues in social, cultural, political, and organizational contexts are then briefly sketched. This discussion draws upon, for example, the ideas from Reber (1989) regarding the unconscious mind and Murray (1938) regarding basic needs or desires, among others, in addition to Simon's own ideas.

What is Rationality?

Many theories and models in the social sciences dealing with human behavior assume that human individuals can be reasonably described as "rational" beings. For instance, many economics models assume that individuals are (on average) rational – that is, they act to maximize the utility of their decisions, with utility calculated in accordance with their own preferences. It has been argued that desires cannot be measured directly, but they can be measured by the behavior to which they give rise. Utility is believed to represent desire or want. In economics, utility is measured by the price that an individual is willing to pay for the expected satisfaction of desire (Marshall, 1920). Many mathematical properties were attributed to the measurement of utility, for example, the properties of being transitive, being complete, and so on. Often, the utility of a decision is calculated based on a linear combination of its parts (von Neumann and Morgenstern, 1953).

As mentioned above, in the 1950s, Simon proposed his theory of bounded rationality (Simon, 1957), which refined some of these assumptions to account for the fact that perfectly rational decisions (e.g., optimal with regard to a utility function) are often not feasible in practice. Less than fully optimal decisions may be adopted

by individuals in many circumstances; hence the ideas of bounded rationality and satisficing. More specifically, in Simon's bounded rationality theory, rationality of individuals is limited by the information to which they have access, the cognitive limitations of their minds, the cost of gathering information, the cost of processing information, the finite amount of time during which they have to make a decision, and other relevant factors. Because individuals lack the ability and/or resources to arrive at optimal decisions, they may instead apply their rationality after having simplified the situation in question (e.g., the number of choices available). Thus, an individual may be a satisficer; one seeking a satisfactory, but possibly suboptimal, solution rather than the strictly optimal one (Simon, 1957). These ideas of Simon's tried to reflect real human abilities to reason and make decisions on the basis of the then prevailing theories in economics. They have enabled the social sciences to move beyond these classical theories in some way.

While Simon's work has been highly significant for several decades, for me it also brings up the very question of what being "rational" means for humans. There have been various usages of this term. They are often different, sometimes vastly different, in terms of meaning (either in terms of denotation or in terms of connotation). We need to examine these different meanings.

For instance, most commonly, being rational means optimality (or near-optimality) with regard to some criterion, for example, with regard to a well-defined numerical criterion, in the form of a well-behaved utility function with many nice mathematical properties (as in economics).

Alternatively, it may also mean following some kind of logic in thinking, that is, undertaking some form of logic reasoning (e.g., as assumed by some schools of philosophy). In this case, the emphasis is on the process, not on the outcome of a process. Other standards on process of rational thinking may be based on probability theory and statistics, instead of some standard forms of logic.

Beyond these meanings above, the term may also be used to indicate a variety of other human tendencies, for example, following some moral or religious principles (Weber, 1915). Furthermore, it sometimes may even mean following some emotional or affective forces (Weber, 1915).

Clearly, rationality could mean a lot of different things to different people. Given the diversity of the meanings of this term, at a minimum, an updated, better understanding of this notion is needed. In particular, we need to take into full consideration the reality of the human mind,

given the advances in understanding the human mind in recent decades since Simon proposed his ideas regarding rationality. The human mind determines human behavior (at least as a proximal cause).

Simon did try to take into account the reality of the human mind (human psychology) as much as he could at the time. But he may not have gone far enough in terms of taking into full consideration its complexity and subtlety in his (otherwise seminal) work. The major ideas that Simon advocated, of bounded rationality and satisficing, seem to be just variations of the same old theme; that is, they appear still inextricably tied to the optimization or near-optimization of an objective criterion (e.g., a utility function).

Simon described ways in which rationality might be made more realistic cognitively: limiting types of utility functions; taking into account costs of gathering and processing information; and so on. Simon further suggested that individuals might use heuristics to make decisions rather than making decisions only based on strict optimization. There have also been further suggestions of the differences between economic rationality and ecological rationality (Gigerenzer and Selten, 2002). These theories may need to be developed more, but many developments since have helped to fill some gaps.

However, as I see it, one fundamental shortcoming remains, which is that the bounded rationality theory (as well as many other related theories since) was still founded on the basis of optimization or near-optimization of a criterion function, objectively defined and well behaved. Simon may not have taken into full consideration the complexity and subtlety of the human mind.

In a related vein, the research program in symbolicist AI, based on a similar rationalist approach and as advocated by Simon himself early on, has stumbled and has since met with serious criticism (e.g., by Dreyfus and Dreyfus, 1987, and many others). Many have found that it is hard to capture human intuition in a symbolic, rationalist way (Dreyfus and Dreyfus, 1987). Symbolic AI and the rationalist approach that it embodies have also suffered from a number of other shortcomings (see, e.g., Sun, 1994 and Haugeland, 1985 for further details).

This brings up the following pertinent question: How 'rational' should humans be along this line, that is, in terms of optimality (even just a limited kind of optimality)?[1] The costs of being 'rational' – time, memory, and other resources – certainly need to be taken into consideration (as already discussed by Simon, 1957, but also see Dayan, 2014 and Trimmer and Houston, 2014). But the highly useful role of

intuition in the human mind also needs to be understood and appreciated, as has been discussed, for example, by Dreyfus and Dreyfus (1987), Sun (1994), Helie and Sun (2010), and Kahneman (2011). Moreover, the indispensable role of affect/emotion in the human mind has also been argued by many, including Damasio (1994), LeDoux (1996), and Morris and Keltner (2000); see also Simon (1967). A similar point may also be made about the role of aesthetic judgment in the human mind. Additionally, all other aspects of the unconscious mind need to be taken into full consideration too. For instance, it has been shown that goals triggered by situational cues may pursue their own fulfillment without conscious intervention (Huang and Bargh, 2014). Traditions, customs/habits, moral/religious beliefs, and so on can also play a significant role in human behavior and in human decision making (Weber, 1915). Some of these aspects may treat process as being more important than outcome.

Some may argue against the points above. For instance, it may be argued that some aspects, such as emotion, have evolved throughout human evolutionary history to respond optimally or near-optimally to various commonly encountered, existentially significant situations, and thus the notion of 'rationality as optimization' is applicable to these aspects (such as emotion) to a large extent.

In this regard, it is important to point out a number of limitations to such an argument. First, evolution does not always lead to optimal or near-optimal solutions (Trimmer and Houston, 2014); for example, it is possible to get stuck in local maxima. Second, situational characteristics change from historical contexts during which evolution occurred to present-day contexts. So, even if optimal solutions were found by evolution in the past, present-day situations may render them irrelevant (either in terms of mechanisms or in terms of parameters of mechanisms). Third, long-run optimality may or may not translate into situational optimality (Trimmer and Houston, 2014). Thus, using optimality as the framework to analyze human behavior may be post hoc and haphazard, and can often be misleading (see also Dayan, 2014 for other related issues).

Similar arguments may be made with regard to the optimality of human intuition resulting from evolution. Similar counter-arguments as the above may also be made with regard to the lack of direct links between intuition and quantifiable optimality (even when evolution is taken into consideration). Furthermore, if 'rationality' is emphasized over intuition, the contribution of human intuition may be reduced or diminished, and human intelligence suffers as a result (Dreyfus and

Dreyfus, 1987; Helie and Sun, 2010). This may have serious societal consequences. One possible societal consequence may be the diminished ability, or even inability, to utilize often powerful human intuition in social, cultural, political, and organizational processes (Sun and Naveh, 2004).

Even when something can be formulated as a utility function that can be optimized, there are many more issues that need to be addressed. If one needs to define a utility function that is capable of being optimized (or capable of being optimized to some extent at least), one has to consider many trade-offs, for example: how reward and punishment are weighted with respect to each other; how much long-term or short-term payoffs are emphasized or discounted; how risks are tolerated with respect to possible payoffs; and so on. Of course, in a post hoc way, there are too many possibilities of constructing a utility function to be optimized, which, however, may not shed much light on a deep understanding of either the psychological or the sociocultural issues involved.

So, what does it mean to be rational after all? Is it about optimizing a however defined, inevitably complex function (due to taking into account all important psychological facets discussed thus far), with many assumptions (as pointed out above), which may be intractable after all, in order to guide actions? Or is it about defining some post hoc function to explain decisions and actions, often on a case-by-case basis? Either way it does not seem too appealing an intellectual project to undertake, individually or collectively, either to guide actions or to explain actions after the fact.

Some may argue that rationality does not necessarily mean optimization of a simple, well-behaved utility function. This is certainly true. However, in reality, an emphasis on rationality often leads to attempts to construct relatively simple (and sometimes simplistic) utility functions (without addressing all important psychological facets) and to optimize them. Furthermore, if utility functions are put aside, there is not much to the notion of rationality any more. To put it another way, there may be too many different meanings associated with the notion (as enumerated earlier), but their intersection may be nearly empty. This is an unfortunate paradox.

At a higher level, historically speaking, it may also be argued that rationality is a sociocultural construction that, as a normative notion, helps to sustain particular socioeconomic paradigms (Foucault, 1977). Modernity is fundamentally about order – about 'rationality' (e.g., objective functions) and rationalization that create order out of chaos. The

assumption has been that creating more 'rationality' (e.g., objective, optimizable functions) is conducive to creating more order, and that the more ordered a society is, the better it will function. Rationality may fundamentally lie in a conception of how a society should function and its power relationships. In this regard, see also the critiques by Foley (1998) and Ellul (1964), in terms of creating a system that subordinates the natural world.

So, taking all of these points into consideration, it does not seem 'rational' to include that many disparate things under the umbrella of one term – rationality. Furthermore, I suspect that, after taking all these points into full consideration, perhaps some kind of mixture may become appealing or even desirable – some combination of 'rationality' and 'irrationality'. Such combinations may be important, not only to understanding the human mind, but also in considering social, cultural, political, and organizational implications of the particulars of the human mind (see, e.g., Sun, 2012).

Max Weber, to his credit, defined a long time ago four different types of rationality (Weber, 1915). The first type, instrumental rationality, is related to attaining particular ends, 'rationally pursued and calculated', based on expectations of the behavior of other human beings or objects in the environment. This notion is the closest to the current common conception of rationality. The second type is value/belief-oriented, where action is undertaken for some ethical, aesthetic, religious or other motive, independent of whether it will lead to success. The third type is affective, determined by an individual's specific affect, feeling, or emotion, on the borderline of what he considered meaningfully oriented. The fourth type is traditional or conventional, determined by habituation. Weber emphasized that it was unusual to find only one of these types, and combinations were common (Weber, 1915).

Given these diverse meanings of the term (as identified by Weber and as the discussion so far shows), I would suggest that maybe it is finally time to put away the notion of 'rationality', which is either somewhat simplistic (e.g., as commonly defined in economics) or overloaded (e.g., as multiply defined by Weber and others, discussed above).[2] Instead, a more nuanced perspective of how the human mind works should be developed, and various particulars of the human mind (human psychology) should be more comprehensively explored. On that basis, some better, more precise notions may emerge. To that end, I discuss below some important aspects of the human mind, especially the unconscious mind, and then I describe a computational cognitive architecture that embraces these aspects. These aspects, I should emphasize,

are important contributors to seeming 'irrationality', which need to be taken into consideration in understanding the human mind and its social implications (Sun, 2012).

Unconscious Mind and Rationality

It has been well argued that there is a distinction between implicit and explicit processes (or between intuitive and reflective processes) in the human mind (Reber, 1989; Sun, 2002; Evans and Frankish, 2009). This distinction may be one of the most important distinctions with regard to the human mind. There are many two-system views, or dual-process theories, endorsing this distinction currently available. They may have significant implications for understanding rationality, because implicit processes are known to show many 'irrational' characteristics (Sun, 2002, 2015; Evans and Frankish, 2009; Reber, 1989).[3]

One such two-system view was proposed early on in Sun (1994). In the dual-process account of Sun (1994), there are two levels (i.e., two types of processes of the mind): the implicit (unconscious) versus the explicit (conscious). The two levels encode somewhat similar or overlapping content. But they encode their content in different ways, with symbolic and subsymbolic representation used, respectively, at the two levels. Symbolic representation is used by explicit processes at one level, and subsymbolic (connectionist, distributed) representation is used by implicit processes at the other (Sun, 1994, 2002). One type of representation is computationally explicit and the other computationally implicit, due to the intrinsic computational properties of these two types of representation (see Sun, 1994, 2002 for details). Different mechanisms are therefore involved at these two levels due to the difference in representation. It was further argued in Sun (1994) that these two different levels could work together synergistically, complementing and supplementing each other (through both 'rationality' and 'irrationality'). This may be, at least in part, the reason why evolution led to these two separate levels.

Currently, various dual-process theories (two-system views) seem to have captured popular imagination (e.g., Kahneman, 2011). However, although the distinction is important, the terms that have been used to characterize it (e.g., intuitive thinking and reflective thinking) are often loaded and ambiguous. To develop a more fine-grained and comprehensive understanding, it is important to avoid conceptual and terminological ambiguity and strive for theories that are more concrete and precise.

To this end, I have presented a more nuanced framework on these two types of processes and their interaction (see Sun, 2002, 2015; Sun and Wilson, 2014). Based on this framework, I have argued that we need to treat implicit processes (with their 'irrational' characteristics) as an integral part of human thinking, reasoning, and decision making, not as an add-on or an auxiliary (see Sun, 1994, 2002). We also need to explore implicit processes in a variety of domains, and their interaction with explicit processes (e.g., Sun *et al.*, 2005). It is important to emphasize implicit processes, given that there has already been an overwhelming amount of research on explicit processes. Along this line, a computational cognitive architecture embodying the theoretical framework has been developed that addresses, in a mechanistic, process-based sense, various issues relevant to this distinction and beyond (Sun, 2002, 2015). (I will discuss the cognitive architecture and related issues in the next section.)

Moreover, this framework recognizes that, beyond implicit cognition (which shows some 'irrational' characteristics), social motivation, emotion, moral instinct, and so on (which show even more 'irrational' characteristics) play central roles in defining human nature, and thus human rationality (and irrationality). These aspects of the mind are mostly implicit, operating beneath conscious awareness (Sun, 2015). But they are at the core of human psychology, and thus they constitute the very essence of human nature.

For instance, Murray (1938) described various basic motives (needs or desires) of humans, more or less universal across cultures. Subsequent work has demonstrated their validity and universality (e.g., Reiss, 2004). Many of these motives are socially oriented, dealing with various social situations, either for competition or for cooperation. The pursuit of goals on that basis, sometimes in an implicit (unconscious) way, has also been explored, as reviewed by Huang and Bargh (2014). They are often not 'well behaved' as presupposed by assumptions used in deriving relatively simple utility functions (e.g., von Neumann and Morgenstern, 1953). Therefore, they are generally not incorporated into utility functions.

Even though some of these needs or desires can conceivably be figured into a utility function that can be optimized or satisficed (however complex and cumbersome they may become), the fact is that they have not been, at least not readily and adequately, up to this date (except for some very simple cases, e.g., Fehr and Gintis, 2007). This fact may point to their possibly 'irrational' nature.

However, they can nevertheless be incorporated into a detailed, comprehensive model of the mind, that is, a computational cognitive

architecture. I will describe such a theoretical model (cognitive architecture) later. Contrary to relatively simple utility functions, what such a theoretical model suggests is that the conventional notion of 'rationality' (e.g., based on optimization or near-optimization of well-behaved utility functions) does not adequately capture the complexity and reality of the human mind, nor the principles by which human society is organized.

Herbert Simon once said, 'Anything that gives us new knowledge gives us an opportunity to be more rational' (Spice, 2000). But, how 'rational' should one be? What is the proper (rational?) mixture of implicit and explicit processes in normal human functioning (Sun, 2002)? For example, for creative problem solving, one needs to rely on intuition to a significant degree (Helie and Sun, 2010; Kahneman and Klein, 2009). After all, to be somewhat 'irrational' is to be human. The human mind necessarily involves a lot of implicit, intuitive, instinctual, motivational, and emotional processes – all of them may be seemingly 'irrational' to some degree.

Herbert Simon also pointed out that 'Technology may create a condition, but the questions are what do we do about ourselves. We better understand ourselves pretty clearly and we better find ways to like ourselves' (Spice, 2000). In my opinion, this point applies well to understanding implicit, emotional, or otherwise 'irrational' processes. It is fundamentally important to appreciate and harness these 'irrational' processes in the human mind and in social, cultural, political, and organizational thinking.

Cognitive Architectures for Broadening Rationality

I will now describe a model of the mind that incorporates these various aspects of the human mind sketched above. This model comes in the form of a computational cognitive architecture, the idea of which was originally advocated by Simon's longtime collaborator Allen Newell (see, e.g., Newell, 1990). Computational models, such as computational cognitive architectures, provide concreteness and precision that may be helpful in clarifying ideas and testing possibilities when studying the human mind – an idea advocated by Simon, Newell, and many others (see, e.g., Fum *et al.*, 2007; Sun, 2008). In particular, a computational cognitive architecture is a comprehensive computational model that describes a wide range of psychological functionalities in a generic way, focusing on architectural issues, and may be applied to detailed explorations of various specific domains.

I will sketch an overall picture of the cognitive architecture CLARION, which is centered on the ideas discussed above, without going into too much technical detail. CLARION has been described meticulously and justified on the basis of voluminous psychological data previously (Sun, 2002, 2015).

CLARION captures a wide variety of psychological processes computationally. It is thus made up of a number of subsystems, including the action-centered subsystem (the ACS), the non-action-centered subsystem (the NACS), the motivational subsystem (the MS), and the metacognitive subsystem (the MCS).

Each of these subsystems consists of two levels of representations and their associated mechanisms and processes, as discussed earlier (Sun, 2002, 2015). Generally speaking, in each subsystem, the top level encodes explicit processes (using symbolic representation) and the bottom level encodes implicit processes (using subsymbolic, distributed connectionist representation, consisting of microfeatures; Rumelhart *et al.*, 1986). This representational difference accords with what was stipulated earlier. Also as indicated earlier, the bottom (implicit) level captures many 'irrational' characteristics of the mind (Sun, 2002, 2015).[4]

In this regard, two orthogonal dichotomies are important for CLARION and thus need to be pointed out: the procedural versus the declarative, and the implicit versus the explicit (Sun, 2015). Procedural processes are included in the ACS, while declarative processes are included in the NACS. Both the ACS and the NACS contain the bottom and top (implicit and explicit) levels. See Figure 5.1.

More specifically, the ACS is responsible for procedural processes, that is, for controlling actions, regardless of whether the actions are for external physical movements or for internal mental operations (e.g., executive control). Among the two types of procedural processes, implicit procedural processes are carried out by Backpropagation neural networks (Rumelhart *et al.*, 1986) at the bottom level of the ACS (which captures some 'irrational' characteristics). Explicit procedural processes, on the other hand, are carried out by explicit 'action rules', at the top level of the ACS.

The NACS is responsible for declarative processes, that is, for maintaining and utilizing declarative knowledge for information and inferences. Implicit declarative processes are carried out by associative memory networks (e.g., Hopfield type neural networks; Rumelhart *et al.*, 1986) at the bottom level of the NACS (which also captures some 'irrational' characteristics). Explicit declarative processes are carried out by explicit 'associative rules', at the top level of the NACS.

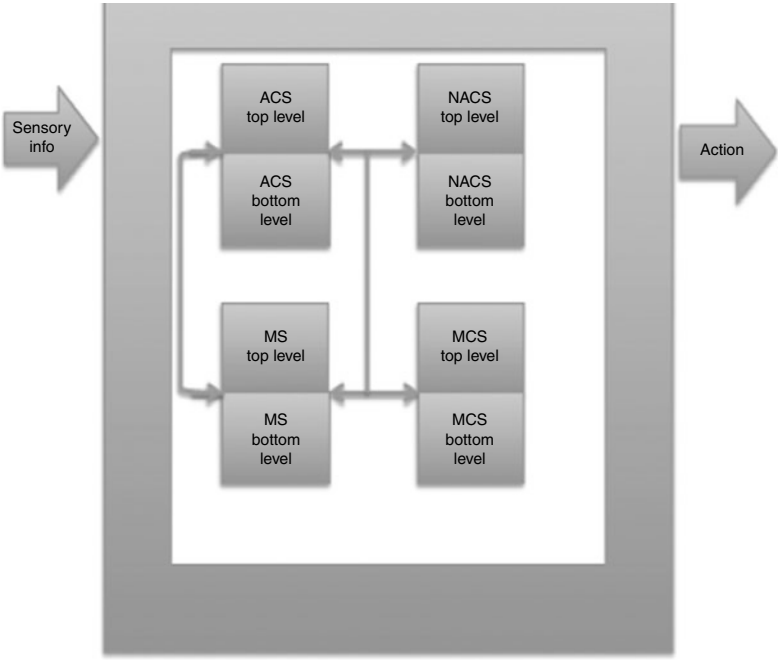


Figure 5.1 The CLARION cognitive architecture with four subsystems
The CLARION cognitive architecture with four subsystems. The major information flows are shown with arrows. ACS stands for the action-centered subsystem. NACS stands for the non-action-centered subsystem. MS stands for the motivational subsystem. MCS stands for the metacognitive subsystem. See the text for explanations. See Sun (2015) for technical details of the subsystems.

The MS provides underlying motivations for perception, action, and cognition (in terms of providing impetus and feedback). Implicit motivational processes are carried out by Backpropagation neural networks for activation of drives (for capturing basic human motives). Murray (1938) developed the idea that a set of basic motives determined human behavior (to a significant extent at least), and this idea has been developed and tested by others (e.g., Reiss, 2004). Based on such ideas, a set of basic drives was posited in the MS of CLARION, as shown in Table 5.1. Some of these drives are approach-oriented (aimed at obtaining positive rewards), while others are avoidance-oriented (aimed at avoiding negative results); see Table 5.2. On the other hand, explicit motivational processes within the MS are centered on explicit goal representation. Goals are determined (mostly) based on drives and in turn

Table 5.1 A list of primary drives in CLARION and their brief specifications

Drives	**Specifications**
Food	The drive to consume nourishment.
Water	The drive to consume liquid.
Sleep	The drive to rest.
Reproduction	The drive to mate.
Avoiding Danger	The drive to avoid situations that have the potential to be harmful.
Avoiding Unpleasant Stimuli	The drive to avoid situations that are physically (or emotionally) uncomfortable or negative in nature.
Affiliation & Belongingness	The drive to associate with other individuals and to be part of social groups.
Dominance & Power	The drive to have power over other individuals.
Recognition & Achievement	The drive to excel and be viewed as competent.
Autonomy	The drive to resist control or influence by others.
Deference	The drive to willingly follow or serve a person of a higher status.
Similance	The drive to identify with other individuals, to imitate others, and to go along with their actions.
Fairness	The drive to ensure that one treats others fairly and is treated fairly by others.
Honor	The drive to follow social norms and codes and to avoid blame.
Nurturance	The drive to care for, or attend to the needs of, others who are in need.
Conservation	The drive to conserve, to preserve, to organize, or to structure (e.g., one's environment).
Curiosity	The drive to explore, to discover, and to gain new knowledge.

Table 5.2 Approach-oriented versus avoidance-oriented drives

Approach drives	**Avoidance drives**	**Both**
Food	Sleep	Affiliation and belongingness
Water	Avoiding danger	Similance
Reproduction	Avoiding unpleasant stimuli	Deference
Nurturance	Honor	Autonomy
Curiosity	Conservation	Fairness
Dominance and power		
Recognition and achievement		

lead to behavior. So, ultimately, drives determine (to a large extent) the behavior of an individual (e.g., Sun and Wilson, 2014).

The MCS is responsible for dynamically monitoring and regulating the operations of the other subsystems. Implicit metacognitive processes

are carried out by Backpropagation neural networks at the bottom level, while explicit metacognitive processes are carried out by explicit rules at the top level (Reder, 1996).

Within each subsystem, the two levels interact (involving both 'rationality' and 'irrationality' in a sense). The interaction between the two levels includes bottom-up and top-down activation. Bottom-up activation is the *explicitation* of implicit information, through the activation of nodes (representing concepts) at the top level by corresponding nodes (representing microfeatures) at the bottom level. Top-down activation is the *implicitation* of explicit information, through the activation of (microfeature) nodes at the bottom level by corresponding (concept) nodes at the top level.

The interaction between the two levels also includes bottom-up and top-down learning. Bottom-up learning means implicit learning first and explicit learning on that basis. That is, implicit knowledge may be learned implicitly (through trial-and-error learning within the neural networks at the bottom level), and then may be explicated to form explicit knowledge (at the top level). Top-down learning means explicit learning first and implicit learning on that basis. That is, explicit knowledge may be explicitly learned, and may then be assimilated into implicit processes gradually.

The interaction between the two levels also includes the integration of the processing results from the two levels. For example, within the ACS, the two levels may cooperate in action decision making, through the integration of the action recommendations from the two levels of the ACS. In this integration, the relative emphasis of implicit versus explicit processing may be determined by the MCS taking into consideration a number of situation-specific factors (Sun, 2015).

Besides implicit and explicit processing as sketched above, CLARION also embodies other psychological phenomena, such as emotion and affect. According to CLARION, emotion is not a standalone mechanism, but emergent results of the interaction of a whole range of mechanisms within the ACS, the MS, the MCS, and the NACS (e.g., Wilson and Sun, 2014). The same may be said about moral judgment, aesthetic judgment, religiosity, and so on.

In this regard, it should be mentioned that, according to CLARION, these aspects have a lot to do with basic human motives. For example, emotion is viewed as resulting from essential human motives and their possible fulfillment (Wilson and Sun, 2014). When a situation is encountered, it triggers reactive affect, positive or negative, based on motivations (i.e., the activation of drives within the MS, which leads to affects generated by the MCS). On that basis, appraisal takes place

(mainly within the NACS), which analyzes the situation in accordance with a number of relevant dimensions. The results of the appraisal help to regulate the system and to guide actions (by the ACS). See Wilson (2012) for details of the emotion model within CLARION.

However, despite the apparent similarity to classical theories of rationality in terms of attributing preferences or aversions to basic motives (i.e., desire or want; Marshall, 1920), no attempt has been made here to reduce complex psychological mechanisms and processes to a well-behaved utility function. It can be argued that such an attempt may not be productive (or even possible) and may instead obscure a deeper exploration necessary for the true understanding of human nature or important social issues in relation to human nature.

Society and Rationality

Social, cultural, political, and organizational implications of some characteristics of the human mind have been explored in Simon's work. Likewise, social implications of the implicit–explicit distinction, emotion, and other related psychological aspects as conceived by CLARION, have also been explored, albeit more recently. I will describe some work in this direction.

In studying large-scale social phenomena, it is often difficult or impossible to run laboratory experiments on them. So they need to be investigated through alternative means, including through multi-agent social simulation. In this regard, cognitive social simulation – multi-agent social simulation based on detailed cognitive/psychological models – can be helpful (Sun, 2006). There are still relatively few social simulations with detailed computational models of psychological processes involved. However, cognitive architectures and other computational psychological models can help social sciences research, including social simulation, by taking into account details of human 'rationality' and 'irrationality.' For instance, studies have shown that both explicit and implicit processes play important roles in social, cultural, political, and organizational processes (e.g., Sun and Naveh, 2004).

On the basis of the CLARION cognitive architecture, a number of cognitive social simulations have been carried out, taking into consideration the implicit–explicit distinction or other aspects of the human mind touched upon earlier. These cognitive social simulations included:

- Role of cognition in organizational decision making
- Patterns of growth of academic science in relation to cognition

- Survival of tribal society in relation to cognitive and motivational factors
- Moral judgments and ethical norms
- Emotion and political behavior

Of course, the simulations conducted thus far did not cover the whole scopes of these topics (e.g., they did not address all aspects of organizational decision making). But the relevance of the CLARION framework to social, cultural, political, and organizational research has been shown through these simulations.

As a simple example, look into an organizational decision task (from Carley *et al.*, 1998). In this task, there is an object in the airspace. An organization must determine its status: whether it is friendly, neutral or hostile. In terms of organizational structures, there are two types: teams and hierarchies. In terms of information accessible by each agent, there are two varieties: distributed access, and blocked access (Sun and Naveh, 2004).

A cognitive social simulation was undertaken using CLARION. The importance of different cognitive capacities and parameters in affecting organizational performance was investigated. The results of the simulation closely accorded with patterns of human data across these conditions (organization × information); see Sun and Naveh (2004) for details. Furthermore, it was shown that effects of cognitive parameters were significant on organizational performance; for instance, a certain proportion of implicit processing helped to improve organizational performance. Such results may be outdated by now, but they are still of historical interest because they pointed to the possibilities very early on of combining realistic human psychology, social science research, and computational modeling.

Beyond this simple example, other work has been carried out within the CLARION framework that investigated these and many other aspects of human psychology, such as emotion and motivation, and their social, cultural, political, and organizational implications. See, for example, Naveh and Sun (2006), Sun and Naveh (2007), Sun and Fleischer (2012), Wilson and Sun (2014), and Sun (2006).

It would also be interesting to link cognitive social simulation (and the CLARION cognitive architecture within) to Foucault's ideas regarding knowledge and power (Foucault, 1977). According to Foucault, 'rationality' serves as a normative notion, and helps to sustain particular socioeconomic paradigms. In the context of cognitive social simulation,

the normative notion serves to highlight certain aspects of human psychology while downplaying some others through social and cultural means. Within the CLARION cognitive architecture as situated in a social context (e.g., running within a social simulation), social processes may impinge on processes within an individual. Through social interaction, an individual's psychological processes are adjusted, altered, and fitted to a sociocultural milieu. Such changes towards 'rationality' may lead to advantages in certain situations, but they may also cause problems in some other settings. One example, as discussed in Sun (2013), would be moral judgments based on utilitarian principles versus those based on fundamental moral principles (deontology, rooted in basic human motives as discussed earlier). The conflict between the two types of principle may often be strong (e.g., Cushman *et al.*, 2012). Modernity is concerned with order and rationalization that creates order, which, however, has its negative side, as discussed by, for example, Foucault (2006), Ellul (1964), and others.

Concluding Remarks

Exploring implicit processes and their interaction with explicit processes helps us to better understand human 'rationality' and 'irrationality'. Motivation, emotion, metacognition, and so on are also important contributing factors to human 'rationality' and 'irrationality', and therefore need to be taken into full account.

Due to the fact that issues involved in implicit processing, motivation, emotion, and other psychological aspects are complex, a detailed computational cognitive architecture may go a long way in helping to disentangle these issues. The CLARION cognitive architecture addresses, in a mechanistic, process-based sense, such important issues (Sun, 2002, 2015). Of course, in using a cognitive architecture to explore these issues, theoretical interpretations and assumptions are inevitably involved. Therefore, further empirical and theoretical work is needed for the sake of better validation.

The relevance of these psychological aspects to social, cultural, political, and organizational issues has also been demonstrated in the literature. As stressed throughout this chapter, the understanding of the true human condition in relation to 'rationality' may have serious implications for society.

Generally speaking, in order to fully explore human 'rationality' and 'irrationality', a much more nuanced view of how the human mind works should be developed, possibly through a combination of various

methodologies, including modeling using a computational cognitive architecture. I suspect that when we give up the outdated notion of 'rationality as optimization', we may open the door wide for new conceptions of the human mind and new understanding of the true human condition to emerge and develop. In this way, we may bring Herbert Simon's program to its logical conclusion. While Simon's work shook the foundation of the classical views on rationality, it behooves us to try to rebuild a new structure (and possibly to dismantle much of the old structure along the way).

Notes

1. When in quotes, 'rationality' denotes those conceptions based on optimization or near-optimization of some kind.
2. Note that I am only talking about problems with the notion of rationality when understanding the human mind and by extension understanding the aggregate of the human mind are involved.
3. For example, these characteristics include inappropriate reliance on similarity, inappropriate effects of priming, seemingly random generation of ideas, and so on.
4. The distinction between implicit and explicit processes and the representational difference between symbolic-localist and distributed representation have been argued many times before, so I will not repeat the arguments here (see Reber, 1989; Sun, 1994, 2002, Rumelhart *et al*., 1986; and so on).

References

Carley, K. M., Prietula, M. J. and Lin, Z. (1998). Design versus Cognition: The Interaction of Agent Cognition and Organizational Design on Organizational Performance. *Journal of Artificial Societies and Social Simulation*, 1 (3).

Cushman, F., Gray, K., Gaffey, A. and Mendes, W. B. (2012). Simulating Murder: The Aversion to Harmful Action. *Emotion*, 12 (1), 2–7.

Damasio, A. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Grosset/Putnam.

Dayan, P. (2014). Rationalizable Irrationalities of Choice. *Topics in Cognitive Science*. 6 (2), 204–28.

Dreyfus, H. and Dreyfus, S. (1987). *Mind Over Machine: The Power of Human Intuition*. New York: The Free Press.

Ellul, J. (1964). *The Technological Society*, trans. John Wilkinson. New York: Random House.

Evans, J. and Frankish, K. (eds) (2009). *In Two Minds: Dual Processes and Beyond*. Oxford: Oxford University Press.

Fehr, E. and Gintis, H. (2007). Human Motivation and Social Cooperation: Experimental and Analytical Foundations. *Annual Review of Sociology*, 33, 43–64.

Foley, D. K. (1998). Introduction (Chapter 1). In *Barriers and Bounds to Rationality: Essays on Economic Complexity and Dynamics in Interactive Systems*, ed. Peter S. Albin. Princeton: Princeton University Press.

Foucault, M. (1977). *Discipline and Punish*. London: Allen Lane.
Foucault, M. (2006). *The History of Madness*. London: Routledge.
Fum, D., Del Missier, F. and Stocco, A. (2007). The Cognitive Modeling of Human Behavior: Why a Model is (Sometimes) Better than 10,000 Words. *Cognitive Systems Research*, 8, 135–42.
Gigerenzer, G. and Selten, R. (2002). *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press.
Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press.
Helie, S. and Sun, R. (2010). Incubation, Insight, and Creative Problem Solving: A Unified Theory and a Connectionist Model. *Psychological Review*, 117 (3), 994–1024.
Huang, J. and Bargh, J. (2014). The Selfish Goal: Autonomously Operating Motivational Structures as the Proximate Cause of Human Judgment and Behavior. *Behavioral and Brain Sciences*. 37,121–75.
Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.
Kahneman, D. and Klein, G. (2009). Conditions for Intuitive Expertise: A Failure to Disagree. *American Psychologist*, 64, 515–26.
LeDoux, J. (1996). *The Emotional Brain*. New York: Simon and Schuster.
Marshall, A. (1920). *Principles of Economics* (rev. edn). London: Macmillan. reprinted by Prometheus Books.
Morris, M. W. and Keltner, D. (2000). How Emotions Work: The Social Functions of Emotional Expression in Negotiations. *Research in Organizational Behaviour*, 22,1–50.
Murray, H. (1938). *Explorations in Personality*. New York: Oxford University Press.
Naveh, I. and Sun, R. (2006). A Cognitively Based Simulation of Academic Science. *Computational and Mathematical Organization Theory*, 12 (4), 313–37.
Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.
Reber, A. (1989). Implicit Learning and Tacit Knowledge. *Journal of Experimental Psychology: General*. 118 (3), 219–35.
Reder, L. (ed.) (1996). *Implicit Memory and Metacognition*. Mahwah, NJ: Erlbaum.
Reiss, S. (2004). Multifaceted Nature of Intrinsic Motivation: The Theory of 16 Basic Desires. *Review of General Psychology*, 8 (3), 179–93.
Rumelhart, D., McClelland, J. and the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Cambridge, MA: MIT Press.
Simon, H. A. (1957). A Behavioral Model of Rational Choice. In *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: Wiley.
Simon, H. A. (1967). Motivational and Emotional Controls of Cognition. *Psychological Review*, 74 (1), 29–39.
Simon, H. A. (1991). Bounded Rationality and Organizational Learning. *Organization Science,* 2 (1), 125–34.
Spice, B. (2000). CMU's Simon Reflects on How Computers will Continue to Shape the World. *Pittsburgh Post-Gazette*, A-10–A-11. October 16, 2000.
Sun, R. (1994). *Integrating Rules and Connectionism for Robust Commonsense Reasoning*. New York: John Wiley and Sons.
Sun, R. (2002). *Duality of the Mind: A Bottom-up Approach Toward Cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sun, R. (ed.) (2006). *Cognition and Multi-Agent Interaction*. New York: Cambridge University Press.
Sun, R. (ed.), (2008). *The Cambridge Handbook of Computational Psychology.* New York: Cambridge University Press.
Sun, R. (ed.) (2012). *Grounding Social Sciences in Cognitive Sciences*. Cambridge, MA: MIT Press.
Sun, R. (2013). Moral Judgment, Human Motivation, and Neural Networks. *Cognitive Computation*, 5 (4), 566–79.
Sun, R. (2015). *Anatomy of the Mind*. New York: Oxford University Press. In press.
Sun, R. and Fleischer, P. (2012). A Cognitive Social Simulation of Tribal Survival strategies: The Importance of Cognitive and Motivational Factors. *Journal of Cognition and Culture*, 12 (3–4), 287–321.
Sun, R. and Naveh, I. (2004). Simulating Organizational Decision-making Using a Cognitively Realistic Agent Model. *Journal of Artificial Societies and Social Simulation,* 7 (3).
Sun, R. and Naveh, I. (2007). Social Institution, Cognition, and Survival: A Cognitive-Social Simulation. *Mind and Society*, 6 (2), 115–42.
Sun, R., Slusarz, P. and Terry, C. (2005). The Interaction of the Explicit and the Implicit in Skill Learning: A Dual-Process Approach. *Psychological Review*, 112 (1), 159–92.
Sun, R. and Wilson, N. (2014). Roles of Implicit Processes: Instinct, Intuition, and Personality. *Mind and Society*, 13 (1), 109–34.
Trimmer, P. C. and Houston, A. I. (2014). An Evolutionary Perspective on Information Processing. *Topics in Cognitive Science*. 6 (2), 312–30.
Von Neumann, J. and Morgenstern, O. (1953). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
Weber, M. (1915). *Konfuzianismus und Taoismus*. English translation published by Free Press: New York. 1968.
Wilson, N. (2012). Towards a Psychologically Plausible Comprehensive Computational Theory of Emotion. Ph.D Dissertation, RPI, Troy, NY.
Wilson, N. and Sun, R. (2014). Coping with Bullying: A Computational Emotion-Theoretic Account. In *Proceedings of the Cognitive Science Society Annual Conference*. Austin: Cognitive Science Society.

6

Boundedly Rational Decision-Making under Certainty and Uncertainty: Some Reflections on Herbert Simon

Mark Pingle

Introduction

Our collective rationality became more bounded on February 9, 2001. Herbert Simon emphasized we humans are cognitively constrained, and those constraints impact our decisions. Yet, Herbert Simon's mind was less constrained than most of our minds. Because of his exceptional thinking and writing, the constraints binding many disciplines have been relaxed. Consequently, those disciplines have become more rational, and less. The purpose of this chapter is to recognize how our collective rationality has been enhanced by the work of Herbert Simon, and related work, on decision-making.

Decision-Making as a Process

Our rationality is bounded by our limited cognitive capacities. This readily recognizable fact should make a theorist uncomfortable about assuming unbounded rationality. It made Herbert Simon uncomfortable. 'The expressed purpose of Friedman's principle of unreality (or as-if hypothesis)', Herbert Simon said, 'is to save Classical theory in the face of the patent invalidity of the assumption that people have the cognitive capacity to find a maximum' (Archibald, *et al.,* 1963). 'The unreality of premises,' Simon continued, 'is not a virtue in scientific theory but a necessary evil – a concession to the finite computing capacity of the scientist'.

Ignoring the bounds to rationality is convenient because it allows a decision problem to be specified as a mathematical optimization

problem. Environmental factors can be parameterized using variables, so a change in one of the environmental variables will change the set of alternatives available to the decision-maker. One can then delineate cause and effect relationships between elements of the decision environment and the optimal decision.

Herbert Simon's critique of unbounded rationality was twofold. First, while all theories are abstractions, the explanatory power of a theory will tend to decrease as the premises of the theory are less representative of reality. Second, we should not just care about the ability of a theory to predict, but we should also care about the assumptions of the theory. The assumptions which underlie our theory are part of our explanation of how the world works, so our explanation will lack credibility to the extent that our assumptions lack realism. Simon (1955, p. 99) sought to 'replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and computational capacities that are actually possessed'.

One of his important insights was that real-world decision processes will tend to involve sequential search and adaptation. Search theory (e.g., Wilde, 1964) informs us that sequential search has a fitness advantage over the simultaneous sampling of alternatives because the knowledge obtained from the current search choice can be used to select the location of the next alternative more effectively. Simon's (1955) bounded rationality model combines a sequential examination of alternatives with a predetermined *satisficing* goal for deciding when to stop incurring deliberation cost and accept an alternative as a choice.

If decision-makers understand cognitive limitations imply deliberation costs, which bind them from achieving what Simon (1976) called 'substantive rationality' (i.e., optimality), then they should also understand that no particular decision procedure will be best for all contexts. 'Procedural rationality' (Simon, 1976) involves coping with cognitive limitations and their implied deliberation costs, within a specific environment, by applying reason in some way. What is reasonable, or procedurally rational, can vary by decision-maker and by context.

Under bounded rationality, then, understanding choice is not just a matter of relating changes in the decision environment to changes in the location of the optimal choice. It involves relating changes in the decision environment to the decision-maker's cognitive abilities and to the decision-maker's available set of decision heuristics or processes. Because the choice is the outcome of the decision process selected, procedural rationality is reasoning applied at the level of selecting the

decision method more than reasoning applied to making the choice itself (Barros, 2010).

Recognizing Bounded Rationality: What Can We Learn?

It is not useful to critique unbounded rationality if recognizing bounded rationality does not bear fruit. What fruit can we pick?

First, we can explain different modes of decision behavior as reasonable ways of coping with deliberation costs (Day and Pingle, 1996). Trial and error, and more sophisticated decision methods, allow a decision-maker who has neither developed a habit nor has another person to imitate or obey, to sort through alternatives and find a good choice by incurring a deliberation cost. A good habit, by definition, will provide a good choice while minimizing deliberation costs. Thus, habit can outperform more consciously rational, more consciously deliberative, choices. Imitation and obeying an authority (e.g., tradition) can similarly provide good choices without much deliberation cost, depending upon the conditions.[1] Hunch or a random choice is risky, but can outperform other methods when the cost of those other methods is relatively high and when there is little difference in the quality of the various alternatives.

The word heuristic is typically reserved for describing a particular approach, imperfect and simplified, to making a choice. Rationality, in some form, is embedded in the heuristic. When we observe human behavior, especially decision behavior, we are mostly observing heuristics. Sometimes the heuristics are consciously chosen, but often they are unconsciously executed. In any case, they tend to respect Herbert Simon's notion that, 'Human rational behavior . . . is shaped by the scissors whose two blades are the structure of the task environment and the computational capabilities of the actor' (Simon, 1990, p. 7).

Gigerenzer and Goldstein (1996, 2002) particularly emphasize the shaping of heuristics by Simon's two blades. They define a 'fast and frugal heuristic' as one that is '(a) ecologically rational (i.e., they exploit structures of information in the environment), (b) founded in evolved psychological capacities such as memory and the perceptual system, and (c) fast, frugal, and simple enough to operate effectively when time, knowledge, and computational might are limited' (2002, p. 75). This perspective suggests much of human behavior can be explained as heuristics that have evolved so people can operate effectively in various environments given their cognitive limitations.

Herbert Simon's work also suggests that the existence of organizations and their particular forms can be explained as evolved responses to bounded rationality. 'The bounded rationality of humans', Simon (1991, p. 37) observed, 'does not allow us to grasp the complex situations that provide the environments for our actions in their entirety . . . One dimension of simplification is to focus on particular goals, and one form of focus is to attend to the goals of an organization or organization unit.' Not only does the organization itself provide focus, but Simon also notes that different positions within a firm orient people toward different specific useful goals. 'Behavior is very much a function of position.'

Simon (1991, p. 38) also emphasized that 'organizations, through the authority mechanism, provide a means for coordinating the activities of groups of individuals in ways that are not always easily achieved by markets'. Standard organizational procedures limit acceptable behaviors, but enhance productivity by providing a more stable and predictable environment. While pay incentivizes workers, organizational authority can further enhance productivity by mitigating free-riding and directing workers toward particular organizational goals. Authority can also 'inculcate individuals with organizational pride and loyalty' (Simon, 1991, p. 36). Just as developing a 'we' versus 'they' mentality motivates sports teams, it can motivate organizational teams, including in profit and non-profit oriented firms (Simon, 2002c).

Simon (2002a) recognizes decomposability as an organizational feature that allows people to cope effectively with bounded rationality. A more complex system is decomposable if it can be decomposed into smaller, relatively independent subsystems. Decomposability allows boundedly rational individuals to gain expertise and efficiency by specializing in relatively simple tasks. Efficient complexity is not obtained by developing the cognition of individuals from simple to complex. Rather, it is obtained by combining simple subsystems where individuals maintain limited and focused knowledge within a particular subsystem. Because people process information serially, 'a central design feature of any human communication system should be conserving attention' (Simon, 2002b), effective organizations limit information that is exchanged between subunits, so each unit receives necessary information, but not extraneous information that distracts the subunit from its focal purpose.

Hayek (1945) emphasizes the role of the price system in helping people cope with bounded rationality. He described the basic economic problem as 'a problem of the utilization of knowledge which is not given to anyone in its totality' (Hayek, 1945, p. 519). Changes in the

environment often require people to change their behaviors, and Hayek felt that only decentralized decision-making could properly respond and utilize the specialized 'time and place' knowledge possessed by individuals. To Hayek, the price system is a communication 'marvel', reducing the complexity of human interactions, which the boundedly rational individual could not hope to understand, into a change in price, which can be understood. 'The most significant fact about this system is the economy of knowledge with which it operates, or how little the individual participants need to know in order to be able to take the right action' (Hayek, 1945, pp. 526–7).

Simon contrasts the role of organizations to the role of the market, and offers some critique of Hayek. Simon (1991, p. 40) notes that 'prices perform their informational function when they are known or reasonably predictable', but 'uncertain prices produced by unpredictable shifts in a system reduce the ability of actors to respond rationally'. Simon also emphasizes that organizations arise because markets do not transform all necessary information into the market price. Uncertainty about product quality, uncertain delivery times, and varying product quality can make vertical integration preferable to market purchases. In summary, the organizational economy exists alongside the market economy because an organization is, at times, a more effective means for coping with bounded rationality than are markets and the price system.

In addition to organizations, Herbert Simon recognized that culture more generally is shaped by the fact that certain cultural norms help people cope with bounded rationality. Fernandes and Simon (1999) find that 'identification based on professional, ethnic or other characteristics can cause individuals to apply problem-solving strategies that match the goals or norms of the group'. Social relationships can preserve learning (Simon, 2000a) and reduce decision inefficiencies associated with bureaucracy (Simon, 2002a). Group loyalty also allows collaborative ventures to enhance efficiency beyond individual abilities, and helps focus effort toward a goal (Simon, 2002c). Hayakawa (2000) expounds on the view that low-cost heuristics to complex choice problems are sought and found in the life styles of social groups.

Indirectly, Simon (1993) relates altruism to bounded rationality. Defining altruism as behavior that reduces the actor's fitness while enhancing the fitness of others creates the question, why would anyone be altruistic? Part of Simon's (1993, p. 156) answer is docility: 'the tendency to depend upon the suggestions, recommendations, persuasion, and information obtained through social channels as a major basis for choice'. We are docile and submit to others because we are wise

enough to know we do not know it all. Without consciously choosing to do so, we tend to become loyal to a group that helps us cope with our bounded rationality. The altruistic willingness to sacrifice for the group follows. Simon (1993, p. 160) indicates 'altruism, especially altruism derived from group loyalties', will play a major role in an economics that 'describes the world in which we actually live'.

Simon (2000b, p. 25) emphasized 'rational behavior in the real world is as much determined by the "inner environment" of people's minds, both their memory contents and their processes, as by the "outer environment" of the world on which they act, and which acts on them'. This implies an important role for psychology and even physiology. For example, Simon (1967) presented the idea that emotions serve to interrupt cognitive processing and redirect it toward high-priority concerns. Kool *et al.* (2010) provide experimental evidence that people are biased toward conserving cognitive resources when making decisions, and link greater cognitive demands to heightened activity in the prefrontal cortex of the brain. Camerer *et al.* (2005) emphasize the ability of the brain to conserve cognitive resources by automating processes, and note that emotional reactions are associated with a different part of the brain than basic cognition and appear to prompt the mind for a decision as opposed to prompting contemplation. They also note the brain's tendency to simplify the world by categorizing.

Bounded Rationality and Uncertainty

Reviewing the bounds to rationality, Herbert Simon (2000a, p. 247) notes: 'The practical empirical limits to computation typically come into play long before the logical and mathematical limits do.' Because people must cope with their bounded rationality, and will do so in different ways in different contexts, we cannot expect one parsimonious model of decision-making to explain human behavior. Or, as Simon (2000a, p. 251) put it, 'Once one introduces into the SEU [subjectively expected utility] maximization Eden the snake of boundedness, it becomes difficult to find a univocal meaning of rationality, hence a unique theory of how people will, or should, decide. Economics, and the social sciences generally, will never have the certainty of natural science.'

An optimal choice cannot involve a deliberation (or optimization) cost because the same choice with no deliberation cost would be better. Thus, a decision-maker cannot both make an optimal choice and know the choice is optimal. It is possible that an unconscious habit has evolved to make an optimal choice. It is possible for imitation or

submission to an authority or hunch to produce an optimal choice. However, knowing the choice you have made is better than other alternatives requires costly deliberation. That is, the boundedly rational decision-maker must always possess some uncertainty about the quality of a given choice (Day and Pingle, 1991; Pingle, 2006).

McKinney and Van Huyck (2007, p. 626) define performance uncertainty as 'the uncertainty a player has about his ability to implement a substantively rational strategy'. Using the game *Nim*, where subjects can reason with more or less depth, they find human subjects vary in their depth of reasoning and typically misjudge their own ability to reason in the form of overconfidence. Experience reduces the overconfidence, or performance uncertainty.

Heiner (1983, p. 562) similarly contends that 'The presence of a C–D (competence-difficulty) gap will introduce uncertainty in selecting most preferred alternatives.' Environmental variables determine the difficulty of the decision problem, and variables internal to the decision-maker determine the decision-maker's competence. Bounded rationality implies a gap between competence and difficulty. Whereas the more standard approach is to assume the competence of the decision-maker arises to meet the problem difficulty. Heiner (1983, p. 561) alternatively proposes behavioral rules 'arise because of uncertainty in distinguishing preferred from less-preferred behavior'.

Heiner's theory is consistent with Herbert Simon's (2000a, p. 251) statement that: 'Perhaps the simplicity we should look for, in place of unattainable classical rationality, will come as we study empirically and concretely . . . how human beings actually adapt to the very severe limitations on their computational powers.' In Heiner's (1983, p. 567) theory, behavior is predictable because 'an agent's repertoire must be limited to actions which are adapted only to relatively likely or "recurrent" situations'. Heiner (1983, p. 573) also explains social institutions as evolved 'social rule-mechanisms for dealing with recurrent situations faced by agents'.

In addition to behavioral rules and social institutions, emotions may be a mechanism boundedly rational people use to cope with uncertainty. Expected utility theory is consequentialist, meaning the presumption is that people cognitively assess the consequences of possible choice alternatives. Lowenstein *et al.* (2001) developed an idea related to Simon's (1967) that emotions serve to redirect cognitive resources in uncertain situations. Their 'risk-as-feelings-hypothesis' recognizes that emotions often conflict with cognitive evaluations. While cognition is more fundamentally human and can override emotion, emotional

responses are more universally exhibited by all animals, serving to alert the animal to danger and risk of the unknown.

Camerer *et al.* (2005) present evidence that emotion in the form of fear is traceable to the amygdala in the brain, and that emotion in the amygdala can be overridden by cognitive 'cortical inputs'. They offer emotion or 'affect', as opposed to consequentialist cognition, as a possible explanation for the existence of bubbles, gambling and particular responses to terrorism. That is, in terms of coping with uncertainty, rationality may not only be bounded by limited cognitive capacity, but it may also be bounded by a physiological and psychological tendency to respond emotionally rather than cognitively to certain types of uncertainty.

Bounded Rationality and Probability

Given the focus here on Herbert Simon, it is worth noting that the very existence of uncertainty and the associated need for probability theory may be because rationality is bounded. 'If every event and phenomenon which occurs in the world has an antecedent cause of some sort', Crovelli (2012, p. 166) notes, 'then we are forced to say that probability is a measure of human ignorance or uncertainty about the causal factors at work in the world. Man's uncertainty in such a world could only stem from his inability to comprehend or account for all of the relevant causal factors at work in any given situation.'

Of course, the bounded rationality view of uncertainty is not the only view. Mulligan (2013) contrasts the perspectives on uncertainty of Ludwig von Mises and Richard von Mises. Ludwig saw the universe as deterministic, but saw probability theory as necessary because the complexity of the universe is beyond our understanding. Even if we could understand the determinism, Ludwig recognized that our memory subjectively samples past experience, so our understanding of the past is partly subjective and partly probabilistic. Alternatively, Richard viewed uncertainty as a physical property of the universe, meaning the universe is naturally probabilistic not deterministic.

Dunn (2001) distinguishes 'fundamental uncertainty' from the uncertainty generated by bounded rationality. While bounded rationality generates uncertainty because of the limited cognitive capacities of agents, fundamental uncertainty exists because the future has not yet been determined. Because humans are creative, the way the world works now is not be the way it will work in the future. Fundamental uncertainty exists because there is no way a decision-maker can know how innovations will alter the decision environment.

Probability and probability theory have been devised by we boundedly rational humans as a means of representing uncertainty. Crovelli (2009, p. 10) defines probability as 'a numerical measure of uncertainty . . . a subjective numerical statement of man's beliefs about the operant causes in the world'. The theory of probability, he notes, provides a methodology for evaluating probabilities. Mulligan (2013, p. 314) describes a probability as a heuristic, much like a market price, providing the decision-maker with high quality information at a much lower cost than would have to be incurred if complete knowledge were pursued.

Gilboa *et al.* (2009) describe three approaches to representing uncertainty with probability. The *classical approach* implements the *principle of insufficient reason* and presumes each outcome is equally likely. The *frequentist approach* presumes the likelihood of an event can be represented by its past empirical frequency. The *subjective approach* presumes that numerical probability is a measure of a degree of belief, constrained to satisfy certain conditions.

There is an element of bounded rationality in each approach to probability. The classical approach involves identifying categories of outcomes such that there is no reason to think one outcome is more likely than another. There may actually be deterministic physical forces which, if understood, would allow a gambler to predict with certainty which number on a six-sided die would arise, just as there may be a reason for a landlord to rent to one of six tenants. However, a lack of knowledge may imply insufficient reason for making a distinction, so all outcomes are treated as equally likely.

Whereas the classical approach may be applied *ex ante*, Mulligan (2013) notes that the relative frequency approach can only be applied *ex post*, or after some history. Like Mulligan, Gilboa *et al.* (2009) note that the relative frequency approach is subjective in that the decision-maker has flexibility regarding how categories are chosen. Crovelli (2012) emphasizes that to categorize event 'classes' one must assume the world is causally deterministic. The boundedly rational decision-maker knows he lacks knowledge, but nonetheless implicitly assumes there is some process determining when an event goes into one class and when into another. Categorization is a heuristic, substituting for knowledge, which assumes similar events are generated in the same manner. Of course, as Gilboa *et al.* note, the relative frequencies obtained depend upon the categorization, and cases may not be identical nor independent. Just as a boundedly rational decision-maker must arbitrarily choose a choice method because of deliberation cost, so a boundedly rational

decision-maker must arbitrarily choose category classes when using the relative frequency approach to probability because limited cognition precludes knowing which cases are truly identical.

Fishburn (1986, p. 335) describes the theory of subjective probability as an attempt 'to make precise the connection between coherent dispositions toward uncertainty and quantitative probability'. Mulligan (2013, p. 322) notes that this perspective is consistent with Keynes, who viewed probability as 'a rationally justified degree of belief, objectively derived by logic'. Mulligan critiques Keynes for implying all should come to the same belief, primarily because knowledge is so diversely spread across the population. Probabilities for single cases must be subjective when the classical approach does not provide insight, nor history provide relative frequencies. Two fighters, for example, might not be subjectively judged equally likely to win.

Just as adaptation may lead the boundedly rational decision-maker to a decision method that suits an environment, so adaptation can lead the boundedly rational decision-maker to a probability that can be considered objective. Gilboa *et al.* (2009, p. 2) define objective probability as being one with which 'most reasonable people would agree'. They argue decision-makers can use past data to improve categorization so the relative frequency approach will provide probabilities with which people will agree. Mulligan (2013) views the relative frequency approach as more objective because it is verifiable, and argues the social norm for a probability will cluster around the classical and frequency definitions when they are mutually reinforcing, such as for coin tossing and games of chance.

There is much evidence that people are boundedly rational when it comes to subjectively formulating a probability to represent uncertainty. Falk and Wilkening (1998) review experiments with children that demonstrate probability requires higher level cognition. Children aged 4 or 5 years old do exhibit 'a glimmer of probability understanding'. A common error, which is almost completely eliminated by age 11, is for the child to prefer a prospect with more chances over a prospect with fewer chances but higher probability. When children consistently perform incorrectly, it is not always because they do not know the correct rule. Rather, it can be because 'they have greater faith in [another rule] or find it easier to apply' (Falk and Wilkening, 1998, p. 1351).

People exhibit conjunctive and disjunctive fallacies in formulating probabilities. Probability theory states that if A is a subset of B, then the probability of A cannot exceed that of B, but people predictably and systematically violate this conjunction rule. The disjunction rule is also violated, which says neither the probability of A nor the probability

of B can be bigger than the probability of A or B. Bar-Hillel and Neter (1993) present a general explanation, borrowing from Kahneman and Tversky (1983), that indicates a bound on rationality relative to categorization. They distinguish basic categories which are mutually exclusive (e.g., chairs, tables, beds), from those which may be constructed as a disjunctive higher order category (e.g., furniture) and conjunctive lower order categories (e.g., leather chair, king-size bed). The conjunctive fallacy occurs when the lower order category is judged to be more probable than the basic category because it is perceived to be more representative of the real-world population. The disjunctive fallacy occurs when a basic category is judged to be more probable than the higher order category because the basic category is more representative of the real-world population. That is, one bound on rationality is the tendency to judge events more likely when they correspond to our preconceived judgments.

Isaac and Brough (2014) review evidence on how subjective categorization choices can bias probability judgments, and then focus on category size. The *alternative outcomes bias* occurs when an irrelevant cue about category size changes a probability judgment. A *partition dependence bias* occurs when the total probability changes when a category is partitioned into a number of mutually exclusive subcategories. Isaac and Brough demonstrate that increasing the size of a category can increase the perceived total probability, a *category size bias*. As an application, they suggest grouping preventable hazards with unpreventable hazards to increase the total hazard probability, which, because of the category size bias, might lead more people to avoid the risky behavior (e.g., smoking).

Barron and Ursino (2013) report a difference in the response to risk when the information is described versus experienced. People tend to overweight small probabilities in decisions from description, while they tend to underweight small probabilities in decisions from experience. The underweighting from experience is perceived to occur because experience generates a small sample, which causes the bias.

Juslin *et al.* (2009, p. 871) make the point that the 'coherence rules of probability theory are themselves only of heuristic value – that is, valuable only insofar as they help to produce decisions with better average return'. Thus, they are not necessarily preferable to other heuristics that are incoherent. They use the fact that people commonly combine different effects by weighting each effect and adding the products because it is intuitive and simple. They demonstrate that errors in judgment are compounded when Bayesian type multiplication and division is applied. Thus, simpler heuristics may provide better guides than more complex probability theory rules. Alternatively, Costello and Watts (2014) show

that their model, which assumes people follow probability theory in their probabilistic reasoning but make random mistakes, can explain many observed biases in probability judgment.

Not only might a particular heuristic not be best for arriving at a probability judgment, it 'may not be rational to choose a single probability measure, a choice that is bound to be arbitrary' (Gilboa *et al.* 2009, p. 9). Ellsberg's (1961) experiments demonstrate people do not always behave as if a subjective probability measure summarizes their beliefs. Andersen *et al.* (2012) present a method for estimating subjective beliefs and show a subjective probability may best be characterized by a probability distribution rather than as a single number. In general, ambiguity may characterize the uncertainty people perceive more so than risk.

Karelitz and Budescu (2004) examine how probabilities are communicated verbally. They find people prefer to express uncertainties verbally, use diverse language to describe uncertainty, and vary in their numerical interpretations of linguistic terms. This implies that likelihoods are often not well communicated. Karelitz and Budescu (2004, p. 26) conclude: 'Except in very special cases, all representations of uncertainties are vague to some degree in the minds of the originators and in the minds of the receivers.' Yet, they note decision-makers must resolve this vagueness in some way because 'one can have imprecise opinions but cannot take imprecise actions'.

Outcomes versus Probabilities and Bounded Rationality

Jeske and Werner (2008) present neurological evidence that people distinguish outcomes from probabilities when they make decisions under uncertainty. Neural functioning can be categorized as cognitive or affective. Cognition involves conscious deliberation, while affect involves an emotional response. The cognitive system seems to be more sensitive to probabilities, while the affective system seems to be more sensitive to outcomes. Outcomes and probabilities also seem to activate different areas of the brain. Functional magnetic resonance imaging has documented that the subcortical nucleus accumbens is activated by an anticipated outcome and the cortical mesial prefrontal cortex is activated by an anticipated probability. 'The brain is much more responsive to changes in gain size than to equivalent changes in probability' (Jeske and Werner, 2008, p. 52).

In seeking to explain why people seem to be more sensitive to changes in outcomes than changes in probabilities, Jeske and Werner (2008,

p. 62) refer to Herbert Simon's (1967) bounded rationality idea that attention is a scarce resource: 'Information about probability might be useless without the corresponding information on outcome, but not necessarily vice versa. This inherent asymmetry alone might explain the overall finding that participants spent more time looking at outcomes.'

Weber (1994) notes that, to be able to explain anomalies, more recent models of decision-making under uncertainty have deviated from the assumption that the subjective probabilities a decision-maker associates with outcomes are independent of the outcomes. Weber labels the models she reviews as *configural*, meaning the weight the decision-maker gives to an outcome is not just the probability of the outcome itself but the weight is dependent upon the rank of the outcome in the configuration of possible outcomes. Lowenstein *et al.* (2001, p. 276) note that 'One of the most robust observations in the domain of decision making under uncertainty is the overweighting of small probabilities, particularly those associated with extreme outcomes.'

Cumulative prospect theory (Tversky and Kahneman, 1992) can be considered a culmination of the effort to modify subjective expected utility to be able to account for behaviors which subjective expected utility cannot explain (e.g., loss aversion, equity premium puzzle, why gamblers also buy insurance). As Weber (1994, p. 234) explains, cumulative prospect theory kept the original value function of prospect theory, which is concave for gains and convex and steeper for losses, but replaced the prospect theory decision weighting function with Quiggin–Yaari rank dependent transformation of cumulative probabilities. The rank dependent transformation allows optimism and pessimism to enter the decision.

Figure 6.1 illustrates what cumulative prospect theory can accomplish. The concavity of value function $V(x)$ for gains implies risk aversion, which one would typically expect. The unweighted $U(x)$ curve in the figure represents how expected utility would value a prospect that yields $x_{max} = 10$ with probability p or $x_{min} = 0$ with probability $1 - p$. For all probabilities p, risk aversion implies $V(x) > U(x)$, so the decision-maker prefers the certain outcome x to taking the chance that yields either x_{max} or x_{min}. This explains many situations, but it cannot explain gambling. The weighted $U(x)$ curve presents a situation where the decision-maker gives extra weight to lower probabilities. As shown in Figure 6.1, this extra weight can imply $U(x) > V(x)$ when the probability p for the outcome x_{max} is low (so the expected utility is near the origin). That is, it is possible for the decision-maker to prefer taking a chance (e.g., gamble) to a certain outcome with equal expected value.

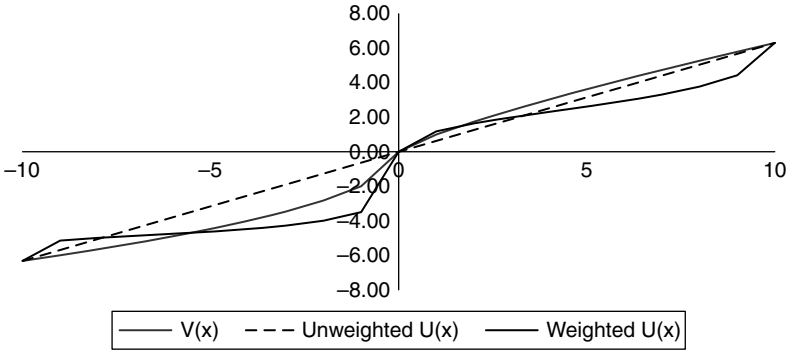


Figure 6.1 Cumulative prospect theory

For losses, the convexity of value function $V(x)$ implies loss aversion. The unweighted $U(x)$ curve in the figure represents how expected utility would value a prospect that yields either $x_{max} = 0$ with probability p or $x_{min} = -10$ with probability $1 - p$. For all probabilities p, loss aversion implies $U(x) > V(x)$, so the decision-maker prefers taking the chance that yields either x_{max} or x_{min} to accepting the certain loss x. The problem with pure loss aversion is it precludes an explanation for insurance purchases. The weighted $U(x)$ curve presents a situation where the decision-maker again gives extra weight to lower probabilities. When the probability p (of no loss) is high, which we would expect in the typical insurance situation, we see in Figure 6.1 that $V(x) > U(x)$, means the decision-maker will prefer buying insurance to taking the chance.

Despite its generality, cumulative prospect theory is not entirely satisfying relative to Herbert Simon's interest in developing theory which 'describes the world in which we actually live.' First, there is much evidence that people perceive uncertainty as ambiguity (i.e., imprecise representations of probability) as opposed to risk (i.e., precise representations of probability). Second, the assumption that the utility function is convex for losses, while it makes the theory work, rejects the law of diminishing marginal utility.

Figure 6.2 illustrates an alternative to prospect theory, based upon the alpha minimax expected utility of Gilboa and Schmeidler (1989). The theory is represented by the value function

$$V(x) = [1 - \lambda][pu(x_{max}) + [1 - p]u(x_{min})] + \lambda[\alpha u(x_{max}) + [1 - \alpha]u(x_{min})] \quad (1)$$

where there are two possible outcomes x_{max} and x_{min}. The degree of ambiguity is captured by λ. When $\lambda = 1$, there is total ambiguity and the

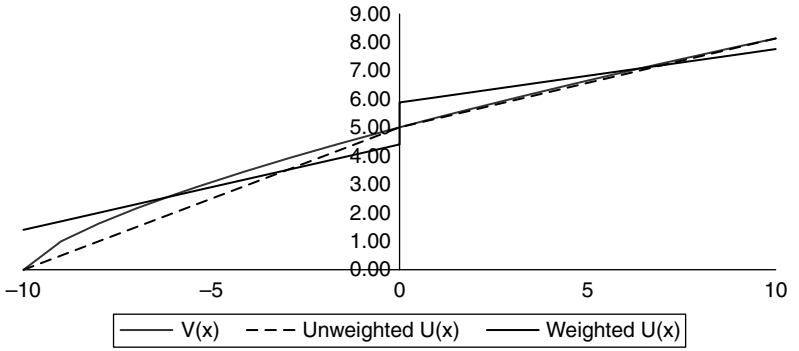


Figure 6.2 Alpha minimax expected utility theory

prospect is valued using the Arrow-Hurwicz (1972) criterion $\alpha u(x_{max}) + [1 - \alpha]u(x_{min})$, where α denotes the degree of optimism. When $\lambda = 0$, there is no ambiguity and the prospect is valued using expected utility $pu(x_{max}) + [1 - p]u(x_{min})$.

In comparing the value function $V(x)$ in Figure 6.2 to that in Figure 6.1, notice diminishing marginal utility is maintained for both positive and negative prospects in the alpha minimax theory. This implies the decision-maker is risk averse for both positive and negative prospects. For cumulative prospect theory, the convex utility function in the loss domain is necessary to explain loss aversion. However, this is not necessary with alpha minimax theory. With some ambiguity (i.e., $\lambda > 0$) and some optimism (i.e., $\alpha > 0$), there is weight in the value function placed upon x_{max}. When p is high, as it would be in the typical insurance case, then risk aversion predominates and, as shown in Figure 6.2, the certain $V(x)$ value is greater in the loss domain than the weighted $U(x)$ value. However, when p is small enough, the weight $U(x)$ value is less than the certain $V(x)$ value, implying the decision-maker would rather take the chance and try to avoid the loss than accept the certain loss. That is, loss aversion can be explained without assuming a convex utility function.

Similarly, in the gain domain, taking a chance (i.e., gambling) can be explained by a combination of ambiguity and optimism. When the probability of the x_{max} outcome is high, as might be the case in the typical stock purchase, Figure 6.2 indicates the decision-maker will prefer the certain outcome to taking a chance (e.g., prefer bonds to stocks, unless there is an equity premium). Alternatively, when the probability of the x_{max} outcome is low, as it would be in the typical gambling situation,

gambling may be preferred, as in the case shown where the optimism level for the decision-maker is set moderately high.

Conclusions

Social science seeks to understand how people behave as they interact. Choice theory, or explaining the choices people make, is core to social science. Herbert Simon will remain an intellectual giant among social scientists because he helped extend our understanding of choice, and his good work motivated many other good scholars to follow in his footsteps.

As Herbert Simon and those following him have recognized bounded rationality, social science has progressed in terms of understanding decision-making processes, how people make relatively good choices in spite of their cognitive limitations, and why organizational and cultural forms are what they are. Abstracting from cognitive limitations may be useful, but Herbert Simon helped us see the progress social science could make by moving beyond this abstraction.

There is still room for improvement, particularly with regard to understanding how boundedly rational decision-makers cope with uncertainty. It may be people behave as if they subjectively form a probability distribution over possible states and then maximize expected utility. However, Herbert Simon encouraged us to move beyond this as-if world to model decision processes that people have the cognitive capacity to use. Moving from the assumed probabilistic sophistication it takes to construct a probability distribution to assuming ambiguity is a good step taken more recently, and this has borne some fruit. However, it is likely that more understanding will be achieved as researchers further seek, like Herbert Simon did, to construct models of decision-making which are consistent with the cognitive capacities people actually possess.

Note

1. See Pingle (1995) and Pingle (1997) for experiments on imitation and obeying authority that demonstrate both the effectiveness and ineffectiveness of these behaviors.

References

Andersen, S., Fountain, J., Harrison, G. W., Hole, A. and Rutström, E. (2012). Inferring Beliefs as Subjectively Imprecise Probabilities. *Theory and Decision* 73, 161–84.

Archibald, G. C., Simon, H. A. and Samuelson, P. (1963). Discussion. *American Economic Review* 53, 227–36.

Arrow, K. J. and Hurwicz, L. (1972). An Optimality Criterion for Decision-making under Ignorance. In *Uncertainty and Expectations in Economics: Essays in Honour of G. L. S. Shackle*, ed. C. F. Carter and J. L. Ford. Oxford: Basil Blackwell.

Bar-Hillel, M. and Neter, E. (1993). How Alike is it versus How Likely is it: A Disjunction Fallacy in Probability Judgments. *Journal of Personality and Social Psychology* 65 (6), 1119–31.

Barron, G. and Ursino, G. (2013). Underweighting Rare Events in Experience Based Decisions: Beyond Sample Error. *Journal of Economic Psychology* 39, 278–86.

Barros, G. (2010). Herbert A. Simon and the Concept of Rationality: Boundaries and Procedures. *Brazilian Journal of Political Economy* 30 (3), 455–72.

Camerer, C., Loewenstein, G., and Prelec, D. (2005). Neuroeconomics: How Neuroscience Can Inform Economics. *Journal of Economic Literature* 43 (1), 9–64.

Costello, F. and Watts, P. (2014). Surprisingly Rational: Probability Theory Plus Noise Explains Biases in Judgment. *Psychological Review* 121 (3), 463–80.

Crovelli, M. (2009). On the Possibility of Assigning Probabilities to Singular Cases, Or: Probability is Subjective Too! *Libertarian Papers* 1 (26), 1–18.

Crovelli, M. (2012). All Probabilistic Methods Assume a Subjective Definition for Probability, *Libertarian Papers* 4 (1), 163–74.

Day, R. H. and Pingle, M. (1991). Economizing Economizing. In *Behavioral Decision-Making: Handbook of Behavioral Economics,* Vol. 2B, ed. R. Frantz, H. Singh, and J. Gerber, 509–22. Greenwich, CT: JAI Press.

Day, R. H. and Pingle, M. (1996). Modes of Economizing Behavior: Experimental Evidence. *Journal of Economic Behavior and Organization* 29, 191–209.

Dunn, S. (2001). Bounded Rationality is not Fundamental Uncertainty: A Post Keynesian Perspective. *Journal of Post Keynesian Economics* 23 (4), 567–87.

Ellsberg, D. (1961). Risk, Ambiguity and the Savage Axioms. *Quarterly Journal of Economics* 75, 643–69.

Falk, R. and Wilkening, F. (1998). Children's Construction of Fair Chances: Adjusting Probabilities. *Developmental Psychology* 34 (6), 1340–57.

Fernandes, R. and Simon, H. A. (1999). A Study of How Individuals Solve Complex and Ill-Structured Problems. *Policy Sciences* 32, 225–45.

Fishburn, P. (1986). The Axioms of Subjective Probability, *Statistical Science* 1 (3), 335–58.

Gigerenzer, G. and Goldstein, D. G. (1996). Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review* 103 (4), 650–69.

Gigerenzer, G. and Goldstein, D. G. (2002). Models of Ecological Rationality: The Recognition Heuristic, *Psychological Review* 109 (1), 75–90.

Gilboa, I., Lieberman, O. and Schmeidler, D. (2009). On the Definition of Objective Probabilities by Empirical Similarity. *Synthese*, 172 (1), 79–95.

Gilboa, I. and Schmeidler, D. (1989). Maxmin Expected Utility with Non-Unique Prior. *Journal of Mathematical Economics* 18, 141–53.

Hayakawa, H. (2000). Bounded Rationality, Social and Cultural Norms, and Interdependence via Reference Groups. *Journal of Economic Behavior & Organization* 43, 1–34.

Hayek, F. A. (1945). The Use of Knowledge in Society. *American Economic Review* 35 (4), 519–30.

Heiner, R. A. (1983). The Origin of Predictable Behavior. *American Economic Review* 73 (4), 560–95.

Isaac, M. and Brough, A. (2014). Judging a Part by the Size of Its Whole: The Category Size Bias in Probability Judgments. *Journal of Consumer Research* 41 (2), 310–25.

Jeske, K.-J. and Werner, U. (2008). Impacts on Decision Making of Executives – Probabilities versus Outcomes. *Journal of Neuroscience, Psychology, and Economics* 1 (1), 49–65.

Juslin, P., Nilsson, H. and Winman, A. (2009). Probability Theory, Not the Very Guide of Life, *Psychological Review* 116 (4), 856–74.

Kahneman, D. and Tversky, A. (1983). Extension versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment, *Psychological Review* 90 (4), 293–315.

Karelitz, T. and Budescu, D. (2004). You Say 'Probable' and I Say 'Likely': Improving Interpersonal Communication with Verbal Probability Phrases. *Journal of Experimental Psychology* 10 (1), 25–41.

Kool, W., McGuire, J. T., Rosen, Z. B. and Botvinick, M. M. (2010). Decision Making and the Avoidance of Cognitive Demand. *Journal of Experimental Psychology: General* 139 (4), 665–82.

Lowenstein, G., Weber, E., Hsee, C. and Welch, N. (2001). Risk as Feelings, *Psychological Bulletin*, 127 (2), 267–86.

McKinney, N. Jr. and Van Huyck, J. (2007). Estimating Bounded Rationality and Pricing Performance Uncertainty. *Journal of Economic Behavior & Organization* 62, 625–39.

Mulligan, R. F. (2013). The Enduring Allure of Objective Probability. *Review of Austrian Economics*, 26, 311–27.

Pingle, M. (1995). Imitation versus Rationality: An Experimental Perspective on Decision-making. *Journal of Socio-Eonomics,* 24 (2), 281–315.

Pingle, M. (1997). Submitting to Authority: Its Effect on Decision-making. *Journal of Psychology,* 18, 45–68.

Pingle, M. (2006). Deliberation Cost as a Foundation for Behavioral Economics. In *Handbook of Contemporary Behavioral Economics: Foundations and Developments*, ed. Morris Altman. Routledge: New York and Abingdon, 340–55.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics,* 69, 99–118.

Simon, H. A. (1967). Motivational and Emotional Controls of Cognition. *Psychological Review,* 57, 386–420.

Simon, H. A. (1976). From Substantive to Procedural Rationality. In *Method and Appraisal in Economics,* ed. S. J. Latsis. Cambridge: Cambridge University Press, 129–48.

Simon, H. A. (1990). Invariants of Human Behavior. *Annual Review of Psychology,* 41, 1–19.

Simon, H. A. (1991). Organization and Markets. *Journal of Economic Perspectives,* 5 (2), 25–44.

Simon, H. A. (1993). Altruism and Economics. *American Economic Review,* 83 (2), 156–61.

Simon, H. A. (2000a). Review: Barrier and Bounds of Rationality. *Structural Change and Economic Dynamics,* 11, 243–53.

Simon, H. A. (2000b). Bounded Rationality in Social Science: Today and Tomorrow. *Mind & Society,* 1, 25–39.
Simon, H. A. (2002a). Near Decomposability and the Speed of Evolution. *Industrial and Corporate Change,* 11, 587–99.
Simon, H. A. (2002b). Organizing and Coordinating Talk and Silence in Organizations. *Industrial and Corporate Change*, 11 (3), 611–18.
Simon, H. A. (2002c). We and They: The Human Urge to Identify with Groups. *Industrial and Corporate Change,* 11 (3), 607–10.
Tversky, A. and Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5, 297–323.
Weber, E. (1994). From Subjective Probabilities to Decision Weights: The Effect of Asymmetric Loss Functions on the Evaluation of Uncertain Outcomes and Events, *Psychological Bulletin,* 115 (2), 228–42.
Wilde, D. J. (1964). *Optimum Seeking Methods*. Englewood Cliffs, NJ: Prentice-Hall.

Part II
Models

7

Herbert Simon and Agent-Based Computational Economics

Shu-Heng Chen and Ying-Fang Kao

Introduction

Herbert Simon was a quintessential interdisciplinary scholar who made pioneering contributions concerning the notion of bounded rationality, built models based on it, and made important advances in understanding complex systems. His importance in the field of artificial intelligence, which was in turn the inspiration of agent-based computational economics (ACE), is discussed in detail in Chen (2005). Among all the Nobel Laureates in Economics, there are at least three whose work has been acknowledged by the ACE community. They are Friedrich Hayek (1899–1992), Thomas Schelling (1921–), and Elinor Ostrom (1933–2012). The last two worked directly on ACE. Schelling's celebrated work on the segregation model is considered one of earliest publications on ACE (Schelling, 1971). Ostrom contributed to the development of empirical agent-based models (Janssen and Ostrom, 2006). Hayek did not work on ACE, but the connection of his work to ACE has been pointed out by Vriend (2002).

We believe that there is a strong connection between the development of ACE and Herbert Simon and that his influence on ACE is no less, if not more, profound than the previous three. However, to the best of our knowledge, there seems to be no single document that, from a holistic perspective, addresses this linkage explicitly.[1] We conjecture that the burgeoning of ACE was too late for the time of Simon, who died in 2001. Even so, it still surprises us that so few attempts have been made to connect Simon and ACE, particularly considering that the latter was founded on artificial intelligence and cognitive psychology, the two pillars to which the former contributed substantially.

In this chapter, we attempt to explore and identify the connections between Simon's contributions and the development of ACE. We concentrate on his influence on the conception of an individual within an economic or social system, his philosophy regarding how social systems are organized and understood, and finally how the underlying rules that govern social interactions can be unearthed by the investigator, in this case a social scientist. We also suggest ways in which future developments within ACE can be geared to be more Simonian in character and closer to his vision.

The rest of the chapter is structured as follows. The next section provides an overview of the setting in which we place our arguments, which we then divide into three main departments: individuals, complex systems, and the epistemology of ACE. In the third section the modeling of software agents is discussed in light of Simon's bounded rationality is discussed. Various aspects of complex systems are included in the fourth section. In the fifth section we elaborate on ACE's potential as an alternative to neoclassical economics. We conclude the paper in the last section.

Setting

Simon's contributions to behavioral economics and artificial intelligence are composed of many different, original ideas, all of which are grounded, either explicitly or implicitly, in the theories and models of computation. On the other hand, ACE routinely studies rule-following agents, computing within an environment that can best be considered as a complex adaptive system. ACE, as a research program, can be considered as mounted on, at least, *four pillars*: individuals (decomposition), interaction, aggregation, and learning (adaptation). Each of these pillars has an important computational component in its characterization. With this background it is fairly evident that computation, and consequently simulation, can be one possible anchor from which one can attempt to explore the influence of Simon's legacy on the development of ACE. Since we are exploring the intellectual links between Simon and ACE, it may be instructive to be aware that Simon did not emphasize all of these pillars uniformly. For instance, he placed more emphasis on the characteristics of economic entities (agents, institutions) as being boundedly rational (Simon, 1957, 1976) and adaptive (Simon, 1996b), and on features such as the near decomposability of complex systems (Simon, 1962, 1995). He appears to have focused less on the interaction aspect, except for his celebrated contributions on stochastic models

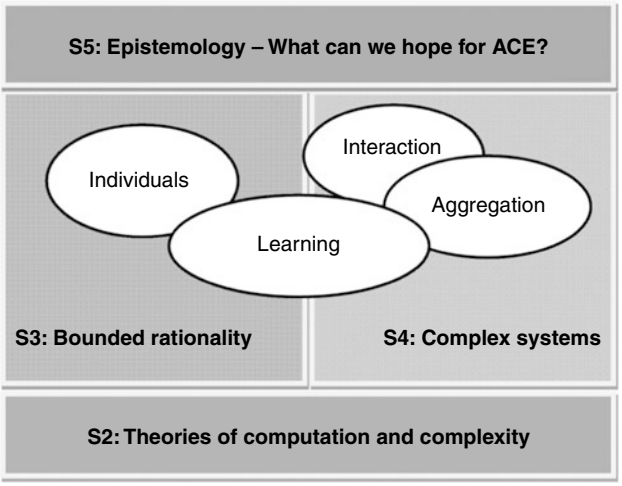


Figure 7.1 Four pillars of ACE placed on the foundation of Simon

(Simon, 1955b) and the resulting aggregate distributions underpinned by preferential attachment (see the section on Preferential Attachment). However, despite the varying degrees of emphasis laid down by Simon, all of the above ideas seem to have had an important and direct impact on ACE.

Figure 7.1 illustrates the perspective we are using to organize and develop this chapter. Of these four pillars learning seems to be the only one at the intersection of the ideas from both bounded rationality and complex systems. Learning is an important feature that is relevant for the components within the system (individuals) and also for the system as a whole. Learning, or evolution, becomes critical when the environment in which the agents live and act is ever-changing and complex. In such environments, optimal survival rules that are suitable for all agents at all times simply do not exist. Simon has always had the view of economic agents as boundedly rational organisms acting in complex environments (Simon, 1959).

There has never been any direct conversation between Simon and ACE, at least ACE in the way we understand it today. This is perhaps due to the timing of the development of the latter. However, Simon has commented on many building blocks of ACE, such as various ways of modeling boundedly rational individuals, the form of complexity in social systems, learning and the evolution of complex systems, the stochastic mechanisms or models that could explain stylized aggregate

distributions, and so on. In this study, we are also interested in identifying the less direct and the possibly unidentified legacies of Simon with regard to ACE. These hidden links include Simon's epistemological point of view towards simulation and his chunking theory (modularity, in its modern form). Besides, Simon's methodological pursuits also provide a hint for carving out a vision for the potential roles of ACE as it evolves, which could include market (more broadly, institutional) design (Marks, 2006), and act as a mode of hypothesis discovery. We argue that these could be important directions for the ACE community to explore to broach new and interesting frontiers.

Computation-Theoretic Underpinning

Given the general framework, as depicted in Figure 7.1, we begin from the bottom. The common thread that connects Simon and ACE is the view of how economic agents and systems should be meaningfully modeled and studied, based on the advantages offered by digital computers. A brief survey of the origins of ACE could be useful to highlight a common ground, which happens to be the theories of computation and complexity. Four distinct, yet interconnected, origins of ACE are identified in a recent survey by Chen (2012). They are market, cellular-automaton, economic-tournament, and experimental economic origins. Except perhaps for market origin, we can find Simon's direct or indirect influence on ACE through the other three origins. In particular, we would like dwell a little more on the cellular-automaton origin.

One important, and probably the earliest, example of the use of cellular automata in social sciences is Schelling's segregation model (Schelling, 1971), which is built upon a checkerboard topology, also known as a checkerboard model. Albin (1975, 1998) explores the connection between Schelling's checkerboard model and the cellular automata tradition, which, in turn, places ACE on its computational-theoretic foundation, underpinned by notions like Turing computability and Wolfram's computational irreducibility. Schelling showed, via many illustrations (Schelling, 1978), how interdependent decisions can lead to unexpected social phenomena, even though the individuals follow simple, or even simplistic, and identical rules.

If we trace the connection between cellular automata and economics, one goes all the way up to John von Neumann, whose pioneering study on self-reproducing automata (von Neumann *et al.*, 1966) laid out a general theory, along with his important contributions to general equilibrium theory and the theory of games in economics. Von Neumann

did not apply cellular automata to studying social or economic problems. Despite being the originator of the theory of cellular automata, von Neumann did not seem to have explored possible direct economic applications, as seen, for example, in Albin (1998, p. xv). Furthermore, the pioneering work by Thomas Schelling on checkerboard models was not *a priori* motivated by automata theory and hence may be viewed as serendipitous one-shot play rather than being a systematic intent to advance a new paradigm. These considerations lead us to place Peter Albin as a pioneer in endowing ACE with a general computation-theoretic underpinning. Albin (1982), reprinted as chapter 2 of Albin (1998), may be considered as one of the early articles to address computability issues in economics. Later on in his preface in Albin (1998) and the introductory chapter authored by Duncan Foley, they proposed a *Chomsky-Wolfram synthesis* as a framework to address complexity in economics. In this effort they were trying to find a thread passing through John von Neumann, Alan Turing, Noam Chomsky, Kurt Gödel, John Conway, and Stephen Wolfram. The thread, called the *automata-theoretic foundation* of economics also nicely connects computer science, linguistics, and dynamical systems.

Albin (1992) applies Wolfram's one-dimensional elementary cellular automata to his proposed spatial (network) prisoner's dilemma game. The class of the spatial prisoner's dilemma game not only provides the simplest explanation of the prevalence of cooperative behavior, but, more importantly, it also provides the first illustration that ACE models can be richly studied in the light of the theory of automata and the associated hierarchy of complexity.

Simon's awareness of the automata theory and its possible implications for the social sciences can be dated back to a very early stage of its development. It turns out that von Neumann in fact presented his work *The General Theory of Automata* at a session during the meeting of the Econometric Society held on September 5, 1950 at Harvard University, where Simon was a discussant.[2] Simon made important remarks concerning hierarchies of rationality and their connection with cellular automata. From Simon's discussion it is evident that social scientists did show interest in drawing an analogy between automata and social organisms at least as early as 1950, much earlier than Schelling's checkerboard models. It is worth noting that Simon's comment dates back to before his proposal of bounded rationality and his theorem proving machine (Newell *et al.*, 1958). Given this background, we explore more of the content of Simon's ideas that are intertwined with ACE in the rest of this chapter.

Agents as Programs: Bounded Rationality

First and foremost, the influence of Herbert Simon on ACE will be apparent once we understand the way agents in an economic system are conceived. Arguably, bounded rationality is one of the most famous terms coined by Simon, and this has been developed in many directions and has managed to acquire many different interpretations over the years. From its definition, as shown in Simon (1957), it states that human minds may often fail to solve problems to the level that is objectively optimal. Impressed by human ability to solve difficult problems, Simon began painstakingly to observe the *processes* of human thinking and devised computer programs to explain the qualitative and quantitative data that he gathered.

This initiative happened during the wave of the cognitive revolution in the mid-1950s and the 1960s, which is also considered to have marked the birth of artificial intelligence. The results of this research project by Simon gradually developed into the theory of human problem-solving, in which information processing systems (IPS) are models that characterize problem solvers in various domains (Newell and Simon, 1972). This approach lies at the foundation of information processing psychology – an important branch of cognitive science, knowledge engineering, and domain expertise in modern computer science. Simon's firm belief seems to be that we have to open the black box of decision-making and understand its procedural (algorithmic) aspects. Only then can we begin to appreciate human wisdom and the complexity of human societies.

LISP and Genetic Programming

We argue that it might be fruitful to associate bounds of rationality with the complexity of algorithms (or procedures) that human, or software, agents can handle and will potentially apply to solve the problems that they encounter. Simon himself has implicitly acknowledged, although with some caveats, this interpretation of bounded rationality through the concept of computational complexity.[3] In any case, the central idea in Simon's conception of an economic agent is that of a problem-solver; his emphasis was on understanding *how*; in other words, the procedural aspects. This demand for the transparency of agents is actively answered in ACE in various forms, such as simple programmed agents, entropy-maximizing agents (zero-intelligence agents), human-written programmed agents, and autonomous agents, which constitute a long

glossary of artificial agents with transparent behavioral rules (Chen, 2012).

In the context of the problem-solver analogy, it is appropriate to discuss Logic Theorist (Newell *et al.*, 1958), which is the very first realization of IPS. Simon viewed problem-solving in general as being analogous to theorem proving, where one starts with a current state and tries to search for the paths to reach the target states with the assumptions or propositions that are available to him/her. Logic Theorist was programmed with Information Processing Language, a kind of list processing. This language is strongly motivated as a practical implementation of the lambda calculus, or the recursive function theory, that was developed in the 1930s by Alonzo Church (1903–95), Stephen Kleene (1909–94) and Alan Turing (1912–54). It was later formally introduced as LISP by John McCarthy (1927–2011).

The syntax of LISP has very broad usage to represent problems in many different domains. The universality of LISP is briefly mentioned in the lecture notes of Newell and Simon's Turing Award (Newell and Simon, 1976). One of the successful examples utilizing LISP that Simon developed is the elementary perceiver and memorizer (EPAM) (Feigenbaum and Simon, 1984; Gobet and Simon, 1996). It is an important model in the theories of expert systems that can be used to explain the evolution and the organization of the associative memory of human subjects.

It is worth noting that genetic programming (GP) (Koza, 1992, 1994), which is used in ACE to model autonomous agents, is essentially inspired by the LISP environment (Figure 7.2). The connection between GP and Simon, however, goes beyond the syntax of LISP. Automatic theorem proving, which motivates problem solvers in Simon's idea of rationality and computation, also motivates a notion of autonomous agents in ACE. The latter point has been elaborated in Chen (2012). GP is a tool for autonomous learning or the evolution of programs (can be rules, strategies, or recipes) without any external intervention. It, therefore, equips artificial agents (program solvers) with a novelty-discovering or chance-discovering capability so that they may constantly exploit the surrounding environment without external intervention, which in turn may also cause the surrounding environment to change or react, and the cycle may continue indefinitely.

Genetic programming, as one of the most powerful models of autonomous agents, has been widely acknowledged in the ACE literature (Duffy, 2006; Chen, 2008). Applications of genetic programming

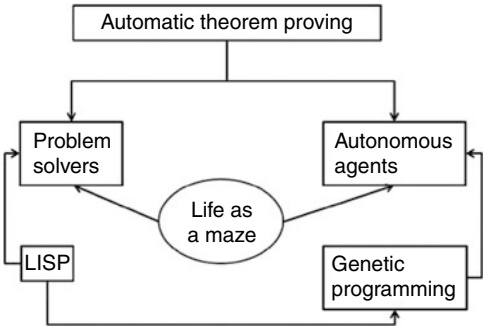


Figure 7.2 LISP and genetic programming

to modeling the constant search for better strategies or products have been illustrated in various ACE models, including double auction markets (Chen and Tai, 2003; Chen and Yu, 2011), artificial stock markets (Kampouridis *et al.*, 2012a,b), and oligopolistic competition with product innovation (Chie and Chen, 2013, 2014).

The modeling principle that leaves artificial agents a larger degree of autonomy arises because the problem introduced by the environment is frequently not well defined, and may vary with agents' perceptions of the problem. In this case, external intervention is neither necessary nor feasible, and leaving agents to go wild on their own is the proper way of modeling this process as they are placed in a jungle or maze, searching for 'truth' or proving a 'theorem'. This jungle is everywhere in life, but a good theoretic model that exemplifies such a complex and perplexing environment is not often seen. In this regard, Simon's story, *Apple* (Simon, 1991), provides an illuminating demonstration, a subject to which we now turn.

Environment as a Maze

Two features of human problem-solving (or decision-making, in general) that are recurrent in the discussions by Simon are *representations* and *procedures*, that is, *what* and *how*. Representation is the subjective description of the problem or the solution (goal) by the decison-making organism (not the observer!); a procedure, on the other hand, provides a sequence of actions which the problem-solver can follow to reach the desired solution. One of the common and important characteristics is that the representations and the procedures are not constant over time, but are dynamic (and are shaped by the perceptions of the

environment), even if the exact same problem is encountered by the problem-solver. The following quotation from Simon (1996b) describes the idea concisely:

> The external environments of thought, both the real world and long-term memory, undergo continual change. In memory the change is adaptive. It updates the knowledge about the real world and adds new knowledge. It adds new procedures that contribute to the skills in particular task domains and improve existing procedures. A scientific theory of human thinking must take account of this process of change in the contents of memory. (Ibid, p. 100)

Simon's story of *Apple* serves as the best material for illuminating this idea. This story tells us that the problem space in which we try to search for an answer may often be too enormous for us to have a complete picture of it. With limited time, energy, and memory to explore the environment in its totality, our tastes and goals are, in turn, shaped and altered from time to time during the process of search (either arbitrary or directed). The message of this metaphor is that even if the environment in which one is acting is absolutely static, we can end up having very different knowledge and representation of it, thereby creating different sub-problems for ourselves and developing different strategies to survive.

Hugo, the ordinary and solitary man in Simon's *Apple* story, who lives in the 'castle' all his life, portrays our everyday decision-making in environments whose entire picture is often not known to us. Hugo desires several aspects of daily life, such as preferred food, aesthetic surroundings, and comfort, while the limitation of resources he suffers is very rigid – his awake time, the energy he has before he collapses due to hunger, his memory, and a notebook on which he writes down the history of his adventure in detail.[4] A well-trained economist might soon formulate this problem in terms of optimizing multiple objectives subject to constraints. Simon realized very early on that the allocation or search problem based on marginal analysis does not work at all (Simon, 1983). Instead, Hugo is formally characterized as a set of rules that are gradually shaped by Hugo's understanding of the castle.

How is this related to the characterization of agents in the models of ACE? We address this question by teasing out the role of GP in these models. It is worth pointing out that the time-variant procedures used by agents to solve problems can be interpreted as a kind of evolutionary algorithm. If one subscribes to this interpretation, then its

relation with GP is quite straightforward. The evolutionary computation as demonstrated by GP allows us to capture the phenomenon of agents' changing (evolving) representations with the change in their employed modules or chunks,[5] and accordingly the change in their survival strategies in the light of possible changes in their preferences (the fitness criteria). The notion of adaptivity that Simon discussed is strongly inherited by GP in the context of economic agents. In this vein, the literature on modeling innovation in the light of consumers' changing preferences using GP is very relevant here (Chen and Chih, 2007; Chen and Wang, 2011).

Selectivity, Satisficing, and Aspiration Levels

In his Nobel prize lecture Simon (1979) spoke of the processes that people use to make difficult decisions and solve complex problems:

> Selectivity, based on rules of thumb or 'heuristics', tends to guide the search into promising regions, so that solutions will generally be found after search of only a tiny part of the total space. Satisficing criteria terminate search when satisfactory problem solutions have been found. (Ibid, p. 507)

In this section, our discussion is based upon the notion of intelligence (heuristics) – a major theme in Simon's research – in *characterizing different agents*. We do so against the backdrop of bounded rationality that is increasingly gaining acceptance as the appropriate way of characterizing agents even within mainstream economics. In particular, we focus on the *characterization of intelligence* in terms of *selectivity*, *satisficing*, and *aspiration levels*, the key components of bounded rationality as propounded by Simon, and see the development of ACE models in this light.

Selectivity

The idea of selectivity becomes important in situations when boundedly rational agents are acting in complex environments, where the problem space is huge and the agents are cognitively constrained. Humans' cognitive constraints, such as short-term memory or working-memory capacity, allow them to process a limited portion of information to which they can get access, and can only effectively deal with a limited number of alternatives at a time (Miller, 1956). In both cases – that is, a large problem space and limited cognitive capacity, selectivity helps them to deal with the difficulties of decision-making. While

these are the reasons why selectivity may be important, we also need to understand how such a selection is executed. One way for human beings to be effective in their selection is to apply familiar chunks.

In ACE, chunks can be acquired by autonomous agents through information encapsulation and compression, and genetic programming can allow us to model such capability of autonomous agents. The *automatic defined terminals* (ADTs), as proposed in Chen and Chih (2007) in their ACE model of product evolution, is an example. By searching for and encapsulating useful chunks, autonomous agents can compress the knowledge that they have acquired incrementally into simple but effective decision rules. Perhaps this avenue of exploration will develop GP in such a way that it can evolve fast and frugal heuristics (Gigerenzer and Selten, 2001; Gigerenzer, 2004). In fact, in some ACE models, genetic programming has already been applied to enable autonomous agents to develop their decision heuristics in the form of *evolutionary decision trees* (Kampouridis *et al.*, 2012a).[6]

It is worth mentioning that, partially in the light of the recent developments in cognitive experimental economics, some ACE models have explicitly considered the cognitive constraints (working-memory constraints) of autonomous agents. For example, the parameter *population size* of genetic programming or genetic algorithms has been chosen as a proxy for the working-memory capacity or, simply, intelligence, of agents (Casari, 2004; Chen and Tai, 2010). By setting artificial agents with different population sizes, their working-memory capacity, therefore, becomes heterogeneous. The consequences of heterogeneity in a human's cognitive capability can then be simulated by using these models jointly with human-subject experiments. This is entirely in the spirit and the vision of Simon (2000, pp. 34–6).

Satisficing and the Aspiration Level

Satisficing is the other intimately related notion which is of importance to ACE. Satisficing, as opposed to optimization, is the objective of a boundedly rational agent according to Simon (1955a). This objective is achieved by looking for good enough solutions, which in turn are judged by *aspiration levels*. Intuitively, satisficing is more general than optimization, because one can aspire to find the best. Satisficing is a natural consequence of a limited computational capacity and is also a common characteristic of various decison-making organisms.[7] In the story *Apple* (see section, Environment as a Maze), Hugo, with his set of heuristics (rules), is an example of a satisficing agent, whose

sub-problems can be seen as: *what to focus on*, *how to evaluate*, and *when to stop*.

A prototype of a computer agent (a machine with built-in mechanisms) behaving in a satisficing manner, and placed in an uncertain environment, can be found in Newell (1955). Newell's program appears to be a more concrete format of what Simon proposed (1955a, 1956), including his story *Apple*. Newell's intention (1955) is to program the computer to *learn* to play good chess. Chess is one of Simon's favorite theoretical settings, because its problem space is as massive as Hugo's castle and yet bounded; in fact, its possible states can be entirely derived from the rules of the game.

In Newell's program, the machine decides which action to take based on answering questions that are arranged in a *goal structure*, and the program is characterized by a set of rules. Learning happens when the set of rules changes over time.

In an act of deciding what to do, there are a few sub-problems that need to be answered first (an act of divide and conquer); they are problems related to consequences, horizons, evaluations, and alternatives (Newell, 1955). The key to answering these questions lies in a thorough understanding of heuristics, aspiration levels, and sets of rules that the program (computer agent) can use. The architecture of this program coincides with the idea of list processing and genetic programming, which suggests that the satisficing procedures can be more sensibly brought into ACE models.

In fact, the satisficing behavior has been extensively included in many ACE models, in particular, the recent advent of agent-based macroeconomic models. In these models, the behavioral adjustments of households and firms, ranging from consumption, pricing, production, and employment to wages, are not based on the pursuit of optimizing a specific target function, but are based on some satisficing criteria, normally formed as *threshold-based* rules or *routine-based rules* (Raberto *et al.*, 2008; Cincotti *et al.*, 2012). In fact, using what we did yesterday as a default except when some unusual conditions are met may be quite familiar; this habitual heuristic may be considered a kind of fast and frugal heuristics.

The adjustment of the aspiration level is one of the key components of satisficing behavior. Simon (1955a) perceives the aspiration level to be tied to the cost of search. This dynamics of the aspiration level has also been incorporated into the ACE literature through its acceptance of prospect theory in general (Mueller and de Haan, 2009; Cincotti *et al.*, 2010) and *reference points* as an important decision anchor (Hommes,

2011). In addition, recent developments in the field of ACE indicate that *social preference* can be another determinant of the dynamics of aspiration levels (Chen and Gostoli, 2012; Delli Gatti *et al.*, 2011; Zschache, 2012).

Having reviewed the connections between Simon and ACE at the level of agents (the left block in the middle layer of Figure 7.1), we now shift our focus to that of the system as whole (the right block).

Complex Systems: Modularity

The complex system approach to the social sciences is another aspect of Simon's legacy that has had an influence on ACE. One property that makes Simon's idea of complex systems unique is that of *hierarchy*, which is in fact a ubiquitous property of many complex systems. We think there is a need to raise the question as to whether Simon's idea of complex systems might have more of an impact on ACE today.

The way in which one can explore and understand a complex system is in itself nontrivial, especially if cognitively constrained agents are engaging in such a task. From an observer's point of view, we try to identify the key forces that govern the behavior of entities in the system. However, from a stakeholder's perspective, his limited capacity may only allow him to act locally with a few relevant entities. No matter which perspective we choose to adopt, a recurrent property that aids in understanding and coping with complex systems is that of modularity.

Modularity is also a modern terminology that refers to Simon's idea of hierarchy in complex systems (Callebaut and Rasskin-Gutman, 2005). The word modularity, although not coined or used by Simon, is an adequate and a broad concept that encompasses two distinct ideas promoted by Simon, namely, *near decomposability*[8] and *chunking*.

However, there is no clear distinction between the two mentioned in the literature. Therefore, we would like to emphasize that near decomposability and chunking are two sides of the same coin (modularity): chunking is the bottom up concept of the development of a complex being; and near decomposability is a top-down perspective for finding ways to simplify problems. Both of these notions have connections to hierarchy, and consequently to modularity. See Figure 7.3 for an illustration. Near decomposability is the unifying hypothesis that helped Simon to understand, as an observer, a variety of complex systems, such as physical systems, symbolic systems, human minds, and social systems (Simon, 1962, 1995, 1996a).

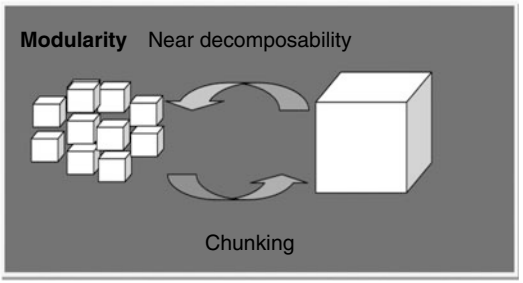


Figure 7.3 The relationships of modularity, near decomposability, and chunking

Chunking, on the other hand, involves assembling several symbols (pieces of information) into a unit for recognition and operation in problem-solving tasks. A chunk can further be joined with other chunks to form a bigger chunk. The ability to chunk is believed to be an important skill for any effective performance of complex tasks. The role of chunking (interpreted as recognizing and classifying patterns) in complex games, such as chess and *Go*, is well documented (Newell and Simon, 1972; Gobet *et al.*, 2004; Kao, 2013). We now explore both aspects – chunking and near decomposability - in greater detail with a focus on their relevance to ACE.

Complex Systems and Near Decomposability

The economic world that we study is often interconnected, complex and not fully decomposable. This in part renders the superiority of ACE, which focuses on the interactions among the actors in a system, and makes it a prominent alternative approach to conventional economic analysis (Stiglitz and Gallegati, 2011). However, when linkages among the actors are not equally strong, the weak feedbacks can be ignored in the short run. In that case, the description and the simulation of the model can be simplified. For builders of ACE, near decomposability can help make simulations better structured and a little smarter, following the principles of parsimony in modeling.

When sectors are dependent on each other, the analysis becomes more difficult. This depends on the degree of coupling that exists between sectors, given the specific change that is being investigated. One noticeable observation is that the interaction between two sectors is not necessarily symmetric (Goodwin, 1947). This asymmetric interaction is referred to as *unilateral coupling* and is related to the idea of *causal* ordering, which is investigated in Simon (1952, 1953) and Simon and

Rescher (1966). The notion of a nearly decomposable system plays the role of reducing the difficulty and complication of analyzing coupled systems, a good property of parsimony that science looks for (Simon, 2001).

Analogous to an economic system solving a problem, Simon considers the phenomena of thinking and the evolution of organisms themselves as problem-solving (Simon, 1995, 1996a). The key connection according to him is that the mind and organisms have a hierarchical structure.[9] If the structure is nearly decomposable, we can investigate the system with a certain degree of isolation. The speed of convergence – that is, finding a solution – *within* any subsystem is faster than it is *between* subsystems. Therefore, if the system is nearly decomposable, only the output of a certain subsystem will influence other subsystems. Furthermore, a nearly decomposable structure has certain implications for the nature of the evolution of the system as a whole. In Simon (2002), it is concluded that the organisms with a nearly decomposable structure, no matter how complex they are, will evolve faster than the organisms that have an indecomposable structure. Again, if evolution is viewed as a process of finding solutions, organisms with nearly independent components will need less time to find the solution than the ones without nearly independent components. Having discussed the importance of near decomposability in the study of complex systems in general, we now focus on the specific role of near decomposability within ACE.

Why or Why Not ACE Needs Near Decomposability

Taking the point of view of reductionism, all observed phenomena are emergent properties that trace their origin all the way back to quark. However, as a social scientist who tried to understand complex social phenomena, Simon was of the view that decomposition to a neural level was more than enough. In fact, Davis (2013) brings out the idea of decomposability to the emergence of ACE and suggests that the basic unit of simulation is not the individual, but the rules (bits) embedded in the individuals.

Simon's suggestion was to view the relations between each level of the system in a hierarchical fashion (Simon, 1962). Such relations are recurrent across the board according to Simon, regardless of whether they are between neurons and the mind, mind and behavior, individuals and organizations, or organizations and social systems. He observed that each of these systems, which are potentially complex, are governed by similar modular structures. We find that the ACE literature, as it stands

now, lacks a consensus on how best to characterize different institutions within models of economic and social systems. A comprehensive review of the ACE literature may lead one to find that most agent-based models developed so far are confined to only two levels, such as the individual-market hierarchy and the firm-market hierarchy.

The recent development of agent-based macroeconomics does provide a three-level hierarchy, individual–market–aggregate. Nonetheless, even in this case, these levels are given exogenously. For example, firms are not endogenously formed through individuals. Hence, by and large, it may be fair to say that we have not seen agent-based economic models as being able to demonstrate the evolving hierarchical, near decomposable, systems as an essential characterization of complex adaptive systems. One may anticipate that the agent-based models of organizations, specifically those focusing on internal organizations, may have some behavioral algorithms to form hierarchies endogenously. However, as Chang and Harrington (2006) have shown in their survey article, such work was, and still is, rare.

On the other hand, ACE has started to take inspiration from neuroeconomics (in particular the dual system hypothesis) to build software agents (Chen, 2014). It is also arguable for the ACE community to question whether the hierarchical fashion in a complex *social* system can be unambiguously defined. Information flow among actors is perhaps the essence of social computations. Due to breakthroughs in the Internet and platforms of social interaction, the format of information flow has gone beyond the conventional understanding that is underlined by institutions (for example, departments and colleges), and hence has transcended what Simon could have imagined in his time.

Social scientists might argue that in today's world, with the revolution in information technology, everyone is almost fully connected (via the Internet) and thus near decomposability might not be valid any more. So, one might argue that the complex social systems we are capable of studying today, like in ACE, have managed to go beyond the demands of near decomposability in terms of structure, and thereby have, in some respects, advanced beyond Simon's vision. However, it is important to remember that while social media provides new means for people to connect with each other, faster and in a less costly way, the question of purposiveness behind the establishment of such a connection or a formation of a social network is not yet fully understood. At the very least, we may say that the importance of near decomposability for ACE remains inconclusive.

Chunking Theory

Chunking theory has important applications in the organization and characterization of knowledge. Some of Simon's pupils and colleagues in the area of computer science had elaborated this approach in detail in relation to the field of *expert systems*, with particular reference to human cognition and memory, such as, Feigenbaum and Simon (1984); Gobet and Simon (1996, 2000); Gobet *et al*. (2004). The central questions of this research program involve understanding how human beings develop expertise in a specific, complex field or a task.

Following the Miller tradition (Miller, 1956), Simon was aware that the size or organization of chunks, rather than the number of chunks, matters for the excellent performance of agents. However, the natural result that happens to an expert who is in a field for over ten years is that there are more than 50,000 chunks that are accumulated over time (see, for example, Simon, 1996b, p. 89). If the size of short-term memory is really small and similar across human minds, then it is possible to conjecture that experts will further group the isolated chunks into bigger chunks and only retrieve the sub-chunks when necessary. In other words, experts may organize their chunks in a better and more complex fashion, and at the same time evolve better heuristics to access sub-portions if and when necessary. A piece of evidence can be found in Simon and Schaeffer (1992) which demonstrates the expert–novice differences in the task of memorizing chess board configuration. The experiment shows that an expert can reproduce a configuration from a famous game (around 25 pieces) quite correctly with only five seconds of staring time. Unsurprisingly, the novice can, at best, retrieve four to five arbitrary pieces. However, when both of them are presented with a random and 'meaningless' configuration, the expert and the novice perform equally badly. This immediately shows that the expert is not smarter, in terms of the ability to memorize, but has the ability to recognize the meaningful chunks on the board in a very short time.

One of the immutable laws in an ever-changing world is that individuals never cease to learn, and what Simon suggests is that they learn by chunking or modularizing. In fact, for a complex adaptive system to grow or to improve, the ability to chunk can be seen as a necessary property. To handle huge computer programs well, one requires a modular design, and to become an expert in a particular domain, one needs to have modular thinking and modular memory to retain an immense amount of symbols and information, despite the severe limitations of memory capacity.

Going back to the *Apple* story (see section, Environment as a Maze), Hugo definitely gains more understanding about the castle, and the more he knows the pickier he becomes. We know that he is learning to acquire tastes; however, it is not clear how he gradually prefers one kind of bread over others. His experience and memory play a heavy role in shaping his tastes. It is quite obvious that his behavior cannot be interpreted well in terms of Bayesian learning (a kind of optimal learning), because the state of affairs that Hugo experiences is neither fixed nor infinite.

It is important to note that GP also uses a similar, modular structure in its encapsulation of knowledge (Roberts *et al.*, 2001). Although chunking theory has been applied to agent-based modeling in other fields, for example, linguistics (Liang and Zhao, 2005), it has not yet gained popularity within the ACE community. We believe that the notion of modularity in genetic programming is, perhaps, the right direction to explore this property (Chen and Wang, 2011).

Scientific Discovery and Market Design

After seeing the connections between Simon and ACE at both individual and system levels, we now move to the top layer of Figure 7.1 and address the connection between the two in a simulation from an epistemological viewpoint. The layout of this section is first summarized in Figure 7.4. We begin by asking: what is the mode of unearthing new knowledge that ACE has to offer? How is it different from the other approaches that already exist, and can we know more? To answer this question, we need to focus on an important aspect of Simon's contributions to the philosophy of science and, in particular, to the logic of scientific discovery (Simon, 1977; Langley *et al.*, 1987).

Social sciences that empirically examine the complex interaction of entities are often categorized as sciences of induction. On the other hand, orthodox economic theorizing that is underpinned by axioms, assumptions and infinite iterations, is best approximated as a branch of applied mathematics, where the possible outcomes of the enquiry are obtained from deduction. While deduction and induction are the two familiar types of reasoning, one has to realize that agent-based computational modeling and simulation constitute neither a method of deduction (theory), nor a method of induction (statistical inference). The distinction from the usual deduction and induction has been acknowledged by economists and social scientists (Axelrod, 1997; Axelrod and Tesfatsion, 2006; Gallegati and Richiardi, 2011; Halas, 2011).

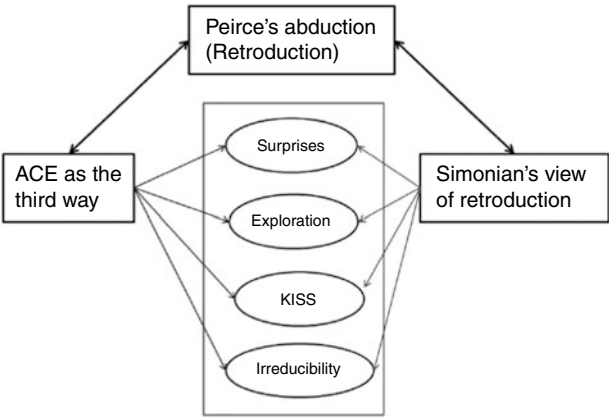


Figure 7.4 ACE and Simon through Peirce

Simon did notice the limitation of normal induction:

> Students are always told that they can't run a successful experiment if they don't have a hypothesis… I believe that is a very bad criterion for the design of experiments… If you look down the list of outstanding discoveries in the physical sciences or the biological sciences – look at Nobel awards in those fields – you will note that a considerable number of the prizes are given to people who had the good fortune to *experience a surprise*. (Simon, *et al.*, 1992, p. 22; italics added)

At this point, agent-based simulations are related to Simon's comment since some emergent phenomena coming out of agent-based simulations bring us novelties and surprises, which inspire us to make hypotheses of these observations (Figure 7.4). In this sense, some social scientists, such as Gallegati and Richiardi (2011) and Halas (2011), also relate agent-based social simulation to what Charles Peirce (1839–1914) called *abduction* or *retroduction*. Peirce advocated that there is a unique type of logical reasoning beyond deduction and induction, which he called abduction and suggested that it was the logic of discovery (Peirce, 1997). While, for many philosophers of science, abduction is treated as a part of induction, Peirce forcefully distinguished between the two by indicating that induction is about the test of an established hypothesis using observations, and that abduction is about the formation of the hypothesis.

While Simon did not directly address agent-based modeling as the third way of doing science, the connection between Simon and Peirce on abduction should not go unnoticed. Simon (1998) explicitly supports the view that the aim of science is to discover, and the process of discovery is neither in the form of pure deduction, nor pure induction. However, Peirce's original notion of abduction is not entirely clear and its vagueness has invoked many objections. Simon (1973) actually clarifies the notion of Peirce's *retroduction*:

> Peirce coined the term 'retroduction' as a label for the systematic processes leading to discovery . . . It is the aim of this paper to clarify the nature of retroduction, and to explain in what sense one can speak of a 'logic of discovery' or 'logic of retroduction'. (Simon (1973, pp. 471–2)

Simon (1973) presented two examples to demonstrate the retroductive process of law discovery, recoding a sequence of letters and concept attainment. He stated that 'A *law discovery process*' is a process for recoding, in parsimonious fashion, sets of empirical data. This passage can be understood by reading Simon (2001), where he elaborates that parsimony is the criterion for choosing among possible explanations. This idea has the agreement of what is stated as being the 'best explanation' in the following quotation:

> Peirce's abduction is now generally identified with a more developed and refined version called inference to the best explanation (Harman, 1965; Lipton, 2004), which seems to solve the problem of both what hypothesis we draw from available data, as well as why we prefer that particular hypothesis. (Halas, 2011)

Hence, abductive reasoning is close to hypothesis discovery. When a phenomenon is observed, the first question is to find hypotheses that can explain the phenomenon (what), but there is always more than one explanation, so the next question is to find the best explanation (why). The simplicity principle or the parsimony principle underlying scientific discovery, also known as *Occam's razor*, as suggested by Simon, has a great influence on ACE practice. In the ACE community, the parsimony principle has received an even more romantic name, the KISS (*keep it simple, stupid*) principle, as originally proposed by Robert Axelrod (Axelrod, 1997).[10]

The cellular-automaton models (see section on Computation-Theoretic Underpinning), such as the Schelling segregation model (Schelling, 1971), the Albin Spatial Prisoners' dilemma model (Albin, 1992), and the Keenan-O'Brien local oligopolistic competition model (Keenan and O'Brien, 1993), show how complex patterns can be formed using simple agents interacting with each other in a social network by following simple rules. The key message of these models is twofold. First, complex unpredictable patterns can emerge from very simple homogeneous interacting behavior. Second, a small change in the individual rule may fundamentally change the nature of the system dynamics from a lower hierarchy of complexity to a higher hierarchy of complexity. This also motivates the name, the 'edge' of chaos, that is, a slight change of the rule on this edge will result in either a stable or a chaotic pattern. The unexpected complexity of the behavior of these simple rules leads us to suspect that *complexity in nature may be due to similar simple mechanisms*.

Epstein and Axtell (1996) was probably the first study to introduce an agent-based model in the study of big history. In Epstein and Axtell (1996), the fundamental collective behaviors, such as group formation, cultural transmission, combat, and trade, are seen to emerge from the interaction of individual agents following a few simple rules. In agent-based financial models, it is found that the models with simple heterogeneity and simple rules, in particular variations of the fundamentalist–chartist model, are sufficient to replicate a number of stylized facts. A complex extension of this model may gain additional explanatory power, but, so far, this power has not been well exploited (Chen *et al.*, 2012). In addition, the simple model makes the later econometric estimation much more feasible.

Maybe, the most prominent example is the simple device, the zero-intelligence agents (Gode and Sunder, 1993). The zero-intelligence device is actually the application of the *maximum entropy approach* to agent-based modeling (Chen, 2012). The capability of this approach to replicate complex financial dynamic systems shows that some aggregate phenomena generated from human-agent systems with complex motives and behavioral rules can be rather well approximated by a system with simple agents characterized by simple motives and simple rules. In a sense, it indicates that adding more complex strategies to the agent-based models may have little macroscopic effect if these complex strategies may interact in such a way that they annihilate each other's forces. It is this possibility that prompts many of us to

think about a general physical system which is equipped with the most rudimentary forces but can overarch several seemingly unrelated social phenomena, for example, from pedestrian counter flow to the Schelling segregation model (Vinković and Kirman, 2006), to the El Farol Bar problem (minority games), and then to financial markets.

Preferential Attachment

The search for a universal pattern underlying different disciplinary phenomena is in line with the pursuit of the parsimony principle. Efforts in search of the universal pattern have been elucidated by Simon (1955b), who tried to identify a class of distributions applicable to rather extensive social and natural phenomena. These distributions include two skewed distributions, which are frequently cited in the ACE community, one being the Pareto distribution of income and the other the Zipf distribution of the frequency of the occurrence of words. Simon's pioneering work provided an empirical foundation for one kind of universality which motivates physicists to work on economics or the social sciences.

The skewed distribution studied by Simon has been constantly followed and extended by others in the economic literature and, recently, also pursued by the ACE community. The development of this literature can be roughly characterized by three directions. First, skewed distributions are found to be applicable to many more economic variables. In addition to income and wealth, they have also been applied to firm size, asset returns, city size, firm returns, innovation size, and so on (Gabaix, 2008). The second direction concerns the statistical or econometric techniques chosen to identify the appropriate skewed distribution among many possibilities. In addition to the frequently cited Pareto and Zipf distributions, there are lognormal and Yule distributions, plus many generalizations of them that are often considered by the ACE community. These distributions may look similar when simply scanned. Therefore, the distinction among them requires deliberate statistical analysis (Gallegati *et al.*, 2006).

One important reason for distinguishing different skewed distributions is that they may be associated with different underlying mechanisms. An example shown by Simon is that, depending on the underlying stochastic process, i.e., whether or not a steady introduction of new firms is involved,[11] one can have either a Yule distribution or lognormal distribution (Simon and Bonini, 1958). Therefore, the third development in this line is to build a theory or offer an explanation

that underlies these distributions. The mechanism proposed by Simon is a cumulative advantage mechanism, which is based on earlier work by a British statistician Udny Yule (1871–1951). Later on, this mechanism, also known as preferential attachment, was applied to form the scale-free network by Albert-László Barabási and Réka Albert (Barabási and Albert, 1999), and had a great influence on the literature of complex networks in general (Mitzenmacher, 2004; Gabaix, 2008) and on ACE models specifically (Alfarano and Milakovic, 2009; Alam and Geller, 2012; Cederman, 2002; Page, 2012).[12]

Simulation and Design

There are two related ways in which simulation can provide new knowledge – one of them obvious, the other perhaps a bit subtle. The obvious point is that, even when we have correct premises, it may be very difficult to discover what they imply. All correct reasoning is a grand system of tautologies, but only God can make direct use of that fact. The rest of us must painstakingly and fallibly tease out the consequences of our assumptions. (Simon, 1996b, p. 15).

The uniqueness of ACE, as the third way of doing science, is its use of simulation as a primary tool for discovery. In other words, if we don't run a simulation, we simply cannot be assured what may happen. This property is coined as *computational irreducibility* by Stephen Wolfram (Wolfram, 2002). Wolfram argues that a new kind of science can be considered a paradigm shift toward the study of computationally irreducible phenomena. If one applies this irreducibility characterization to economics or the social sciences, one can equally perceive a new kind of economics or a new kind of social science; for example, see Borrill and Tesfatsion (2011):

> Not only in practice, but now also in theory, we have come to realize that the only option we have to understand the global properties of many *social systems* of interest is to build and run computer models of these systems and observe what happens. (Ibid., p. 230; italics added.)

As also mentioned in Vriend (1995):

> we are interested in those regularities that cannot be deduced from the built-in properties of the individual agents or some other microeconomic aspect of the model; at least not by any argument which is substantially shorter than producing that regularity by running the simulation itself.' (Ibid, p. 212)

As a footnote to this quote, Vriend added:

> Clearly, the emergent behavior and self-organization *are* a function of the underlying configuration. The relevant point is, however, the following. Given a certain model with a certain parametrization, can one reason, that is, without running a simulation, *which* functions of the parametrization the outcomes are? (Ibid, p. 228)

Through agent-based modeling and simulation, one can navigate the territory of computational irreducibility, and explore both 'known unknowns' and 'unknown unknowns.' This endeavor may be useful for market design and policy design. Since market design is not just about a set of rules operating the market, but also involves agents' behavior under these rules, either individually or collectively through their interactions, it will not be hard to be convinced that market processes can generally be computationally irreducible. If so, market design as a science can benefit from the involvement of agent-based modeling and simulation. Currently, ACE have been applied to financial markets, electricity markets, fish markets, housing markets, school admission systems, national lottery design, tax evasion, futures markets, and labor markets. It is a matter of time to see when state of the art will advance into reality.

For Simon, simulations, or viewing economic entities' behavioral rules as computer programs, are ways to explore problems and find out possible solutions:

> The use of computer simulations will also enable economics to build realistic theory of firm that will go far beyond the traditional production function and short- and long-run cost curves into characteristics of organization structure and human motivation and their consequences for the decision-making process. (Simon, 2000, p. 36)

ACE can be seen as a continuation of Simon's quote. It will provide a realistic theory of markets, hierarchies, and networks by bringing light to deep darkness in the sea of complexity.

Concluding Remarks

Based on what we have reviewed in this chapter, it is clear that Simon's connection to ACE is probably more comprehensive than

that of Thomas Schelling, Elinor Ostrom, or Friedrich Hayek. Yet this connection has been much ignored by the ACE community, and, sometimes, has been simplified to just bounded rationality. In this chapter, from a computational-theoretic underpinning, to artificial cognitive and psychological agents, to complex systems, and further to the epistemology of simulation (Figure 7.1, once again), we give a more thorough systematic treatment, by extending Simon's one-dimensional connection with ACE into a multi-dimensional one.

Of course, by connection, we carefully mean that the current state of ACE was developed largely outside of Simon's influence. However, establishing this connection can still be useful. For example, many ACE models tend to ignore their computational theoretical underpinnings, and underestimate the complexity of the ACE models as abstract machines, such as Turing machines. This ignorance and the subsequent ignorance of the undecidability property may lead us to be overconfident of the effectiveness of validation and robustness checks when an effective algorithm to perform these jobs does not even exist (Velupillai and Zambelli, 2011).

In addition, by connection we also mean that ACE may not fully stand on the same side as each argumentation made by Simon. Whether the system has to be near decomposable to be scientifically interesting is not immediately clear for ACE. Presumably, every single unit (decision maker) can depend upon every other decision unit without being guaranteed a fixed sub-structure. The network connecting them may constantly evolve (Davis, 2013), which makes it very challenging to identify near decomposable subsystems. However, at this point, ACE cannot formally address this question because, as we mentioned earlier, most ACE models have only two layers. A truly hierarchical ACE model is yet to be seen. Despite that, by standing on Simon's shoulders, we hope that ACE, when it reaches maturity, can provide useful knowledge for extracting gems from complexity.[13]

Notes

1. John Davis (Davis, 2013) has recently made an attempt to place agent-based models in the context of Simon (1962). The essence of Davis (2013) will be reviewed later in this chapter.
2. Simon's discussion, along with all other abstracts of papers presented at the meeting, are in *Report of the Harvard Meeting, August 31–September 5, 1950*, *Econometrica*, 19, (1), 55–72. The one-page discussion was later collected by Simon in Simon (1977), chapter 4.1.
3. See Velupillai and Kao (2014) for the details.

4. Hugo's notebook is an analogy of long-term memory, in our view. After he collected enough data in the notebook, he began to identify patterns from history. For example, he even started to infer the correlation between the color of the wall and the food presented on the table. In order to make his choice slightly more effective, he needs to organize his database for good inductions. Simon seemed to imply that inference is a very natural action tht human beings acquire in the course of decision making.
5. In genetic programming, these primitives are known as terminals or functions. Hence, technically speaking, the agents' set of terminals and functions may change over time, which helps them gain a different representation of the problem surrounding them, even though the environment remains unchanged.
6. There are a large number of ACE models built upon evolutionary computation algorithms, including genetic algorithms (GAs) and genetic programming (Chen, 2002). These algorithms may be considered biased search in an immense space, which is close in spirit to Simon's selectivity. In GAs, chunks are known as *building blocks*. The implicit parallelism applied to evaluate a large number of building blocks allows us to identify the promising search area, instead of a blindly random search.
7. For a formal discussion of this idea, the interested reader is referred to Velupillai (2000, 2010a, 2010b).
8. When the data are organized in a matrix, then decomposability can be understood with a rigorous mathematical definition. A square matrix A is said to be decomposable if there exists a permutation matrix P such that

$$PAP^T = \begin{bmatrix} B & 0 \\ C & D \end{bmatrix}$$

   Otherwise A is indecomposable. If 0 is replaced by a small amount o, then A is nearly decomposable.
9. Such a structure also appears in programs. Koza made it very clear in the introduction of *Genetic Programming II* (1994) that the *automatically defined function* can successfully solve many complex problems, especially when three-step hierarchical problem solving (divide and conquer) is activated. The three steps are decomposing, solving the sub problems, and solving the original problem. This is squarely within Simon's approach to problem solving.
10. Having said that, we must also point out the opposition to this principle. The interested reader is referred to the special issue on 'The Methodology of Simulation Models' of the *Journal of Artificial Societies and Social Simulation,* 12 (4), 2009. Furthermore, even though the simplicity principle is generally known as the *minimum description length* (MDL) *principle* and may be regarded as a *generalized maximum likelihood principle* (Rissanen, 1989), one should be ready to accept any 'surprise' that the ACE model may offer, and one such kind of surprise is the inconsistency between the micro motives and macro behavior (Schelling, 1978). In fact, given the observed aggregate phenomenon, by the simplicity principle, the most compelling hypothesis is the one that is consistent between the micro and the macro level. Nothing can be simpler than linear scaling-up. However, if we do so, we are back

to the mainstream representative-agent approach to economics, and are no longer doing ACE. Hence, what makes the Schelling model intriguing is that the observed segregation phenomenon can actually emerge from a group of people who can each be tolerant of different kinds (ethnicities) of people.

11. The stochastic process governing the birth of new firms is described in section 2 of Simon (1955b), in particular, assumption 2.
12. The preferential attachment rule is an intuitive behavioral rule for new nodes (newcomers, immigrants) in forming personal networks with existing nodes (local residents). Basically, newcomers will consider who are the most important persons in the town and attach higher probabilities to connecting with them. In the Barabási-Albert scale-free model, the importance is measured by the number of connections. Hence, the nodes that have already been connected extensively will attract more newcomers than those who are less connected.
13. The authors would like to thank Professor K. Vela Velupillai, without implicating him in their errors, for his inspiration and influence that has caused them to think deeply about themes advanced by Herbert Simon. The authors are also grateful to Dr Ragupathy Venkatachalam for many of the discussions and suggestions that he provided to improve the chapter. The research support in the form of the Ministry of Science and Technology (MOST) Grant, MOST 103-2410-H-004-009-MY3, is gratefully acknowledged.

References

Alam, S. J. and Geller, A. (2012). Networks in Agent-Based Social Simulation. In *Agent-Based Models of Geographical Systems*, ed. A. J. Heppenstall, A. T. Crooks, L. M. See and M. Batty. Heidelberg: Springer.

Albin, P. S. (1975). *The Analysis of Complex Socioeconomic Systems*. Lexington: Lexington Books.

Albin, P. S. (1982). The Metalogic of Economic Predictions, Calculations and Propositions. *Mathematical Social Sciences* 3 (4), 329–58.

Albin, P. S. (1992). Approximations of Cooperative Equilibria in Multi-person Prisoners' Dilemma Played by Cellular Automata. *Mathematical Social Sciences, 24* (2), 293–319.

Albin, P. S. (1998). *Barriers and Bounds to Rationality: Essays on Economic Complexity and Dynamics in Interactive Systems*. Princeton: Princeton University Press.

Alfarano, S. and Milakovic, M. (2009). Network structure and *N*-dependence in agent based herding models. *Journal of Economic Dynamics & Control,* 33, 78–92.

Axelrod, R. (1997). Advancing the Art of Simulation in the Social Sciences. In *Simulating Social Phenomena*. Heidelberg: Springer, 21–40.

Axelrod, R. and Tesfatsion, L. (2006). Appendix A: A Guide for Newcomers to Agent-Based Modeling in the Social Sciences. *Handbook of computational economics 2*, 1647–59.

Barabási, A.-L. and Albert, R. (1999). Emergence of Scaling in Random Networks. *Science,* 286 (5439), 509–12.

Borrill, P. L. and Tesfatsion, L. (2011). Agent-Based Modeling: The Right Mathematics for the Social Sciences? In *The Elgar Companion to Recent Economic Methodology*, ed. J. B. Davis. Cheltenham: Edward Elgar, 228–54.

Callebaut, W. and Rasskin-Gutman D. (eds) (2005). *Modularity: Understanding the Development and Evolution of Natural Complex Systems*. Cambridge, MA: MIT Press.

Casari, M. (2004). Can Genetic Algorithms Explain Experimental Anomalies? *Computational Economics,* 24 (3), 257–75.

Cederman, L.-E. (2002). Agent-Based Modeling in Political Science. *The Political Methodologist* 10 (1), 16–22.

Chang, M.-H. and Harrington, J. E. (2006). Agent-Based Models of Organizations. In L. Tesfatsion and Judd, K. L. (eds), *Handbook of Computational Economics*, Vol. 2, ch. 26. Amsterdam: Elsevier, 1273–337.

Chen, S.-H. (2002). *Evolutionary Computation in Economics and Finance*, Volume 100. Heidelberg: Springer.

Chen, S.-H. (2005). Computational Intelligence in Economics and Finance: Carrying on the Legacy of Herbert Simon. *Information Sciences*, 170, 121–31.

Chen, S.-H. (2008). Computational Intelligence in Agent-Based Computational Economics. In *In Computational Intelligence: A Compendium*, ed. J. Fulcher and L. C. Jain. Heidelberg: Springer, 517–94.

Chen, S.-H. (2012). Varieties of Agents in Agent-Based Computational Economics: A Historical and an Interdisciplinary Perspective. *Journal of Economic Dynamics and Control,* 36, 1–25.

Chen, S.-H. (2014). Neuroeconomics and Agent-Based Computational Economics. *International Journal of Applied Behavioral Economics*, 3 (2), 15–34.

Chen, S.-H., Chang, C.-L. and Du, Y.-R. (2012). Agent-Based Economic Models and Econometrics. *The Knowledge Engineering Review,* 27 (02), 187–219.

Chen, S.-H. and Chih, B.-T. (2007). Modularity, Product Innovation, and Consumer Satisfaction: An Agent-Based Approach. In *Intelligent Data Engineering about Automated Learning, IDEAL 2007, LNCS 4881*, ed. H. Yin, P. Tino, E. Corchado, and W. Byrne. Heidelberg: Springer.

Chen, S.-H. and Gostoli, U. (2012). Coordination in the El-Farol Bar problem: The Role of Social Preferences and Social Networks. In *2012 IEEE Congress on Evolutionary Computation (CEC)*, 1–8. IEEE.

Chen, S.-H. and Tai, C.-C. (2003). Trading Restrictions, Price Dynamics and Allocative Efficiency in Double Auction Markets: Analysis Based on Agent-Based Modeling and Simulations. *Advances in Complex Systems,* 6 (3), 283–302.

Chen, S.-H. and Tai, C.-C. (2010). The Agent-Based Double Auction Markets: 15 Years on. In *Simulating Interacting Agents and Social Phenomena*, ed. K. Takadama, C. Cioffi-Revilla, and G. Deffuant. Heidelberg: Springer, 119–36.

Chen, S.-H. and Wang, S. G. (2011). Emergent Complexity in Agent-Based Computational Economics. *Journal of Economic Surveys,* 25 (3), 527–46.

Chen, S.-H. and Yu, T. (2011). Agents Learned, but Do We? Knowledge Discovery Using the Agent-Based Double Auction Markets. *Frontiers of Electrical and Electronic Engineering in China*, 6 (1), 159–70.

Chie, B.-T. and Chen, S.-H (2013). Non-Price Competition in a Modular Economy: An Agent-Based Computational Model. *Economia Politica,* XXX (3), 273–99.

Chie, B.-T. and Chen, S.-H (2014). Competition in a New Industrial Economy: Toward an Agent-Based Economic Model of Modularity. *Administrative Sciences,* 4 (3), 192–218.

Cincotti, S., Raberto, M. and Teglio, A. (2010). Credit Money and Macroeconomic Instability in the Agent-Based Model and Simulator EURACE. *Economics: The Open-Access, Open-Assessment E-Journal,* 4, 1–32.

Cincotti, S., Raberto, M. and Teglio, A. (2012). The EURACE Macroeconomic Model and Simulator. In *Agent-based Dynamics, Norms, and Corporate Governance. The proceedings of the 16th World Congress of the International Economic Association, Volume 2*. Basingstoke: Palgrave Macmillan.

Davis, J. B. (2013). The Emergence of Agent-Based Modeling in Economics: Individuals Come Down to Bits. *Filosofia de la Economia,* 1 (2), 229–46.

Delli Gatti, D., Desiderio, S., Gaffeo, E., Cirillo, P. and Gallegati, M. (2011). *Macroeconomics from the Bottom-up*, *Volume 1*. Heidelberg: Springer.

Duffy, J. (2006). Agent-based models and human subject experiments. In *Handbook of Computational Economics*, ed. L. Tesfatsion and K. L. Judd. Amsterdam: Elsevier, ch. 19, 49–1011.

Epstein, J. M. and Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, D.C.: Brookings Institution Press.

Feigenbaum, E. A. and Simon, H. A. (1984). EPAM-like Models of Recognition and Learning. *Cognitive Science,* 8 (4), 305–36.

Gabaix, X. (2008). *Power Laws in Economics and Finance*. Technical report, National Bureau of Economic Research.

Gallegati, M., Keen, S., Lux, T. and Ormerod, P. (2006). Worrying Trends in Econophysics. *Physica A: Statistical Mechanics and its Applications,* 370 (1), 1–6.

Gallegati, M. and Richiardi, M. G. (2011). Agent Based Models in Economics and Complexity. In, *Complex Systems in Finance and Econometrics*, ed. R. A. Meyers. Heidelberg: Springer, 30–53.

Gigerenzer, G. (2004). Fast and Frugal Heuristics: The Tools of Bounded Rationality. In *Blackwell Handbook of Judgment and Decision Making*, ed. D. J. Koehler and N. Harvey. Oxford: Blackwell.

Gigerenzer, G. and Selten, R. (eds) (2001). *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press.

Gobet, F., de Voogt, A. and Retschitzki, J. (2004). *Moves in Mind: the Psychology of Board Games*. New York: Psychology Press, Taylor & Francis.

Gobet, F. and Simon, H. A. (1996). Templates in Chess Memory: A Mechanism for Recalling Several Boards. *Cognitive Psychology,* 31 (1), 1–40.

Gobet, F. and Simon, H. A. (2000). Five Seconds or Sixty? Presentation Time in Expert Memory. *Cognitive Science,* 24 (4), 651–82.

Gode, D. K. and Sunder, S. (1993). Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality. *Journal of Political Economy,* 101 (1), 119–37.

Goodwin, R. M. (1947). Dynamical Coupling with Special Reference to Markets Having Production Lags. *Econometrica,* 15 (3), 181–203.

Halas, M. (2011). Abductive Reasoning as the Logic of Agent-Based Modelling. In *Proceedings of the 25th European Conference on Modelling and Simulation*, ed. T. Burczynski, J. Kolodziej, A. Byrski, and M. Carvalho. European Council for Modelling and Simulation.

Hommes, C. (2011). The Heterogeneous Expectations Hypothesis: Some Evidence from the Lab. *Journal of Economic Dynamics and Control*, 35 (1), 1–24.

Janssen, M. A. and Ostrom, E. (2006). Empirically Based, Agent-Based Models. *Ecology and Society*, 11 (2), 37.

Kampouridis, M., Chen, S.-H. and Tsang, E. (2012a). Market Fraction Hypothesis: A Proposed Test. *International Review of Financial Analysis*, 23, 41–54.

Kampouridis, M., Chen, S.-H. and Tsang, E. (2012b). Microstructure Dynamics and Agent-Based Financial Markets: Can Dinosaurs Return? *Advances in Complex Systems,* 15 (supp02).

Kao, Y.-F. (2013). *Studies in Classical Behavioural Economics*. Ph D thesis, University of Trento, Italy.

Keenan, D. C. and O'Brien M. J. (1993). Competition, Collusion, and Chaos. *Journal of Economic Dynamics and Control,* 17 (3), 327–353.

Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* Cambridge, MA: MIT Press.

Koza, J. R. (1994). *Genetic Programming II: Automatic Discovery of Reusable Programs*. Cambridge, MA: MIT Press.

Langley, P., Simon, H. A., Bradshaw, G. L. and Zytkow, J. M (1987). *Scientific Discovery: Computational Explorations of the Creative Processes*. Cambridge, MA: MIT Press.

Liang, Y.-H. and Zhao, T.-J. (2005). Distributed English Text Chunking Using Multiagent Based Architecture. In *MICAI 2005: Advances in Artificial Intelligence*, *Lecture Notes in Computer Science*, ed. A. Gelbukh, A. de Albornoz, and H. Terashima-Marin. Berlin Heidelberg: Springer, 752–60.

Marks, R. (2006). Market Design Using Agent-Based Models. In *Handbook of Computational Economics*, ed. L. Tesfatsion and K. L. Judd. Amsterdam: Elsevier, ch. 27, 1339–80.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review,* 63 (2), 81–97.

Mitzenmacher, M. (2004). A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics,* 1 (2), 226–251.

Mueller, M. G. and de Haan, P. (2009). How Much Do Incentives Affect Car Purchase? Agent-Based Microsimulation of Consumer Choice of New Cars – Part I: Model Structure, Simulation of Bounded Rationality, and Model Validation. *Energy Policy*, 37 (3), 1072–82.

Newell, A. (1955). *The Chess Machine: An Example of Dealing with a Complex Task by Adaptation*. Technical Report P-620, The Rand Corporation.

Newell, A., Shaw, J. C. and Simon, H. A. (1958). Elements of the Theory of Human Problem Solving. *Psychological Review*, 65 (3), 151–66.

Newell, A. and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs: Prentice-Hall.

Newell, A. and Simon, H. A. (1976). Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM,* 19 (3), 113–26.

Page, S. E. (2012). Aggregation in Agent-Based Models of Economics. *The Knowledge Engineering Review*, 27 (2), 151–62.

Peirce, C. S. (1997). *Pragmatism as a Principle and Method of Right Thinking: The 1903 Harvard Lectures on Pragmatism*. Albany: SUNY Press.

Raberto, M., Teglio, A. and Cincotti, S. (2008). Integrating Real and Financial Markets in an Agent-Based Economic Model: an Application to Monetary Policy Design. *Computational Economics,* 32 (1–2), 147–62.

Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry, Volume 511*. Singapore: World Scientific.

Roberts, S. C., Howard, D. and Koza, J. R. (2001). Evolving Modules in Genetic Programming by Subtree Encapsulation. In *Genetic Programming, 4th European Conference, EuroGP 2001 Lake Como, Italy, April 18–20, 2001 Proceedings*, ed. J. M. Marco, T. P. L. Lanzi, C. R. A. G. Tettamanzi, and W. B. Langdon, 160–75.

Schelling, T. C. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1 (2), 143–86.

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. New York: North.

Simon, H. A. (1952). On the Definition of the Causal Relation. *The Journal of Philosophy*, 49 (16), 517–28.

Simon, H. A. (1953). Casual Ordering and Identifiability. In *Studies in Econometric Method*, ed. W. C. Hood and T. Koopmans. New York: Wiley London: Chapman & Hall.

Simon, H. A. (1955a). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, 69 (1), 99–118.

Simon, H. A. (1955b). On a Class of Skew Distribution Functions. *Biometrika*, 42 (3/4), 425–40.

Simon, H. A. (1956). Rational Choice and the Structure of the Environment. *Psychological Review*, 63 (2), 129–38.

Simon, H. A. (1957). *Models of Man*. New York: Wiley.

Simon, H. A. (1959). Theories of Decision-Making in Economics and Behavioral Science. *The American Economic Review*, 49 (3), 253–83.

Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society, 106* (6), 467–82.

Simon, H. A. (1973). Does Scientific Discovery Have a Logic? *Philosophy of Science, 40* (4), 471–480.

Simon, H. A. (1976). From Substantive to Procedural Rationality. In *Method and Appraisal in Economics*, ed. S. J. Latsis. Cambridge: Cambridge University Press, 129–48.

Simon, H. A. (1977). *Models of Discovery and Other Topics in the Methods of Science*. Dordrecht: Reidel.

Simon, H. A. (1979). Rational Decision Making in Business Organization. *The American Economic Review*, 69 (4), 493–513.

Simon, H. A. (1983). *Reason in Human Affairs*. Oxford: Basil Blackwell.

Simon, H. A. (1991). *Models of My Life*. Cambridge, MA: MIT Press.

Simon, H. A. (1995). Near Decomposability and Complexity: How a Mind Resides in a Brain. In *The Mind, the Brain, and Complex Adaptive Systems*, *Volume XXII of Santa Fe Institute Studies in the Sciences of Complexity*, ed. H. J. Morowitz and J. L. Singer. Boston: Addison-Wesley.

Simon, H. A. (1996a). Machine as Mind. In *Machines and Thought – The Legacy of Alan Turing*, ed. P. Macmillan and A. Clark. Oxford: Oxford University Press, 1 (5), 81–101.

Simon, H. A. (1996b). *The Sciences of the Artificial* (3rd edn). Cambridge, MA: MIT Press.

Simon, H. A. (1998). Discovering Explanation. *Minds and Machines*, 8 (1), 7–37.

Simon, H. A. (2000). Bounded Rationality in Social Science: Today and Tomorrow. *Mind & Society*, 1 (1), 25–39.

Simon, H. A. (2001). Science Seeks Parsimony, Not Simplicity: Searching for Pattern in Phenomena. In *Simplicity, Inference and Modelling: Keeping it*

Sophisticatedly Simple, ed. A. Zellner and H. A. Keuzenkamp. Cambridge: Cambridge University Press.
Simon, H. A. (2002). Near Decomposability and the Speed of Evolution. *Industrial and Corporate Change*, 11 (3), 587–99.
Simon, H. A. and Bonini, C. P. (1958). The Size Distribution of Business Firms. *The American Economic Review*, 607–17.
Simon, H. A., Egidi, M. and Marris, R. L. (1992). *Economics, Bounded Rationality and the Cognitive Revolution*. Cheltenham: Edward Elgar Publishing.
Simon, H. A. and Rescher, N. (1966). Cause and Counterfactual. *Philosophy of Science*, 33 (4), 323–40.
Simon, H. A. and Schaeffer, J. (1992). The Game of Chess. In *Handbook of Game Theory with Economic Application, Volume 1*, ed. R. J. Aumann and S. Hart (). Amsterdam: Elsevier.
Stiglitz, J. E. and Gallegati, M. (2011). Heterogeneous Interacting Agent Models for Understanding Monetary Economies. *Eastern Economic Journal*, 37, 6–12.
Velupillai, K. V. (2000). *Computable Economics*. Oxford: Oxford University Press.
Velupillai, K. V. (2010a). *Computable Foundations for Economics*. London: Routledge.
Velupillai, K. V. (2010b). Foundations of Boundedly Rational Choice and Satisficing Decisions. *Advances in Decision Sciences 2010*, 16 pages.
Velupillai, K. V. and Kao, Y.-F. (2014). Computable and Computational Complexity Theoretic Bases for Herbert Simon's Cognitive Behavioral Economics. *Cognitive System Research*, 29–30, 40–52.
Velupillai, K. V. and Zambelli, S. (2011). Computing in Economics. In *The Elgar Companion to Recent Economic Methodology*, ed. J. Davis and W. Hands. Cheltenham: Edward Elgar.
Vinković, D. and Kirman, A. (2006). A Physical Analogue of the Schelling Model. *Proceedings of the National Academy of Sciences*, 103 (51), 19261–5.
von Neumann, J., (completed by Burks, A. W.), 1966). Theory of Self-Reproducing Automata. *IEEE Transactions on Neural Networks*, 5 (1), 3–14.
Vriend, N. J. (1995). Self-organization of Markets: An Example of a Computational Approach. *Computational Economics*, 8 (3), 205–31.
Vriend, N. J. (2002). Was Hayek an ACE? *Southern Economic Journal*, 68 (4), 811–40.
Wolfram, S. (2002). *A New Kind of Science, Volume 5*. Champaign: Wolfram Media.
Zschache, J. (2012). Producing Public Goods in Networks: Some Effects of Social Comparison and Endogenous Network Change. *Social Networks*, 34 (4), 539–48.

8

Simon's (Lost?) Legacy in Agent-Based Computational Economics

Marco Castellani and Marco Novarese

In 1960 Herbert Simon proposed the use of computer simulations in order to enrich and improve economic modeling (Clarkson and Simon, 1960). Computer science was among the many fields he contributed to. Therefore, his interest in computer simulations is not surprising. Simon is often quoted in the agent based simulation literature, where bounded rationality is highly recognized and used as a starting point, and yet his approach to simulation seems to be quite neglected: the website jstor.org shows very few citations of Clarkson and Simon's paper; a specialized journal like JASSS (*Journal of Artificial Societies and Social Simulation*) never quotes that paper; very few economists working in the field seem to know the paper; and no one has pursued his methodological approach.

This chapter aims to understand the reason for this apparent paradox while reviewing the variety of approaches and interests Simon pursued. It will be shown, again, that Simon's aims and methodology were different from even heterodox economists. The two points may be linked. Simon was interested in solving problems, often practical problems. For this reason there was no need for only one approach. Depending on the problem, he pursued what he thought was the best approach. Simon was a true polymath.

The (Ab)use of Bounded Rationality

The role played in economics by Herbert Simon with his *pars destruens*, mainly by opposing the fictitious model of full rationality, is certainly beyond any doubt. The basic building blocks of his standpoint, from the unrealistic assumptions of subjective utility theory to the uselessness of the maximization process, have long been discussed within

non-standard economics debates. On the other hand, the course of his *pars construens* was certainly questionable within the social sciences. Simon mainly developed the notion of bounded rationality within the empirical field of social sciences. From this perspective, the core of Simonian work was used as a reference model to explain human behavior, given that an agent's computational power is limited by specific constraints.

Such an approach, however, does not necessarily imply only observing the reduction of cognitive capabilities through which agents face decision setting. In this direction, many authors have specifically pointed out that there is a peculiar side view of Simon's work that cannot be rejected. This essential feature concerns the information processing system and its use for the purpose of decision-making. What is this key point? The core of Simon's proposal is that the nature of a task environment cannot be considered as static, unless over-simplifications – such as those related to the idea of *steady state conditions* – are considered. It is certainly easy to imagine that nothing in the outside world changes while the decision maker examines each alternative, grounding this process on the available information, and then she/he chooses the best option. If this is the typical condition in each chosen scenario, then we could see optimizing reasoning as the correct decision-making process. The only limitation would lie in not being able to assess all the available evaluation criteria. However, if we embrace a *parallel processing style*, it would be possible to estimate each usable alternative and eventually consider the best one according to our expected utility. Many works in microeconomics have developed just that side of bounded rationality, supposing that such a version of bounded rationality could be operationalized by constrained optimization models.

However, the core of the Simonian proposal is to contrast with the idea of parallel processing, because the mechanism of the aspiration levels has been specifically proposed for its capability to adapt to the dynamic nature of the external world. The main purpose of the aspiration mechanism is its ability to adapt to real dynamic environments, frequently defined as 'on-line' in decision-making literature (Gavetti and Levinthal, 2000). So, when an option satisfies the fixed aspiration level, on the basis of a series of complex cognitive processes, then that option may be chosen. The evaluation path is not parallel but sequential, in the sense that it is not possible to implement an option which was analyzed in previous periods without facing a loss of significance of such an evaluation. For this reason constrained optimization as an expression of bounded rationality does not make any sense (Gigerenzer, 2001a;

2001b). Aspiration levels, therefore, still represent the core of Simon's approach, since they can properly fit the dynamic nature of decision setting and, at the same time, they can be adjusted upward and downward depending on the complexity of the task environment (when it is easy to meet satisfactory options they will be increased and, conversely, they will be reduced).

Also in the light of these considerations, it seems proper to look at all the approaches that are more focused on heuristics as operational implementations of Simon's aspiration levels, in particular those deeply concerned with the dynamic complexity of the environment, such as the, so-called, 'fast and frugal' ones (Goldstein and Gigerenzer, 2002).

Simon from Empirical Observations to Experimental Evidence

The main methodological approach used by Herbert Simon to propose his model of human behavior was mainly empirical. His seminal works about the, so-called, *homo administrativus* were built upon observations of how real agents make both routine and strategic choices (Simon, 1947). In this research, Simon always tried to propose alternative models with unambiguous features, starting from real human skills. This is the main reason why Simon's models were also complex, since he was always seeking for the key feature of perception and information processing. The capability of humans for framing a task environment by defining the main traits was a crucial topic in all empirical observations (Simon, 1983), as far as understanding the role of memory, selective attention, cognitive representation, perception, reasoning by analogy, and so on. The case of chess players' behavior is highly significant in this regard. The key aspect of the experience of a skilled player is the ability to read the situation. Moreover, the way a particular configuration of pieces is recognized and connected to the agent's own past experience is crucial. For this reason a suitable model should include an overall representation of the chess board, the configuration of which must be fixed at each step, starting from the perception and recognition of each piece. The next step should be about how information caught by eye movement is then processed and retained in short-term memory. Players have no possibility of building up a comprehensive representation of the board, so they have to use their selective attention when connecting different pieces. So, it is quite clear that all these issues need to be empirically studied in order to understand what the key points of cognition are. Consequently, different tools and techniques are required to

model all these processes, for example technical eye tracking, protocol analysis, and so on (Ericsson and Simon, 1993). Simon, after addressing the importance of case-study technique (Simon, 1992), was also explicit in outlining the need for economists to use interviews as a crucial tool for understanding the real world:

> If you are trying to understand what firms are and how they operate, you will learn a lot from this kind of very detailed study of the processes of decision . . . Of course, we should not stop with five firms. Biologists have described millions of species of plants and animals in the world, and they think they've hardly started the job. Now, I'm not suggesting that we should go out and describe decision-making in a million firms; but we might at least get on with the task and see if we can describe the first thousand. That doesn't immediately solve the aggregation problem, but surely, and in spite of the question of sampling, it is better to form an aggregate from detailed empirical knowledge of a thousand firms, or five, than from direct knowledge of none. But the latter is what we have been doing in economics for too many years. (Simon, 1992, p.20)

In addition, the core of experimental economics virtually rejects such a multiple way of investigation, at least if we look at the most influential journals, where economics papers are mainly directed to compare and test different models. This is not surprising, since experimental economics was primarily developed to refute neoclassical models and to improve alternatives, while remaining within the analytical purpose. For example, Kahneman and Tversky (1979) criticized the expected utility theory, but by using experimental contexts which seem quite similar to the standard ones (to show that people behave differently). Conversely, Simon's approach to laboratory experiments was motivated by an interdisciplinary viewpoint (Kulkarni and Simon, 1990), not surprising given his scientific open-mindedness. This methodological discrepancy is in addition to the epistemological long-term issue of what should be validated/falsified or not. From the neoclassical standpoint, which is strongly based on hypothetico-deductive reasoning approach, it makes no sense to start the investigation process from the empirical regularities that are observed (in a laboratory or in the outside world), even more so if such behaviors are derived from different hybrid techniques. This standpoint led many mainstream scholars to the strongest criticism of Simon's viewpoint, reinforced by an explicit rejection of being both empirically grounded and strongly interdisciplinary. This apodictic

view was then added by mainstream economics to the lack of attention of bounded rationality on the basic conceptual level (the satisficing process and the mechanism of aspiration levels are rarely mentioned in microeconomic handbooks).

Simon and Simulations

The previous sections have highlighted the high level of complexity in Simon's model of action, due to the variety of cognitive processes involved. If this feature of Simon's *pars costruens* has often been misinterpreted in the theoretical field, even for strictly utilitarian concerns, similar conclusions can be drawn about the simulation methodology. There are several reasons for this phenomenon, but the strongest one lies in the main target for which the simulation method is developed. If we consider the purpose of the, probably more influential, simulative approach in the social sciences today compared to that of agent based models (ABMs), following the Santa Fe Institute, we note that such models have been largely favored over time for a clear and defined focus on the characteristics of agent interactions, more than on agents' cognitive architecture and their behavioral model.

The main reason behind this methodological standpoint is that complex social phenomena can emerge even from very simple agent behavior. This concept is developed in the current debate as the well-known KISS principle ('keep it simple stupid', with some possible variations). Despite some remarkable exceptions, therefore, the emergentist 'from the bottom up' logic is typical of ABMs, which are driven towards this principle of economic methodology so they can focus better on using different forms of interaction as explanandum. In this view it is not surprising that Simon is mainly quoted in a more general and argumentative way (Sallans *et al.*, 2003) than in a substantive one, even in the ABM literature.

We believe there is an interesting topic that could make this methodological mixture more understandable, and that could better exploit Simon's methodological thinking. This proposal involves connecting the analytical procedures of ABMs and their simulations with experimental evidence. The suggestion, which may not be supported by a large number of contemporary simulation scientists, is very simple, since it requires the use of experimental data to model the agents placed at the center of the simulations. This setting could deeply affect the agent micro-foundations. It is not meant to simply complicate the agents' cognitive model and is not only for theoretical reasons, but primarily

because there should be empirical and experimental studies that confirm the robustness of such behavioral processes. This micro-empirical foundation may be crucial when an ABM is supposed to fit with typical social science problems, such as being compared to what happens in real life and in concrete social task environments.

References

Clarkson, G. P. A., and Simon, H. A. (1960). Simulation of Individual and Group Behaviour. *The American Economic Review*, 50 (5, Dec.), 920–32.

Ericsson, K. A., and Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.

Gavetti, G., and Levinthal, D. (2000). Looking Forward and Looking Backward: Cognitive and Experiential Search. *Administrative Science Quarterly* 45 (1): 113–37.

Gigerenzer, G. (2001a). Rethinking Rationality. In G. Gigerenzer and R. Selten (eds), Cambridge, MA: MIT Press, 1–12.

Gigerenzer, G. (2001b). The Adaptive Toolbox. In G. Gigerenzer and R. Selten (eds) *Bounded Rationality – The Adaptive Toolbox*. Cambridge, MA: MIT Press, 37–50.

Goldstein, D. G., and Gigerenzer, G. (2002). Models of Ecological Rationality: The Recognition Heuristic. *Psychological Review* 109 (1): 75–90.

Kahneman, D., and Tversky, A. (1979). Prospect Theory: An Analysis of Decisions Under Risk. *Econometrica* 47 (2): 313–27.

Kulkarni, D., and Simon, H. A. (1990). Experimentation in Machine Discovery. In *Computational Models of Scientific Discovery and Theory Formation*, ed. J. Shrager and P. Langley. San Mateo, CA: Morgan Kaufmann.

Sallans, B., Pfister, A., Karatzoglou, A. and Dorffner, G. (2003). Simulation and Validation of an Integrated Markets Model. *Journal of Artificial Societies and Social Simulation* 6 (4).

Simon, H. A. (1947). *Administrative Behavior*. New York: Macmillan.

Simon, H. A. (1983). *Reason in Human Affairs*. Oxford: Basil Blackwell.

Simon, H. A. (1992). Colloquium with H. A. Simon. In *Economics, Bounded Rationality and the Cognitive Revolution,* ed. M. Egidi and R. Marris. Aldershot: Elgar, 8–36.

9

From Bounded Rationality to Expertise

Fernand Gobet

Introduction

Historically, a pervasive assumption in the social sciences, in particular economics, is that humans are perfect rational agents. Having full access to information and enjoying unlimited computational resources, they maximize utility when making decisions. As is well known, Herbert A. Simon rejected this assumption, calling it a fantasy for two main reasons. First, the complexity of the environment makes it impossible for humans to have full access to information. Second, a number of important restrictions impede the human cognitive system, such as limited attention and slow learning rates. Therefore, humans display only a bounded rationality and must satisfice, i.e., make decisions that are good enough, but not necessarily optimal.

Research into expertise has contributed to the question of rationality in two important ways. First, to what extent can some of the very best among us – super experts – approximate full rationality? Second, by what means do experts, at least in part, circumvent the constraints imposed by bounded rationality?

This chapter takes the shape of a fugue, with the themes of bounded rationality and expertise first played in the background of personal recollections, and then elaborated with a more formal survey of Simon's research into expertise. The themes are played a third and final time with a discussion of the heuristics (rules of thumb) proposed by Simon for achieving a successful career in science.

Becoming an Expert: A Personal Recollection

My collaboration with Herbert A. Simon lasted over ten years, including six years spent at Carnegie Mellon. While I was working on my PhD thesis on chess players' memory, I secured a research fellowship from the Swiss National Science Foundation to work with him. The qualifications I listed in my introductory letter to Simon were rather limited, a first degree in psychology and the title of International Chess Master. Simon, who probably saw an opportunity to reactivate research he carried out on chess expertise in the 1960s and 1970s (see below), but which had been dormant since, accepted to host me.

Meeting the Man

I can still recall our first meeting on a beautiful morning in January 1990. His office was welcoming, but also rather disorganized, with stacks of papers and books hiding his desk. The meeting was short but cordial, and Simon gave me advice about life and housing in Pittsburgh, and briefly talked about the projects he was currently involved in.

The second meeting was my first real scientific discussion with Simon. It was actually a shock. In a polite and friendly way, Simon demolished the research line I had in mind for my PhD, which was to elicit a chess grandmaster's knowledge of a small and specific domain (rook and pawn endgames), and to build a program implementing this knowledge. The aim, inspired by research on expert systems, was to compare the amount of procedural (knowing how) and declarative (knowing that) knowledge. Simon found that the project was not realistic enough ('A player like Kasparov will give you lectures on rook endgames for several days; what are you going to do with all these data?'). In addition, he thought that the project would dovetail better with the research of his colleague John Anderson. I can still feel the panic that invaded me when he told me this, as it was an invitation to sever collaboration before the end of the first meeting! In the discussion that followed, he made it clear that he would prefer a project directly linked to the chunking theory he had developed with Chase in the 1970s to account for chess expertise (Chase and Simon, 1973a, 1973b). This influential theory, and in particular the computer model MAPP (Memory-Aided Pattern Perceiver; Simon and Gilmartin, 1973) that implemented it in part, had been severely criticized, and Simon wanted to improve on it. Thus, my first lesson was that Simon, while open to other ideas, was very selective about the research lines he invested time in, and made sure that they addressed his central interests.

Luckily, my back-up project precisely addressed the issues that Simon mentioned. However, it was less developed than my first choice, and I spent a rather anxious week trying to improve on it. The idea was to carry out a series of experiments on chess players' memory, focusing on the amount of information that they could memorize after a brief presentation, and the extent to which this information was encoded with specific cues about spatial localization. The experimental results would be simulated by a program based on MAPP and the idea of chunking. With hindsight, it is obvious that this second project, which combined experiments with computer modeling, fitted Simon's scientific approach much better. In addition, MAPP was a variation of EPAM (Elementary Perceiver and Memorizer; Simon and Feigenbaum, 1964), a general theory of learning embodied as a computer program, which played a central role in Simon's approach to expertise and bounded rationality more generally.

My visit was supposed to last for one year. However, the privilege of working with Simon and the exciting, interdisciplinary CMU environment convinced me to extend it, and I eventually stayed for six years. I would meet Simon face-to-face for one hour a week on average. In general, I would decide on the agenda, which typically included our research on chess memory and a discussion of some of his other research projects, not only in psychology but also in other fields, mostly artificial intelligence and philosophy.

'What new data do we have about chess today?' Simon would often ask at the beginning of our meetings. Sitting on a simple chair, with no desk between him and his guest, he had an informal and welcoming style. He was immensely curious about new phenomena, and took obvious pleasure in analysing data, always looking for hidden patterns. He was outstanding at making sense out of complex data sets, but would also often ask me to re-analyse data, by using better representations for example.

Meetings with him were alternatively dense and relaxed, focused and wide-ranging. His informal style made me sometimes forget that I was talking to one of the greatest minds and one of the last true humanists of the twentieth century. While open to new ideas, he also would immediately, albeit elegantly, rebut bad arguments. However, even then, he managed to refute them in a way that did not make you feel too dejected – even though more often than not, your idea had been irremediably torn apart.

In spite of his age, Simon was extremely energetic. Not only did he have an active program of research, but he also had a normal teaching

load and was deeply involved with the politics of CMU. (Imagining his energy when he was in the prime of his life was a rather depressing exercise.) For him, research was a hobby, and working 80 hours a week seemed reasonable. Of course, he expected his collaborators to show a similar dedication to research, including their working hours. This expectation was a logical application of his work on expertise, where he suggested that, in most domains, a minimum of 10,000 hours of practice and study was necessary to become an expert (Simon and Chase, 1973).

Simon was exigent and expected his collaborators to share his passion for work. Working with him was extremely rewarding, and most of his students and collaborators did show strong commitment, although few managed to work for 80 hours a week consistently. In a very elegant and efficient way he taught us much about science, mainly that scientific research can be an exciting voyage of discovery.

Another striking characteristic of Simon was his curiosity; the title of one of his talks was 'The cat that curiosity couldn't kill.' He had mastered an astonishing range of domains, from piano to chess to entomology, and was interested in all sorts of topics, including non-scientific questions. Once he had decided to find an answer to a question – sometimes a trivial question – he would work on it obsessively; it was as if heuristic search was a goal in itself. This, of course, is a powerful personality trait for a researcher to have.

Simon was also very generous with his students and colleagues, and always ready to support us in difficult moments. When some arcane immigration regulation allowed me to be employed by Carnegie Mellon University but not to be paid, he did not hesitate to support me financially for a couple of months. His generosity went beyond the academic world and he gave the money associated with his Nobel Prize to charity.

Collecting Data

As noted above, our collaborative work focused on chunking theory and on the empirical difficulties it faced. In their seminal work, Chase and Simon (1973a, 1973b) had analysed in great detail the way players of different skill levels recalled briefly-presented chess positions. Their hypothesis was that experts in chess – and other domains – encode domain-specific material as *chunks*: groups of items that are related perceptually and semantically. In line with bounded rationality, a chunk takes a relatively long time to store in long-term memory, but it can be accessed rapidly afterwards, in a few hundred milliseconds.

However, Simon had been tormented by an anomaly in the data ever since the publication of his work with Chase. Although the master used larger chunks than the weaker players, as predicted by the theory, he also used more chunks, which was inconsistent with the hypothesis that all players had the same limited short-term memory capacity. In addition, several experiments with individuals trained to remember long lists of digits presented for one second each (Chase and Ericsson, 1982), clearly indicated that information could be encoded more rapidly into long-term memory than postulated by theory. Simon hypothesized that experts in chess and in the digit-span task used *retrieval structures*, that is, long-term memory schemas allowing them to encode information rapidly. What eluded him was the exact nature of these structures.

Our replication of Chase and Simon's experiment improved on it in two important ways: we had a much larger sample; and we collected more reliable data by using a computer program rather than physical boards and pieces. The results clearly showed that masters possessed and used much larger chunks than had been found in the original study (Gobet and Simon, 1998). In addition, critically, the number of chunks replaced did not differ between skill levels, all players used three chunks on average. Simon was very pleased with these results, not only because they corrected the anomaly found in Chase and Simon's study, but also because they were consistent with data he had collected a few years earlier on the memory for Chinese ideograms (Zhang and Simon, 1985). This interest in numerical parameters of cognition (another is eight seconds to learn a chunk) was a signature of Simon's scientific style and a logical consequence of the hypothesis of bounded rationality.

Support for the three-chunk capacity of visual short-term memory was also found in a series of experiments where we tried to establish the limit of chess experts' memory. In these experiments, chess players had to memorize not only one chess position presented for a short amount of time (typically five seconds), but also several of them, each containing around 25 pieces (Gobet and Simon, 1996d). This task was very difficult, and only masters could cope with it. Interestingly, there seemed to be a limit of around three or four boards – again, the magical number three!

Simon was intrigued by this task, and convinced me to conduct an experiment with his favorite research method, to study a single subject in great detail. He also persuaded me to volunteer as the subject, I was an International Chess Master after all! In a longitudinal experiment that lasted nearly two years (with a few interruptions), I practiced memorizing as many positions as possible, several times a week. The software I had written, together with a large database of chess positions, allowed

me to carry out a well-controlled experiment, despite being both the experimenter and the subject. This experiment produced a huge amount of data (every single action and its timing was automatically recorded), which, unfortunately, we were slow to analyse (For preliminary analyses, see Gobet and Simon, 1996d, Gobet, 2013). I improved from four boards to a personal record of ten boards with a minimum of 80% accuracy on each board, but was never able to go beyond. At the beginning of our weekly meetings, Simon would ask me about my progress – and often my lack of progress – and we used these data to develop mechanisms for retrieval structures in chess in what became the *template theory* (Gobet and Simon, 1996d).

In another experiment, we studied the effect of modifying chess boards by mirror image reflections (Gobet and Simon, 1996a). The results provided useful information on the way chunks are encoded by chess players. In particular, they showed that information about location is encoded in chunks. Another experiment systematically varied the presentation time from one second to 60 seconds (Gobet and Simon, 2000). The results confirmed, with a visuo-spatial task, a key parameter that Feigenbaum and Simon (1984) had estimated with a verbal task, that it takes around eight seconds to create a new chunk in long-term memory. While these results were important, it was another result, which I had mentioned in passing and rather anecdotally, that captured Simon's imagination. World champion Garry Kasparov had played several matches against national teams, facing up to eight grandmasters or masters simultaneously. In most cases, he had won these matches. Crucially, computing Kasparov's performance in these matches showed that he played, on average, at a level that still placed him in the six best players in the world at the time (Gobet and Simon, 1996c). For Simon, this was a spectacular illustration of the role of pattern recognition and selective search in expert decision making, of how experts can (partly) overcome the limits imposed by bounded rationality on their cognitive abilities.

Building Computer Models

One exciting aspect of my collaboration with Simon was the development of several computer models. It was also a very challenging experience, not only for technical reasons, but also because, whatever the beauty of one's model, the moment of truth is whether the model accounts for the experimental data – and experimental data are ruthless. The first version of CHREST (Chunk Hierarchy & REtrieval STructures) combined MAPP with PERCEIVER, a model simulating chess players' eye

movements (Simon and Barenfeld, 1969). One improvement over the earlier models was that chunks were not input by the modeler, but were learned autonomously as a function of the positions that the model had seen. Eye movements played an important role in learning and, in turn, the chunks that had been learned were essential in directing future eye movements.

The key contribution of CHREST was closely to link mechanisms for perception, learning, and memory, and to provide mechanisms for the concept of a retrieval structure. Our first hypothesis was that chess players' retrieval structures are similar to the structures used by individuals specialized in memorizing digits; such structures are generic and can be used with any kind of material as long as it is taken from the domain of expertise. This version of CHREST was able to simulate several empirical data successfully, but also suffered from some serious weaknesses. In particular, it overestimated recall performance with random positions, and could not replicate the experimental results obtained with modifying chess boards by taking their mirror image.

After much trial and error, we reached the conclusion that chess players' retrieval structures were specific; that is, they could be used only when the board contained some specific patterns. In a sense, these templates were more similar to the schemas discussed in psychology and artificial intelligence than to the structures identified in the digit-span task. The modified version of CHREST accounted for a wide range of data concerning eye movements, recall performance with diverse types of positions (game positions, random positions, positions modified by mirror image), number and type of errors made, and type of chunks used. An important contribution of the model, which goes beyond chess, was that it provided mechanisms explaining how schemas are constructed automatically, including the way variables and default values are built. CHREST was later applied to other domains of expertise and to the simulation of first language acquisition, which can be considered as a kind of expertise (Gobet *et al.*, 2001).

An interesting episode in the development of CHREST is worth mentioning, since proving Simon wrong was extremely difficult. When simulating the recall of random positions, the second version of CHREST systematically predicted that there should be a skill effect, although it was much smaller than with game positions. This prediction was contrary to Chase and Simon's (1973a) result, where the master performed as badly as the weaker players with random positions. In fact, this lack of skill difference had become a standard result in psychology, found in most textbooks. Simon first thought that there were bugs in my

program. However, after much double-checking and many replications, it was clear that the effect was genuine; even in random positions, a version of the model knowing many chunks is more likely than a version knowing fewer chunks to recognize, by chance, a pattern of pieces on the board. When we ran new experiments and collated all studies in the literature that had used random positions as a control condition, it became clear that the effect was also present with humans. In most experiments, the effect was small and statistically non-significant, but it was reliable when all experiments were combined (Gobet and Simon, 1996b). Ironically, these results provided strong support for chunking theory, since they are difficult to explain with other theories of expertise (Gobet, 1998).

CHREST was developed at the same time as two other models based on EPAM. The first one accounted for results in categorization, and in particular highlighted the role of strategies in that task (Gobet *et al.*, 1997). The second explained how an individual with a normal short-term memory capacity managed, after intense practice and training, to memorize 106 digits, each presented for one second only (Richman *et al.*, 1995, Richman *et al.*, 1996). In the second part of my stay at CMU, these models were discussed during near-weekly meetings of the EPAM group, which also included Howard Richman, Jim Staszewski, and Shmuel Ur. These meetings were very lively and included a considerable amount of brainstorming, and while sometimes lacking structure, they offered a productive environment for exploring various aspects of EPAM. Simon was active in these discussions, but was non-directive.

Bounded Rationality and Expertise

The kind of simple tasks typically studied in psychology can only go so far to identify the properties of human cognition, including strategies and cognitive invariants. By studying much more demanding tasks, research into expertise offers a unique window into cognition (Gobet, 2015), and, in particular, how humans cope with complex environments. As a first approximation, it is possible to divide Simon's research on expertise into three periods: problem solving (until the mid-1960s); perception and memory in chess (late 1960s to mid-1970s); and broadening the horizon (from the late 1970s to Simon's death).

Problem Solving

Simon's interest in expertise is apparent in his early books, such as *Administrative Behavior* (Simon, 1947a) and *The Technique of Municipal*

Administration (Simon, 1947b), where the issue is how organizations can best use expert skills, particularly with respect to decision making. Simon argues that hierarchies offer the best structure to do so, as they allow decisions to be made in the part of the organization where they are most useful.

During the 1950s, Simon started to study expertise empirically through his work on chess problem solving. Dutch chess master and psychologist Adriaan de Groot had shown that chess players' search is highly selective; among the numerous moves possible in a position, they look only at a handful (De Groot, 1946). Indeed, the stronger the player, the narrower the search behavior, inferior options being rarely considered. To flesh out the mechanisms that allowed moves to be generated so selectively, Newell and Simon (1965, 1972) carefully analysed the verbal protocol of a single player trying to find the best move in a given position. Several characteristics of move generation were identified. For example, if the analysis of a move leads to a positive evaluation, then this move is further investigated. If the evaluation is negative, then a different move is examined.

The research led to several simulation models, which are of considerable interest since they represent a direct attempt to model concepts borrowed from Simon's theory of bounded rationality. As is well known, Simon strongly advocated the use of formal models in the social sciences. While he originally used mathematical methods, such as differential calculus, he had noted essential limitations with them and concluded that other techniques were necessary. This, of course, led to the development of artificial intelligence and computer modeling, tools that were not, for many years, distinct in his mind. For the study of expertise, computer modeling has clear advantages: theories are precisely stated; clear predictions can be made, both quantitatively and qualitatively; and simulations can examine the structure of the task environment.

NSS, a chess program developed by Newell, Shaw and Simon (1958) uses goals such as maintenance of material balance and control of the center. Based on these goals, two move generators are engaged: the first generates possible moves in the problem situation; the second generates moves that are possible during look-ahead search (some search is carried out to evaluate the suitability of the proposed moves). NSS directly implements the concept of *satisficing* by playing the first move that is evaluated above an aspiration threshold. The program demonstrated that it is possible to choose reasonable moves with very small search trees (less than 100 positions). However, the quality of its play was low.

Selective search is also demonstrated by MATER (Baylor and Simon, 1966). This was achieved by limiting search to forced moves and moves that minimized the number of options available to the opponent. The program played good chess in positions that contained a forced checkmate combination; however, it was limited to this sub-domain of chess.

Science is, of course, a kind of expertise, and Simon devoted considerable attention to this topic. In his early writings about scientific discovery and creativity, he was mostly interested in technical and philosophical aspects (e.g., most of the essays collected in Simon, 1977). However, following their work on problem solving, Newell, Shaw and Simon (1962) speculated in the late 1950s and early 1960s about how the human mind can be creative, and whether this could be described objectively and explained scientifically. Their answer was that creativity is a special case of problem solving, and thus can be studied with the same conceptual tools.

Perception and Memory in Chess

Perhaps Simon's central contribution to the study of expertise was made in trying to answer another question studied by De Groot (1946), how do perception and memory mechanisms allow masters to understand the gist of a position rapidly, in a matter of seconds? As we saw earlier, Chase and Simon (1973a, 1973b) developed a means to identify chunks and conducted a series of clever experiments on chess players' perception and memory. In addition, they proposed mechanisms not only accounting for these experiments but also explaining selective search. The central ideas of their chunking theory are clearly linked to Simon's concept of bounded rationality. Experts' cognition suffers from a number of limitations (e.g., limited-capacity short-term memory and slow learning rates), and these limitations are assuaged by knowledge. Building knowledge predominantly consists of acquiring a large number of chunks, which are both perceptual and semantic units. These units are linked to possible actions, forming *productions*; for example, in chess, if a file is open, occupy it with a rook. Thus, pattern recognition makes it possible to demonstrate expertise despite strict limits on computational capabilities. A strong implication of the theory is that expert intuition is essentially pattern recognition. Another implication is that the best way to explain expertise in chess, and in other domains, is to use the formalism of a production system (i.e., a system specifying how productions are used for solving problems).

As noted earlier, several computational models were developed by Simon and his colleagues to account for perception and memory in chess. PERCEIVER (Simon and Barenfeld, 1969) simulated the eye movements of a chess player, using some of the mechanisms present in MATER, in particular, to compute piece movement. The program was built to take issue with the claim of gestalt psychology that perception is holistic and complex. Using local and simple mechanisms, PERCEIVER was able to convincingly reproduce the eye movements of one player in a given position. The key assumption was that perceptual information relates to attack/defense relations between pairs of pieces or between a piece and a square. One limit of the study was that only one position was presented to test the validity of the program.

MAPP (Simon and Gilmartin, 1973) incorporated some of the mechanisms postulated by chunking theory and, like EPAM, uses a discrimination net to organize information. A discrimination net is a hierarchical network of nodes where features of the objects to learn are tested to determine what new information should be added to the existing hierarchy. MAPP uses two parameters that strongly limit its cognitive abilities: short-term memory can store only seven chunks; and learning a new chunk takes eight seconds. During the presentation of a chess position, MAPP tries to recognize chunks in long-term memory and, when information is successfully identified, pointers to those chunks are placed in short-term memory. During the reconstruction phase, MAPP simply unpacks the information provided by these chunks.

MAPP was able to simulate the recall performance of a strong amateur, but not of a master. Using mathematical extrapolations from the computer simulations, Simon and Gilmartin concluded that from 10,000 to 100,000 chunks (50,000 as a first approximation) are necessary to reach expert level in chess and in other domains. Later simulations suggested that the number might be as large as 300,000 (Gobet and Simon, 2000).

Broadening the Horizon

During this final period, Simon both revisited old research topics and studied new domains of expertise. One new topic was novice and expert differences in solving physics problems, which Simon studied using verbal protocols (Bhaskar and Simon, 1977, Larkin *et al*., 1980). The results showed that novices tend to search backward, from the goal to the givens of the problem, while experts tend to search forward, from the givens to the goal. However, when problems are difficult, experts revert to backward search. Regardless of difficulty, experts use heuristics to

reduce the amount of search and draw on more efficient representations of the problem. The importance of representations is also clear in economics, where experts are better at developing multiple representations, for example verbal and diagrammatic representations (Tabachnek-Schijf *et al.*, 1997).

To account for these empirical results, Simon built several computational models, implemented as production systems. One model accounted for problem solving in thermodynamics, a semantically rich domain (Bhaskar and Simon, 1977). An important offshoot of this work was SAPA, a program that semi-automatically coded verbal protocols. Another production system, ABLE, provided mechanisms accounting for how novices become experts in solving physics problems (Larkin and Simon, 1981), including the change from backward search to forward search. An interesting aspect of ABLE is that it can use declarative statements to derive new results; in turn, these results can be used to solve new problems. Understanding task instructions and problem descriptions is, of course, crucial for building internal representations of problems and solving them, not least because different instructions lead to different representations. This process had been modeled by Hayes and Simon (1974) with respect to several variants of the puzzle known as the Tower of Hanoi.

In economics, Simon and colleagues developed CaMeRa, a model simulating visual reasoning and the way experts combine different kinds of representation (Tabachnek-Schijf *et al.*, 1997). When solving problems, CaMeRa can interact with external representations, such as diagrams. It uses several formalisms: a parallel system accounting for low-level vision; a semantic network storing semantic knowledge; and a production system used for problem solving.

This period also saw a return to the study of scientific discovery. Together with several collaborators, Simon developed a number of computer programs able to simulate famous discoveries in the history of science (Langley *et al.*, 1987, Bradshaw *et al.*, 1983). For example, a computer program called BACON re-discovered several scientific laws, including Kepler's third law of planetary motion, using the same data as those available in the original discoveries. Heuristics made it possible for the program to search selectively through the space of possible equations; interestingly, experiments with students confirmed that they use the same heuristics (Qin and Simon, 1990). Another program, KEKADA, was able to design experiments, change theory as a function of the results, and then design new experiments; it was able to simulate Krebs' discovery of the urea cycle in 1932. In his autobiography, Simon (1991)

suggests that these programs would be a good start for somebody trying to simulate his own scientific creativity!

In the first part of this chapter, I described the work that Simon undertook with chess during this period. At the same time we developed CHREST to account for chess expertise, Simon worked on EPAM-IV (Richman *et al.*, 1995), a model accounting for superior digit memory. This model is particularly relevant in the context of bounded rationality, since it shows how experts can strategically compensate for structural limits in their cognitive architecture, in this case, short-term memory. The task under study was the digit-span task, where one has to memorize a sequence of digits that are dictated rapidly (typically, one second each). While most of us can remember only about seven items, some individuals were trained to recall many more. For example, DD, the human subject simulated by Richman and colleagues, recalled up to 106 digits. DD used different kinds of mnemonics (techniques aimed at increasing one's memory), and a sophisticated semantic knowledge of numbers (historical dates, typical running times), to produce such an incredible feat. In line with previous research, he also used retrieval structures (Chase and Ericsson, 1982), structures that enable a rapid storage in long-term memory. DD's behavior and performance are obviously hard to explain with chunking theory, and consequently spurred Simon on to develop a model accounting for these results.

Just like CHREST, EPAM-IV combines chunking mechanisms with the notion of a retrieval structure. The difference, however, is that the retrieval structures postulated in the digit-span are acquired deliberately and consciously. The model specifies, in great detail, structures and mechanisms for short-term memory and long-term memory, and the way chunks, semantic knowledge, and retrieval structures are acquired through learning. Each cognitive process has a time cost, which makes it possible to simulate DD's performance with great precision. The simulations showed that the model successfully accounts for how DD acquired expertise in this domain; indeed, the model was able to capture his development both qualitatively and quantitatively.

Bounded Rationality and Heuristics

During his career, Simon had amassed considerable evidence that human rationality is bounded. Bounded rationality does not mean that humans are irrational, but that humans are rational within the confines of their computational capabilities. An important means to reach this rationality is by taking advantage of the statistical structure of the task

environment and extracting regularities through learning. The importance of learning is amply supported by data from expertise research and is perhaps most apparent in the central role played by pattern recognition in expertise. However, using such regularities is not enough, it is still important to carry out search through the space of promising options. As systematic search is not possible due to the limits of human cognition (in particular the limits of short-term memory), searching through the problem space must be selective and guided by heuristics. Again, research into expertise supports this hypothesis and selective search is a constant theme, as has been illustrated several times in this chapter.

In order to provide a complete theory of bounded rationality, including the way experts manage to circumvent their limited computational capabilities, one needs to provide mechanisms both for pattern recognition and search. The models developed by Simon to account for expertise provide examples of such mechanisms. Selective search is sometimes guided by pattern recognition, and it is sometimes guided by heuristics. Thus, to become an expert it is essential to acquire powerful heuristics, and experts in science are no exception.

In writings and talks he would give around the CMU campus, Simon provided heuristics to help students and colleagues succeed in science. Many showed how selective attention can be a powerful tool to deal with the limits imposed by bounded rationality. Here is a small sample of Simon's heuristics that have impressed me (for additional examples, see Valdes-Perez, 2002, Langley, 2002). 'What is worth doing is worth doing badly. Carry out your research diligently, but not more so than necessary.' This is a direct application of the notion of satisficing, optimal solutions are out of reach and one has to content oneself with good-enough solutions. To Simon, 'a PhD thesis is only a progress report.' This is a particularly useful heuristic since many students are paralyzed by the myth that a PhD thesis has to make a major contribution to its field of research. In some cases, this paralysis is so severe that the PhD is never finished. How did Simon find the time to master numerous scientific domains? 'Your time is precious. Don't waste it by reading newspapers and watching TV. If something really important happens in the world, you'll know it through your friends.' Some heuristics dealt with the content of research: 'Choose important but also realistic research questions;' and 'Play with your knowledge, explore unexpected connections.' Perhaps his most powerful heuristic was to be surrounded by collaborators and friends. I was very fortunate that Simon used this heuristic with me.

References

Baylor, G. W. and Simon, H. A. (1966). A Chess Mating Combinations Program. *1966 Spring Joint Computer Conference*. Boston.

Bhaskar, R. and Simon, H. A. (1977). Problem Solving in Semantically Rich Domains: An Example from Engineering Thermodynamics. *Cognitive Science,* 1, 193–215.

Bradshaw, G., Langley, P. W. and Simon, H. A. (1983). Studying Scientific Discovery by Computer Simulation. *Science,* 222, 971–75.

Chase, W. G. and Ericsson, K. A. (1982). Skill and Working Memory. *The Psychology of Learning and Motivation,* 16, 1–58.

Chase, W. G. and Simon, H. A. (1973a). The Mind's Eye in Chess. In *Visual Information Processing,* ed. W.G. Chase. New York: Academic Press.

Chase, W. G. and Simon, H. A. (1973b). Perception in Chess. *Cognitive Psychology,* 4, 55–81.

De Groot, A. D. (1946). *Het Denken van den Schaker,* Amsterdam, Noord Hollandsche.

Feigenbaum, E. A. and Simon, H. A. (1984). EPAM-like Models of Recognition and Learning. *Cognitive Science,* 8, 305–36.

Gobet, F. (1998). Expert Memory: A Comparison of Four Theories. *Cognition,* 66, 115–52.

Gobet, F. (2013). Chunks and Templates in Semantic Long-term Memory: The Importance of Specialization. In *Expertise and Skill Acquisition: The Impact of William G. Chase,* ed. J. J. Staszewski. New York: Psychology Press.

Gobet, F. (2015). *Understanding Expertise,* London, Palgrave Macmillan.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I. and Pine, J. M. (2001). Chunking Mechanisms in Human Learning. *Trends in Cognitive Sciences,* 5, 236–43.

Gobet, F., Richman, H. B., Staszewski, J. J. and Simon, H. A. (1997). Goals, Representations, and Strategies in a Concept Attainment Task: The EPAM Model. *The Psychology of Learning and Motivation,* 37, 265–90.

Gobet, F. and Simon, H. A. (1996a). Recall of Random and Distorted Positions. Implications for the Theory of Expertise. *Memory and Cognition,* 24, 493–503.

Gobet, F. and Simon, H. A. (1996b). Recall of Rapidly Presented Random Chess Positions is a Function of Skill. *Psychonomic Bulletin and Review,* 3, 159–63.

Gobet, F. and Simon, H. A. (1996c). The Roles of Recognition Processes and Look-ahead Search in Time-constrained Expert Problem Solving: Evidence from Grandmaster Level Chess. *Psychological Science,* 7, 52–5.

Gobet, F. and Simon, H. A. (1996d). Templates in Chess Memory: A Mechanism for Recalling Several Boards. *Cognitive Psychology,* 31, 1–40.

Gobet, F. and Simon, H. A. (1998). Expert Chess Memory: Revisiting the Chunking Hypothesis. *Memory,* 6, 225–55.

Gobet, F. and Simon, H. A. (2000). Five Seconds or Sixty? Presentation Time in Expert Memory. *Cognitive Science,* 24, 651–82.

Hayes, J. R. and Simon, H. A. (1974). Understanding Written Instruction. In *Knowledge and Cognition,* ed. L. W. Gregg. Hillsdale, NJ: Erlbaum.

Langley, P. (2002). Heuristics for Scientific Discovery: The Legacy of Herbert Simon. In *Models of a Man: Essays in Memory of Herbert A. Simon*, ed. E. Augier and J.G. March, J. G. Boston: MIT Press.

Langley, P., Simon, H. A., Bradshaw, G. L. and Zytkow, J. M. (1987). *Scientific Discovery,* Cambridge, MA, MIT Press.

Larkin, J., Mcdermott, J., Simon, D. P. and Simon, H. A. (1980). Expert and Novice Performance in Solving Physics Problems. *Science,* 208, 1335–42.

Larkin, J. H. and Simon, H. A. (1981). Learning through Growth of Skill in Mental Modeling. *Proceedings of the Third Annual Conference of the Cognitive Science Society.* Berkeley: Cognitive Science Society.

Newell, A., Shaw, J. C. and Simon, H. A. (1958). Chess-playing Programs and the Problem of Complexity. *IBM Journal of Research and Development,* 2, 320–35.

Newell, A., Shaw, J. C. and Simon, H. A. (1962). The Process of Creative Thinking. In *Contemporary Approaches to Creative Thinking, ed.* Gruber, H. E., Terrell, G. and Werheimer. New York: Atherton Press.

Newell, A. and Simon, H. A. (1965). An Example of Human Chess Play in the Light of Chess-playing Programs. In *Progress in Biocybernetics, ed.* Weiner, N. and Schade, J. P. Amsterdam: Elsevier.

Newell, A. and Simon, H. A. (1972). *Human Problem Solving,* Englewood Cliffs, NJ, Prentice-Hall.

Qin, Y. and Simon, H. A. (1990). Laboratory Replication of Scientific Discovery Processes. *Cognitive Science,* 14, 281–312.

Richman, H. B., Gobet, F., Staszewski, J. J. and Simon, H. A. (1996). Perceptual and Memory Processes in the Acquisition of Expert Performance: The EPAM Model. In *The Road to Excellence, ed.* Ericsson, K. A. Mahwah, NJ: Erlbaum.

Richman, H. B., Staszewski, J. J. and Simon, H. A. (1995). Simulation of Expert Memory with EPAM IV. *Psychological Review,* 102, 305–30.

Simon, H. A. (1947a). *Administrative Behavior,* New York, NY, Macmillan.

Simon, H. A. (ed.) (1947b). *The Technique of Municipal Administration* (3rd edn), Chicago: International City Managers' Association.

Simon, H. A. (1977). *Models of Discovery and Other Topics in the Methods of Science,* Dordrecht, Reidel.

Simon, H. A. (1991). *Models of My Life,* New York, Basic Books.

Simon, H. A. and Barenfeld, M. (1969). Information Processing Analysis of Perceptual Processes in Problem Solving. *Psychological Review,* 7, 473–83.

Simon, H. A. and Chase, W. G. (1973). Skill in Chess. *American Scientist,* 61, 393–403.

Simon, H. A. and Feigenbaum, E. A. (1964). An Information Processing Theory of Some Effects of Similarity, Familiarity, and Meaningfulness in Verbal Learning. *Journal of Verbal Learning and Verbal Behavior,* 3, 385–96.

Simon, H. A. and Gilmartin, K. J. (1973). A Simulation of Memory for Chess Positions. *Cognitive Psychology,* 5, 29–46.

Tabachnek-Schijf, H. J. M., Leonardo, A. M. and Simon, H. A. (1997). CaMeRa: A Computational Model of Multiple Representations. *Cognitive Science,* 21, 305–50.

Valdes-Perez, R. E. (2002). Personal Recollections from 15 Years of Monthly Meetings. In *Models of a Man: Essays in Memory of Herbert A. Simon, ed.* Augier, E. and March, J. G. Boston: MIT Press.

Zhang, G. and Simon, H. A. (1985). STM Capacity for Chinese Words and Idioms: Chunking and Acoustical Loop Hypothesis. *Memory and Cognition,* 13, 193–201.

10

Multiple Equilibria, Bounded Rationality, and the Indeterminacy of Economic Outcomes: Closing the System with Institutional Parameters

Morris Altman

A critical point made by behavioral economists from a wide set of methodological perspectives is that individuals typically do not make decisions that are consistent with conventional economic theoretical norms of rational behavior. This is true of those building on the errors and biases, or heuristics and biases, approach derived from the research of Kahneman and Tversky (Kahneman, 2003, 2011; Thaler and Sunstein, 2008), and those building upon the bounded rationality approach introduced by Herbert Simon (1978, 1979, 1987; Altman, 1999; Gigerenzer, 2007; Smith, 2003). Such 'irrational' behavior from the perspective of the mainstream is considered to be inefficient or sub-optimal. And sub-optimal outcomes should not be able to survive; that is to fail the test of the survival of the fittest. However, different socio-economic outcomes or solutions to the same specific decision problems appear to be consistent with survival in the market place. This is even true of economic outcomes, when firms are not maximizing productivity. Both low and high productivity firms can survive simultaneously in the market. Moreover, ethical or socially considerate firms, and other-giving and empathic individuals, can also survive and persist, even if such behavior is often considered to be sub-optimal and irrational from the perspective of conventional economic wisdom.

It was just such apparent anomalies that Herbert Simon attempted to address through the concept of multiple equilibria, set in contrast with the more mainstream focus on convergence towards some optimal

and unique equilibrium.[1] From this perspective, both inefficient and efficient economic entities can persist over time in equilibrium. Therefore, survival and existence need not be, in any way, indicative or proof of uniqueness, optimality, or efficiency in outcomes or decision-making processes.

It is important to note that conventional and dominant economic methodology that deduces optimality and efficiency, and even uniqueness, from survival and existence is derived from the methodological paradigm articulated by Milton Friedman (1953; see also Alchain, 1950). He maintained that survival is proof of optimality and efficiency in both outcomes and decision-making processes. One can deduce from outcomes – from survival – that individuals or economic agents behave in a particular and unique fashion – optimally and efficiently.

One can also infer causality from results – the firm has survived because it currently exists – so that individuals or economic agents must have behaved in a particular and optimal fashion and such behavior generated the observed outcome (survival). One need not investigate, empirically, how individuals actually behaved with the regards to pertinent decisions and processes. Related to this, one can construct the normative statement that the surviving firms are optimal and efficient. Surviving firms of the moment must be optimal and efficient because they exist. No empirical investigation need be made as to the actual state of the firm because of the allowed-for deductions predicated upon what is assumed, willy-nilly, about the surviving or, more precisely, current firms. The possibility of multiple equilibria is assumed away in Friedman's representation of the market economy. It is assumed that real markets function in a fashion that results in optimality and efficiency without testing this assumption (Altman 1999).

Any observed outcome A is assumed to be efficient and, it is further assumed to be caused by B. Hence if A, then B. There can be, by assumption, no sub-optimal outcome A′, A″, A″′, or A^n that can exist simultaneously with A. Nor can there be alternative behaviors or processes, B′, B″, B″′, or B^n that can yield some optimal outcome A^0.

$$A \Leftrightarrow B$$

$$B \Rightarrow A$$

A and B are both highly correlated, with B causing A. All other possibilities are eliminated by assumption. A and B represent two stable and causally connected equlibria.

Although, thus far, emphasis has been on the firm, the unique equilibrium–optimality approach has also been applied to consumer

behavior. A strict but common interpretation of consumer sovereignty assumes that consumers make choices that are optimal and welfare maximizing not only from their own perspective, but also from the perspective of a calculating omnipotent decision maker.

Building on Simon, I examine and model alternative reasons for why non-conventional or non-neoclassical decision outcomes can survive in both the medium and long run, even while relatively more efficient options and entities exist (hence, multiple equilibria). This is largely ignored in the conventional economics literature, which focuses on logical time as opposed to historical time when analysing various possible decision-making outcomes. Also discussed is how optimal outcomes can be achieved through alternative means – different organizational and related decision-making processes can generate the same output, with the possibility of persistent multiple process equilibria.

Overall, this chapter presents a modeling framework that captures and allows for the existence of multiple equilibria in its many dimensions. Methodologically, this provides us with a broader theoretical lens, which incorporates the possibility that convergence towards a unique equilibrium is not necessary in the short, medium, or long run, one size does not fit all. The evolutionary process does not necessarily generate the most economically efficient and socially optimal outcomes. Hence, one should not so easily deduce optimality from the existence of a phenomenon, especially when relatively more efficient phenomena/outcomes also exist simultaneously.[2]

What appears to be a unique equilibrium output might only be a point in a set of multiple equilibria. But this possibility can only be evident from a modeling framework (effectively a data search engine) that allows for and emphasizes the significance of multiple equilibria. Otherwise, this possibility and reality would be simply assumed away, with serious theoretical and policy consequences. The approach taken in this paper is part of the behavioralist tradition of Herbert Simon, wherein one's theory is synergistically and dialectically derived from one's observation of pertinent aspects of reality.

One may begin this multiple equilibria discourse, building on Simon, with two critical quotes from Simon in this domain. In the first quote Simon focuses on the possibility of sub-optimal equilibria. As long as all competitors are at least equally sub-optimal and inefficient, equilibrium can be achieved in a sub-optimal state (Simon, 1997 p. 283):

> In the biological world at least, many organisms survive that are not maximizers but that operate at far less than the highest achievable efficiency. Their survival is not threatened as long as no

> other organisms have evolved that can challenge the possession of their specific niches. Analogously, since there is no reason to suppose that every business firm is challenged by an optimally efficient competitor, survival only requires meeting the competition. In a system in which there are innumerable rents, of long-term and short-term duration, even egregious sub-optimality may permit survival.

Simon (1978, p. 4) elaborates on this notion of sub-optimal equilibrium, moving towards the notions of multiple equilibria, and the possibility of achieving these through different means:

> The point may be stated more formally. Functional arguments are arguments about the movements of systems toward stable self-maintaining equilibria. But without further specification, there is no reason to suppose that the attained equilibria that are reached will be global maxima or minima of some function rather than local, relative maxima or minima. In fact, we know that the conditions that every local maximum of a system be a global maximum are very strong (usually some kind of 'convexity' conditions). Further, when the system is complex and its environment continually changing (that is, in the conditions under which biological and social evolution actually take place), there is no assurance that the system's momentary position will lie anywhere near a point of equilibrium, whether local or global. Hence, all that can be concluded from a functional argument is that certain characteristics (the satisfaction of certain functional requirements in a particular way) are consistent with the survival and further development of the system, not that these same requirements could not be satisfied in some other way. Thus, for example, societies can satisfy their functional needs for food by hunting or fishing activities, by agriculture, or by predatory exploitation of other societies.

There two important and distinct multiple equilibria scenarios that are important to consider. One relates to multiple equilibria in outcomes, the other relates to multiple equilibria in processes. To the extent that such multiple equilibria can persist for a reasonable length of historical time, one cannot infer the simple fact that, for 'survival' and existence, that which exists is necessarily the most economically efficient outcome or process to achieve a particular outcome. One causal inference often made in conventional economics is that one can deduce

economic efficiency from survival. Only the efficient (or in a softer sense, the relatively efficient) can survive. Hence survival implies, if not proves, economic efficiency. The same may be said of the inferred causality between unique particular processes yielding economic efficiency. However, the possibility of multiple equilibria should force the analyst to think carefully about such simple causal inferences, where other possible and reasonable explanations for survival are available and are, moreover, consistent with the evidence.

This formulation of multiple equilibria is related to David Hume's is–ought problem or fallacy, articulated in the *A Treatise of Human Nature* (2014 [1738], p. 576). This raises the problem of individuals deducing what ought to be from what is, and attributing particular causes to that which exists in a particular moment in time. Moreover, it is assumed that what exists is normatively ideal because it exists. But because these deductions are not empirically based, according to Hume, they represent fallacies. At best, these propositions represent testable hypotheses. One cannot impute anything in particular from the reality of existence of a phenomenon. This is, of course, exactly what Friedman does, flowing from his assumption that markets must generate optimal and efficient outcomes. Simon rejects this form of empirically empty causal and normative analyses. The concept of multiple equilibria allows for different understandings of existing phenomena, inclusive of the conventional economic interpretation of events.

The possibility of multiple equilibria can take on different forms, largely conditional upon an organism or economic entity's ability to survive in the short to long run. When survival does not require optimality or efficiency, multiple equilibria are possible across a set of differentially efficient or inefficient economic entities. Moreover, survival is consistent with a range of behaviors (processes and decisions), none of which need generate economic efficiency, and none of which need take the form of the severe calculating behavior of *homo economicus* (economic man).

Multiple Equilibria in Consumption

As a prelude to this discussion it is important to note that a large array of choices are not subject to any market discipline and therefore cannot be forced, even in theory, to converge on a unique equilibrium (Altman, 2005b). For example, acts of altruism (choices which generate an immediate reduction in income or wealth with no high probability of income or wealth returns on such expenditures), need not negatively impact on

the survival of the donor. The same holds true of expenditure on more expensive 'ethical' products or tipping for service, which can have an ethical component. Such acts simply reduce the income or wealth of the individual, but need not negatively impact on their current health, life expectancy, or capacity to procreate. Indeed, such acts of 'giving' are consistent with an individual realizing their preferences and, thereby, are consistent with such acts increasing their wellbeing or utility, as measured by the individual.

Hence, one might have a multiplicity of equilibria with regards to non-wealth or income maximizing decisions (controlling for risk), each a function of the preferences of the individual. Each individual could have multiple sustainable choices. Also, there could be multiple sustainable choices across several individuals. These multiple equilibria could be derived simply from the differential preferences of individuals. There is a choice set that is not sustainable. This would be one where an individual's choices reduced their level of material wellbeing below what was required for survival – a highly unlikely scenario. Even here the individual can survive if he or she is vested in a family or community that supports or subsidizes choices that are not sustainable on an individualized basis. So, even when those choices are made they can represent a sustainable equilibrium, part of a wider spectrum of equilibria, when part of a sustainable group or community.

A possible prediction of unique equilibrium with regard to consumer choice would flow logically from the assumption that individuals are wealth or income maximizers. If one assumes that all economic agents are wealth maximizers, by backward induction, one could derive a unique equilibrium for each individual consistent with income and wealth maximizing, controlling for risk. Deviations for such benchmarks could than be identified as unstable equilibria to be dissipated through the choices of 'rational' wealth and income maximizing individuals or economic agents. However, this type of scenario is not generally consistent with the actual behavior of individuals. Most decision-makers, in the domain of consumer choice, are not unconditional wealth or income maximizers. There is an array of individualized preferences, only some of these (a subset) consistent with the wealth or income maximizing assumption of conventional economics. So long as preferences aren't homogeneous with regard to wealth and income maximization, one should not expect or predict a unique equilibrium. For such a unique equilibrium would not be consistent with the decision maker's utility satisficing or maximizing behavior.

Multiple Equilibria in Production: An Introduction

At first glance, a unique equilibrium, or the convergence to a unique equilibrium, in the domain of production might appear to be a reasonable proposition, at least when modeling immediate and longer run scenarios. But this type of prediction hinges upon an array of unreasonable behavioral and institutional assumptions. As referenced above, a key point made by Simon is that multiple equilibria are pervasive and that it is crucial to explain and model this reality. A critical assumption is that market forces should force all firms into being optimally efficient for reasons of survival. Alternatively, a milder assumption put forth by Alchain (1950) is that market forces would ensure that only the relatively most efficient firms survive. An even stronger hypothesis is that economic agents are hardwired to be wealth or income maximizers, such that they behave in a fashion consistent with firms being optimally efficient. All these assumptions imply a unique equilibrium in terms of economic efficiency. All surviving firms should, therefore, be either efficient or relatively efficient, or converging to this unique equilibrium. But the conditions allowing for multiple equilibria are pervasive. Hence, one might argue that the default modeling assumption should be multiple equilibria inclusive of inefficient points within a set – a distribution of economic entities that ranges from efficient to inefficient. The type of distribution would be an empirical question. In the unique equilibrium approach all firms should by bunched together at the efficient end of the distribution. In the multiple equilibria approach, the distribution would be spread all over the efficiency–inefficiency spectrum. One could have a uniform distribution, with firms spread equally across the efficiency–inefficiency spectrum. Alternatively, firms could be normally distributed around some level of inefficiency, or one could even have two normal distributions, with one set of firms being efficient and the other being relatively inefficient.

An important starting point of this analysis is to understand that being cost competitive does not require that firms are economically efficient. This point is elaborated upon by Leibenstein (1966, 1979) in his discussion of x-efficiency theory. Moreover, product price need not be directly linked to the extent of economic efficiency – an argument advanced by Altman (1996, 2005a, 2005b, 2008). In this case, economically efficient firms need not be low-priced-product firms and low-priced-product firms need not be economically efficient. In the latter case, market forces, per se, cannot guarantee efficiency. In addition, it

is sustainable to have a wide array of firms, in terms of different degrees of efficiency, in a relatively uncompetitive environment. However, as I shall argue, even a competitive environment is not sufficient either to guarantee economic efficiency or a unique equilibrium – such as when only the relatively most efficient firms survive.

Since the extent of competitiveness and market forces is critical to conventional wisdom's assumption that economies converge towards an efficient equilibrium, it is important to appreciate the extent to which market forces can be mitigated, and often are. To the extent that economic entities can be protected from market forces, higher cost firms that charge higher prices to compensate for maintaining a 'normal' rate of return can survive over time, in the long run. The extent of protection afforded to firms can determine the extent to which relatively higher cost firms survive in the market in the long run. And, differential protection across firms yields multiple equilibria in terms of firms that are characterized by higher to lower unit costs of production for the same product. Protection can take on many forms inclusive of subsidies, tariffs, protective rules, and regulations. One should note that, dynamically, subsidies and tariffs, in the short run, can contribute to the development of efficient firms and sectors. Be this as it may, once one introduces protection there need not be convergence towards one unique efficient equilibrium and one should expect differences in equilibria across firms and nations contingent upon levels of protection. Therefore, at any given point in time, surviving firms should not, necessarily, be expected to be economically efficient.

Multiple Equilibria in Production: X-Inefficiency and Managerial Slack

This brings us directly to a discussion of x-efficiency theory and its pertinence to a discussion of convergence towards a unique equilibrium and multiple equilibria. A key point made by Leibenstein (1966, 1979) is that when product markets are not highly competitive this generates not only the standard allocative inefficiencies (which are more of a macro phenomenon), which tend to be relatively small, but also what he refers to as x-inefficiencies in production. Leibenstein breaks with the conventional wisdom, arguing that economic agents do not automatically maximize productivity, given the constraints that they face, most significant of which would be capital, labor, and technology. He argues that a key component of productivity is how hard and smart economic agents work. This translates into the assumption that effort inputs per

unit of labor time are a variable in the production function and in the utility function of economic agents. Leibenstein focuses on effort discretion on the part of managers and owners – agents at the top end of the firm's decision-making hierarchy. This is unlike the conventional modeling of the firm, where it is assumed that effort inputs are fixed at a minimum, if not maximized. If decision-makers maximize their utility by reducing effort levels (quantity and quality dimensions), effort diminishes from some optimum or fixed level, thereby reducing firm productivity and increasing unit cost of production. Here one would have an instance of managerial slack. The difference between what firm productivity is when effort is, in some sense, maximized and its actual level of productivity is a measure of x-inefficiency.

The direct relationship between productivity and average cost is illustrated below, where average cost can be given by the following equation, which assumes a very simply economy where labor is the only costed input (Altman 2001). If labor is only one of a number of inputs this does not affect the general direction of the argument.

$$AC = \frac{w}{\left(\frac{Q}{L}\right)} \tag{1}$$

AC is average cost; w is the wage rate or, more generally, the unit cost of inputs; (Q/L) is the average product of labor; Q is total output; and L is labor input measured in terms of hours worked. Anything that reduces productivity, such as managerial slack will, *ceteris paribus*, increase average cost (AC). This assumes that w remains constant in the face of changes to average cost.

Another way to visualize this arguement is as follows:

$$\Delta e \eth \Delta(Q/L) \eth \Delta AC \tag{2}$$

Changes in effort input (e) yield changes in labor productivity (Q/L) yield changes in average costs (AC). Maximizing effort input, maximizes average product and, thereby, minimizes average cost.

This takes the initial protection scenario deeper into the black box of the firm. Here higher costs are explicitly modeled as a function of the preferences of the firm's decision-makers, where these preferences are not in sync with conventional assumptions of profit maximization. These higher costs need not be a product of diminished returns or outdated technology, they could be a product of the choices made by decision-makers. In this case, given protection, one cannot expect

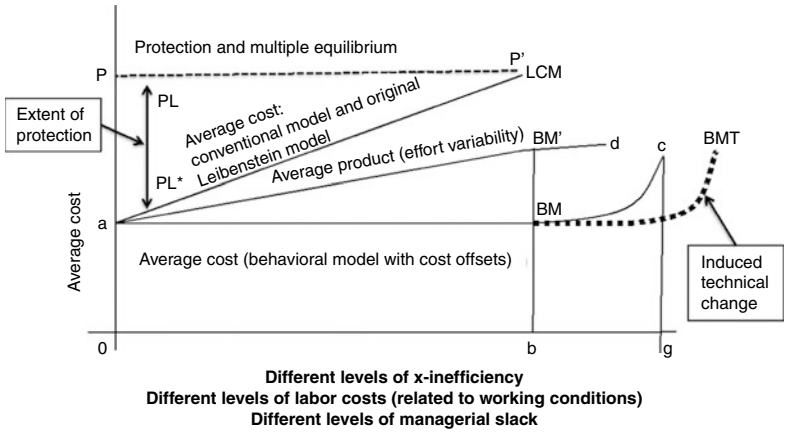


Figure 10.1 Multiple equilibria in production

convergence to an efficient equilibrium. Moreover, one would predict multiple equilibria in terms of the extent of x-inefficiency and average cost, given different levels of effort inputs. These multiple equilibria are sustainable with different levels of protection afforded across firms and across countries. Changes in the level of protection can be expected to yield changes in the level of x-inefficiency across firms and countries.

Some of these points are illustrated in Figure 10.1. In the Leibenstein narrative, managerial slack increases average cost by reducing labor productivity, and this is given by *aLCM*. With no managerial slack, the firm becomes x-efficient, given by point a. X-inefficient firms can survive on the market as consequence of protection, given by protection curve, *PP'*. The extent of protection is given by the vertical distance between line segment *PP'* and *aLCM*. Note, for example, that the protection required by the most x-inefficient firm is given by the difference in average cost between the latter and the average cost in the most x-efficient firm, *Pa* in this case. The smaller the difference in average cost between a given x-inefficient firm and the most x-efficient firm, the less protection that is required. Such protection allows for multiple equilibria across firms from the most to the least x-efficient firms.

Leibenstein maintains that such x-inefficiency is a product of quasi-rational behavior since managerial slack deviates from neoclassical economic norms. But I would argue that, since managerial slack is utility maximizing from the perspective of the economic agents in question, it is rational. However, such choice behavior is socially sub-optimal as it

reduces productivity and output from what it might otherwise be when effort in maximized. Moreover, such individually rational but socially sub-optimal choice behavior can represent a stable equilibrium. This would be a case of *rational inefficiency*.

Multiple Equilibria in Production: X-Inefficiency and Agency

One can take this argument one step further (Altman, 1996, 1999, 2002, 2005a, 2005b). Evidence suggests that there is causal relationship between labor costs, inclusive of all aspects of the overall work environment, and productivity, with variations in effort input being an important intermediate variable. This relates to principle-agent issues, where a particular resolution to principle-agent problems need not yield x-efficiency in production, especially in a conflictual work environment. The empirically based prediction here would be that improvements in the work environment yield higher effort levels (quantity and quality dimensions) and this yields improvements in productivity, while a poorer work environment yields lower effort levels. Of course, changes in productivity affect average cost, *ceteris paribus*. *Ceteris paribus* is Leibenstein's original scenario, as only effort input is allowed to vary, as illustrated in Equation 1. But if labor costs are allowed to vary, and are causally connected to changes in productivity through variations in labor's effort input, average cost need not change as effort inputs, and therefore productivity, varies.

Because of this positive relationship between labor costs, effort input, and productivity, increases in labor productivity serve to offset increases in labor costs, while reductions in labor productivity serve to offset reductions in labor costs. These cost offsets can be enhanced to the extent that technological change is induced by higher labor costs and lower labor costs impede technological change. This type of induced technological change is motivated by pressure to remain competitive in the face of rising or relatively high labor costs (Altman, 2009).

To the extent that changes in labor costs are just offset by changes in labor productivity, it is possible for average cost to remain constant along an array of labor costs – low to high. In this case, there would be an array of levels of labor productivity consistent with constant average cost. Here again there is no unique equilibrium and multiple equilibria exist with regard to sustainable levels of productivity and therefore sustainable levels of x-inefficiency. But, unlike in the initial Leibenstein modeling scenario, such multiple equilibria would even be consistent

with no product market imperfections or protection to support the relatively inefficient economic entities. Moreover, there is no market imperative for convergence to take place towards some particular efficient equilibrium. Each point along an array of different productivity levels is sustainable since average cost is fixed across this array.

A critical point here is that even in the conventional economic ideal modeling scenario, where product markets are perfectly competitive, and even when market pressure is at its most severe, there need not be any market imperative towards a unique efficient competitive equilibrium. Therefore, competitive markets are not a sufficient condition to achieve or force convergence on a unique efficient equilibrium in the realm of production.

Some of these points are illustrated in Figure 10.1. In the conventional narrative, x-efficiency is assumed, and increasing wages or labor costs drive up average costs, given by, *aLCM*. However, given x-inefficiency and the causal relationship between labor cost and the related work environment, effort input, and productivity, changes in labor costs need not have any impact on average cost until effort input is effectively maximized, given by *aBM*. Past point *BM*, labor cost increases by more than increases in effort input and related increases in labor productivity, wherein average cost increases with increases in labor cost. This can be discerned from average productivity curve, *aBM'd*. Thereafter, increases in labor cost can incentivize firms to engage in technical change (Altman, 2009), shifting the average cost function to the right to *aBMT*, for example. But along *aBM*, there are multiple equilibria across firms with different levels of x-efficiency, different levels of labor cost, and different incentive environments. These various (non-unique) equilibria are sustainable in the long run, given the cost offsets, positive and negative, allowed for through effort variability and induced technical change.

In this scenario, different levels of efficiency (multiple equilibria) are contingent upon different incentive environments within the firm that induce different levels of effort inputs, therefore, different levels of x-inefficiency and different levels of technical change (Altman, 2002, 2009). Related to this, different preferences among decision-makers towards their employees can affect the level of labor productivity. An array of such preferences can yield an array of levels of labor productivity. But these might all be consistent with a given level of average cost. For example, employers who favor higher wages and an improved work environment might not see an increase in average cost if this results in compensating increases in productivity; employers who favor lower wages and a poor or even deteriorating work environment need not

witness a reduction in average cost if this causes compensating decreases in labor productivity (Altman, 2002). Also, one might have a cooperative organization where higher wages and improved working conditions are part of the mission of the organization. Where labor costs, effort input, and labor productivity are highly and positively causally correlated, one cannot glean from the survival of firms, even in perfectly or highly productive product markets, that such firms are efficient. In this multiple equilibria scenario, one has to delve into the black box of the firm to determine the extent to which the firm is relatively efficient and the type of in-firm processes which give rise to the level of efficiency that characterizes a particular firm (Cyert and March, 1963).

Multiple Equilibria in Production: Some Related Scenarios

Related to the above modeling scenario, one can interrogate a number of propositions that flow from the conventional wisdom, a crucial one being that ethical behavior by the firm should result in the firm's demise, unless it is protected from market forces. Ethical behavior might take the form, for example, of improving working conditions within the firm or working towards making one's plant more environmentally friendly. This should increase production costs, *ceterus paribus*. Ethical firms would therefore have to be protected from the competitive threat posed by the relatively lower cost of less ethical firms. But this assumes that firms are already x-efficient and technological change is not induced (Altman 2001, 2002, 2005b, 2009). But in this scenario, there can be no efficient unique equilibrium since the higher cost ethical firms, through protection, could co-exist with their relatively unethical counterparts. If x-efficiency cannot be taken as a prior, and protection is not sufficient, then ethical firms can be expected find ways of improving productivity to remain competitive. This might not always meet with success. However, the appropriate prediction would be that ethical considerations can be expected to induce increased productivity.

Given the pervasiveness of multiple equilibria, one would expect that different processes would be in place across firms to achieve outcomes consistent with firm survival where some of these processes are consistent with x-efficiency and others are not. Moreover, efficient outcomes might be achievable through an array of different processes. For example, cooperative organizations (worker or consumer owned) would achieve efficiency through different means than a privately or investor owned firm. But evidence suggests that even investor owned firms tend to achieve efficiency through cooperative labor–management–owner processes as opposed to conflictual–non-cooperative processes (Altman,

2002). Therefore, one would predict multiple equilibria in this domain as well. There are ownership forms that can achieve x-efficiency. There is no one size fits all process model as to how efficiency is achieved within firms – a key point of behavioral approaches to the firm. And, this fits well with a multiple equilibria approach to modeling.

Historical and Logical Time

Another area of concern with regard to multiple equilibria is in the domain of historical time as opposed to the logical time in which conventional economics is largely vested. In logical time, the focus is on the determinants of equilibrium and how one moves from disequilibrium to equilibrium in particular markets, where the *ceteris paribus* assumption holds. The argument presented here is that when one models equilibrium scenarios, it is most probable that there will not be a unique equilibrium. Rather, there will be multiple equilibria, even when highly competitive product markets prevail. This interrogates the conventional wisdom at its core.

But this particular focus abstracts from the process by which equilibrium is achieved. In the real economy movement towards some equilibrium, including multiple equilibria, takes place over historical time and it can take considerable time to move towards an endpoint. During this process, one would expect that, taking a snapshot at a given point in time (a moment), there would be an array of firms that would be inefficient, even if the conventional hypothesis holds that in equilibrium all economic entities need to converge towards a unique equilibrium. Existence or survival, at any given point in time does not imply, in itself, that economic efficiency prevails. This is contrary to strong interpretations of the efficient market hypothesis, that firms should always be efficient.

Moreover, the extent of such moment, or snapshot, inefficiencies can be a positive function of the extent to which inefficiency can persist even in equilibrium – when sustainable inefficiencies are greater. This simply reinforces the notion that to determine whether existing firms are efficient, one has to examine the black box of the firm to determine the extent of economic inefficiency.[3]

Rent-Seeking and X-Efficiency

One last point is worthy of consideration. When one considers economic efficiency or x-inefficiency, one is largely investigating the extent

to which firms are x-efficient. However, x-efficiency in production, does not imply that such x-efficient firms are contributing positively to an economy's overall growth performance.

One example of this would be rent-seeking firms – firms whose objective is to earn income by transferring into their own coffers the income of others. This can be achieved by coercion or through rules and regulation that facilitate such rent-seeking behavior. These are not productive or wealth generating activities. Rather, they are ventures in income and/or wealth redistribution. And, to the extent that the macro-institutional environment encourages rent-seeking behavior, more investors will move into this domain as opposed to productive economic activities (North, 1990, 1994).

Rent seekers can be perfectly x-efficient, but serve to reduce the wealth of nations. Even if all existing rent-seeking firms converge towards a unique efficient equilibrium, one may not deduce from this a signal that society is maximizing its real income or real growth rate. Quite the opposite might be taking place. Economic efficiency at a local level does not necessarily translate into economic efficiency at the societal level.

Conclusion

A multiple equilibria analytical framework affords an alternative and more scientifically robust modeling scenario than does the modeling assumption of convergence towards a unique equilibrium that is both economically efficient and imputes unique processes by which such a unique equilibrium is achieved. In the multiple equilibria scenario, a unique equilibrium represents but one possible outcome. A multiple equilibria framework and narrative were championed by Herbert Simon as the more realistic and scientifically appropriate modeling worldview with which to tackle real world socio-economic issues. From Simon's perspective and approach to behavioral economics and scientific analysis, economic theory must be related (induced) from the stylized facts of life, which is the case in multiple equilibria scenarios. The multiple equilibria analytical template forces one to go beyond superficial analyses, which starts with the assumption of a unique and efficient equilibrium in consumption and production and in the process(es) to achieve this equilibrium outcome. This is driven by market forces and the hardwiring of the economic agent or decision maker. The possibilities of sustainable alternative choices in consumption, production and decision-making processes are assumed away. This falls victim to the

ought-is fallacy of assuming that that which is (exists, survives) is both rational and efficient because it exists.[4]

The pervasive success of the unique equilibrium approach is its simplicity, and its consistency with conventional economic worldviews that market forces should generate efficient outcomes. Developing upon Simon's insights, I model the conditions whereby multiple equilibria in choices, processes, and outcomes are sustainable in consumption and production. These conditions are reasonable given the structure of real world economics and the behavioral characteristics of human decision-makers. Therefore, the existence of particular choice sets, organizational forms, or processes should not be taken as proof of efficiency or uniqueness. One has to delve further into the black box of the firm and the household to make a determination of whether particular choices and outcomes are efficient or, in some sense, optimal.

Some of these arguments are highlighted in Figures 10.1 and 10.2. In Figure 10.2, multiple equilibria are linked to both consumption and production outcomes and to decision-making and organizational processes. In the consumption domain causality is linked importantly to differences in preferences (this needs to be controlled for real income and relative prices). In the production domain causality is linked to market forces, differential preferences (among economic agents), bargaining power (affects decisions, the decision-making process, and thereby outcomes), the legal environment, and customs (inclusive of norms, culture). Also of importance are multiple equilibria in human resource

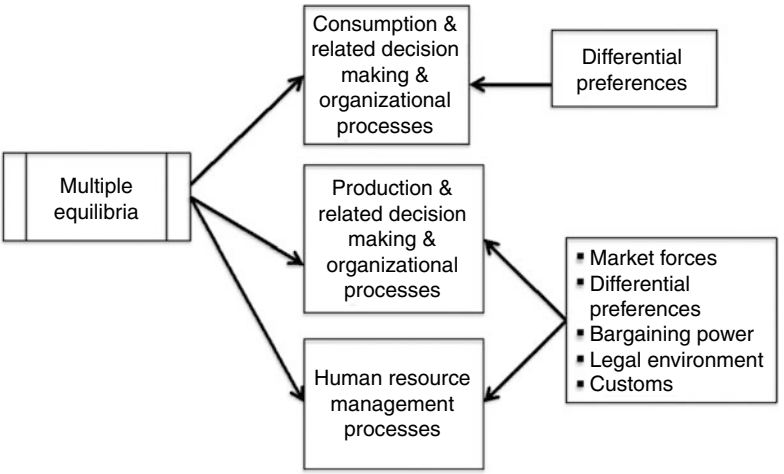


Figure 10.2 Multiple equilibria in consumption and production

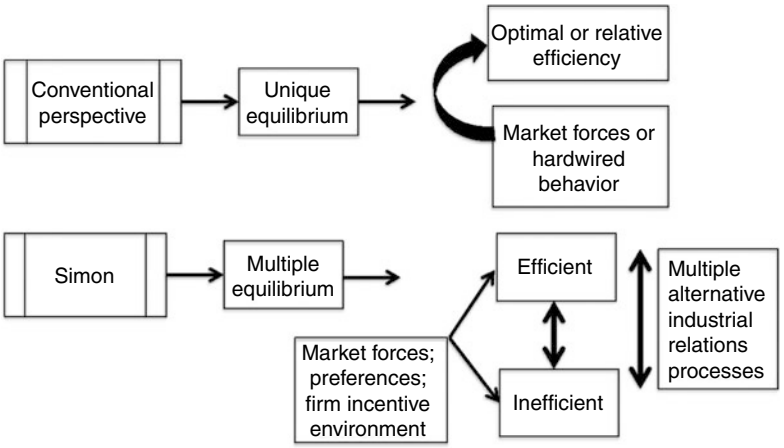


Figure 10.3 Perspectives on equilibrium states

management processes, which are causally linked to the same causal variables as production outcomes.

Figure 10.3 highlights some of the key differences between the conventional and Simon (multiple equilibria) modeling approaches. In the former, unique and efficient equilibria are causally related to market forces and hardwired behavior. Whereas, multi equilibria are causally linked to a spectrum of sustainable outcomes and processes.

A fundamental difference between these two different analytical approaches is that, unlike the efficient unique equilibrium approach, from a multiple equilibria perspective, one does not presume uniqueness and efficiency as a given. These are only possibilities. This approach is also consistent with rationality in decision-making as no presumption is made that any particular sustainable outcome and decision-making process is irrational because it deviates from a particular conventional economic norm. A multiple equilibria approach of this type opens the door to richer, more nuanced, and causally robust analyses of economies, building upon how actual economies function and evolve through historical time.[5]

Notes

1. Game theorists recognize the existence of multiple equilibria and even the existence of sub-optimal Nash, Prisoner's Dilemma-type, equilibria. But this is more in tune with the existence of a multiplicity of possible equilibria that are achievable where only one of the possible set is actually realized.

2. For aspects of the science perspective supportive of multiple equilibria see: Gould and Eldredge, 1977; Gould, 1997a, 1997b. More generally see, Hodgson, 2004.
3. Both Robinson (1980) and North (1994) have written on the importance of historical in contrast to logical time for economic analysis.
4. Related to this, extreme versions of ecological efficiency (a concept pioneered by Hayek) suggest that outcomes and choices, even if they are inconsistent with conventional economic norms, are not only rational (or smart), but are also efficient. Why? Because these choices and outcomes have passed the test of survival.
5. The author wishes to thank Roger Frantz and Louise Lamontagne for their comments and suggestions.

References

Alchain, A. A. (1950). Uncertainty, Evolution and Economic Theory. *Journal of Political Economy*, 58, 211–21.

Altman, M. (1996). *Human Agency and Material Welfare: Revisions in Microeconomics and their Implications for Public Policy*. Boston, Dordrecht, London: Kluwer Academic Publishers.

Altman, M. (1999). The Methodology of Economics and the Survivor Principle Revisited and Revised: Some Welfare and Public Policy Implications of Modeling the Economic Agent. *Review of Social Economics*, 57, 427–49.

Altman, M. (2001). When Green Isn't Mean: Economic Theory and the Heuristics of the Impact of Environmental Regulations on Competitiveness and Opportunity Cost. *Ecological Economics*, 36, 31–44.

Altman, M. (2002). Economic Theory, Public Policy and the Challenge of Innovative Work Practices. *Economic and Industrial Democracy: An International Journal*, 23, 271–90.

Altman, M. (2005a). Behavioral Economics, Rational Inefficiencies, Fuzzy Sets, and Public Policy. *Journal of Economic Issues*, 34, 683–706.

Altman, M. (2005b). Reconciling Altruistic, Moralistic, and Ethical Behavior with the Rational Economic Agent and Competitive Markets. *Journal of Economic Psychology*, 26, 732–57.

Altman, M. (2008). Behavioral Economics, Economic Theory and Public Policy. *Australasian Journal of Economic Education*, 5, 1–55.

Altman, M. (2009). A Behavioral-Institutional Model of Endogenous Growth and Induced Technical Change, *Journal of Economic Issues,* 63, 685–713.

Cyert, R. M. and March, J. C. (1963). *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.

Friedman, M. (1953). The Methodology of Positive Economics. In *Essays in Positive Economics*. Chicago: University of Chicago Press.

Gigerenzer, G. (2007). *Gut Feelings: The Intelligence of the Unconscious*. New York: Viking.

Gould, S. J. (1997a). Darwinian Fundamentalism. *The New York Review of Books*, June 12, 34–7.

Gould, S. J. (1997b). Evolution: The Pleasures of Pluralism. *The New York Review of Books*, June 26, 47–52.

Gould, S. J. and Eldredge, N. (1977). Punctuated Equilibria: The Tempo and Mode Of Evolution Reconsidered. *Paleobiology*, 3, 115–51.
Hodgson, G. M. (2004). *The Evolution of Institutional Economics: Agency, Structure and Darwinism in American Institutionalism*. London and New York: Routledge.
Hume, D. (2014 [1738]). *A Treatise on Human Nature*. Some Good Press. Kindle file.
Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, 93, 1449–75.
Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Strauss & Giroux.
Leibenstein, H. (1966). Allocative Efficiency vs. 'X-efficiency'. *American Economic Review*, 56, 392–415.
Leibenstein, H. (1979). A Branch of Economics Is Missing: Micro-Micro Theory. *Journal of Economic Literature,* 17, 477–502.
North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press.
North, D. C. (1994). Economic Performance Through Time. *American Economic Review*, 84, 359–68.
Robinson, J. (1980). Time in Economic Theory. *Kykos,* 33, 219–29.
Simon, H. A. (1978). Rationality as a Process and as a Product of Thought. *American Economic Review*, 70, 1–16.
Simon, H. A. (1979). Rational Decision Making in Business Organizations. *American Economic Review*, 69, 493–513.
Simon, H. A. (1987). Behavioral Economics. In *The New Palgrave: A Dictionary of Economics*, ed. J. Eatwell, M. Millgate and P. Newman. London: Macmillan, 266–7.
Simon, H.A. (1997). *Models of Bounded Rationality. Empirical grounded economic Reasons*, Vol. III, Cambridge, MA, MIT Press.
Smith, V. L. (2003). Constructivist and Ecological Rationality in Economics. *American Economic Review*, 93, 465–508.
Thaler, R. H., and Sunstein, C. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven and London: Yale University Press.

11
Organizational Decisions in the Lab

Massimo Egidi

A Short Introduction on the Origins of a Bounded Rationality Approach

Bounded rationality is a label that covers the most important advancements of Herbert Simon's scientific production. His fundamental contributions to cognitive psychology and to the theory of problem-solving were developed jointly, each being nurtured by the discoveries emanating from the other discipline. I will briefly review some steps on the path to the creation of the theory of bounded rationality, in order to introduce the issue of organizational decision-making and associated laboratory experiments.

From the very start, Simon built the idea of bounded rationality on close observation of the behavior of employees and managers in large organizations. In *Administrative Behavior*, published in 1947, he came to the realization that an organization's internal mechanisms, insofar as they are characterized by division of labor and cooperation, are the product of a complex activity of goal achieving. Important progress in this direction was achieved in the 1950s through some empirical analyses of managerial decisions that he conducted at the Graduate School of Industrial Administration of Carnegie Mellon. Among them, of primary interest is the research he conducted jointly with Cyert and Trow in which they realized that beyond routine decisions, managers make non-repetitive decisions that require solving problems in ill-defined conditions. (Cyert, Simon and Trow, 1956, p. 238)

The field analysis of problem-solving sowed the seeds of the bounded rationality theory. In *Organizations* (1958) March and Simon moved forward from the notion of problem-solving as individual activity to the notion of organizational problem-solving, with a clear recognition of the evolutionary processes of organizational adaptation and

organizational learning within business corporations. The identification of these processes was enhanced by the assumption that the division of labor can be considered a collective problem-solving activity. Thus, the development of a deeper theory of problem-solving became crucial in explaining human decisions and for the creation of new ideas in theory of organization – in particular, the notion of organizational routines within business firms – and of their evolution.

At the time he was finishing his work on *Organizations*, Simon began his collaboration with Allen Newell. *Human Problem-Solving*, which they published together in 1972, is a bridge between computation, artificial intelligence, and cognitive psychology. Simon went beyond the notion of computation as a human activity that relates means to ends, replacing it with the notion of *symbolic manipulation* and deepening the various connected mental abilities: memorization, evocation, categorization, abstraction, judgment.

The Dual Model of Reasoning and the Chunking Theory

In parallel with organizational theory and artificial intelligence, cognitive psychology became Simon's fundamental area of inquiry. A characteristic feature of his research style throughout his life was to maintain a close connection between his experiments on the psychology of cognition and the creation of the theory of problem-solving. Chess was selected as an important task environment for this research:

> As genetics needs its model organisms, its *Drosophila* and *Neurospora*, so psychology needs standard task environments around which knowledge and understanding can cumulate. Chess has proved to be an excellent model environment for this purpose. About a decade ago in the pages of this journal, one of us, with Allen Newell, described the progress that had been made up to that time in using information-processing models and the techniques of computer simulation to explain human problem-solving processes . . . A part of our article was devoted to a theory of the processes that expert chess players use in discovering checkmating combinations . . . a theory that was subsequently developed further, embodied in a running computer program, MATER, and subjected to additional empirical testing. (Simon and Chase 1973, p. 394)

With the exception of Alfred Binet's research (1894a, 1894b), whose importance was not recognized for a long time, and Djakow, et al.

(1927), psychological studies on the performance of chess players arose from De Groot's experiments in short-term memory during the 1960s. De Groot (1965) found that master chess players could reconstruct the position of pieces on a chessboard after only five seconds of study. He found a vast superiority of masters over weaker players in recalling meaningful information, which was not attributable to the masters' superior memory capacity:

> This result could not be attributed to the masters' generally superior memory ability, for when chess positions were constructed by placing the same numbers of pieces randomly on the board, the masters could then do no better in reconstructing them than weaker players. Hence, the masters appear to be constrained by the same severe short-term memory limits as amassed through years of constant practice. (Chase and Simon 1973, pp. 55–6)

De Groot attributed the above-capacity performance of masters to an *ability to classify groups of pieces as instances of familiar playing categories* and accordingly concluded that the key to expertise does not lie in any superior general processing abilities, but rather in *domain-specific skills*. These skills provide masters with a fast, unconscious, perceptual capacity of process board configurations.

As is well known, in his influential paper 'The magical number seven, plus or minus two,' George Miller (1956) estimated the capacity of short-term memory to be limited to about seven information units, defined as chunks; where a chunk is understood as information stored in long-term memory that is recalled as a single perceptual unit.

Chase and Simon (1973) proposed, as an order of magnitude estimate, the figure of 50,000 chunks – in this context familiar patterns of pieces – in the memories of chess masters.[1] The amount of information that must be processed simultaneously can overload a player's finite amount of working memory. The problem is to explain how masters avoid overload and use a huge amount of information, efficiently and without effort, given the limits of short-term working memory.

In 1973, Chase and Simon developed a *chunking theory* proposing that, in the course of acquiring their skill, chess players stored chunks in long-term memory corresponding to *patterns* of pieces. Each chunk consists of a small pattern that recurs frequently in the chess positions encountered while playing. In this approach, therefore, experts acquire and memorize *domain-specific* knowledge. This stored knowledge is, to some extent, organized; long-term memory is supposed to hold cognitive schemata

that vary in their degree of complexity and automation. It follows that human expertise comes from knowledge stored in these schemata, and the ability to engage in reasoning is related to the elements that have been organized in long-term memory:

> One key to understanding chess mastery . . . seems to lie in the immediate *perceptual processing*, for it is here that the game is structured, and it is here in the static analysis that the good moves are generated for subsequent processing. Behind this perceptual analysis, as with all skills . . . lies an extensive cognitive apparatus amassed through years of constant practice. What was once accomplished by slow, conscious deductive reasoning is now arrived at by fast, unconscious perceptual processing. It is no mistake of language for the chess master to say that he 'sees' the right move. (Chase and Simon, 1973, p. 56)

Two mechanisms characterize a chess master's decision:

> The first mechanism is recognition of cues in chess positions that evoke information from the expert's memory about possible moves and other implications of familiar recognized patterns of pieces. The second mechanism is planning by looking ahead at possible moves, possible responses by the opponent, possible responses to those responses, and so on. (Gobet and Simon, 1996b, p. 3)

The advancements proposed by Simon and colleagues then led to the discovery that the architecture of thinking is characterized by a complex interaction between the automatic and fast recall of the elements stored in the long-term memory and the conscious process of symbolic manipulation over the mental items. Simon and colleagues went further in this direction by trying to identify the size of chunks and to better qualify the process through which individuals can manage stored knowledge, despite the limits of short-term memory.

Automaticity

The chunking model has not only generated considerable empirical work, but has also been challenged on several grounds.[2] I will not review such criticisms, which do not challenge the general pillars on which the theory rests.

Automaticity implies that the process of recall is not under the conscious control of the individual, and therefore can be biased by

Table 11.1 The Einstellung experiment: data used in the problems

Problem	Given jars of the following size			Obtain the amount
	A	B	C	
1	29	3		20
2 Einstellung 1	21	127	3	100
3 Einstellung 2	14	163	25	99
4 Einstellung 3	18	43	10	5
5 Einstellung 4	9	42	6	21
6 Einstellung 5	20	59	4	31
7 Critical 1	23	49	3	20
8 Critical 2	15	39	3	18
9	28	76	3	25
10 Critical 3	18	48	4	22
11 Critical 4	14	36	8	6

distortion. Prior to Simon's research on chess, Luchins (1942), and Luchins and Luchins (1950), conducted experiments with subjects exposed to mathematical problems that had solutions at different levels of efficiency. The authors showed that subjects, having identified a simple solution to a task in a given context, may automatically and systematically apply the solution to other contexts, where it proves to be suboptimal. This process is called *mechanization of thought*.

In one of the most cited experiments (Luchins, 1942), participants had three jars of varying sizes and an unlimited water supply, and were asked to obtain a required amount of water. Everyone received a practice problem. People in the experimental group then received five problems (problems 2–6 in Table 11.1) prior to critical test problems (7, 8, 10, and 11). People in the control group went straight from the practice problems to problems 7–11. (Table 11.1). Problems 2–6 were designed to establish a set, called *einstellung*[3], for solving the problems with the same formula (using jars B-A-2C as a solution).

People in the experimental group were highly likely to use the *einstellung* solution on the critical problems even though more efficient procedures were available. In contrast, people in the control group used direct solutions that were much more simple (Table 11.2).

The experiments demonstrate that once a mental computation, deliberately performed to solve a given problem, has been repeatedly applied to solve analogous problems it may become *mechanized*. Mechanization

Table 11.2 The Einstellung experiment: solutions to the critical problems

Critical Problem	Einstellung solution B-A-2C	Direct Solution
7	$49-23-3-3=20$	$23-3=20$
8	$39-15-3-3=18$	$15+3=18$
10	$48-18-4-4=22$	$18+4=22$
11	$36-14-8-8=6$	$14-8=6$

enables individuals to pass from deliberate effortful mental activity to partially automatic, unconscious, and effortless mental operations.[4]

Moreover, the experiment also shows that when subjects have identified a solution to a task in a given context, they automatically transfer it to contexts in which the solution is inappropriate. This is a clear example of a *distortion* due to the automaticity of the process.

More recently, the dualism between the automatic process of recall and the effortful process of symbol manipulation has been deeply explored. The dual view, now widely accepted among psychologists, is based on the evidence that a large part of neural activity is related to *automatic* processes, which are faster than conscious deliberations and which occur with little or no awareness of effort. As within Simon's analysis, thinking is supposed to be composed of two different cognitive processes: on the one hand a controlled, deliberate, sequential, and effortful process of mental manipulation of items; on the other, a non-deliberate, automatic, effortless, and fast process of eliciting mental items from long-term memory.

According to Schneider and Shiffrin (1977), automatic processing is activation of a *learned sequence* of elements in long-term memory that proceeds without subject control, without stressing the capacity limitations of the system, and without necessarily demanding attention. Note that once a rule has been memorized in long-term memory, its automatic retrieval and use may happen in contexts that are not necessarily identical to the situation in which it originated. This happens in Luchins' experiments, in which the rule recalled in a context where it is not appropriate leads to a suboptimal solution.

This issue is a central point illustrated by Kahneman in his Nobel Lecture (2002), where he holds that the different *accessibility* of items may give rise to biases in judgment that can be corrected by deliberate reasoning. We will see in the next section the importance of accessibility in understanding the different learning ability of players and the biases in their activity of solving problems.

Organizational Routines and Individual Skills

Simon's findings on the cognitive processes of chess players allowed important improvements in the understanding of the features of decision-making within organizations. The great advancements he achieved in the field of the psychology of decision-making were essential for the creation of artificial models of decision and for applications in organizational contexts. Simon's empirical research on the cognitive properties of decision-making has been flourishing for a long time, thanks to his use of an artificial context (chess game) suitable for discovering the properties of *individual* decision-making. Only in 1994, was an artificial context created for studying *organizational* decision-making. Since then, organizational studies have been focused on field experiments, leaving untouched the area of laboratory experiments. The first relevant attempt in this direction was due to Michael Cohen, who created a game in the laboratory to explore the emergence of rules of coordination, or *organizational routines*:

> One line of recent work in psychology has developed in a way that nicely reinforces traditional organization theory views of routine. Work on procedural memory in human individuals has shown that it has distinctive properties. It is centered on skills, or know-how, rather than on facts, theories, or episodes (know-what), which seem to be more the province of an alternate, 'declarative', memory system. Procedural memory differs from declarative in its long decay times, and greater difficulty of transfer and of verbalization. This fits nicely with properties of routines observed in the field and in the laboratory (Egidi, 1994; Cohen and Bacdayan, 1994). And it appears to provide a firmer foundation in individual psychology for the characterization found in Nelson and Winter of routines as 'tacit' and highly stable analogs of individual skills. (Cohen *et al.*, 1996, p. 667)

In 1994 Michael Cohen and Paul Bacdayan created a card game called Target The Two (TTT) to explore in the laboratory the emergence of rules of coordination. The game is played by pairs of players who must cooperate in order to achieve a shared goal. The characteristics of the game also make it suitable for the study of team decision-making. In fact, the pairs of individuals who play this game do so in a context that displays certain features typical of teamwork.

On each run, the two players receive a random distribution of six cards (2♥, 3♥, 4♥, and 2♣, 3♣, 4♣. When the experimenter deals the six cards, each player has a card in hand that cannot be seen by the other

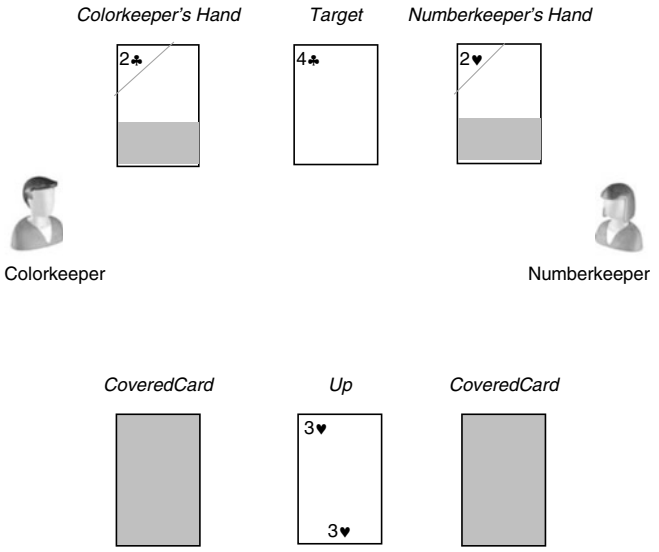


Figure 11.1 The board for Target the Two

player (Figure 11.1 indicates this one-sided visibility with a turned-up corner, below the headings *Colorkeeper's Hand* and *Numberkeeper's Hand*); beyond these two cards, two cards are face-up and two cards are face-down. Thus any player can see two cards beyond the card in his hand. The ultimate object of each hand is to maneuver the 2♥ into the area marked 'Target.'

A player moves by exchanging the card in his hand with one of the cards on the board (or the player can pass).

The players are subject to certain restrictions on moves. The player on the left, called Colorkeeper, may exchange with the target area card only if the *color* in the target is preserved. The player on the right, Numberkeeper, may exchange with the target only if the action preserves the *number* in the target area. Exchanges with board areas other than the target are not restricted. At every hand, Colorkeeper moves first.

Nelson and Winter note:

> In an intriguing experimental study, Cohen and Bacdayan (1994) demonstrated connections between the characteristics of skill memory at the individual level and some widely noted phenomena associated with organizational routines. In their study, subjects used playing cards in a two-person cooperative game of moderate

> complexity . . . As subjects played the game repeatedly, they became more and more efficient at recognizing the needed moves and making them quickly . . . Many results suggest that the micro-foundations of our routine-based perspective do reflect the realities of human physiology and cognitive functioning. At the least, this is true for routines that involve a substantial amount of skilled behavior at the individual level. (Nelson and Winter, 2002, p. 32)

By studying TTT in a computer game version, I found that the game permits two strategies; both allow players to achieve the goal, but one of the two strategies is more efficient than the other, depending upon the initial card distribution (Egidi, 1994).

Call the strategies S_1 and S_2; assume for simplicity that in the beginning there is a black card (n♣ n=2,3,4) in the target area, and that the goal card is 2♥, the two strategies can be defined by the following simple rules:

S_1: first Colorkeeper searches for the 2♣ and moves it into the target, and then Numberkeeper searches for the 2♥ and moves it into the target.

S_2: first Numberkeeper looks for a red card (with the same number as the one in the target) and moves it into the target; then Colorkeeper looks for and moves the 2♥ into the target.

It is clear that both strategies require coordination between the two players and that both players must jointly learn and choose the same strategy, otherwise they cannot achieve the goal. Moreover, as happens in Luchins' problem, for some initial card distributions S_1 dominates S_2 (i.e., S_1 requires fewer moves than S_2 to achieve the goal); for other initial distributions S_2 dominates S_1 (i.e., S_2 requires fewer moves than S_1 to achieve the goal).

Call A the set of the card distributions for which $S_1(A)$ dominates $S_2(A)$; analogously, call B the set of distributions for which $S_2(B)$ dominates $S_1(B)$. If the starting distribution of cards belongs to A, a *rational* player should activate strategy $S_1(A)$, because A dominates strategy B and vice versa for the cards distributions where B dominates A.

To check if players were able to discover both strategies and to apply them in a *fully rational* way, in an experiment with Narduzzo (Egidi and Narduzzo, 1997), we organized a tournament for two groups of players, P_A and P_B. During the first part of the tournament (the training phase) the P_A group was exposed to a set of starting configurations which could

be more easily played using strategy A than strategy B; and, vice versa, group P_B was exposed to starting distributions more easily playable using strategy B than strategy A. After the training phase, both groups were exposed to the same sequence of starting configurations. We observed the emergence of a persistent differentiation in player behavior. The group of players exposed to a set of configurations that led to easily learning one strategy continued to use it more frequently in the second part of the tournament, and symmetric behavior arose in the other group (see Figure 11.2).

In the second part of the tournament half of the starting distribution of the cards were of type A, and half of type B, therefore the sequence of runs of a *rational* pair of players should be characterized by the activation of S_1(A) and S_2(B) in the same proportion. However, within both groups a subset of players emerged that used the strategy they had learned in the training phase for all runs in the tournament, even when this strategy was dominated by the other.

We defined the behavior of these subgroup pairs as *fully routinized*: each pair of the subgroups was locked into one strategy and was unable to pay attention to different solutions, namely to learn again during the tournament. On the other hand a limited subgroup of pairs of players were *fully rational*, i.e., exhibited the ability to play the best strategy in the right context. This group played S_1(A) correctly half of the time and S_2(B) in the other half; all other pairs were influenced by the

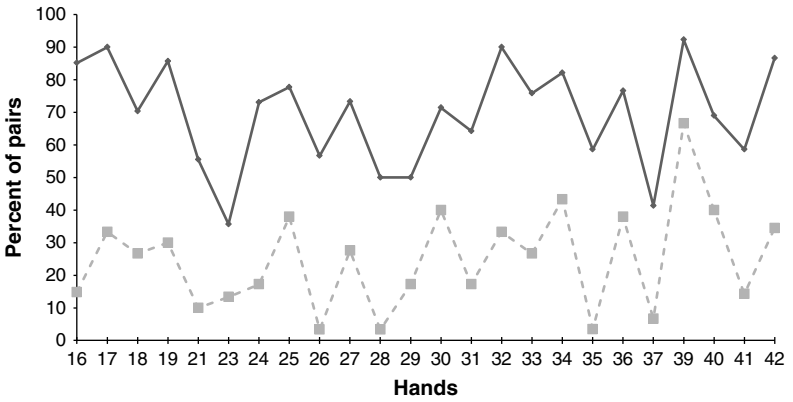


Figure 11.2 The dotted line shows how many times (in percentage) the group P_a played the strategy S_2 after the 16th hand, while the bold line shows how many times (percentage) the group P_b played the strategy S_2

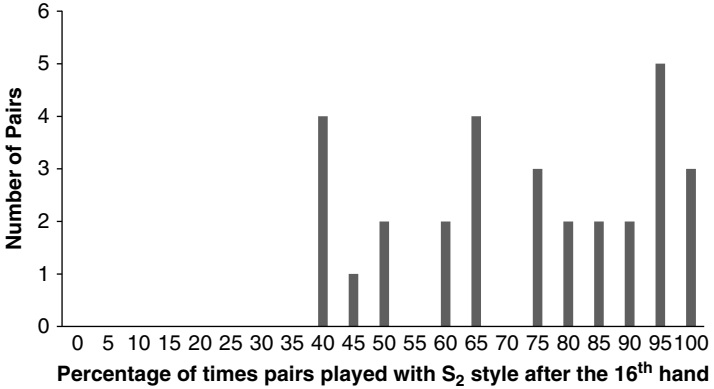


Figure 11.3 The vertical axis shows how many pairs played the strategy S_2 after the 16th hand x% of the times

initial 'imprinting' and played the strategy they had learned during the training phase more times that was rationally required (See Figure 11.3.).

These findings lead to two main different explanations about the persistence of routinized behaviors. The first explanation refers to the complexity of the cooperation between players. It is based on the fact that cooperation without verbal communication is a highly difficult task and it may easily fail if there are many different strategies the two players must jointly select without speaking each other. On the other hand, if both players persist in adopting the same strategy in all conditions, this facilitates the task of cooperating, even though the result is suboptimal.

The persistence of fully routinized behaviors could thus be interpreted on the basis of the difficulty of cooperating without verbal communication and, more precisely, the difficulty the players have in learning about each other in order to achieve a common knowledge of the available strategies. From this perspective the complexity of the organizational rules that prevent players from moving towards the discovery of new strategies can be seen. This fits neatly with the main statements of the literature that emphasize the importance of *sharing mental models* with members of a cooperating team (Orasanu and Salas, 1993). According to this literature, it is the sharing of mental models that enables each member of a team to synchronize his functions with the actions and decisions of his colleagues. A crucial role in cooperation is played by reciprocal expectations concerning each partner's strategic choice. It is on the basis of these expectations that each member of the team coordinates his

decisions and actions with other members. In our case, the two members of a cooperating pair must know the two alternative strategies and assume that their partner has the same knowledge; a complex situation that easily leads to failure, while if there is only one strategy all complexity in coordination vanishes.

The second explanation of the persistence of routinized behavior regards bounded rationality. Players exhibit cognitive limits in the discovery of strategies, insofar as after the discovery of one strategy their attention is led by the key elements of that strategy: this makes it more difficult to further pursue the discovery. This seems to fit perfectly with the dualism between the automatic retrieval of memorized items and the conscious elaboration of these items in short-term memory. In this view, the routinized behavior of a pair originates from the features of the *individual* learning process.

Thus a question arising from this experiment is whether the lock-in of many players into a routinized behavior is a process generated primarily by the organizational complexity, as the first point seems to suggest, or whether it emerges from some characteristic of the human individual learning process, as is implied by the second.

To disentangle the two alternatives, I prepared an experiment in which every player played the roles of both Numberkeeper and Colorkeeper. The sequences of the starting boards were the same as in the Egidi–Narduzzo experiment, to allow for a full comparison between the experiments.

The results showed the rise of persistent differentiation in player behavior (see Figures 11.4 and 11.5). The group of players exposed to a set of configurations which led more easily to one strategy continued to use it more frequently in the second part of the tournament, and symmetric behavior arose in the other group.

Moreover, in both groups there emerged a subset of players locked-in to one strategy alone, i.e., groups of players who, after the training phase, adopted a strategy once and for all, and insisted on using it even when hands could not be played efficiently with the strategy adopted.

We have therefore the experimental evidence that, also in the context of individual action, players may be trapped into a suboptimal strategy insofar as they used the same set of rules of action, even when they were inefficient, and were unable or unwilling to find alternative rules of action.

The experiment fully matches Luchins and Luchins (1950) previous experiments on the "mechanization of thought"; as we have seen, they show how a process of controlled reasoning – typically composed of

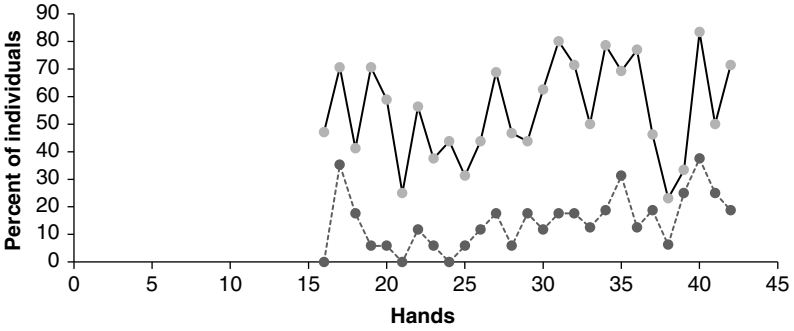


Figure 11.4 The dotted line shows how many times (in percentage) the group P_a played the strategy S_2 after the 16th hand, while the bold line shows how many times (percentage) the group P_b played the strategy S_2

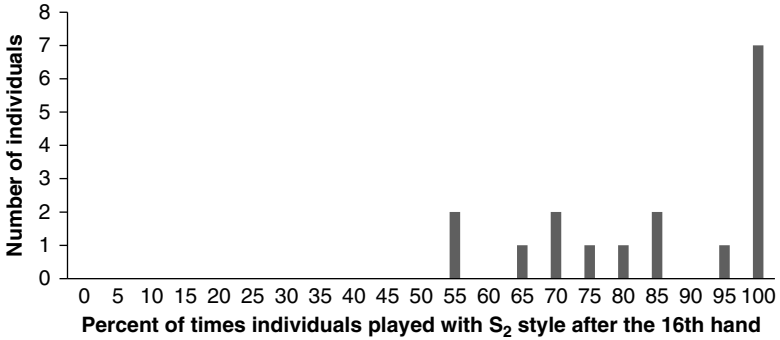


Figure 11.5 The vertical axis (x) shows how many pairs played the strategy S_2 after the 16th hand x percent of the time

effortful and deliberate mental operations – comes to be substituted by an effortless process of automatic and unconscious thinking. In our last experiment, the reason why some players stay locked-in to one single strategy cannot be explained by the difficulty of cooperating, due to the complexity of organizational rules, for the simple reason that cooperation does not exists in this context. Here the explanation must be found in the cognitive properties of the individual reasoning process, and precisely in the mechanization of thought. The discussion in the next section will thus consider the consequences of mechanization of thought on players' behavior.

Differentials of Accessibility

As we have seen before, the studies of chess provide evidence that during problem-solving activities the problem's complexity may generate a mental overload. Chunking, through mechanization of thoughts, reduces short-term memory load and mental effort. Experiments of neuroimaging (Haier *et al.*,1992) confirm this process. When too many symbolic manipulations are required to explore the alternatives, players may be unable to create a comprehensive internal model of the actions required to play optimally. They explore only a limited part of the strategy space, and memorize them. The mental overload provides an explanation for chunking, since it prevents players from achieving a full exploration of the problem's space. While the two different solutions of the game are very simple and easily representable in abstract form, different players come up with different solutions in relation to the history of their exploration of the space of the solutions.

We have created different perspectives for the two groups, each exposed to different sequences of starting configurations in such a way that they become familiar with a different set of configurations, all easily solvable with the same simple strategy. In that way, we have *artificially* created a familiarity with a subset of the game's configurations and its related strategy.

When information comes to mind, it might not necessarily be the right information, nor the most suitable for the problem at hand:

> Highly accessible features will influence decisions, while features of low accessibility will be largely ignored. Unfortunately, there is no reason to believe that the most accessible features are also the most relevant to a good decision. (Kahneman, 2002, p. 459)

In both experiments the players' attention was artificially manipulated. In fact during the preliminary training phase players were exposed to card distribution that led them to learn one strategy more easily than the other. After the training phase the key cards of *one* strategy were familiar to the players, and therefore more accessible than the key cards of the alternative strategy. Thus the *differential accessibility* explains in a simple way why a large number of players continued to use the familiar strategy even when inefficient: they simply did not pay attention to the key card of the unfamiliar strategy.

Of course, a question arises about the reasons why some players repeatedly employ the same strategy, while others are able to learn new

ones. One central element in beginning to answer this question is the role of automatic processes, in *directing players' attention*. If players are more familiar with a strategy, say S_1, the strategy directs their attention and search to the cards 2♣ and 2♥; if they are more familiar with S_2 this strategy directs their attention towards different key cards. Thus if strategy S_1 is more accessible to a player then S_2, he simply does not pay attention to the cards that should permit him to activate S_2.

As Kahneman notes:

> The acquisition of skills gradually increases the *accessibility* of useful responses and of productive ways to organize information, until skilled performance becomes almost effortless. This effect of practice is not limited to motor skills. A master chess player does not see the same board as the novice (Kahneman, 2002, p. 453)

It follows that if one strategy comes to the mind of a player more easily than another, the effect is to free the memory load, direct attention, and filter 'relevant' information.

In consequence, it could reasonably be expected that the time taken to discover the second strategy varies in relation to the extent to which the first strategy governs the player's attention. The ease with which the first strategy comes to the mind of a player can help or hinder the discovery of the second. Two conflicting processes are active: one the one hand the automatization of one strategy reduces the mental load permitting the exploration of new alternatives; on the other hand, automatization implies high accessibility to the familiar strategy and this, in turn, leads the attention to the strategy's key cards.

Once a player is familiar with one strategy and has reduced the mental load for solving the task, he may search for different strategies *if his attention is not strictly focused on the key cards of the familiar strategy*. We can thus argue that a key element in understanding what triggers the search for a new strategy is the strength with which the routinization directs the attention of the players, preventing a better solution being found. This point has been clearly discussed by Bilalic *et al.* (2008, 2010) in recent papers. They show that having found one solution, expert chess players were looking for a better one. But their eye movements showed that they continued to look at features of the problem related to the solution they had already envisaged. So there are contrasting elements that prevent or lead to the discovery of a new solution and it is clear that some of these elements are not the product of a conscious deliberation. An advancement in this respect is due to Schuck *et al.*

(2015). Through multivariate neuroimaging analyses they show that before the spontaneous change to an alternative strategy, the medial prefrontal cortex encoded information that was irrelevant to the current strategy but necessary for the new one.

More research is certainly required in this direction; a better understanding of the different degrees of accessibility of memorized items will prove extremely valuable in order to discover the triggering elements of search and innovation.

A key component in understanding the emergence of search and discovery is related to the opposition between strategy exploitation and exploration. This opposition was first highlighted by Jim March (1991) and later discussed by Levinthal and March (1993). They suggest that during the adjustment to changing external conditions, organizations can be trapped into using strategies which only prove efficient in the short run. Organizations may not be able to jump out of the trap and reorganize themselves in a more efficient way when the external conditions change. As we have already seen, the same happens in TTT and thanks to the experimental data we can distinguish between the individual and the organizational elements that govern routinization and innovative behavior.

The notion of 'cognitive traps', then, can be considered in both individual (Schuck *et al.*, 2015) and organizational contexts, and the problem of increasing the understanding of the connections between the micro (neural and psychological) approach and the 'macro' (organizational data from the field) approach is a challenging problem to pursue.

When a Problem is Really Difficult: Limits for Artificial Intelligence?

So far, we have seen a number of aspects of the progress in artificial intelligence, psychology of cognition, and theory of organizations that have resulted from Simon's research. Before concluding it might be interesting to identify some areas of major difficulty to further progress.

In *Human Problem-Solving* (1972) Newell and Simon described the characteristics of an artificial system able to solve problems, the *General Problem Solver*. The ideas for the basic mechanisms of the General Problem Solver were derived from a careful analysis of the masters' playing characteristics during the performance of a problem-solving task.

General Problem Solver fully develops a computational model for representing and solving problems based on simple basic ideas; every

problem can be represented by a tree consisting of nodes (representing the different states of the problems), and branches (representing the relations between nodes). In chess the nodes are the chessboards and the branches are the possible moves; by applying the permitted moves to the initial board, then to the boards of the possible responses by the opponent, and so on repeatedly, we expand the tree of possible states of the game. As is well known, the size of this tree goes beyond the practical limits of computability of every modern computer. Even if it has been demonstrated that a winning strategy exists in chess, it has thus far proved impossible to compute.

It follows that a player, or an artificial player, must limit the exploration of the space of problems, and select only a fraction of the strategy. Through his extensive research on the psychology of chess masters, Simon discovered that the search of the skilled player is guided by *heuristics*, which permit the exploration of future chessboards to be restricted to a small tree of possibilities (usually less than 100). The study of heuristics paved the way for the creation of the first artificial player.

One of the most widely known heuristics to solve problems suggests decomposing it into parts that can be solved separately. The decomposition process is repeatedly applied to each sub-problem until elementary sub-problems, easily solvable, are identified. The efficiency of the General Problem Solver model requires the overcoming of two obstacles: one is the problem of reducing and simplifying the problem space; the other is the decomposition and modularization of this space.

The latter issue has received noticeable attention in recent years in a series of works that study the properties of different decompositions and, in particular, the trade-off generated by problem decompositions. On the one hand the more a problem is decomposed into smaller modules the more easily it can be solved and the more such modules can be re-used in similar problems. On the other hand, finer decompositions are more likely to separate interdependent elements into different modules, which are therefore less likely to generate optimal or high quality solutions (Marengo *et al.*, 2005; Egidi, 2007).

The first issue, namely the problem of reducing and simplifying the problem space is a remarkably difficult one, as Simon recognized:

> From the problem-space approach there emerged the idea that size of problem space (number of branches at each node and depth of search to a solution node) was a principal determinant of problem difficulty. This is certainly true (and mathematically demonstrable) if problems are solved by random trial-and-error search. However, it

> *does not explain why some problems with quite small problem spaces are difficult for intelligent people*. The Missionaries and Cannibals (Hobbits and Ores) problem has a problem space of only 16 nodes, and monster problem versions of the three-disk Tower of Hanoi problem, only 27 nodes. Both problems are known to be difficult for human subjects who encounter them for the first time. The Tower of Hanoi problem has long been a major task environment for work in problem-solving. (Kotovsky, *et al.*, 1985, pp. 247–8, italics added)

Simon claimed that the performance of all reasoning systems crucially depends on *problem representation*: the same problem might be easy or difficult, depending on the way we describe it. In *The Sciences of the Artificial* (1996), he discusses the representation problem in the context of design and suggests that problem-solving can be read as *changes of representation*, 're-framing' of a problem in a way that makes the solution easier.[5] It is also worth noting that there are very close connections between the representation of a problem and its decomposability. In a sense, an effective representation of a problem is the one which somehow generates an effective decomposition, i.e., one that strikes an effective balance in the above mentioned trade-off between the ease and re-usability granted by finer decompositions and optimality of the solution.[6]

I will not go further in this direction, but would draw attention to the opinion expressed by Allen Newell in 1965 when he noted that the greatest 'limitation of the current stock of ideas about problem-solving' is that 'we do not yet have any useful representations of possible representations.'

Although researchers have acknowledged the importance of alternative representations of a problem – not least due to renewed attention after Kahneman and Tversky's discovery of the framing effect, in which different representations of the same problem give rise to different decisions – little investigation has been carried out in this area. The efforts required in building up an artificial mechanism that may solve problems through changes of representations are still in their infancy.

The extraordinary achievements of Herb Simon in understanding rationality and, more importantly, human reasoning have generated an extremely successful class of artificial intelligence models.[7] Nevertheless, there exist problems that either cannot be solved or cannot be efficiently solved by an artificial intelligence program. However, despite the computational limits, within Simon's heritage the idea of using chess and more generally artificial games as the appropriate environment in which

to cumulate knowledge on the psychology of human decisions reveals great potential for new advancement in understanding organizational behavior.[8]

Notes

1. As they note, this is a magnitude roughly comparable to that of natural language vocabularies of the college-educated people.
2. The template theory (Gobet and Simon, 1996a) was proposed as a refinement of the chunking theory, to respond to the criticisms. It retains the idea that chunks, which are recursively made of (sub) chunks, are indexed by a hierarchical discrimination network, but suggests that frequently encountered chunks develop into higher-level structures (templates) with slots allowing rapid long-term memory encoding.
3. '*Einstellung* – habituation – creates a mechanized state of mind, a blind attitude toward problems; one does not look at the problem on its own merits but is led by a mechanical application of a used method' (Luchins, 1942).
4. I will use the terms *mechanization* and *routinization* interchangeably.
5. A comprehensive discussion of the state of the art in problem representation is provided by Eugene Fink (2002).
6. See Marengo (2015) for a detailed discussion of the properties of representations and the derived solvability of a puzzle studied by Simon himself, i.e., the Tower of Hanoi.
7. A well-known example is Deep Blue, a computer program developed by IBM, which won its first game against a world champion on February 10, 1996, when it defeated Garry Kasparov in game one of a six-game match.
8. This chapter originates from a presentation at the XX *Organization Science* Winter Conference 'The Conversation Continues: Reflecting and Building on the Work of Michael Cohen', Steamboat Springs, Colorado February 6–9, 2014. I am grateful to many participants for their useful observations. I am also grateful to Fernand Gobet for valuable suggestions regarding this chapter.

References

Bilalic, M., McLeod, P. and Gobet, F. (2008). Why Good Thoughts Block Better Ones: The Mechanism of the Pernicious Einstellung (Set) Effect. *Cognition*, 108 (3).

Bilalic, M., McLeod, P. and Gobet, F. (2010). The Mechanism of the Einstellung (Set) Effect: A Pervasive Source of Cognitive Bias, *Current Directions in Psychological Science*, 19(2) 111–15.

Binet, A. and Henneguy, L. (1894a). *Psychologie des grands calculateurs et joueurs d'échecs*. Paris: Hachette.

Binet, A. (1894b). La mémoire des joueurs d'échecs qui jouent sans voir. *Travaux du Laboratoire de Psychologie Physiologique des Hautes Études*, 2, 32–8.

Chase, W. G. and Simon, H. A. (1973). Perception in Chess. *Cognitive Psychology*, 4, 55–61.

Cohen, M. D. and Bacdayan, P. (1994) Organizational Routines are Stored as Procedural Memory: Evidence from a Laboratory Study. *Organization Science*, 5 (4), 554–68.

Cohen, M. D., Burkhart, R., Dosi, G. Egidi, M., Marengo, L., Warglien, M. and Winter, S. (1996). Routines and Other Recurring Action Patterns of Organizations: Contemporary Research Issues. *Industrial and Corporate Change*, 5 (3), 653–98.

Cyert, R. M., Simon, H. A. and Trow, D. B. (1956). Observation of a Business Decision. *Journal of Business*, 29, 237–48.

De Groot, A.D. (1965). *Thought and Choice in Chess*. The Hague: Mouton.

Djakow, I.N., Petrowski, N.W., and Rudik, P.A. (1927). *Psychologie des Schachspiels*. Berlin: de Gruyter.

Egidi, M. (1994). Routines, Hierarchies of Problems, Procedural Behaviour: Some Evidence from Experiments, IIASA working Paper WP-94-58. In *The Rational Foundations of Economic Behaviour*, ed. K.J. Arrow, E. Colombatto and M. Perlman., London: Macmillan, 114, 303–33.

Egidi, M. (2007). Decomposition Patterns in Problem Solving. In *Cognitive Economics: New Trends, Contributions to Economic Analysis*, vol. 280, ch. 1, part I, Decisions and Beliefs, ed. R. Topol and B. Walliser. New York : Elsevier.

Egidi, M. and Narduzzo, A. (1997). The Emergence of Path Dependent Behaviors in Cooperative Contexts. *International Journal of Industrial Organization*, 15 (6), 677–709.

Fink, E. (2002). *Changes of Problem Representation: Theory and Experiments*. Berlin: Springer-Verlag.

Gobet, F. and Simon, H. A. (1996a). Templates in Chess Memory: A Mechanism for Recalling Several Boards. *Cognitive Psychology*, 31, 1–40.

Gobet, F. and Simon, H. A. (1996b). The Roles of Recognition Processes and Lookahead Search in Time-constrained Expert Problem Solving: Evidence from Grandmaster Level Chess. *Psychological Science*, 7, 52–5.

Haier, R. J., Siegel, B. V., MacLachlan, A., Soderling, E., Lottenberg, S. and Buchsbaum, M. S. (1992). Regional Glucose Metabolic Changes after Learning a Complex Visuospatial/Motor Task: A Positron Emission Tomographic Study. *Brain Research*, 570(1–2): 134–43.

Kahneman, D. (2002). *Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice*. Nobel Prize Lecture, December 8.

Knez, M. and Camerer, C. (1994). Creating Expectational Assets in the Laboratory: Coordination in 'Weakest-Link' Games. *Strategic Management Journal*, 15, 101–19.

Kotovsky, K., Hayes, J. R. and Simon, H. A. (1985). Why are Some Problems Hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, 17, 248–94.

Levinthal, D. A. and March, J. G. (1993). The Myopia of Learning. *Strategic Management Journal*, 14, 95–112.

Luchins, A. S. (1942). Mechanization in Problem-Solving. *Psychological Monograph*, 54, 1–95.

Luchins, A. S. and Luchins, E. H. (1950). New Experimental Attempts in Preventing Mechanization in Problem-Solving. *The Journal of General Psychology*, 42, 279–91.

March, J. G. (1991) Exploration and Exploitation in Organizational Learning. *Organization Science*, 2(1), Special Issue: Organizational Learning: Papers in Honor of (and by) James G. March, pp. 71–87.

March, J. G. and Simon, H. A. (1958). *Organizations*. New York, NY: Wiley.
Marengo, L. (2015). Representation, Search and the Evolution of Routines in Problem Solving. Forthcoming in *Industrial and Corporate Change*.
Marengo, L., Pasquali, C. and Valente, M. (2005). Decomposability and Modularity of Economic Interactions. In *Modularity: Understanding the Development and Evolution of Complex Natural Systems*, ed. W. Callebaut and D. Rasskin-Gutman. The Vienna Series in Theoretical Biology. Cambridge, MA: MIT Press, 835–97.
Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *The Psychological Review*, 63 (1).
Nelson, R. R. and Winter, S. G. (2002). Evolutionary Theorizing in Economics. *The Journal of Economic Perspectives*, 16 (2), 23–46.
Newell, A. (1965). Limitations of the Current Stock of Ideas about Problem Solving. In *Electronic Information Handling*, ed. A. Kent and O. Taulbee. Washington, DC: Spartan Books.
Newell, A. and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.
Orasanu, J. and Salas, E. (1993). Team Decision Making in Complex Environments. In *Decision-making in Action: Models and Methods*, ed. G. Klein, J. Orasanu, R. Calderwood and C. Zsambok. Norwood, NJ: Ablex.
Schneider, W. and Shiffrin, R. M. (1977). Controlled and Automatic Human Information Processing: I. Detection, Search, and Attention. *Psychological Review,* 84 (1).
Schuck, N. W., Gaschler, R., Wenke, D., Heinzle, J., Frensch, P. A., Haynes, J. and Reverberi, C. (2015). Medial Prefrontal Cortex Predicts Internally Driven Strategy Shifts, *Neuron*, 86, 331–40.
Simon, H. A. (1947). *Administrative Behavior, a Study of Decision-Making Processes in Administrative Organizations*. New York: Macmillan.
Simon, H. A. (1996). *The Sciences of the Artificial*. Cambridge, MA: MIT Press.
Simon, H. A. and Chase, W. G. (1973). Skill in Chess: Experiments with Chess-playing Tasks and Computer Simulation of Skilled Performance Throw Light on Some Human Perceptual and Memory Processes. *American Scientist*, 61 (4), 394–403.

Part III
Milieux

12

Simon on Social Identification: Two Connections with Bounded Rationality

Rouslan Koumakhov

Introduction

Social identification is one of Herbert Simon's most recurrent themes. Starting with *Administrative Behavior* (hereafter, *AB,* Simon, 1947/1997), he investigated that theme in his scientific work on an impressive number of occasions. Perhaps it is section 3, entitled Perception and Identifications, of chapter 6 (Cognitive Limits on Rationality), in March and Simon (1958/1993), that symbolizes Simon's main concern in this issue, its connection with human rationality and emphasis on every individual's multiple *belongings* to social groups (in the broad sense, i.e., from primary groups, to formal organizations, to the whole of society). From this general standpoint, 'identification with groups is the major selective mechanism controlling human attention in organizations (and elsewhere)' (Simon, 1993, p. 137). Accordingly, social identification is a process that allows people to stabilize their anticipations, and to coordinate perceptions and interpretations of reality. While this tendency to identify with groups appears necessary to build and maintain social systems, it also leads to mimetic opinions and behavioral conformity.

Compared with the notion of bounded rationality, however, Simon's analysis of identification was only taken up to a limited extent by the social and human sciences that he so strongly influenced. Because his analysis is complex and appeals to major concepts developed in related disciplines, this begs the question of the exact place of social identification in Simon's account of decision process and social interaction. My argument is that, in this account, not only is there a strong connection between bounded rationality and social identification, but also that

such a connection implies value systems and cognitive representations. Considered in this manner, the problem of identification is central to Simon's decision-making and social theory, with its focus on mental states and understanding reality.

Antecedents: Identification and Role Theory

Historically, the concept of social identification mainly emerged within the framework of role theory under a double influence. One appeals to Cooley (1902) and Mead (1934), but it can also be traced to other pragmatists like James (1925) and Dewey (1930). The second influence is from the structuralist tradition in anthropology (Linton, 1936) and sociology (Merton, 1938). From at least the 1930s, role theory – with the related concept of identification – proliferated in social sciences, including many seminal works (Katz and Schanck, 1938; Sarbin, 1943; Lindesmith and Strauss, 1949; Newcomb, 1950; Foote, 1951; Parsons, 1951; Merton, 1957; Turner, 1956; Gouldner, 1957, 1958; Becker and Geer, 1960).[1] In spite of various theoretical differences, most of the authors define roles as *socially expected (collections of) behavioral patterns*. In this view, roles prescribe behavior conforming to conventional norms, and thus enable the stabilization of actors' anticipations and of social institutions. Correspondingly, social identification is explicitly or implicitly considered as identification with roles or, at least, as implying role-taking.

Simon seems to have assimilated this approach, especially in its structuralist version, when he published *AB* in 1947, including a long chapter entirely devoted to organizational identification. In very simple terms, Simon's standpoint may be formulated as follows: 'agreement of an employee to join an organization is essentially agreement to accept a role' (Simon, 1955, p. 44). Accordingly, the role corresponds to a set of organizational behavioral rules. Accepting one's own role also means accepting the general distribution of roles, which helps to stabilize mutual expectations and make organizational behavior predictable. As March and Simon (1958/1993) expressed it later, 'not only is the role defined for the individual who occupies it, but it is known in considerable detail to others in the organization who have occasion to deal with him' (p. 22).

More generally, each social system is a 'system of interlocking roles' (Simon, 1958: 53) so that 'the "institutions" of a society may be regarded as rules specifying the roles that particular persons will assume in relation to one another under certain circumstances' (*AB*: 183). This also

applies to political systems: 'A political régime prescribes appropriate behavior rôles to its participants; these rôles include appropriate actions to constrain any particular participant (or small group of participants) who departs from his rôle . . . To each individual in a political régime, consequently, the régime looks exceedingly stable as long as he expects other individuals to support it' (Simon, 1953b, p. 510; *régime* and *rôle* are in the original).

It is striking, however, that, in Simon's account of identification with groups, the use of the role concept appears to be rather secondary. It would be a mistake to explain this by the fact that role theory was still relatively novel at the time when *AB* or Simon *et al.* (1950/1991) offered quite elaborate views on social identification. Thus, while there is close continuity between these books and Simon's further investigations of identification (whether involving the notion of roles or not), such investigations do not seem to be influenced by later developments of role theory, like those using the symbolic-interactionist framework (on the concepts of role and role-identity developed within this framework, see, for example, Turner, 1962; Callero, 1986). My interest in this point is as follows: because the structuralist tradition was primarily interested in norms and duties derived from positions in social groups, and because the Cooley-Mead pragmatist tradition was less concerned by the individual himself than by his socialization and his public self[2], the role theories mentioned above rarely address directly the question of the cognitive or subjective motivational mechanisms leading to and sustaining role-taking.[3] As Foote (1951) notes, 'roles as such do not provide their own motives' (p. 14).[4]

Simon provides an answer to this question, because he departs in two important ways from purely role-theory approaches to identification, both departures being related to the bounded rationality assumption.

Social Identification 1: Value System as a Cognitive and Moral Filter

The first departure appears in Simon's emphasis of the close relationship between the concept of identification and that of common values, or value systems. Accordingly, 'through identification, organized society imposes upon the individual the scheme of social values in place of his personal motives' (*AB*, p. 295). The idea applies to other social settings. Thus, organizational identification is the process whereby 'the (organizational) values gradually become "internalized" and are incorporated into the psychology and attitudes of the individual participant'

(*AB*, p. 278). Similarly, when analyzing the immediate level of everyday interaction represented by working groups within formal organizations, Simon *et al.* (1950/1991) insist that 'individual members tend to *identify* with the group so that they interpret the values accepted by the group as their own values. The group becomes an extension of the individual's self' (p. 95).

Strictly speaking, this value-oriented idea of identification does not contradict the role-oriented idea. That the internalization of common values leads to identification with a social entity, and to various degrees implies role-taking (or vice versa), is a point on which most structuralists agree (e.g., Merton, 1938). Similarly, from Simon's standpoint, role, as a decision premise, relates behavior to a given value (provided that such a value is formulated in the operational form). What is, however, distinctive about this standpoint is that it shifts the focus from the behavioral component of identification (as developed within the framework of role theory) to both cognitive and moral components.

With cognitive components, this shift is due to the connection that Simon establishes between social identification and bounded rationality. The main point is that such a connection is mediated by group values. Thus, 'it is clear that attention may narrow the range of vision by selecting particular values . . . Identification, then, has a firm basis in the limitations of human psychology in coping with the problem of rational choice' (*AB*, p. 288). It is by relating intentions to a limited number of values that social identification influences an actor's perceptions and, therefore, appears as a solution for restricted human cognitive abilities. These limited values define a set of objectives that are not necessarily the best in reality, but are commonly believed to be right and/or rational and, from this socially recognized, or conventional, angle appear to be the best. To the extent that individuals consciously or blindly accept such values–beliefs as a result of belonging to a given group, they accept corresponding objectives as their own. In other words, by relating values to bounded rationality, Simon stresses the motivational function of social identification in the decision process.

Note that emphasis on the cognitive aspects of this motivation follows on from one of the main (and, probably, one of the most underestimated) tenets of Simon's account of decision-making, that accepting group values results from the actor's own psychology.

> The rational individual is, and must be, an organized and institutionalized individual. If the severe limitations imposed by human psychology upon deliberation are to be relaxed, the individual must

> in his decisions be subject to the influence of the organized group in which he participates. His decisions must not only be the product of his own mental processes, but also reflect the broader considerations to which it is the function of the organized group to give effect. (*AB*, p. 111)

The tenet, that the boundedly rational individual is necessarily organized and institutionalized, is a major consequence of the Simonian value-focused approach to identification. This approach establishes the following connection that can be called *Connection 1*: *bounded rationality – common values – social identification*. Characterized in these terms, identification is not limited to role playing. Roles, as sets of behavioral rules and expectations, limit the choice of actions with regard to given goals, while values (even if they also influence the choice of actions) limit the choice of the goals. *Connection 1* means that internalizing values mainly involves mental processes, especially intentions, while role playing refers to resulting behaviors.

Focusing on values has another implication; these cognitive components of identification closely interact with *moral* ones. An illustration is provided by what Simon regards as the most important common values influencing the decision process within formal organizations: efficiency; organizational and sub-unit goals; fairness standards; professional standards; and respect of status and authority systems or broader social goals (full employment, promotion of trade, and so on). In varying degrees, these values offer configurations of rational and ethical imperatives.

It may be objected that moral imperatives do not really apply to efficiency, which is a merely technical criterion. Simon's argument, however, is that efficiency does involve various ethical concerns. Simon *et al.* (1950/1991) take the example of the US Post Office, which can be easily generalized to private corporations. They note that efficiency can be seen from at least two different viewpoints. From the ordinary citizen's viewpoint (and from the viewpoint of Post Office executives), the expected result is a speedy mail service, the work of a postman being just the neutral means. From the viewpoint of the postman, however, the expected result is his salary and/or various kinds of non-monetary satisfaction, the mail service being the neutral means. As Simon *et al.* (1950/1991) emphasize, it is the citizen's viewpoint 'that is almost always taken in traditional discussions of efficiency . . . But it is simply a reflection of the particular values of our society, as already explained, and not the intrinsic property of the organizational system'

(p. 499). That is, what seems to be an issue of *objective* rationality at work also appeals to the prevailing values of the society. Individual interpretations of the notion of efficiency depend therefore upon how people internalize *community* values or, in other words, identify with community.

Social Identification 2: Accepting a Cognitive Framework

Simon's second departure from role theory can be stated as follows: when people identify with a social group, 'actually, their representations of the world change. The change is not simply in values but also in what they know and believe' (Simon, 1999, p. 113). Simon clearly appealed to mental representations – including the actor's *knowledge* and *beliefs* – from the beginning of 1950s, though related ideas had already been presented in *AB*. Correspondingly, social role is a set of behavioral rules, which do not fully determine individual actions: 'The fact that behavior is structured in roles says nothing, one way or another, about how flexible or inflexible it is' (Simon, 1991, p. 127), so that an actor can make his boundedly rational choice by a more or less deliberate interpretation of these rules. This corresponds to Simon's general view that focusing on behavior does not lead to an appropriate account of the multiple factors influencing decision-making and social interaction. In this view, to provide such an account, theory needs to make *decision premises* its central notion. Social roles thus correspond to only some of the decision premises. Among others, there are 'premises about the state of the environment based directly on perception, premises representing beliefs and knowledge' (Simon, 1963, p. 742). This regular return to knowledge and beliefs, viewed as main components of mental representations influencing decision premises, is part of Simon's explicitly formulated criticism of behaviorism in social sciences and has strong links to his own proactive role in the cognitive revolution of the 1950s and 1960s.

To study mental representations (with corresponding knowledge and beliefs), Simon's decision theory uses the following concept:

> The limit of human understanding in the presence of complex social structures leads human beings to construct simplified maps (i.e., theories or models) of the social system in which they operate, and to behave as though the maps were the reality. To the extent that such maps are held in common, they must be counted among the internal constraints on rational adaptation. (Simon, 1952, p. 1135)

What Simon describes here as simplified maps, models, or theories of reality – and also as states of mind, frames of reference or cognitive frameworks (Simon 1952, 1959, 1999; Simon *et al.*, 1950/1991; March and Simon, 1958/1993) – imply three major features.

(i) This mental construction is a necessary consequence of bounded rationality. Simplified maps/cognitive models replace reality for an individual by selecting information and by providing a biased picture of the world. That is, rational thinking – viewed as the actual-life process, as what happens in people's minds and not in terms of modeling individual *objective* choice – does not deal directly with reality, but only with a mental image, which gives a very incomplete and inexact reproduction of reality. Decision-making is thus possible to the extent that simplified maps/cognitive models involve categories, classification schemes and, more generally, criteria allowing actors to identify relevant matters in a given situation. Because simplified maps help actors to define problems, they can thereafter choose corresponding solutions to these problems.

(ii) Insofar as simplified maps (models) are shared, their content (beliefs and knowledge) is largely defined by the institutional environment – viewed in terms of social-group influence. From this standpoint, the process of social identification strongly shapes the mental representations of a group member. As Simon put it, the 'institutional setting . . . provides the framework within which his own mental processes operate' (*AB*, p. 111). In other words, mental representations are largely formed and maintained through the process of social identification. As in the case of internalized values (*Connection 1*), this process applies to groups at various levels, such as professional and local communities, working teams, or society. Insofar as group members adopt the same simplified map (model), they share the same definition of the environment, of problems, and of the knowledge required to deal with them. Correspondingly, such a shared cognitive map, or model, defines what is rational and moral in the ways in which group members solve problems, and is thus a social construction, both cognitive and ethical.

It is in this sense – adaptation to bounded rationality by accepting a common cognitive framework – that 'group identification is a major determinant both of goals (defining the "we") and of knowledge and beliefs, for both are formed in the groups that people associate with in work and leisure' (Simon, 2001, p. 12,785). Various types and levels of social identification (with corresponding values and normative actions) are stabilized because they draw on those common cognitive models that filter information for an individual. That is, social identification 'defines a context or framework for viewing situations' (Simon,

1993, p. 197) and serves thus as a perceptual mechanism. In the real-life context, in the context of limited rationality, such a mechanism guides an individual's attention and leads to a common understanding of the situation.

(iii) Simplified maps, or models, contain value systems. Remember that, in Simon's view, the value approach to identification is strongly marked with moral considerations. Insofar as this approach is integrated within the cognitive framework approach, such a framework defines not only what action is rational in the situation, but also what is right and fair (and what is not). In other words, a common simplified model adopted within a given group offers an ethical outlook on problem setting and problem solving.

This second approach to identification thus establishes the connection that can be called *Connection 2*: *bounded rationality – common cognitive framework – social identification*. Note that simplified models/cognitive frameworks contain value systems so that this *Connection 2* implies or, in a sense, absorbs *Connection 1* above. Correspondingly, the value approach to social identification becomes integrated within the cognitive framework approach.

One of Simon's many illustrations of these relationships between identification and mental representations, describes how bureaucracy interacts with the world of politics and how, in this context, bureaucrats encounter various types of information: 'Filtering this information through its previous beliefs and expectations, interpreting it in the light of legislative mandates and policies, in the light of the values held in the social strata from which its members are drawn, bureaucracy translates and transforms the incoming information flow into actions (and inactions)' (Simon, 1967, p. 94). Bureaucracy thus becomes an information-processing mechanism with values and cognitive frameworks that serve as filters for the profession. This helps Simon to discuss the neutrality of public administrators with regard to the political environment. He focuses on two kinds of values. One is linked to social origin (birthright) and is confirmed through selection processes, for instance, the working-class origin of policemen or the middle-class origin of teachers. The second is linked to professions and are held by experts recruited by organizations to provide technical skills. 'But the professional and specialist bring to the organization not only these skills, but values as well – values that are acquired during professional training and enforced by the desire for professional approval and esteem' (p. 96). The importance of such professional values in bureaucratic decision-making is one of Simon's major topics. Thus, Simon *et al.* (1950/1991) invoke value assumptions involved in standard technical

solutions provided by city planners and architects who are 'often quite explicit in telling us that we *ought* to have plenty of green spaces among our dwellings, or that we *ought* to build our public structures in a modern style rather than an imitation of Gothic or Renaissance' (p. 546). Experts in other professional communities may be less explicit about what basic value preferences they have in common, or even not really understand them. The lack of such understanding is one of the reasons why professionals can be very resistant to change and new concepts.

Another example is that of the forestry department in the public administration. This department, 'and schools of forestry, belong to a single social group that is responsible for forming the values of the future forest ranger during his training as well as supervising his application of those values on the job' (Simon, 1967, p. 97). In this case (and in others; Simon talks about cases of a municipal library or the US Public Health Service), professional and organizational values are very close.

From this investigation of professional values and cognitive frameworks, Simon concludes that members of bureaucracy are *neither neutral agents of the state, nor pure technicians*. Not only do bureaucrats share the same mental representations and related value systems as the political environment they operate in, but also this sharing is reinforced by pressure from their reference groups, especially the professions.

One major implication of establishing *Connection 2* is that it links together issues of cognition to issues of normativity. Shared representations clearly appear as decision premises providing a legitimate basis for the choice of action. It is remarkable that the notion of common cognitive maps/models does not reduce the foundations of human behavior to internalized norms, nor does it provide an account of decision-making and social interaction on the grounds of rational choice theory. Joining a group becomes an answer to limited mental abilities, the normative character of this solution being a consequence of the cognitive properties of any individual. Rationality, in the Simonian understanding (as a means–ends relationship in general), thus implies, among other things, belonging to a group with its rules, values, and knowledge systems. In a sense (to the extent that such maps are held in common), common representations – and thereafter identifications with corresponding groups – are necessary components of human rationality. This implication has several consequences.

Multiple Identifications

To introduce one of these consequences, I take Simon's (1953a) study of the formation of the Economic Cooperation Administration (ECA) in

1948. As a former insider (Simon held a position in this US government agency), he tells the story of how the ECA became responsible for organizing aid to European countries under the Marshall Plan. Though this case is not widely known among social scientists, Simon comes back to it on various occasions (see, for example, Simon *et al.*, 1950/1991; Simon, 1955, 1991, 1993) thereby showing the importance he attaches to this experience. He distinguishes between six major approaches to the ECA's mission and organization, which 'were far from congruent to each other' (1953a, p. 227). Each approach appeared as a mental representation shared by a group of people, broadly corresponding to the administrative unit in which they were employed. Such a representation involved a specific conception of foreign aid, including a comprehension of the general mission, tasks, methods, and, thereby, structure of the ECA. In other words, these alternative approaches suggested different views of the problem so that, as Simon later pointed out, for 'each of these views, a set of organizational roles could be inferred, and each such structure of roles was quite different from the others' (1991, p. 132).

However, the key idea of the ECA story is that these suggested definitions and roles were closely connected to the educational and professional backgrounds of employees, particularly their previous shared experiences within the public administration. Thus, the *commodity-screening approach* derived from the tradition of wartime aid programs supplying specific commodities to allies. Its main proponents came from corresponding government agencies (such as the export licensing unit of the International Trade Administration in the Commerce Department, or the interim-aid unit in the State Department). Similarly, the *balance of trade approach*, mainly supported by professional economists, originated from previous research work on estimating the aggregate needs of each European country (definition of consumption levels, productive capacities, and resulting dollar gap). It was this second problem representation that finally evolved into the ECA's major approach. The mechanisms that led to this approach dominating others are not important for my concern (though they have rich implications for organizational studies). What is important is that the ECA story shows a connection between social identification, here rooted in professional belonging, and understanding the world.

Moreover, the story clearly explains one of the major reasons for Simon's interest in formal organizations. More precisely, any organization appears as a place of multiple interacting identifications, regarded in terms of the assimilation of group values and problem representations. The ECA case emphasizes the importance of alternative

professional affiliations in defining situations and goal setting. It clearly demonstrates how such alternatives give rise to real-life difficulties, because resulting representations (and then actions) are not necessarily compatible with each other. To the extent that each identification implies a stable mental representation of the situation and problem, individuals do more than just passively internalize professional (or other sub-group) values. Rather, they tend to understand organizational life in general, and organizational goals in particular, from the perspective of those values.

Starting with *AB*, Simon analyzes this issue of multiple identifications in terms of value compatibility, sometimes formulated as conformity to mores or customs. For instance, 'a society establishes certain very general values through its basic institutional structure, and attempts to bring about some conformity between these general values and the organizational values of the various groups that exist within it' (*AB*, p. 280). Because such conformity cannot be taken for granted, organizational identification is often in conflict with societal identification. To illustrate this, Simon used a simple example taken from the newspaper. This concerned a confrontation between the California State Engineer, defending the priority of rebuilding rural roads, and the Federal Government, trying to improve highways to military standards. The Engineer regarded his role 'in terms of the value of "civilian need" rather than in the value of "military need" or some composite of both values' so that his 'judgments are *consequences of his organizational identifications*, and that his conclusions can be reached *only if these identifications are assumed*' (*AB*, p. 286, emphasis added).

Simon, however, does not reduce the problem to such incompatibility between organizational and extra-organizational levels of social interaction. He assigns similar – and sometimes even stronger – importance to interaction between identifications with different sub-groups within the organization, and identification with the 'focus organization itself', to use March and Simon's (1958/1993) term. In his view, in most small, face to face, everyday working groups, members 'generally develop a "we" feeling' (Simon *et al.*, 1950/1991), which creates stronger identification than loyalty to the whole organization. In this connection March and Simon (1958/1993) distinguish between three main types of groups – beside the organization itself – and corresponding identifications. These – surely, overlapping – identifications are extra-organizational (profession, family, local community, unions), sub-group (usually organizational sub-units), and task group (class of individuals performing the same task). March and Simon's constant focus on

organizational sub-goals – corresponding to group (especially professional) cognitive frameworks – in contradistinction to organizational goals – corresponding to a broader cognitive framework – is marked with the same idea that group identification usually exerts more powerful influence on mental representations, moral imperatives, task goals, and on conformity to the corresponding behavioral standards of the organization members, than does identification with the focus organization itself.

Simon's insistence that the same organization member has various, often conflicting, social identifications (with appropriate cognitive frames and behavioral expectations) may certainly be regarded as expanding on W. James' idea of multiple selves (1925). Simon's emphasis on the role of working groups refers to this idea: 'Each individual possesses, therefore, a whole hierarchy or pyramid of loyalties, with the loyalties to the smaller, more intimate groups toward the base of the pyramid usually taking precedence over loyalties to the larger groups in case of conflict' (Simon *et al.*, 1950/1991, p. 97). Note that he discusses hierarchy and conflict between loyalties/identifications before further work on contradictory normative expectations in structuralist (especially in terms of roles, e.g., Merton, 1957; Goode, 1960) or symbolic-interactionist (especially in terms of identity salience, e.g., Stryker, 1968) traditions.

More importantly, the way, in which Simon treats the question of multiple identifications, reveals two main concerns of his account of organizations. One appeals to what he views as a major feature of organizations (and many other social entities), that they are systems of 'interlocking' groups (*AB*; Simon *et al.*, 1950/1991; Simon, 1952). This, apparently rather ordinary fact, gives rise to serious decision-making and social interaction problems. In Simon's account, the rationality of an individual's choice in organizational settings depends upon his/her anticipation of other individuals' choices. Conformity to group beliefs – as a normative solution, facilitating the stability of mutual expectations and choice making – does not guarantee successful interaction if these beliefs contradict those adopted by other groups and/or by the focus organization. Each organizational group may establish its own cognitive, moral, and behavioral norms, incompatible with those established by other groups, and thus leading to a lack of coordination between groups. Moreover, for an organization member who simultaneously identifies with some or even many of these groups with conflicting frameworks and interests, the decision process and coordination become rather complex issues. A second, related concern is the interaction between organizational mores and societal mores, such

interaction being regarded as a major topic in organization theory (Simon, 1952). Because the issue of group mores influencing individual choices and behavior becomes an issue of social identifications, the programmatic concern about interaction also becomes a concern about identification with organization and identification with society. Viewed in this way, the question of multiple identifications is at the heart of Simonian decision and social theory.

Self-enforcement and Social Legitimation of Knowledge

Another set of consequences resulting from the cognitively based concept of social identification refers to the idea of knowledge itself. March and Simon (1958/1993) explore knowledge systems in the context of formal organizations. Starting with the individual level of cognition, they note that the simplified mental model – which they also call 'frame of reference' – filters perceptions, retains those that seem compatible, and reinterprets them to avoid incompatibility. Passing from the individual to the organizational level, March and Simon focus on the appearance of new strong selection mechanisms created through various kinds of in-group communication: 'Since these perceptions have already been filtered by one or more communicators, most of whom have a frame of reference similar to our own, the reports are generally consonant with the filtered reports of our own perceptions, and serve to reinforce the latter' (March and Simon, 1958/1993, p. 174). In this view, organizational/social identification is a self-enforcing cognitive mechanism, which selects information and structures it into a set of hierarchical priorities. Selecting information through the common cognitive framework has two related implications, both of which are conventional in character.

The first concerns the phenomenon of *stipulated facts*. Speaking about the limited cognitive abilities of the real-life individual, Simon (1958, p. 54) points out: 'Even the "facts" on which he acts obtain their status as facts by a social process of legitimation, and have only a very indirect and tenuous connection with the evidence of his senses.' March and Simon (1958/1993) expand this idea with regard to formal organizations. Drawing on their account of social identifications as selection mechanisms, they emphasize that accepting information – already selected, summarized, and then transmitted by other organizational sub-group members – is a matter of belief. The recipient of such information should typically 'repose his confidence' in this process of work division and internal communication, 'and if he accepts the communication at all, accept it pretty much as it stands' (p. 187). Moreover,

this collective confidence-belief, which is responsible for considering such information as legitimate, is reinforced during the process of communication insofar as both organization members who directly process information and recipients share identical values and other references. Because of this, the meaning of stipulated facts conforms to (and thereafter reinforces) common reference frames. Facts become conventional phenomena insofar as they are considered as correct within a certain cognitive/moral framework adopted by a social group.

The second implication concerns the phenomenon that Simon (1957) labels *authority of confidence*. Our capacity to evaluate directly the validity and correctness of the knowledge and judgments of experts and technicians is extremely limited, so we rely on shared confidence in the source of such knowledge and judgments. Some of Simon's examples are formulated in a straightforward manner: 'Few persons before taking prescribed medicines ask their physicians for a demonstration of the curative properties of the prescription' (*AB*, p. 181). Similarly, 'if experts in air warfare tell us that the defense of the country requires B-36's, who (except a competing expert) will dare substitute his judgment for theirs?' (Simon *et al.*, 1950/1991, p. 547). It is not merely technical competencies or skills that ultimately legitimate such an authority, but social recognition of those competencies and skills. As in the case of stipulated facts, such recognition relies on common confidence-belief, which is corroborated through work division and internal communication channels. This time, however, confidence-belief applies to one of the major sources of the stipulated facts, 'the one who possesses the credentials of "expertness"' (*AB*, p. 181) and, more generally, people collectively regarded as specialists. In this view, social recognition is rooted in the shared cognitive framework, which validates the information interpreted by a specialist. As March and Simon emphasize, 'the "facts" he communicates can be disbelieved, but they can only rarely be checked. Hence, by the very nature and limits of the communication system, a great deal of discretion and influence is exercised by those persons who are in direct contact with some part of the "reality" that is of concern to the organization' (1958/1993, p. 187). For this reason, the process of legitimizing stipulated facts is used, 'consciously and unconsciously, as a technique for acquiring and exercising power' (ibid.). That is, technical authority, which, at first sight, might seem to draw only on the possession of certain skills and knowledge, becomes a convention grounded in generalized opinion.

This idea of authority of confidence also appears in Simon *et al.* (1950/1991), who discuss leadership that relies on the recognition of

personal capacities rather than on formal status. As they point out, 'the followers are seldom in a position to judge accurately all the personal qualities of the leader. If he can maintain a high level of group loyalty, they will be quite prepared to believe that he is intelligent or well-informed' (p. 104). The legitimacy of such leadership is not provided by the group members' direct comprehension of the leader's capacities but by his/her ability to use strong group identification. In this context, followers identify with the leader – 'he becomes a rallying point, a symbol of the group's oneness' (p. 105) –, which further reinforces identification with the whole group. In this process, group confidence in a leader's real qualities replaces knowledge of such reality.

Financial crises, and the generalized euphoria that usually precedes them, typically provide a good illustration of the judgment process related to Simonian stipulated facts as an important part of identification processes. Conformity of expectations – canonically studied, at least, since Keynes (1936) – implies that investors make money when they correctly evaluate changes in the dominant opinion, even if such opinion is wrong (see also Reinhart and Rogoff, 2001; Shiller, 2008). In other words, legitimacy of information does not result from reality, but from shared opinions, which individuals belonging to the same professional community try to follow. It is remarkable that Simon (1958) applies the idea of stipulated facts in a similar sense: 'Expectations may be based less on the observation of external realities than on the observation of the expectations of others. Although the term "panic" is somewhat old-fashioned, its synonyms play an important role in contemporary business cycle theory' (p. 54).

Financial crises also exemplify authority of confidence, another consequence of identification mechanism. Evidence provided by insiders (e.g., McDonald, 2009; Michaelson, 2009) shows to what extent generalized belief in expertise and blind loyalty to charismatic leaders are key values of the financial community and are anchored in analysts' mental structures. Judgments filtered by such values and structures seem to replace neutral and technical evaluations (on this point, the financial world bears a curious resemblance to that of Simon's bureaucrats) and to be largely responsible for the 2007–8 mortgage crisis.

Conclusion

Simon shows how social identifications provide stability for expectations and behavior by setting value systems and mental frames

that deal with bounded rationality. Identification with groups is thus deeply rooted in cognitive processes. Because such identification implies a strong self-reinforcement socio-psychological mechanism, it helps anticipations to converge and serves to focus the attention of individuals, organizations, professions, and other social entities in both decision and coordination processes.

Notes

1. These references, which correspond to the golden age of role theory, are of course incomplete. For reviews of the literature, see Biddle (1986) or Stryker (2002).
2. On this point, see, for example, Kuhn (1964), Stryker (2002) or Meltzer (2003).
3. I leave aside the question of the extent to which role theory, and especially its pragmatist variations, was marked by *psychological* behaviorism.
4. Indeed, Parsons (1937) demonstrates some interest (intertwined with logical considerations) in cognitive processes, but this earlier work does not contain any explicit concept of social role.

References

Becker, H. S. and Geer, B. (1960). Latent Culture: A Note on the Theory of Latent Social Roles. *Administrative Science Quarterly*, 5, 303–13.

Biddle, B. J. (1986). Recent Developments in Role Theory. *Annual Review of Sociology*, 12, 67–92.

Callero, P. (1986). Toward a Meadian Conceptualization of Role. *Sociological Quarterly*, 27, 343–58.

Cooley, C. (1902). *Human Nature and Social Order*. New York: Scribners.

Dewey, J. (1930). *Human Nature and Conduct*. New York: Modern Library.

Foote, N. (1951). Identification as the Basis for a Theory of Motivation. *American Sociological Review*, 16, 14–21.

Goode, W. (1960). Norm Commitment and Conformity to Role-Status Obligations. *American Journal of Sociology*, 66, 246–58.

Gouldner, A. (1957). Cosmopolitans and Locals: Toward an Analysis of Latent Social Roles—I. *Administrative Science Quarterly*, 2, 281–306.

Gouldner, A. (1958). Cosmopolitans and Locals: Toward an Analysis of Latent Social Roles—II. *Administrative Science Quarterly*, 2, 444–80.

James, W. (1925). *The Principles of Psychology*. New York: Henry Holt.

Katz, D. and Schanck, R. (1938). *Social Psychology*. New York: John Wiley.

Keynes, J. M. (1936). *The General Theory of Employment Interest and Money*. London: Macmillan.

Kuhn, M. H. (1964). The Reference Group Reconsidered. *Sociological Quarterly*, 5, 5-21.

Lindesmith, A. R. and Strauss, A. (1949). *Social Psychology*. New York: Dryden.

Linton, R. (1936). *The Study of Man*. New York: Appleton-Century.

March, J. and Simon, H. A. (1958/1993). *Organizations* (2nd edn) Cambridge: Basil Blackwell.

McCall, G. and Simmons, J. L. (1966). *Identities and Interaction*. New York: Free Press.
McDonald, L. (2009). *A Colossal Failure of Common Sense: The Inside Story of the Collapse of Lehman Brothers*. New York: Three Rivers Press.
Mead, G. H. (1934). *Mind, Self and Society: From the Standpoint of a Social Behaviorist*. Chicago: University of Chicago Press.
Meltzer, B. (2003). Mead's Social Psychology. In *Symbolic Interaction: An Introduction to Social Psychology*, ed. N. J. Herman and L. T. Reynolds. Walnut Creek, CA: AltaMira Press, 38–54.
Merton, R. (1938). Social Structure and Anomie. *American Sociological Review*, 3, 672–82.
Merton, R. (1957). The Role-Set: Problems in Sociological Theory. *American Sociological Review*, 8, 106–20.
Michaelson, A. (2009). *The Foreclosure of America*. New York: Berkley Books.
Newcomb, T. M. (1950). *Social Psychology*. New York: The Dryden Press.
Parsons, T. (1937). *The Structure of Social Action*. New York: McGraw-Hill.
Parsons, T. (1951). *The Social System*. London: Routledge & Paul.
Reinhart, C. and Rogoff, K. (2001). *This Time Is Different: Eight Centuries of Financial Folly*. Princeton: Princeton University Press.
Sarbin, T. (1943). The Concept of Role-Taking. *Sociometry*, 6, 273–85.
Shiller, R. (2008). *The Subprime Solution: How Today's Global Financial Crisis Happened and What to Do about It*. Princeton: Princeton University Press.
Simon, H. A. (1947/1997). *Administrative Behavior* (4thedn). New York: Free Press.
Simon, H. A. (1952). Comments on the Theory of Organizations. *American Political Science Review*, 46, 1130–1139.
Simon, H. A. (1953a). Birth of an Organization: The Economic Cooperation Administration. *Public Administration Review*, 13, 227–36.
Simon, H. A. (1953b). Notes on the Observation and Measurement of Political Power. *Journal of Politics*, 15, 500–16.
Simon, H. A. (1955). Recent Advances in Organization Theory. In *Research Frontiers in Politics and Government*, ed. S.K. Bailey. Washington: Brookings Institution, 23–44.
Simon, H. A. (1957). Authority. In *Research in Industrial Human Relations. A Critical Appraisal*, ed. C. M. Arensberg, S. Barkin, W. E. Cambers, H. L. Wilensky, J. C. Worthy and B. D. Dennis. New York: Harper & Brothers Publishers, 103–15.
Simon, H. A. (1958). The Role of Expectations in an Adaptive or Behavioristic Model. In *Expectations, Uncertainty, and Business Behavior*, ed. M. J. Bowman. New York: Social Science Research Council, 49–58.
Simon, H. A. (1959). Theories of Decision-Making in Economics and Behavioral Science. *American Economic Review*, 49, 253–83.
Simon, H. A. (1963). Economics and Psychology. In *Psychology: A Study of a Science, Volume 6*, ed. S. Koch. New York: McGraw Hill, 685–723.
Simon, H. A. (1967). The Changing Theory and Changing Practice of Public Administration. In *Contemporary Political Science: Toward Empirical Theory*, ed. I. de Sola Pool. New York: McGraw-Hill, 86–120.
Simon, H. A. (1991). Bounded Rationality and Organizational Learning. *Organization Science*, 2, 125–34.
Simon, H. A. (1993). Strategy and Organizational Evolution. *Strategic Management Journal*, 14 (S2), 131–42.

Simon, H. A. (1999). The Potlatch between Economics and Political Science. In *Competition and Cooperation: Conversations with Nobelists about Economics and Political Science*, ed. J. E. Alt, M. Levi and E. Ostrom. New York: Russell Sage Foundation, 112–19.

Simon, H. A. (2001). Rationality in Society. In *International Encyclopedia of the Social and Behavioral Sciences, Volume 19*, ed. N. Smelser and P. Baltes. Oxford: Elsevier Science, 12, 782–6.

Simon, H. A., Thompson, V. A. and Smithburg, D. W. (1950/1991). *Public Administration* (2nd edn) New Brunswick: Transaction Publishers.

Stryker, S. (1968). Identity Salience and Role Performance: The Relevance of Symbolic Interaction Theory for Family Research. *Journal of Marriage and the Family*, 30, 558–78.

Stryker, S. (2002). Traditional Symbolic Interactionism, Role Theory, and Structural Symbolic Interactionism. The Road to Identity Theory. In *Handbook of Sociological Theory*, ed. J. Turner. New York: Kluwer Academic/Plenum Publishers, 211–31.

Turner, R. H. (1956). Role-taking, Role Standpoint and Reference-group Behavior. *American Journal of Sociology*, 61, 316–28.

Turner, R. H. (1962). Role-Taking: Process vs Conformity. In *Human Behavior and Social Process*, ed. A.M. Rose. Boston: Houghton Mifflin, 20–40.

13
Models of Environment

Marcin Miłkowski

Herbert A. Simon is well known for his account of bounded rationality. Whereas classical economics idealized economic agency and framed rational choice in terms of the decision theory, Simon insisted that agents need not be optimal in their choices. They might be mere *satisficers*, i.e., attain good enough goals rather than optimal ones. At the same time, behaviorally as well as computationally, bounded rationality is much more realistic.

One of the most important factors in his theorizing on bounded rationality was the structure of the environment of the agent (Simon, 1956). This might sound surprising today because Simon is all too often classified as one of the proponents of classical, symbolic cognitive science. After all, he favored symbolic models over situated action frameworks (Vera and Simon, 1993). However, already in his 1956 paper he acknowledged that his account of bounded rationality was similar to robotic models built by Grey Walter. Moreover, Simon's (1996) story about the ant that uses the environment to make the navigational task easier has become one of the classic examples for later proponents of the extended mind (Clark and Chalmers, 1998). So why did Simon stress the situatedness of cognition and deny the need to reject symbolic modeling? Was he deluded or self-contradictory?

The purpose of this chapter is to understand the role of the structure of the environment in Simon's work on models of cognition. It will be shown that his modeling methodology includes both internals of information-processing architectures and environmental constraints. The inner architecture is important insofar as it is a constraint on adaptation to the environment, and remains invariant over multiple different environments; hence, it is relevant to explaining behavior in any environment. For this reason, physical symbol systems are to be

understood as both situated and adaptive; otherwise, they cannot be flexible and support cognition. Even if Simon's treatment of symbols remains vague and underspecified, the idea that naturalistic models need to interleave internal and external states remains surprisingly timely.

Bounding Rational Decisions in the Environment

One of Simon's key ideas is bounded rationality. In contrast to classical economics, with its highly idealized instrumental rationality of the *homo economicus*, Simon opted for a behaviorally and psychologically plausible alternative. The classical account of rationality in economics and decision theory relies on the assumption that decisions are determined by preferences over outcomes, and that these outcomes are known and fixed. Decision makers are supposed to maximize their net benefits, or utilities, by making choices that lead to the highest benefit. This model makes decision-making instantaneous, and idealizes away any learning or developmental processes; moreover, the basic factors are the incentives, or the expected utilities of the outcomes (Jones, 1999).

In contrast, Simon claims that organisms 'fall short of the ideal of "maximizing" postulated in economic theory' (Simon, 1956, p. 129). Instead, they adapt well enough to *satisfice*. Real decision-making in limited agents is markedly constrained and enabled by 'the limitations upon the capacities and complexity of the organism' (ibid.). Moreover, the environments possess properties that permit further simplification of the choice mechanisms in organisms. To argue for this claim, he performs a systematic, mathematical exploration of a simple model that approximates ideal rationality in an extremely simple agent. The model is not supposed to reflect real decision-making processes; it simply demonstrates the possibility that agents can make rational decisions thanks to the environmental constraints appropriately reflected in their choice mechanisms. (At the same time, Simon believes that the model does represent human rationality to some extent.)

This theme recurs in Simon's thinking. The idea that the complexity of the environment can help simple agents solve complex tasks is illustrated in *The Sciences of the Artificial* with ant navigation. The ant navigates in the sand; the route taken is quite complex, as he does not foresee the obstacles on his way home:

> He must adapt his course repeatedly to the difficulties he encounters and often detour uncrossable barriers. His horizons are very close, so that he deals with each obstacle as he comes to it; he probes for ways

> around or over it, without much thought for future obstacles. It is easy to trap him into deep detours. (Simon, 1996, p. 51)

Ants are simple agents but the complexity of their behavior reflects the complexity of the environment they find themselves in. The ant is strikingly similar to the simple agent hypothesized by Simon in 1956. Simon (1956, p. 130, n. 8; 1996, p. 52) also points out a striking resemblance to electromechanical tortoises built by W. G. Walter, which had only a tactile sense of the environment.

What the Tortoise Taught Us

The electromechanical tortoises (W. G. Walter 1950b, 1953, 1950a) were one of the milestones in the history of biorobotics (the discipline which explains biological behavior with robots; cf. Webb, 2002) and bionics (the discipline that builds robots inspired by biological solutions). W. G. Walter built robots – called *tortoises* for their looks – as models of animal behavior: environmentally induced exploration (in a robot called jokingly *Machina speculatrix*); and conditional reflexes (in *Machina docilis*). The robots are mobile and react to light, and if their battery is depleted, they return to their charging base. The case of both robots contains a simple switch that is activated whenever a robot bumps into an obstacle; then it changes route. The first robot explores the environment freely, as driven by moderate light; it avoids both darkness and extreme light. The second one also contains a microphone and reacts to auditory stimuli. Interestingly, it can learn by association.

W. G. Walter describes these machines as displaying several features of life, namely:

1. *parsimony*, or economy of structure and function;
2. *speculation*, or the propensity to explore the environment actively rather than to wait passively for something to happen;
3. *positive tropism*, or attraction to certain perceptible variables;
4. *negative tropism*, or avoidance of certain perceptible variables;
5. *discernment*, that is, distinction between effective and ineffective behavior;
6. *optima*, that is, a tendency to seek conditions with moderate and most favorable properties rather than maxima;
7. *self-recognition*.

To be exact, the last feature is a little exaggeration on Walter's part. Basically, the machines he built do not recognize individuals at all; they

merely react to certain environmental conditions but cannot categorize anything. They react to their own light in the mirror because they are sensitive to light, and that induces particular movements in robots, but calling such movement *self-recognition* is a huge stretch.

Moreover, what Walter calls *optima* has, obviously, little to do with satisificing in Simon's sense. If robots are appropriately adjusted, they avoid extreme stimuli. The exact range of stimuli they find attractive is adjustable by the modeler.

Parsimony is the feature that these robots share with the early model of bounded rationality developed by Simon (1956). Walter started the tradition of building minimalistic models of life and cognition (Beer, 1996; Slocum, *et al.*, 2000; Barandiaran and Moreno, 2006). It is a truism today that complex behavior can be a result of fairly simple interactions by simple structures (in the 1950s it was, of course, much more striking.) The tortoises also display minimally *adaptive* behavior in that they are able to return to recharge their batteries without explicitly computing their goals as based on their preferences. If they are rational at all, their rationality is of the bounded kind. Building minimal models of the adaptive brain was the overall goal of Walter and other cyberneticians at that time (Pickering, 2009, p. 7).

However, Walter's approach to modeling is not to be confused with behaviorist black box models. As one of the pioneers of EEG and a neurophysiologist, he was concerned with building biologically plausible models of neurons. Indeed, the robots are driven by electronic, tube based circuits that are (analog) models of neural mechanisms, and they are intended to be more realistic than early leaky cable models (they predate the Hodgkin–Huxley, 1952. model of the neuron). The circuits serve as models of Pavlovian conditioning.

Modern, behavior based robotics is inspired by Walter's models (Brooks, 1991). There is even an unpublished manuscript from 1961 by Walter that describes four reflex behavioral patterns, some of which are more prepotent over others, which is equivalent to modern 'subsumption architecture' in robotics (Boden, 2008, p. 227). However, Walter had no interest, in contrast to Rodney Brooks, in denying that symbolic representations exist. On the contrary, he claimed that brains are devices for recognizing patterns in sensations (Walter, 1953). His theorizing about patterns is vague, but there are two major differences between Walter and Brooks. First, Walter's models do not represent existing animals but are designed as models of neural mechanisms, and not of *behaviors*. Again, Walter's concern was how to open the black box and to understand the brain – in terms of the neuronal activity detected by EEG. Note

that Walter designed a Toposcope, or a device to localize brain waves in different areas. Brooks, in contradistinction to Walter, is interested in modeling behavior via sensing and interaction. His early models do not contain any realistic or plausible models of nerve activity. Second, Walter was concerned with learning and memory. Although he did not use the vocabulary of information theory to describe his models, he did not subscribe to the view that intelligence was to be modeled without representation.

Physical Symbols and the Environment

In the 1980s, many researchers started to stress that cognition is situated, or embedded deeply, within the environment (for recent reviews, see Robbins and Aydede, 2009; Walter, 2013). Vera and Simon (1993) respond to their claims, and argue that the contrast between information-processing models and situated cognition is misconceived.

Already in 1969, Simon proposed replacing the word *ant* with the word *human*, and stressed that adaptive systems do reflect their inner and outer environments. For this reason, one may not neglect the environment when modeling cognitive processes as long as cognition is adaptive. There are, so claims Simon, only a few 'intrinsic' characteristics of the inner environment 'that limit the adaptation of thought to the shape of the problem environment' (Simon 1996, p. 54). To argue for this point, he uses his classical studies on cryptarithmetic, which involves puzzles like this: SEND + MORE = MONEY.

The task is to find a mapping from letter to digits so that the result (MONEY) is a sum of numbers encoded by SEND and MORE. Simon and Newell studied similar tasks in depth and built detailed computer simulations, able to predict the performance of individual subjects (Newell and Simon, 1972). Two lessons are drawn from these experiments by Simon: first, human subjects do not always discover effective (and learnable!) strategies to deal with the problems posed to them; second, human beings have insufficient short-term memory to apply an efficient strategy. In other words, he thinks that the basic constraint of human adaptivity to the problem environment is the seven, plus or minus two 'magic' rule discovered by George Miller (Miller, 1956): people are able to deal with around seven meaningful chunks at one time.[1] He also considers other limitations on memory, to be modeled in his information-processing architecture.

In short, Simon uses a notion of an *inner* environment to talk of the constraints on adaptivity imposed by the cognitive architecture,

which limits the kinds of information-processing that are possible for an organism. Instead of insisting that there are, for example, innate constraints on language syntax that make learning possible, he suggests that learnability depends on general learning constraints of the information-processing architecture. It is in this context that we need to consider the claim that being a physical symbol system is both necessary and sufficient for cognition (Newell, 1980; Newell and Simon, 1976). Namely, symbol-processing architecture is a *physical* constraint on adaptivity, and remains invariant over various kinds of environments that a system finds itself in. For this reason, the architecture remains explanatorily essential for behavior since it explains why behavior is not optimal; it is adaptive only as far its architecture allows. Problem solving in an environment requires searching over a space of possible solutions, and that space is always constrained by the architecture.

There is, therefore, a sense in which Simon's modeling has always been situated, and his impatience with proponents of situated cognition can be easily understood. But, to be exact, he modeled explicitly only the structure of the *inner* environment, and failed to mention that cognitive physical symbol systems need to be adaptive. I will return to these points.

Not only do Vera and Simon criticize the misinterpretation of the information-processing models of cognition, but they also try to pin down the notion of *situated action* (henceforth, SA) approaches to the study of human–human and human–machine interaction. They distinguish two forms of SA: (1) the hard one, which 'is a methodology for investigating human–human and human–machine interaction, always within the full context in which they occur' (Vera and Simon, 1993, p. 11); and (2) the soft one, which 'builds AI systems that incorporate the SA principles of representing objects functionally and interacting with the environment in a direct and unmediated way' (ibid.). They stress that the hard form is wrongheaded as long as it regards the symbolic approach as antithetical to the study of human interaction in the full context; but Vera and Simon consider it complementary. In contrast to some proponents of SA, they think that internal representations are important in information-processing models of behavior but not crucial. The traditional symbolic view:

> must not be interpreted as suggesting that internal representations should be the central focus of investigation in understanding the relation between behavior and cognition. On the contrary, information-processing theories fundamentally and necessarily

> involve the architecture's relation to the environment. The symbolic approach does not focus narrowly on what is in the head without concern for the relation between the intelligent system and its surround. (Vera and Simon, 1993, p. 12)

So how are symbols to be understood? The term *symbol* has become ambiguous in cognitive and computer science over the years. In computer science, the word functions in at least two different meanings. In the first meaning, a symbol is simply a token from an alphabet, or a piece of information that is processed by a computer. For example, it is customary to talk of symbols on the tape of a Turing machine (Cutland, 1980, pp. 53–4). These symbols are purely formal. They need not possess any content nor refer to anything. All computer systems contain symbols in this sense, including connectionist networks, nonconventional computers, and analog computers. This is because one can always identify physical processes that can be in at least two states (the number of states defines the length of the alphabet) in such computers, and this is all that is required to have symbols. For this reason, Vera and Simon (1993) are right to say that connectionist networks have symbols *in this sense*.

A second meaning of symbol draws on LISP, a programming language frequently employed in classical AI, in which a symbol is a pointer to a list structure (Steels, 2008, p. 228). The structure contains the symbol's name, temporarily assigned value, a definition of function associated with this symbol, and the like. One of the manifestos of the physical symbol view, written by a long-time collaborator of Herbert Simon, Allen Newell (1980), often mentions LISP, and Newell's insistence on access and assign operations as essential to symbols definitely gives away his reliance on LISPish ways of thinking.

It is this sense that makes the physical symbol systems hypothesis plausible, and Newell and Simon seem to use the notion this way:

> What makes symbols symbolic is their ability to designate – i.e., to have a referent. This means that an information process can take a symbol (more precisely, a symbol token) as input and gain access to the referenced object to affect it or be affected by it in some way – to read it, modify it, build a new structure with it, and so on . . . In discussing linguistic matters one normally takes as prototypic of designation the relation between a proper name and the object named – e.g., George Washington and a particular man who was once President of the United States . . . In our theory of

> information processing systems, the prototypic designatory relationship is between a symbol and a symbol structure. Thus, X2 is the name of (i.e., designates) the list (A, B, C); so that given a symbol token X2 one can obtain access to the list – for example, obtain its first element, which is A. (Newell and Simon, 1972, p. 74)

But this definition does not license Newell and Simon to say that symbols designate anything *outside* the cognitive system. The inner environment seems to be the only thing that symbols can refer to. Yet both Newell and Simon do think that symbols are useful as long as they can refer to the embedding environment. In Newell's mind, this can be captured *entirely* by the model of the cognitive computation:

> The system behaves in interaction with the environment, but this is accounted for entirely by the operation of the **input** and **behave** operators. The operation of these two operators depends on the total environment in which the system is embedded. (Newell 1980, p. 117)

Vera and Simon define the notion in yet another way:

> We call patterns symbols when they can designate or denote. An information system can take a symbol token as input and use it to gain access to a referenced object in order to affect it or be affected by it in some way. Symbols may designate other symbols, but they may also designate patterns of sensory stimuli, and they may designate motor actions. Thus, the receipt of certain patterns of sensory stimulation may cause the creation in memory of the symbol (say, CAT) that designates a cat (not the word 'cat,' but the animal). (Vera and Simon, 1993, p. 9)

This definition goes beyond LISP but the notion of designation is not explicated further, though it is still ambiguous. How can CAT designate a cat, while still designating patterns of sensory stimuli? However, setting this ambiguity aside, the notion seems to be so general as to be synonymous with the notion of representation. In contrast to how Fodor (1975), Pylyshyn (1981) or (1984) would use the term *symbol*, Vera and Simon's symbols can be either analog or geometric representations. So Vera and Simon's position is that inner representations can play an important role in explaining behavior, and that the structure of the environment does not explain *everything* about behavior. This is not an extreme claim at all.

To summarize, it is unclear what aspect of the *inner* environment is covered by the notion of the symbol, as the notion was elucidated by Simon in various ways. At the same time, the physical symbol system's inner environment is a constraint on possible adaptivity, hence it remains relevant to explanations of behavior in any environment.

The Role of Environment in Symbolic Models

So where is the environment in Simon's models? His classical models, of chess or cryptarithmetic, do not contain anything related to the environment. Vera and Simon (1993) cite symbolic – or rather *representational* – models implemented on robots such as Navlab (Thorpe, 1990). Navlab works in the physical environment, in contrast to pure computational models, and has a rich, representational structure.

In their early models, Simon and Newell could not model sensorimotor interactions, mainly because of the modest empirical evidence both in neurobiology and psychology, and because of the nature of sensorimotor processes since the underlying physiological system is highly parallel (Newell and Simon 1972, p. 796). Their models were incomplete insofar as they just assumed that external visual stimuli played a major role but they did not include processes for scanning and recognizing components visible on displays (ibid., p. 800). The task environment had a crucial role, however, in constraining the problem space for the subject (ibid., p. 790). But there can be no guarantee that all the relevant information in the task environment is reflected in the problem space. However, the studies performed by Newell and Simon do not shed much light on the mechanisms determining the particular problem space used by a problem solver (ibid.).

They stress that only a few characteristics of *inner* environment are invariant over task and subject, and that adaptive devices shape themselves to the environment in which they are embedded. However, why they shape themselves the way they do remains outside the scope of their explanations and the role of the task environment remains largely implicit.

In contrast, the robotic model can be readily understood, not only in terms of its inner environment but also in terms of its interactions with the environment, while robot builders have a relatively good understanding of the role of sensory processing. Simon does not consider robotic models, such as Navlab, electromechanical tortoises, and Brooks's creatures, as antithetical to his approach. These models complement his work on the inner environment.

Conclusion

It is all too easy to classify classical models of problem solving under the umbrella of symbolic cognitivism. Just because such cognitivism focuses on inner cognitive structures, it may seem that Herbert A. Simon and Allen Newell are simply following Fodor's principles of methodological solipsism (Fodor, 1980). But appearances are deceptive. Simon's assumptions contradict methodological solipsism; without the constraints of the task environment, the structure of the problem space is undetermined, and the problem solving cannot proceed. Bounded rationality is enabled and constrained by the environment. Simon is also aware that such environmental constraints are related to sensory and bodily processes, and he acknowledges, for example, Gibson's work on sensory systems (Newell and Simon, 1972, p. 799). But when defending the physical symbol hypothesis, he fails to say that adaptivity is at the heart of cognition; that shaping the symbol to the environment is crucial, even if he stresses this in other contexts.

By framing the process of constraining the cognitive processes by the environment in adaptive terms, which were also at the heart of the idea of bounded rationality, Simon was able to abstract away from the details, which were costly or impossible to acquire, but he embraced such work later on. For this reason, it is no wonder that he considered robotic models complementary to his computational modeling. They are just two sides of the same information-processing coin. Both need to be included in complete, biologically and psychologically realistic models of adaptive behavior.[2]

Notes

1. Simon also stresses that symbolic-level processing is serial, and not parallel. I will set this claim aside as it is not directly relevant to the issue at hand.
2. The author wishes to thank Konrad Talmont-Kaminski and Witold Wachowski for comments on the draft version of this chapter. The work was financed by NCN research grant in SONATA BIS 5 program, under the decision DEC-2014/2014/14/E/HS1/00803.

References

Barandiaran, X. E. and Moreno, A. (2006). On What Makes Certain Dynamical Systems Cognitive: A Minimally Cognitive Organization Program. *Adaptive Behavior* 14 (2) (June 1): 171–185. doi:10.1177/105971230601400208. http://adb.sagepub.com/cgi/doi/10.1177/105971230601400208.

Beer, R. D. (1996). Toward the Evolution of Dynamical Neural Networks for Minimally Cognitive Behavior. In *From Animals to Animats 4: 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, ed. P. Maes, M. Mataric, J. A. Meyer, J. Pollack, and S. Wilson, 4:421–429. Cambridge, Mass.: MIT Press. http://books.google.com/books?hl=en&lr=&id=V3pksEEKxUkC&oi=fnd&pg=PA421&dq=Toward+the+Evolution+of+Dynamical+Neural+Networks+for+Minimally+Cognitive+Behavior&ots=sGkVIusS5r&sig=_nQCFxRTKzzB9pak3AiVga0LDck.

Boden, M. A. (2008). *Mind as Machine: A History of Cognitive Science* (1st edn). Oxford: Oxford University Press.

Brooks, R. A. (1991). Intelligence without Representation. *Artificial Intelligence* 47 (October): 139–59.

Clark, A. and Chalmers, D. J. (1998). The Extended Mind. *Analysis* 58 (1): 7–19.

Cutland, N. (1980). *Computability: an Introduction to Recursive Function Theory*. Cambridge [Eng.]; New York: Cambridge University Press.

Fodor, J. A. (1975). *The Language of Thought* (1st edn). New York: Thomas Y. Crowell Company.

——. (1980). Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology. *Behavioral and Brain Sciences* III (1): 63–72.

Hodgkin, A. L., and Huxley, A. F. (1952). A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve. *Bulletin of Mathematical Biology* (117): 25–71. doi:10.1016/S0092-8240(05)80004-7.

Jones, B. D. (1999). Bounded Rationality. *Annual Review of Political Science* 2: 297–321. doi:10.1146/annurev.polisci.2.1.297.

Miller, G. A. (1956). The Magical Number Seven, plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review* 63 (2) (April): 81–97. doi:10.1037/h0043158. http://www.ncbi.nlm.nih.gov/pubmed/8022966.

Newell, A. (1980). Physical Symbol Systems. *Cognitive Science: A Multidisciplinary Journal* 4 (2): 135–183. doi:10.1207/s15516709cog0402_2. http://www.informaworld.com/openurl?genre=article&doi=10.1207/s15516709cog0402_2&magic=crossref||D404A21C5BB053405B1A640AFFD44AE3.

Newell, A. and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

——. (1976). Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM* 19 (3) (March): 113–126. doi:10.1145/360018.360022. http://portal.acm.org/citation.cfm?doid=360018.360022.

Pickering, A. (2009). *The Cybernetic Brain: Sketches of Another Future*. Chicago; London: Univ. of Chicago Press.

Pylyshyn, Z. W. (1981). The Imagery Debate: Analogue Media versus Tacit Knowledge. *Psychological Review* 88 (1): 16–45. doi:10.1037//0033-295X.88.1.16. http://content.apa.org/journals/rev/88/1/16.

——. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, Mass.: MIT Press. http://psycnet.apa.org/psycinfo/1986-97211-000.

Robbins, P. and Aydede, M. (2009). *The Cambridge Handbook of Situated Cognition*. New York: Cambridge University Press. http://www.neuromech.northwestern.edu/publications/sitcog_ToC.pdf.

Simon, H. A. (1956). Rational Choice and the Structure of the Environment. *Psychological Review* 63 (2) (March): 129–38. http://www.ncbi.nlm.nih.gov/pubmed/13310708.

——. (1996). *The Sciences of the Artificial*. Cambridge, USA, MA: MIT Press. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+sciences+of+the+artificial#0.

Slocum, A.C., Downey, D.C. and Beer, R. D. (2000). Further Experiments in the Evolution of Minimally Cognitive Behavior: From Perceiving Affordances to Selective Attention. In *From Animals to Animats*, 6:430–439. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.43.522&rep=rep1&type=pdf.

Steels, L. (2008). The Symbol Grounding Problem Has Been Solved, so What's Next? In *Symbols and Embodiment: Debates on Meaning and Cognition*, ed. Manuel de Vega, Arthur M. Glenberg, and Arthur C. Graesser, 223–44. Oxford: Oxford University Press. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.78.3463&rep=rep1&type=pdf.

Thorpe, C. (1990). *Vision and Navigation: The Carnegie Mellon Navlab*. Boston: Kluwer Academic Publishers.

Vera, A. H. and Simon, H. A. (1993). Situated Action: A Symbolic Interpretation. *Cognitive Science* 17: 7–48.

Walter, S. (2013). Situated Cognition: A Field Guide to Some Open Conceptual and Ontological Issues. *Review of Philosophy and Psychology* (November 6). doi:10.1007/s13164-013-0167-y. http://link.springer.com/10.1007/s13164-013-0167-y.

Walter, W. G. (1950a). An Imitation of Life. *Scientific American*. doi:10.1038/scientificamerican0550-42.

——. (1950b). AN ELECTRO-MECHANICAL »ANIMAL«. *Dialectica* 4 (3) (September): 206–213. doi:10.1111/j.1746-8361.1950.tb01020.x. http://doi.wiley.com/10.1111/j.1746-8361.1950.tb01020.x.

——. (1953). *The Living Brain*. New York: Norton.

Webb, B. (2002). Can Robots Make Good Models of Biological Behaviour? *Behavioral and Brain Sciences* 24 (6) (December 17): 1033–50; discussion 1050–94. doi:10.1017/S0140525X01000127. http://www.ncbi.nlm.nih.gov/pubmed/12412325.

14

Bounded Rationality, Shared Experiences, and Social Relationships in Herbert A. Simon's Perspective

Stefano Fiori

Introduction

In his autobiography, Herbert Simon writes: 'The most important years of my life as a scientist were 1955 and 1956' (1991a [1996], p. 189). In those years he published two important articles that laid the foundations of his theory of bounded rationality, 'A Behavioral Model of Rational Choice' (1955) and 'Rational Choice and the Structure of the Environment' (1956). One year later, in 1957, Simon wrote a short story, 'The Apple', in which he presented in literary form the scientific results of his 1956 article.

The thesis of this chapter is that 'The Apple' gives interesting insights into Simon's research and that, because of its literary form, it highlights topics which are less apparent in his scientific papers. Hugo, the protagonist of 'The Apple', lives in isolation in a castle, and his story represents how a rationally bounded agent interacts with, and learns from, an environment by choosing alternatives that are not optimal but *satisficing*. The perspective which inspired both Simon (1956) and 'The Apple' would remain essentially unchanged over time, even when Simon examined bounded rationality in light of artificial intelligence and of simulations performed by means of computer programs.

The question from which we begin concerns the implications of an analysis of (bounded) rationality which removes human relationships, as occurs in Simon's scientific paper of 1956 and his short story of 1957. In fact, Hugo reminds us of *homo œconomicus* in the neoclassical

approach, that is, an individual who, given environmental constraints, is concerned solely with her/his needs. The difference with Simon's view is that in the neoclassical tradition s/he is perfectly rational and maximizes her/his utility, while in Simon's view s/he is rationally limited and chooses not the best but a *satisficing* alternative, i.e., an alternative which meets or exceeds certain criteria different from those required by the maximization of utility function.

However, it would be a mistake to state that Simon does not consider social relationships in his work. On the contrary, he took them into account in his early works on administrative and organizational behavior, and even more so in the 1990s, when themes like loyalty, identification with organizational goals, and altruism became the topic of new inquiries. These themes were not elaborated in light of artificial intelligence; rather, they remained connected to the theory of organizations, or they were influenced by other approaches, such as the Darwinian view which Simon took into account for his hypotheses on altruism. Although the two perspectives (the one that emerged in the 1950s and was later developed within artificial intelligence and cognitive psychology, which analyzes bounded rationality at the individual level; and the other, which deals with bounded rationality within administrative systems and organizations) are basically connected, they partially refer to different theoretical tools and use different languages.[1] Their analysis is the subject matter of this paper: the first section compares Simon's (1956) model of bounded rationality with its literary version; the second section examines how bounded rationality, especially in 'The Apple', is represented by starting from the traditional image of an isolated individual; the third section discusses how the paradigm of the isolated individual raises problems relative to the emergence of meanings; the penultimate section shows that Simon dealt with relationships between the individual and society in his approaches to organizational and administrative behavior and in his studies of the 1990s; the final section concludes.

The Apple and the Formal Model of Bounded Rationality

'*Rational Choice and the Structure of the Environment*' is based on the idea that agents have an 'approximate' rationality characterized by 'limited information and limited computational facilities' (Simon, 1956, p. 261). This implies that they choose not optimal, but *satisficing* alternatives, although their behaviors exhibit the capacity to adapt to the environment. In particular, in order to develop his view of bounded rationality,

Simon supposes that the agent is an 'organism' with 'very simple perceptual and choice mechanisms', which can satisfy its need for food and can 'assure a high probability of its survival over extended periods of time' (ibid.). The problem of rational choice consists of choosing a path that allows the organism not to starve, where survival is a satisfactory goal achieved through non-optimal strategies. The environment is a sort of maze in which each branch point represents a choice point that has to be explored in a selective and sequential way. The organism exhibits learning abilities through which it associates clues with the appearance of food; and when it deals with multiple, but conflicting goals, it relies on a mechanism that defines priorities without resorting to the complex calculus of marginal rates of substitutions among different wants. The choice among available alternatives does not need maximization of an utility function, since it suffices that the organism defines its aspiration levels.

In 'The Apple', the literary transformation of the formal model of choice has interesting implications. The simple organism is now Hugo, an individual who lives in a castle with numerous rooms and without contact with the external world. He has not met any other human being except for his mother, who – now deceased – did not know the outside world either. The castle is a maze, where it is impossible to retrace one's steps, and Hugo has to move from room to room seeking food, which unexpectedly appears on a table. In the beginning, Hugo has elementary preferences and basic aspirations levels, but this condition has to change in consequence of learning processes which gradually lead him to develop new tastes. This also implies that he spends most of his time searching for his preferred food. Murals cover the walls of the rooms, but they do not help him to know the outside world deductively, since they are stylized images which do not represent concrete objects, such as trees. Yet, these paintings help him in another way, since 'The long hours spent in examining them developed in him a considerable capacity for understanding and appreciating abstract relations' (1991a [1996], p. 182). Hugo has only one book, the Bible, that he has learned to read thanks to his mother's teaching. Like the murals, the parables and the creation stories narrated by the Bible have an 'abstract symbolic meaning' and, although Hugo could not know the experiences to which those parables and stories referred, by abstraction he 'could usually translate the stories directly back to the propositions they sought to communicate' (ibid.). The search for food, organized as a problem-solving activity, gradually becomes specialized. Information gathered and processed allows Hugo to formulate the 'theory that rooms

decorated in green were more likely to lead to white bread than other rooms, while the color blue was a significant sign that he was approaching some ripe olives' (1991a [1996], p. 184). Finally, Hugo, who has multiple goals, assigns different values to foods, learns to distribute time among different activities, and refines his search so that he can establish his objectives for the day. His search is aimed at achieving only satisfactory, and not optimal, results. In fact, failures moderate Hugo's ambitions, and 'he was satisfied in attaining more modest goals' (1991a [1996], p. 187), that is, in the language of the formal model, he reduces his aspiration levels.

Hugo and the Paradigm of the Individual

Hugo's story represents the characteristics of bounded rationality. However, precisely because of its literary form it emphasizes aspects which do not appear so clearly in the formal model. Hugo is an isolated individual, like Robinson Crusoe, who inspired the paradigm of the *homo œconomicus* of the neoclassical approach and, according to Marx, influenced classical economics. This suggests that bounded rationality, like the rationality of the neoclassical view, can, in principle, be analyzed without reference to social relationships. It can be studied by observing the procedures through which individual tastes and preferences are discovered and elaborated thanks to learning processes, and by focusing on how problem-solving activities develop within an environment, which constitutes a constraint on action. Like Robinson, Hugo did not live with absolutely no contact with civilization. While Defoe's Robinson began his adventure by gathering work tools, weapons, clothes, and gunpowder from the ship wreck (Defoe, [1719] 1965, pp. 69–73), Hugo lived in a world that someone had organized and furnished with chairs, tables, sofas, tablecloths, and murals. Unlike Robinson, someone prepared food for Hugo, but like Robinson he had to discover where the food was, planning courses of action that increasingly enabled him to achieve satisfactory results. Finally, Hugo like Robinson possessed a Bible.

As is well known, Marx maintained that the *Robinsonades* of classical economics and the paradigm of the isolated individual (as a subject analyzable in the absence of social relations and in connection only with her/his needs) were a mystification. The isolated individual is a deceit that conceals the social character of economic relations precisely because it reduces economic behavior to a relation between man and environment, tractable like a technical datum, in which only an instrumental rationality is required in order to render the means–end relation

effective (Marx, [1867] 1977; [1857–8] 1993). Contrary to this view, the story of Robinson Crusoe does not describe a simple relation between man and nature. The commodities, by means of which needs are satisfied, are not mere objects since they embody the social relationships that produced them. Therefore, Robinson begins his story not with simple tools designed to dominate the environment but with commodities that incorporate social relations of the civilized world (Hymer, 1971; Karaöz, 2014).

Evidently, all this is distant from Simon, but it helps to raise questions about the relationships between Hugo and the outside world. In fact, Hugo (or any other simple organism), as an isolated individual, evokes some aspects of the instrumental rationality of the neoclassical *homo œconomicus*, that is, the paradigmatic figure from which Simon intends to distance his model of man.[2]

Simon's model of bounded rationality differs from the neoclassical one, not with respect to the concept of the environment, which in both views constitutes an objective constraint, but because it cannot be dealt with by an Olympian rationality. Rationality, in Simon's terms, exploits the external world to a limited extent, with the consequence that neither maximization of utility nor perfect information is required. Moreover, in time Simon considered bounded rationality to be a characteristic of all symbolic systems which exhibit intelligent behavior, especially human beings and computers (Frantz, 2003, pp. 271–2), which can be dealt with as classes of symbolic structures irrespective of their biological or physical properties. This therefore makes it possible to study rationality 'independently of the details of the biological mechanism' (Simon, 1964, p. 77).[3] In short, while in the mid-1950s bounded rationality was considered a property of any organism which possesses 'very simple perceptual and choice mechanisms' (Simon, 1956, p. 261), it subsequently became a more general condition characterizing any entity able to behave intelligently. Yet, the fact that in 'The Apple' the agent is a man, and not a generic organism, has important implications.

Emergence of Meanings and Experience

Hugo was a very special human being because he had learned to read but had no experience of a human world; neither of human relationships described in the Bible nor of the symbolic meanings attached to those relationships. Moreover, he was unable to relate murals to real objects. As a consequence, 'a large part of the "world outside" [the Bible] talked about was almost meaningless to him'(1991a [1996], p. 182).

Hugo learned the word tree from the Bible but did know the object to which it referred.[4] Therefore, it is not clear how he could relate this token to a real object (the tree, in the physical sense) of which he knew nothing.

The problem can be illustrated by referring to the philosopher of science Hilary Putnam. Suppose that some human beings live on Mars, and that they have never seen or imagined trees. Moreover,

> Suppose one day a picture of tree is accidentally dropped on their planet by a spaceship which passes on without having other contact with them ... All sorts of speculations occur to them: a building, a canopy, even an animal of some kind. But suppose they never come close to the truth.
>
> For *us* the picture is a representation of a tree. For these humans the picture represents only a strange object, nature and function unknown. Suppose one of them has a mental image which is exactly like one of my mental images of a tree as a result of having seen the picture. His mental image is not a *representation of a tree*. It is only a representation of the strange object (whatever it is) that the mysterious picture represents. (Putnam, 1981, pp. 3–4; italics in original)

Hugo did not even have the images that humans who live on Mars possess, since the murals of the castle did not represent concrete objects like trees. However, he knew the word tree, which he attached to the taste of apple, of which he did have experience. In Putnam's reasoning, this knowledge should not be very meaningful, since the word tree uttered on Mars (and in Hugo's castle) cannot refer to the same thing that we refer to, since the word is not correlated with its real object. It is impossible to fix the reference in an isolated world like Hugo's, since something fundamental is lacking. According to Putnam, determination of reference is social and not individual. We perceive, handle, deal with objects of the real world, and 'Our talk of [these objects] is intimately connected with our *non-verbal* transactions' with them (Putnam, 1981, p. 11; italics in original).

Simon seems to be of a different opinion. In fact, he maintains that Hugo elaborated 'abstract relations' by means of the murals, despite the lack of concrete objects with which to associate them. Therefore, 'it must be supposed that he read the creation myths and the parables of the Bible in much the same way – the concrete objects taking on for him an abstract symbolic meaning' (1991a [1996], p. 182). Yet, although

Hugo could know an abstract meaning in the absence of a real referent, understanding biblical mythology is a complex matter, since the referent is not a simple object, like a tree, but a network of social and cultural relations to which the parables refer, and from whose knowledge the understanding of meaning should emerge. Hugo did not have this knowledge, but this did not impede him from learning meanings with a peculiar abstract procedure:

> his way of understanding the Bible was just the reverse of the way in which it was written. The authors of these stories had found in them a means for conveying to humble people in terms of their daily experiences profound truths about the meaning of the world. Hugo deprived of these experiences, but experienced in abstraction, could usually translate the stories directly back to the propositions they sought to communicate. (Ibid.)

This model of rationality implies a purely abstract (that is, not based on experience), but correct, understanding of the meanings both of concrete terms, like tree, and of more complex utterances, which convey biblical teachings and concern social and cultural facts. Experience can be replaced and offset by information dealt with by appropriate procedures of abstraction. This perspective was subsequently developed in Simon's contributions on artificial intelligence, and it would imply definition of logical procedures based on symbolic processing. These procedures require elements like learning, memory, rules of thumb, and sequential problem-solving activities (used for reconstruction of unknown meanings), which lead to the achievement of satisfactory results. In terms of artificial intelligence, the problem consists of processing and manipulating information that regards the environment and its constraints. If the task environment is circumscribed to processing information contained in a book (the Bible), whose symbols are known, and if there is an internal representation of the environment which allows activation of problem-solving processes, then the interpretation of the text is achievable without referring to experience. In this way, (bounded) rationality defines its own procedure by using information, exploring the world, and resolving problems raised by the environment. What it is important to point out is that bounded rationality works in the same way both when an agent explores the physical world of which he has experience, and when he explores a problem (as in the case of interpreting the Bible's parables). For Simon, human beings and computers are rationally limited processors of information,[5]

but humans relate information to experiences, while computers do not, although they can process information which comes from the programmer's experience. In Putnam's perspective, this can be a problem because the information *tree* does not have the same meaning for a man, a Martian, or a computer, since only a human being has experience of trees. And, even more complex is conceiving a cultural universe, like the one described by the Bible, without experience of the human and social affairs that engendered it. In this respect, the term experience regards not a historical event (no one can experience a remote past like the one to which the Bible refers), but more basic elements of human relationships, which render phenomena, such as beliefs, sense of community, religion, emotions, and so on, familiar.

However, it would be erroneous to maintain that experience does not play a role in Simon's story. Hugo did not understand why Eve had sacrificed her life in the Garden of Eden because of an apple. He knew apples, and precisely this experience prevented him from understanding Eve's behavior. Despite the conflict between abstract reasoning and experience of apples,[6] 'He did, in time, learn the answer, but experience and not abstraction led him to it' (1991a [1996], p. 182). By rereading the Bible passage where it is said that 'the tree was good for food, and that it was pleasant to the eyes', the meaning finally became perfectly clear, and the experience of apple ceased to interfere with his powers of abstraction: 'The meaning, he knew now, lay not in the apple, but in him' (1991a [1996], p. 188). Although – Simon says – we do not know exactly what this meaning was, he conjectures that 'Hugo found a meaning not very different from the one I have arrived at, journeying through the maze of my own life' (ibid.). This meaning seems to depend on subjective paths of discovery, rather than on the real properties of an object (an apple).

Experience plays a fundamental, but ambiguous, role. Simon maintains that we can understand what happened to Hugo not by means of abstraction, but by 'empathizing with the trials of journey, *interpreting them in the light of our own experiences*' (ibid., italics added). Human experience constitutes the thread that connects Hugo, Simon, and all human beings, and it makes interpretation of meanings possible. Experience, and not abstract reasoning, enables Simon to identify Hugo's behavior, and enables Hugo to discover the answer to Eve's conduct. We can relate our experience to Hugo's and discover meanings on the basis of familiarity with certain events. But how can Hugo identify others' behavior if he lacks knowledge of human relations like those recounted by the Bible? And what are the consequences of reducing (bounded) rationality to

the manipulation and processing of physical symbols, which humans and computer have in common regardless of processes which require empathy?

Although Simon apparently does not focus on these questions, he took shared experiences into account, especially when he considered loyalty, identification with organizational goals, and altruism. These concepts appeared in early studies on administrative and organizational behavior, and they were developed in the last part of his scientific career. Although they are closely connected to bounded rationality, they are dealt with using tools which do not refer to symbolic processing and seem to cover areas not easily tractable by means of artificial intelligence.

Identification, Loyalty, Altruism, and Other Human Traits

What Simon had in mind when he wrote 'The Apple' seems to be clarified in the premise to Part IV of *Models of Man*, which was published in 1957, the same year in which he wrote the story of Hugo:

> It can be said with equal truth that a theory of rational choice can hardly exist without a theory of organization. Robinson Crusoe, it may be argued, proves the contrary. But an understanding of Robinson Crusoe, however important as a first step, is only a preliminary to an understanding of modern, urbanized man. The characteristic environment of man is constituted not of nature but of his fellows. His rational decision making . . . takes place in social groups including organizations. (Simon, 1957, p. 196)

The Robinson Crusoe fiction is not rejected. It is considered a preliminary, but insufficient, step towards understanding complex societies, where the environment is not nature but social relationships. This perspective can be applied to 'Rational Choice and the Structure of the Environment' and its literary explanation, 'The Apple', since in these writings bounded rationality explores only the physical environment. The passage from the physical to the social environment, however, does not imply a change in how bounded rationality deals with the external world, which in both circumstances is assumed to be a set of constraints that have to be internally represented in order to devise satisfactory strategies. Moreover, in the subsequent pages of *Models of Man* Simon's criticism is directed at the omniscience of 'economic man', to which he opposes the notion of bounded rationality, rather than at the fact

that this paradigm does not consider sociality. In short, Simon emphasizes limits of individual rationality rather than the way in which social relationships influence decision making. In this framework, organizations are evoked, since, given the human limits of knowledge, foresight, and skill, they 'are useful instruments for the achievement of human purposes' (Simon, 1957, p. 199).

It is not in *Models of Man* that Simon deals with relations between individual and society, but in his studies on administrative systems and organizations. In *Administrative Behavior* (1955 [1947]) there appeared themes, like loyalty and identification with the goals of organization, which he would develop in the mature phase of his inquiry. He distinguished personal and organizational decisions and maintained that 'a person identifies himself with a group when, in making decisions, he evaluates the several alternatives of choice in terms of their consequences for the specified group' (1955 [1947], p. 205). Identification is the process by which organizational objectives replace individual objectives, although at that time he considered the psychological bases of identification to be 'obscure' (1955 [1947], p. 218). He also pointed out that when individuals operate in organizations, they behave on the basis of value premises which have a conventional character and refer to social relationships. Therefore, changing the value system implies changes in the interpretation of right or wrong actions (Koumakhov, 2014). Identification of the individual with groups (not only organizations, but also family groups, community groups, professional associations, and subgroups) is also dealt with in *Organizations* (March and Simon, [1958] 1993 pp. 76–101). In this book March and Simon maintain that definitions of the situation 'involve a complex interweaving of affective and cognitive processes' (p. 172), where the 'motivational and affective factors' are about the relation between individual and organizational goals, while cognition covers the limits of the actors' rationality. In other words, human and social relationships within groups and organizations and cognitive limits are complementary conditions to understanding agent choices and behaviors.

These topics would be reconsidered in the 1990s with Simon's analysis of docility and altruism: 'To be docile is to be tractable, manageable, and above all, teachable. Docile people tend to adapt their behavior to norms and pressures of the society' (Simon, 1991b, p. 35), and docility's contribution to human fitness is immediately evident when we consider the characteristics of non-docility, such as intractability, unmanageability, unteachability, and incorrigibility. Docility, which means responsiveness to social influences, induces altruism, and 'is used to

inculcate individuals with organizational pride and loyalty' (p. 36). This engenders identification with 'us' (that is, groups like the family, city, nation, and so on), and separation from 'them' (ibid.). Finally, docility exhibits a cognitive and a motivational component. One implies 'belief legitimated by social processes' and not by empirical self-evaluation, and the other implies 'acceptance on the basis of social legitimation and approval rather than acceptance on the basis of individual drives and motives that were not socially acquired' (Simon, 1992, p. 75). All this leads to a notion of altruism compatible with Darwinism (Simon, 1990b, 1992), which is manifest in loyalty to (and identification with) groups to which individuals belong.

Identification with group goals implies that organizations or other groups are able to convey shared images of reality because they share values and ethical premises. According to Koumakhov (2014, p. 266), this engenders a 'mental representation of reality, which thus becomes a social cognitive construction, a sort of common belief system coordinating individual perceptions'.

Conclusion

According to James March, Simon's co-author, myths, symbols, rituals, and stories are the tools with which meaning is constructed: 'Meaning comes from social interaction and takes both its coherence and its contradictions from its social basis. Interpretations are shared through communication, and their character is transformed through the social process by which they are shared. Social exchange leads a community, group, or organization toward internally shared understandings of experiences' (March, 1994, p. 210). Symbolic meanings pervade decisional processes, and this prompts 'a view that moves meanings to the center of the analysis, rather than one that sees meaning as instrumental to action' (p. 218). Interpretation of the social environment depends on how meanings emerge from myths, symbols, rituals, and stories, and it constitutes a condition to explain decisional processes. March's perspective focuses on problem setting, rather than on problem-solving, since in a sense problem setting precedes problem-solving. Problems are not given, but are constructed by means of interpretation; and this determines how problem-solving comes about.

Simon, as is well known, devoted a great deal of effort to the definition of problem-solving. His models of bounded rationality of the 1950s and subsequent decades, and the literary representation narrated in *The Apple*, reflect the problem-solving perspective. Hugo is

a problem solver who has a limited, instrumental rationality with which to solve problems posed by the environment. The notion of bounded rationality, in a sense, considers problem setting, since, to use Simon's language of the 1970s and 1980s, the subject represents and interprets the environment internally (internal representation), given a problem space in which his problem-solving activities take place. But it does not specify how social mechanisms determine the internal representations. This perspective was somehow taken into account when Simon analyzed organizations and administrative system; that is, when he examined the framework in which an important class of social relationships takes place and involves phenomena like identification, loyalty, adoption of shared value premises, and altruism. The two lines of inquiry – which examine bounded rationality at an individual level and within organizations – are inseparable, and in Simon's work the one is continuously reinterpreted in terms of the other. Nonetheless the reader perceives that they deploy different tools and use different languages. Certainly, both reflect Simon's intention to develop a rigorous empirical science opposed to the *a priori* assumptions of neoclassical economics. Significant in this regard is his warning to economists who use econometrics: 'They will have to venture out into the world itself, like anthropologists who learn the language of natives, speak with them, and observe them' (Simon, 1992, p. 81). Not an isolated individual, but society must be examined, with its relations, languages, and beliefs which confer meaning on the world that we know.

Notes

1. Augier (2000), and Augier and March (2002) point out a 'considerable continuity' in Simon's writings, and highlight his constant effort to clarify the real processes of human decision making; an endeavor which must be viewed in terms of a gradual transformation of his theory. This interpretation is in polemic with Sent's and Mirowski's views (Sent, 2000; Mirowski, 2002), according to which the Cold War, the climate generated by cyborg science, and Simon's experience at RAND Corporation in the mid-1950s exerted a strong influence which culminated in a 'seachange in Simon's interest'. More precisely, Sent writes, this radical change connoted his shift of focus from the analysis of human decision making in organizations and public administration to problem solving analysis (Sent, 2000).
2. Winograd and Flores (1986, p. 22) maintain that Simon does not contest the 'rationalistic tradition', but only the version that implies perfect knowledge, perfect foresight, and optimitization criteria (see also Gardner, 1987, p. 361; Crowther-Heyck, 2005, p. 60).

3. 'Like a modern digital computer's, Man's equipment for thinking is basically serial in organization . . . there is much reason to think that the basic repertoire of processes in the two systems [human thought and computers] is quite similar. Man and computer can both recognize symbols (patterns), store symbols, copy symbols, compare symbols for identity, and output symbols. These processes seem to be the fundamental components of thinking as they are of computation' (Simon, 1976, p. 430). From this point of view, human and computer rationality are bounded because they exhibit limits as regards both computation and the manipulation of symbols (Simon, 1990a, pp. 8, 17; see Newell and Simon, 1972, pp. 54-55).
4. 'You might suppose that the murals on the castle's walls would have helped him to understand this world outside, and to learn the meaning of such simple words as "tree". But the pictures were of little help – at least in any ordinary way – for the designs the castle's muralists had painted on its walls were entirely abstract, and no object as prosaic as tree – or recognizable as such to an inhabitant of the outside world – ever appeared in them' (1991a [1996], p. 182).
5. In terms of artificial intelligence, the problem-solving procedure – which characterizes man and computer – implies having a *generator* of symbol structures, that is, a move generator for potential solutions, and a *test* which evaluates these solutions. A problem will be solved if the generator produces a structure which satisfies the test (Newell and Simon, 1981, pp. 52–53). This means that we know what we want to do, but we do not know how to accomplish it. Therefore, man and computer are not omniscient, and they exhibit the same limitations as regards computation and rationality.
6. The problem was that 'apples seen or tested impeded the abstraction of his thought' (1991a [1996], p. 188).

References

Augier, M. (2000). Models of Herbert A. Simon. *Perspective on Science*, 8, 407–43.

Augier, M. and March, J. (2002). A model scholar: Herbert A. Simon (1916–2001). *Journal of Economic Behavior & Organization*, 49, (1), 1–17.

Crowther-Heyck, H. (2005). *Herbert A. Simon. The Bounds of Reason in Modern America*. Baltimore: John Hopkins University Press.

Defoe, D. ([1719] 1985). *The Life and Adventures of Robinson Crusoe*. London: Penguin Books.

Frantz, R. (2003). Herbert Simon. Artificial intelligence as a framework for understanding intuition. *Journal of Economic Psychology*, 24, (2), 265–77.

Gardner, H. (1987). *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books.

Hymer, S. (1971). Robinson Crusoe and the Secret of Primitive Accumulation. *Monthly Review*, 23, (4), 11–36, available on http://monthlyreview.org/2011/09/01/robinson-crusoe-and-the-secret-of-primitive-accumulation/; access date: August 23, 2015.

Karaöz, U. (2014). The Neoclassical Robinson: Antecedents and Implications. *History of Economic Ideas*, 22 (2), 75–100.

Koumakhov, R. (2014). Conventionalism, coordination, and mental models: from Poincaré to Simon. *Journal of Economic Methodology*, 21 (3), 251–72.
March, J. G. (1994). *A Primer on Decision Making. How Decisions Happen*. New York: The Free Press.
March, J. G. and Simon, H. A. (1993 [1958]). *Organizations* (2nd edn). Cambridge, MA: Blackwell.
Marx, K. (1977 [1867]). *Capital*, *Volume 1*, introduced by Ernest Mandel. New York: Vintage Books.
Marx, K. (1993 [1857–8]). *Grundrisse. Foundations of the Critique of Political Economy.* London: Penguin Books.
Mirowski, P. (2002). *Machine Dreams. Economics Becomes a Cyborg Science*. Cambridge: Cambridge University Press.
Newell, A. and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs: Prentice-Hall.
Newell, A. and Simon, H. A. (1981). Computer Science as Empirical Inquiry: Symbols and Search. In *Mind and Design. Philosophy, Psychology, Artificial Intelligence*, ed. J. Haugeland. Cambridge, MA: MIT Press.
Putnam, H. (1981). *Reason, Truth and History.* Cambridge: Cambridge University Press.
Sent, E-M. (2000). Herbert A. Simon as a Cyborg Scientist. *Perspective on Science*, 8, 380–406.
Simon, H. A. (1955 [1947]). *Administrative Behavior*. New York: Macmillan.
Simon, H. A. (1956). Rational Choice and the Structure of the Environment. Repr. in Simon (1957), 261–73.
Simon, H. A. (1957). *Models of Man*. New York: Wiley.
Simon, H. A. (1964). Information Processing in Computer and Man. In *Economics, Bounded Rationality and the Cognitive Revolutions*, ed. H. A. Simon, M. Egidi, R. Marris and R. Viale. Aldershot: Edward Elgar, 1992.
Simon, H. A. (1976). From substantive to procedural rationality. In *Models of Bounded Rationality. Behavioral Economics and Business Organization, Volume 2*, H. A. Simon Cambridge, MA: MIT Press, 1982.
Simon, H. A. (1990a). Invariants of Human Behavior. *Annual Review of Psychology*, 41, 1-19.
Simon, H. A. (1990b). A Mechanism for Social Selection and Successful Altruism. *Science*, 250, 1665–8.
Simon, H. A. (1991a [1996]). *Models of My Life*. Cambridge, MA: MIT Press.
Simon, H. A. (1991b). Organizations and Markets. *Journal of Economic Perspectives*, 5 (2), 25–44.
Simon, H. A. (1992). Altruism and Economics. *Eastern Economic Journal*, 18 (1), 73–83.
Winograd, T. and Flores, F. (1986). *Understanding Computers and Cognition. A New Foundation for Design*. Norwood: Ablex.

15
Bounded Rationality in the Digital Age

Peter E. Earl

Introduction

One of the great tragedies in economics in the decades since Simon received the 1978 Alfred Nobel Memorial Prize in Economic Sciences is that the uptake of his ideas within the discipline has been either poor or in a partial manner that does not properly capture his vision (as with mainstream models purporting to address bounded rationality). In this chapter I begin by trying to make sense of this situation and then argue that the digital revolution is making it more imperative than ever that economists take up Simon's key ideas – not merely his satisficing view of choice in the face of bounded rationality but also his thinking on artificial intelligence and the evolutionary roles of altruism and system design. The modern economy is undergoing supply side upheavals at the heart of which lie the issues of programmability and modularity. On the demand side, buyers now have to contend with choice problems of extraordinary complexity, whose solutions increasingly rely on social inputs.

A recurrent theme in what follows is that, in the digital age, Simon's (1991, pp. 306–7) Travel Theorem takes on a wider significance. He set out the theorem with reference to what one can hope to learn about something in a good public library, as opposed to making a journey to study it at firsthand for a short period (for example, as a tourist or business consultant). His contention was that if information is all one hopes to obtain, being there is far less efficient that trying to gather it remotely. Hence, if journeys are actually undertaken, they are/should be for reasons other than the gathering of information. In the world of the Internet, webcams, smartphones, Skype, virtual reality experiences, and

so on, Simon's Travel Theorem provides a powerful starting point for asking questions about motivation and economic organization.

Resistance to Simon's Approach to Economics

Despite its increasing openness to modern behavioral economics, the mainstream of economics has not given Herbert Simon's contributions the attention they deserve. But his ideas are nearly invisible even in modern behavioral contributions, despite the fact that at the time Simon received his Nobel award he was viewed as the father of the behavioral approach. Instead, such contributions focus on using knowledge of heuristics and bias as foundations for building models that present a twisted form of constrained optimization as a means of making sense of behavioral 'anomalies' (see Sent, 2004; Berg and Gigerenzer, 2010). The reference point for judging what is an anomaly remains that of conventional rational choice theory not Simon's (1976) procedural rationality. To the extent that bounded rationality is modeled, it presents finite cognitive capacity as an extra constraint within an optimization process.

Modern behavioral economics is thus based on the approach to choice of 2002 Nobel Laureate Daniel Kahneman, rather than Simon's. Economists have mostly ignored modern work by psychologists in the spirit of Simon, most notably the 'fast and frugal decision rules' approach of Gigerenzer and his colleagues (1999). Kahneman's own role in all this has been, at best, disappointing to Simon scholars, for in his ruminations on the lessons of his career (Kahneman, 2011) he only gives attention to Simon's work on seemingly intuitive choices, which complements his own view of fast thinking, and he buries in the endnotes any mention of the Simon/Gigerenzer view of heuristics as necessary and often far from dysfunctional (see further, Earl, 2012). Behavioral economics did not have to be like this. It is possible, if one is willing, simultaneously to embrace both dysfunctional and fast and frugal heuristics within a general heuristics-based view of choice of the kind that Simon proposed. The human brain is not guaranteed to select efficient heuristics to aid choice and, to the extent that seemingly inefficient ones are part of human nature, their presence ought to signify that they once assisted evolutionary fitness.

Simon's contributions have not fared well even within heterodox economics. While the mainstream found his rejection of constrained optimization too radical, heterodox economists, especially those of a post-Keynesian persuasion, have presented Simon's bounded rationality

approach as not radical enough (see Shackle, 1961, pp. 100–1; Dunn, 2000).

This is a consequence of how Simon tried to enhance his chances of making his satisficing framework acceptable to those wedded to the static optimization approach. He began with the standard view of rational choice as involving the discovery of the best solution from among a given set of possibilities. He then presented bounded rationality as arising because of impediments to discovering members of this set and determining whether any of the discovered options is actually the best means for serving the end in question. He could hardly have fared worse had he mounted a full-on critique of the static, equilibrium-focused aspects of rational choice theory and included the creative aspects of problem-solving in his critique of optimizing models. Mainstream economists could question the significance of bounded rationality via Day's (1967) argument that, by repeated experimentation, firms could eventually stumble upon optimal choices even if they were choosing with simple decision rules. Day's critique of satisficing relies on the possibility set being fixed, as Winter (1971) pointed out, whereas the real world is characterized by Schumpeterian creative destruction in which innovation keeps changing what is possible.

In addition to forestalling Day's argument, Simon would at least have won more admirers from heterodox circles if he had not underplayed creative destruction and the significance of the radical/fundamental uncertainty associated with innovative choices. The Schumpeterian view is widely held by heterodox economists and, from the post-Keynesian standpoint, there is no fundamental objection to Simon's view that choices are typically addressed via rules for search and acceptability. Post-Keynesians therefore should have embraced Simon's ideas enthusiastically. The trouble was that Simon seemed insufficiently mindful that some choices are what Shackle (1961) categorized as 'crucial experiments', with decision makers sometimes fearing that surprising new options could become available after they have made commitments to specific durable assets. With Keynes (1937) having spent a few paragraphs suggesting that people cope with situations in which they 'simply do not know' with the aid of simple decision rules and by following others, it is as if most post-Keynesians seem to have decided that Keynes was ahead of Simon and that they have no need for the latter's full-blown research program on satisficing or his later writing (such as Simon, 1992) on the evolutionary role of docile adherence to social norms. Likewise, within institutional economics, the place of rules and routines is emphasized via earlier work by Veblen (1899), despite the

attention also given to Simon by Hodgson (1988) in his manifesto for modern institutional economics.

Simon's work is unlikely to be grasped enthusiastically by mainstream or modern behavioral economists so long as they maintain their core belief that all choices should be viewed as acts of constrained optimization. But well-informed heterodox economists have no such barrier to taking more notice of Simon's work. It is clear that, despite the way he tried to convey the notion of bounded rationality to the mainstream, he was impressed by Schumpeter's way of looking at the world (see, for example, Simon, 1984). Treating Keynes and Veblen as sources of all wisdom does not seem, to an eclectic economist such as myself, to be a wise heuristic for trying to advance knowledge in economics, convenient though it might be in terms of the range of reading that one ends up having to do. But at least heterodox economists largely reject the mainstream view that economic theory should apply to any institutional and historical context. They may thus be open to the arguments that follow.

The Past and Future of Work and Production Systems

There is a sense in which the nature of work has gone in a circle in the 240 years since Adam Smith (1976 [1776]) wrote *The Wealth of Nations*. In Smith's time, the industrial revolution was starting to transform production away from traditional craft-based processes involving self-paced work using specialized expertise. A weaver working at home certainly had to work on average at a fast enough pace to produce enough cloth to pay for the necessities of life but, by working faster or longer, it was possible to make up for a late start or sickness. Central to protecting capacity to earn a living had been the ability to limit entry, typically facilitated by guilds and long apprenticeships, rather as with the professions of the twenty-first century. Where the worker was performing a complex set of operations to make something largely from scratch, there was scope to think how to deal with variability in the quality of materials being used or in the accuracy with which the worker could make various components that needed to be fitted together; if things did not fit first-time, filing them could overcome the problem (Leijonhufvud, 1986; Langlois and Robertson, 1995).

By contrast, the emerging industrial working class toiled for fixed shifts in factories and the pace of their work was frequently dictated by the speed of the machines that they were operating. Toward the end of the nineteenth century, they were increasingly working with

standardized components that were made accurately enough to fit together without any need for modification. This did not merely simplify and cheapen the process of manufacturing, it also increased the viable lifespan of products by making it possible to replace worn parts with identical ones from off the shelf, while standardized interfaces made it much easier for businesses to profit by designing or purchasing bolt-on upgrades. In an important sense, this modularization and standardization of the elements of complex production systems and products was a manifestation of the evolutionary benefits of partial decomposability that Simon (1962) emphasized. This change in the nature of products and production systems opened up the range of choice in production and consumption by permitting the creative construction of novel combinations of existing components with common interfaces (Earl, 2003), further increasing the evolutionary fitness of the component modules. In the digital age, this parts-bin approach to manufacturing applies with both hardware and software, as evidenced by the success of Apple.

But opportunities for exercising creativity were the preserve of entrepreneurs, specialist designers or tinkering, system-building consumers and were denied to the typical factory worker. Around the time that Simon was born, Henry Ford had begun increasing productivity by coupling the removal of any choice of pace with the division (in other words, modularization) of tasks into highly specific routines. The worker's role was programmed to a very high degree, such that there was very little need to engage in creative problem-solving; in effect, the worker had been turned into a part of the machinery, a change brilliantly satirized by Charlie Chaplin in his 1936 movie *Modern Times*. For growing numbers of workers, the major part of their time at work involved no deliberation at all, merely the executing of if–then instructions. The worker was merely a hired hand, not a person with a brain and a wide repertoire of capacities to be called upon to deal with infrequently asked questions.

In the digital age, many workers could, in principle, be working from home rather than grouped in factories or offices, for the things to which they add value are sent to them digitally via the Internet and they, in turn, produce value-added output that can be delivered digitally. Simon's (1991) Travel Theorem is a useful device for understanding the limits to both a return to working from home and the kind of income levels people may hope to earn as the present century unfolds. Although production processes may involve digitally delivered inputs and outputs, being there rather than telecommuting may, depending on the

kind of work, have impacts on productivity. A key issue here, and for the future availability of work, is the distinction that I have already drawn, and which Simon (1958) was among the first to emphasize, between routine-based work and creative problem-solving.

If workers are paid for their output (as is increasingly common for those who find themselves having to operate as self-employed subcontractors) rather than via a fixed salary, then there is no reason for them to require monitoring in the sense of checking that they are actually working. In such cases, and wherever performance-related bonuses form a major part of remuneration, the only reason for traveling to work with or alongside others is because of the impact of face-to-face interaction on their productivity. As many an academic would attest, having to work with a co-author remotely is not the same as working with the same person as an immediate colleague as far as the spontaneous generation of new ideas is concerned. Using Skype is not quite the same as being there if the job entails being creative, even though Simon's Travel Theorem applies as far as the transmission of *given* information is concerned. Likewise, those who work in financial markets may perform better if confined to dealing rooms (where their focus on making deals is enhanced by seeing others busy doing so) or getting early access to significant information, say, when it slips out over lunch, rather than relying on receiving it via email or text messages that those in the know might not think of (or risk) sending.

With spectacular reductions in the price of robotic technology, the scope for workers to earn a living as fully programmed hired hands is rapidly vanishing, and will soon do so even in newly industrializing countries. In the age of artificial intelligence and robots, the ability to command a well-paid job will depend on one's capacity for creative thinking and/or possession of specialized knowledge that enables one to solve problems and address questions whose rarity makes them not viable to program a computer to answer. In a digital world, we must move away from the economist's traditional focus on marginal costs, for such costs are often negligible and what matters are prospective average fixed costs, including the costs of programming (Earl and Mandeville, 2009). If the digital age is a time in which even the skills of general medical practitioners and university-level teachers can mostly be automated, it is not a good time for the less intellectually gifted – unless they can acquire a craft/trade whose tasks are sufficiently varied as to be protected against programming and, ideally, pertain to an activity that cannot be performed remotely, thereby being protected from global competition. For example, because of the variability in housing designs, there may

be good opportunities for those who have the skills needed to renovate bathrooms, whereas anyone hoping to get by as a taxi driver will soon have to contend with a world of GPS- and laser-guided driverless taxis.

In its most general form, Simon's Travel Theorem points to the following question: Why seek a product or service from a human rather than from a programmed system? If the technology for analysing verbal or typed responses to sets of questions that begin 'How can I help you?' is good enough, there may be no discernible difference between dealing with a human call center operative or online chat assistant, versus dealing with an automated system that can answer one's questions. Indeed, the latter might even be a better experience if queuing times are reduced and the computer is 'speaking' using sound samples from a native speaker of one's mother tongue. Moreover, though it may crash occasionally, the automated system will not suffer the lapses of memory or attention that afflict the boundedly rational human worker, and it will not be driven by pride, or other personality related factors, to keep wrestling with a problem that needs to be referred to a specialist to solve. In future, for many applications, it may only be the *superiority* of a programmed system's responses that gives the game away (in terms of a Turing Test) that one is not actually dealing with a person at the other end of the line.

Ultimately, it may be human limitations that provide an enduring basis for humans to be there to assist us with our problems. Attempts to program tasks may be confounded by the tacit knowledge problem introduced to economics by Nelson and Winter (1982); if an expert cannot put into words the essence of a knack that they have – even if it is something they have acquired via a process of learning by doing rather than something that was innate – then we cannot program what they do. As Simon learned via his research on chess masters (Chase and Simon, 1973; Simon and Chase, 1973), the seemingly intuitive behavior of experts arises from the subconscious ability to find matches between the situation at hand and elements of a large set of memorized experiences (see also Frantz, 2005). In principle, the expert's mode of operation is programmable in just the same way that it is possible to program object recognition devices, such as those that can scan vehicle license plates. However, in practice, an expert's judgment capabilities may be impossible to replicate via a computer program due to the expert being unable to articulate, in enough detail, their vast array of stored experiences.

Related to this is the issue of whether one can program skills in asking clients the right questions and judging the credibility of the answers that

they provide. An automated medical diagnosis system might be more reliable than a medical practitioner in making statistically correct inferences from a given answer to a particular question when that answer is received remotely and in a typed or tick the box form. However, the automated system may lack the capacity of a human for 'reading' patients face to face and sensing which further questions to ask or which choice of phrasing, intonation, and body language to employ when asking them.

So long as such skills remain elusive, we will continue to need to go to the doctor, banks will want to see us in person before granting us major loans, police interrogations will not be automated, neither will job interviews, and so on. Indeed, as regards job interviews, there may be major downsides to modern-day attempts to ensure fairness and counteract biases in interviews by rigidly programming them. In such selection processes, all candidates (who typically have been shortlisted on the basis of tick the box selection criteria) face the same set of questions in the same sequence. This attempt to approximate an impersonal mechanized system rules out using a more improvised, play it by ear, approach to interrogating each candidate based on the impressions that expert interviewers are able to start putting together as the interview proceeds. At the very least, it may be wise to design such programmed systems to allow room for unscripted supplementary questions to be raised in the light of answers received to those that have been programmed.

Busy Consumers

In the century since Simon's birth, the leisure hours available to workers have increased substantially. Yet, for many, the modern world is one of feeling harried by time pressure, with too much to do when not engaged in paid work (Linder, 1970; Thompson, 1996). This continues despite the arrival of digital aids for organizing one's life, getting around, and shopping. In the conventional perspective, rising affluence enables consumers to select from an expanded set of options and make an optimal selection among time-saving and time-hungry products. Diminishing marginal utility for particular kinds of products can arise due the crowding effects of the accumulation of durable goods that take time to consume. Such a perspective leads one to predict growth in relative sales of more ephemeral products that are efficient as means of using up income, for example, each CD that one buys potentially stands in the way of listening to those that one already has, for years to come, whereas attending a concert precludes just one evening of listening to

CDs and leaves one only with memories. However, note the need to consider the relevance of Simon's Travel Theorem in this context and recognize that it might be more efficient in terms of time to stream a video recording of the performance via the Internet, which begs the question (explored in Earl, 2001), namely, why, in the digital age, do consumers attend live performances?

A constrained optimization view of time allocation leaves no place for the sense of feeling under pressure due to insufficient time: one simply makes the best use of the time and income that one has. This view also makes no sense in logical terms once we recognize, with Simon (1978, p. 13; see also the discussion in Berger, 1989, pp. 217–23) that attention is a scarce resource with peculiar properties. The question of how much time to spend attending to the allocation of one's attention is an intrinsic part of time allocation problems, as more time spent on this task could result in a better allocation, for example, via forming a better assessment of one's possible future circumstances. It is by operating, as Simon realized, on a problem-solving basis that the human brain avoids the infinite regress entailed in choosing how to allocate time; we immerse ourselves in what we have started doing until jolted into a new focus by something that we are programmed to take note of (such as a warning siren), or by happening to discover an area in which we are (or could be in danger of) failing to meet our aspirations. We attend to goals sequentially, ranking aspirations in order of priority. Repeated inability to meet aspirations will result in them being lowered eventually to more achievable levels but the resilience of aspirations is a key source of a sense of frustration and pressure when they are not being met (Simon, 1959).

In contrast to the optimizing consumer, the real consumer in the modern world feels harried as a result of setting, and resisting changing, aspirations that are more demanding than those set in earlier generations. This is partly the result of the use of social reference standards in setting aspirations, particularly in relation to the opportunities that parents ought to provide for their children in affluent societies (as with the soccer mom phenomenon). Finite capacity to imagine possibilities and compute their implications combines with the truncation of attention to a particular issue to ensure that the consumer is working with plans whose details are only fleshed out sequentially on the run, often in the course of dealing with surprises. Consumers are not capable of making comprehensive sets of contingent choices at the start of any period.

Ambitious aspiration and bounded rationality result in what Thompson (1996) calls a 'juggling lifestyle', particularly for women

trying to combine a career with bringing up children; the consumer is constantly trying to solve problems due to working with insufficient slack to absorb unexpected outcomes/events in one area with ramifications for their ability to meet other goals. The consumer's problem thus has similarities with the origins of chaos in complex, centrally planned economies and mass transit systems. In terms of Simon's (1962) analysis, those who enjoy the luxury of discretion in affluent societies would be wise to build low-stress, resilient lifestyles of independent, or at least nearly decomposable, modules.

Spoilt for Choice

Economics needs to be able to provide insights into how consumers choose when their options are many and varied. Hitherto, adherents to its mainstream have viewed limited access to products and services as the consumer's key problem. The size of the consumer's feasible set limits the level of utility that can be attained, so a bigger range of choice ends up being viewed within mainstream economics as an unambiguously good thing. But from Simon's standpoint the bigger the potential choice set, the bigger the problem of bounded rationality is because there is more information for which to search, more evaluation to be done, and more information to be processed once judgments have been made about what the options that have been discovered have to offer. In short, a bigger feasible set and more differentiation between products increase the challenge that the consumer faces, leading to stress in the process of choice and nagging doubts about the quality of one's choice.

The scale of the consumer's choice problem has changed spectacularly over the past century, but the dominant view of how consumers should be envisaged as choosing has barely moved on. It is essentially still based on the marginal utility framework of the 1870s, augmented via Slutsky's (1915) income and substitution effect framework that built on the indifference curve idea pioneered by Edgeworth (1881). The one size fits all models developed from these foundations have typically presumed that it is perfectly in order to generalize up from a simple two-good diagrammatic version (in intermediate microeconomics courses) to an n-dimensional algebraic version (in more advanced courses and research papers). Simon's work gives reasons to doubt that this is a reasonable representation in the digital age and provides an alternative way forward.

The traditional framework might not have been too wide of the mark a century ago, but now the consumer's choice problem is on a grand

scale. Growth in international trade is one reason, even though, as Keynes (1919, p. 11) pointed out, it was possible at the start of the twentieth century to order products over the phone from anywhere in the world. This precursor to global online shopping was available only to a rich elite and, even by the time Simon was born, the first wave of globalization was already unraveling due to World War I. For much of Simon's life, particularly from the Great Depression onwards, barriers to international trade limited the choice sets that consumers enjoyed. In the last quarter of a century, globalization (fostered by the World Trade Organization and aided by the Internet) has greatly expanded the range of choice that consumers face. If an authorized distributor does not deign to import the product that I want from its overseas manufacturer, I can order it from an overseas supplier. I can also increase the size of my choice set if I can get products more cheaply by importing them directly rather than buying them from a local distributor. (My partner provided an extreme yet mundane example during the writing of this chapter. Instead of buying Magic Eraser cleaning pads for $4 for a pack of two in our local supermarkets, she procured a pack of 200 identical pads directly from China for a mere $8, including postage.)

Technological advances have had arguably even more significant impacts on the nature of the choice problem by presenting consumers with unfamiliar products that may require considerable expertise to choose and to use. Rising real incomes have amplified this by making unfamiliar categories of products accessible to more and more people. When Simon was born, the consumer's choice set was already expanding rapidly: homes were being electrified; electrical appliances were starting to appear; motorization was underway; and consumers were enjoying access to new products that used new materials, such as celluloid. But this was nothing compared to the set of opportunities open to modern consumers in advanced economies, whose lives have been revolutionized by the discovery and application of new plastics, microchips, and digitization.

As Schumpeterian processes of creative destruction change the possibilities available to consumers, the set of products may change gradually in some areas – a module or so at a time – rather than in a revolutionary manner. This makes it easy to update knowledge of low-cost products that are purchased frequently and can be explored without substantial investment being necessary. But with durable items individual consumers are only in the market infrequently, so each time they return they face an unfamiliar landscape. This may be particularly

puzzling if the technology is new and no standard has yet emerged, for example, early plasma and LCD flat screen TVs were joined, and then displaced, by LED TVs, but the latter were rapidly augmented with 3D smart TVs, then super-high-definition TVs. There is massive scope for error for those who decide they are in the market for a particular product but are unaware that is about to fall greatly in price or be rendered obsolete. Assuming that consumers are fully informed is convenient for mainstream economists, but it does not tell us how consumers can or should actually cope with such challenges.

In addition to the general expansion of the typical consumer's choice set there is the widening that occurs as more and more niche products appear. This is facilitated by the combination of parts-bin manufacturing (reusing existing modules from a variety of products) and computer-aided design (which greatly reduces the costs of designing new modules). In contrast to the 1916 Model-T Ford, whose cost was partly contained by limiting the range of variants, the 2016 BMW Mini can be specified in more than a million different option combinations. Moreover, the set of products being offered can now change faster than ever before, based on the same processes; the more technology modules that have been developed, the greater the number of novel combinations that can be created rapidly. At the retail stage, the consumer is not constrained by local stores that stock a limited range, but can shop at mega-marts that offer tens of thousands of product lines, or at online hubs such as Amazon.com that list millions of alternatives.

Digital technology has also widened the choice set in the area of human relationships as online dating sites permit people to do focused search and sampling instead of relying on happenstance, local networks, or arranged marriages. Anyone determined to find The One and avoid ending up in disastrous or dull relationships potentially can find such aids to choice very time-consuming; those who are both wise and likely to be attractive to many will use de-marketing techniques to try to limit the number of expressions of interest with which they have to deal.

There is potential for this explosion in the consumer's range of choice to result in what Iyengar and Lepper (2000) and Schwartz (2005) call 'choice overload'. However, although being spoilt for choice may be demotivating and lead us to choose only not to buy anything at all in a particular area (as argued by Iyengar and Lepper, 2000), most of the time choices do get made, despite the potential for information overload, and despite the fact that choices are made with reference to product characteristics whose number keeps being inflated as products include more

and more features. A Simon-inspired view of how this happens includes the following ingredients:

(i) Consumers approach the problem of choice in a hierarchical manner, only choosing among products that come into similar categories rather than choosing among many classes simultaneously. If necessary, the consumer filters the set under consideration by homing in on more finely grained categories. Retail environments are set out in ways that facilitate this process as well as helping boundedly rational staff.

(ii) Consumers use their experience with brands and/or brand reputations as a means for selecting potentially acceptable suppliers and ignoring unfamiliar suppliers, thereby reducing the scale of their information gathering and processing problem.

(iii) Consumers limit their range of products to consider by focusing on those whose prices are within a budget range that they have selected.

(iv) Consumers use intolerant, non-compensatory decision rules to construct shortlists of potentially acceptable products. Only those products that survive the initial screening in terms of key requirements are examined in further detail. If there are many products that satisfy these requirements and/or many non-core requirements in terms of which they differ, non-compensatory decision rules may also be applied to those that get shortlisted, instead of marginal trade-offs being made to compute overall ratings (for empirical studies of how increasing the range of options leads to switches in favor of non-additive decision rules, see Payne, *et al.*, 1993; and Lenton and Stewart, 2008 – the latter study focuses on the online dating problem that did not even exist at the time of the former).

(v) Consumers outsource not merely supplies of information about what options are available and their characteristics, but also the answer to the normative puzzle in unfamiliar situations of 'What should I want?' In other words, they make use of 'the market for preferences' (Earl and Potts, 2004). They may know what they want, and the extent to which they are willing, if at all, to make trade-offs in terms of general features, such as safety and convenience, but be very unclear on how to choose between alternative technologies that offer different means to such ends. So long as there is no apparent conflict with their broad requirements, they will be open to using recommendations as the basis for their

> choices where they lack knowledge of how the suggested means actually works to achieve their ends. They will use heuristics – such as those for judging source credibility – to resolve conflicting recommendations of what they ought to do to solve the choice problem at hand.

Such processes have been a research focus in marketing since the 1970s. They go against the traditional economics view of well-informed consumers with given preferences carefully making trade-offs at the margin. Limited information gathering and the use of intolerant decision rules appears irrational in terms of the traditional perspective, potentially leading to intransitive behavior and missed opportunities. However, it is inappropriate to judge the quality of choices that consumers make in fluid, complex environments in terms of a rational choice model. In such contexts, optimal choices for a consumer may be elusive *even to economists*, due to changes in the set of options occurring so frequently that it is impossible to know what the set is at any particular point in time, quite apart from the impossibility of knowing how it may change shortly and hence whether or not the optimal strategy is to defer choice.

Even where the product is one that can, broadly speaking, be chosen on the basis of how much it costs in delivering well defined quantitative results (such as a mobile phone connection plan for making calls, sending SMS messages, and accessing the Internet), there may be so many variants on offer, with such complex usage terms that even an economist would only be able to pronounce on what the best choice to serve a particular set of requirements would have been months earlier. Moreover, bounded rationality will limit the quantity and quality of information that economists can extract from consumers about the usage they would hope to get from the product in question; survey respondents will suffer from fatigue in the face of complex questionnaires to find out what their preferences are and may actually have a very poor idea of their usage patterns even in the recent past, let alone in prospect.

Simon (1976) was thus prescient in advocating the idea of using procedural rationality as a basis for making normative assessments of choices; while the optimal choice may be elusive, one might at least be able to identify ways of taking decisions that were usually inferior to others in a particular context. For example, a consumer who is trying to solve the mobile phone service contract choice problem singlehandedly

may make errors due to difficulties in making sense of complex contracts, or arithmetical errors, or merely errors when using a calculator. Spending a lot of time trying to solve the problem may entail major costs, depending on the chooser's circumstances, for example, if one is merely a short-term international visitor to a country and trying to avoid major international roaming charges. In such a case, it would be entirely reasonable to select a pre-paid deal from a major supplier's retail site. For a consumer with longer-term needs in the same market, a procedurally rational strategy might entail choosing a plan ranked highly by more than one comparison website (more than one in order to limit the risk of recommendations being biased by any conflicts of interest that the comparison sites may face). Such websites may be incomplete and out of date but the issue is not whether they can uncover inherently elusive optima; rather, it is whether they save time and money compared with a do-it-yourself attempt at choosing.

Natural Pedagogy and Virtual Shopping

In terms of Simon's Travel Theorem, the consumer has no information-related reason for visiting a bricks and mortar retail site in preference to purchasing online. In the time it takes to do the former and receive in-store information and preference recommendations, the consumer can potentially get better information and recommendations online from websites or social contacts. The latter sources may have a wider range of experience in the market in question and not face the conflicts of interest that a salesperson faces. The willingness of fellow consumers to be suppliers in the market for preferences is vital for dealing with the problem of the division of knowledge in cases where the market would fail to make it viable for a commercial comparison site to operate. That this willingness exists widely is evident each time we find product reviews on YouTube and other websites, or find members of our social networks willing to help us in our quests. We source our information and knowledge from the crowd and, in areas where we have expertise or have learned via experience, we contribute to the crowd-sourced choices of others, many of whom we do not know.

This process relies upon altruism; people give their time to others without receiving any payment. It is hard to make sense of in a conventional utility-maximization framework, for the self-serving economic agent would freeride unless access to such sources is restricted to those

who also make a contribution, which is plainly not so in many cases. Otherwise, it would have to be explained in terms of the narcissistic buzz or warm glow that the economic agent gets from showing off their possessions via a product review. Since most reviews are posted via usernames that do not readily identify the reviewer, it is actually rather difficult to get any kind of celebrity status in this (except insofar as word gets around their social networks about their YouTube performances). The narcissistic reviewer might therefore mainly enjoy a kind of statistical payoff, via being able to see, as with WordPress blogs or YouTube, just how many hits their post has accumulated or, as with Amazon, how many people rate their reviews as useful relative to others posted there. But can such a perspective really account for the great bulk of reviews or inputs to discussion boards?

If we accept that the seemingly altruistic contributions to the market for preferences are indeed not self-serving but reflect some kind of feeling the contributors have that makes them want to help others, we can make sense of this in evolutionary terms, as in Simon's (1992) theory of altruism. Viewed thus, altruism is a human behavioral tendency that was selected by evolutionary processes because it enhanced the fitness of the species. It could be genetically inherited as part of the brain's hardwiring and/or transmitted down the ages via cultural processes that enabled particular groups to thrive because they were not thoroughly selfish. It may thus be viewed as a modern manifestation of what anthropologists Csibra and Gergely (2011) refer to as the human tendency toward 'natural pedagogy'. From this standpoint, a tendency to want to share knowledge with others is part of what being human is all about. This tendency seems connected with our tendency to experience sympathy and do what seems morally right to help others (as emphasized by Adam Smith, 1976 [1759]) – in this case, to help others resolve a puzzle we have dealt with or to avoid an unpleasant experience that we have had.

Conclusion

Mainstream economists believe that their equilibrium/constrained optimization view of their world is applicable via *as if* models to any context at any point in time. This chapter has attempted to demonstrate that, whatever the past merits of the mainstream framework in simpler, slower moving eras, it results in implausible and misleading models if used to understand the modern digital economy. Simon's approach to economics, by contrast, provides ways of making sense of

how this economy functions via routines, aspiration driven satisficing, hierarchical decomposition, modular structures, and altruistic inputs.[1]

Note

1. Some of the arguments and examples in this chapter originated via my involvement in a long-term research project funded by an Australian Research Council Discovery Grant (DP1093840), in which Lana Friesen and I have been studying how, and how well, Australian consumers choose their mobile phone service contracts. Although I am grateful to Lana and our research assistant Christopher Shadforth for many discussions, the usual disclaimer applies.

References

Berg, N. and Gigerenzer, G. (2010). As-If Behavioral Economics: Neoclassical Economics in Disguise? *History of Economic Ideas*, 18 (1), 133–66.

Berger, L. A. (1989). Economics and Hermeneutics. *Economics and Philosophy*, 5 (2), 209–34.

Chase, W. G. and Simon, H. A. (1973). Perception in Chess. *Cognitive Psychology*, 4 (1), 55–81.

Csibra, G. and Gergely, G. (2011). Natural Pedagogy as an Evolutionary Adaptation. *Philosophical Transactions of the Royal Society (B)*, 366, 1149–57.

Day, R. H. (1967). Profits, Learning and the Convergence of Satisficing to Marginalism. *Quarterly Journal of Economics*, 81(2), 302–11.

Dunn, S. P. (2000). Fundamental Uncertainty and the Firm in the Long Run. *Review of Political Economy*, 12 (4), 419–33.

Earl, P. E. (2001). Simon's Travel Theorem and the Demand for Live Music. *Journal of Economic Psychology*, 22 (3), 335–58.

Earl, P. E. (2003). The Entrepreneur as a Constructor of Connections. In *Austrian Economics and Entrepreneurial Studies: Advances in Austrian Economics, Volume 6*, ed. R. Koppl. Oxford: Elsevier, 113–30.

Earl, P. E. (2012). Kahneman's Thinking, Fast and Slow: What You See is Not All There Is. *Prometheus*, 30 (4), 449–55.

Earl, P. E. and Mandeville, T. (2009). The Competitive Process in the Age of the Internet. *Prometheus*, 17 (3), 195–209.

Earl, P. E. and Potts, J. (2004). The Market for Preferences. *Cambridge Journal of Economics*, 28 (4), 619–33.

Edgeworth, F. Y. (1881). *Mathematical Psychics*. London: C. Kegan Paul & Co.

Frantz, R. (2005). *Two Minds: Intuition and Analysis in the History of Economic Thought*. New York: Springer.

Gigerenzer, G., Todd, P. M. and the ABC Research Group (1999). *Simple Heuristics that Make Us Smart*. New York: Oxford University Press.

Hodgson, G. M. (1988). *Economics and Institutions: A Manifesto for a Modern Institutional Economics*. Cambridge: Polity Press.

Iyengar, S. S. and Lepper, M. R. (2000). When Choice is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of Personality and Social Psychology*, 79 (6), 995–1006.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Strauss and Giroux.
Keynes, J. M. (1919). *The Economic Consequences of the Peace*. London: Macmillan.
Keynes, J. M. (1937). The General Theory of Employment. *Quarterly Journal of Economics*, 51 (2), 209–23.
Langlois, R. N. and Robertson, P. L. (1995). *Firms, Markets and Economic Change: A Dynamic Theory of Business Institutions*. London and New York: Routledge.
Leijonhufvud, A. (1986). Capitalism and the Factory System. In *Economics as a Process: Essays on the New Institutional Economics*, ed. R. N. Langlois. New York: Cambridge University Press, 203–23.
Lenton, A. P. and Stewart, A. (2008). Changing her Ways: The Number of Options and Mate-Standard Strength Impact Mate Choice Strategy and Satisfaction. *Judgment and Decision Making*, 3 (7), 501–11.
Linder, S. B. (1970). *The Harried Leisure Class*. New York: Columbia University Press.
Nelson, R. R. and Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, MA: Belknap Press of Harvard University Press.
Payne, J. W., Bettman, J. R. and Johnson, E. J. (1993). *The Adaptive Decision Maker*. Cambridge: Cambridge University Press.
Schwartz, B. (2005). *The Paradox of Choice: Why More is Less*. New York: Harper Perennial.
Sent, E.-M. (2004). Behavioral Economics: How Psychology Made its (Limited) Way Back into Economics. *History of Political Economy*, 36 (4), 735–60.
Shackle, G. L. S. (1961). *Decision, Order and Time in Human Affairs*. Cambridge: Cambridge University Press.
Simon, H. A. (1958). The Role of Expectations in an Adaptive or Behavioral Model. In *Expectations, Uncertainty and Business Behavior*, ed. M. J. Bowman. New York, NY: Social Science Research Council, 49–58.
Simon, H. A. (1959). Theories of Decision-making in Economics and Behavioral Science. *American Economic Review*, 49 (3), 253–83.
Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society,* 106 (6, December), 467–82.
Simon, H. A. (1976). From Substantive to Procedural Rationality. In *Method and Appraisal in Economics*, ed. S. J. Latsis. Cambridge: Cambridge University Press, 129–48.
Simon, H. A. (1978). Rationality as a Process and Product of Thought. *American Economic Review*, 68 (2), 1–16.
Simon, H. A. (1984). On the Behavioral and Rational Foundations of Economic Dynamics. *Journal of Economic Behavior and Organization*, 5 (1), 35–55.
Simon, H. A. (1991). *Models of My Life*. New York, NY: Basic Books.
Simon, H. A. (1992). Altruism and Economics. *Eastern Economic Journal*, 18 (1), 73–83.
Simon, H. A. and Chase, W. G. (1973). Skill in Chess. *American Scientist*, 61 (4), 394–403.
Slutsky, E. (1915). Sulla teoria del bilancio del consonatore. *Giornale degli Economisti*, 51, 1–26.
Smith, A. (1976 [1759]). *The Theory of Moral Sentiments*, ed. D. D. Raphael and A. L. MacFie. Oxford: Clarendon Press.

Smith, A. (1976 [1776]). *An Inquiry into the Nature and Causes of the Wealth of Nations*. ed. A.S. Skinner, A.L. MacFie and W.B. Todd. Oxford: Clarendon Press.
Thompson, C. J. (1996). Caring Consumers: Gendered Consumption Meanings and the Juggling Lifestyle. *Journal of Consumer Research,* 22 (4), 388–407.
Veblen, T. B. (1899). *The Theory of the Leisure Class: An Economic Study of Institutions*. New York: Macmillan.
Winter, S. G. (1971). Satisficing, Selection and the Innovating Remnant. *Quarterly Journal of Economics, 85*(2), 237–61.

16

Herbert Simon and Some Unresolved Tensions in Professional Schools

Mie Augier and Bhavna Hariharan

Introduction

Herbert Simon is recognized for his contributions to fields such as organization theory, economics, cognitive science, artificial intelligence, and psychology, as well as others:[1]

> Organizing a professional school. Is very much like mixing oil with water: it is easy to describe the intended product, less easy to produce it. And the task is not finished when the goal has been achieved. Left to themselves, the oil and water will separate again. So also will the disciplines and the professions. Organizing, in these situations, is not a once-and-for-all activity. It is a continuing administrative responsibility, vital for the sustained success of the enterprise. (Simon, 1967, p. 16)

His paper on the business school as a problem of organizational design (1967) is, although perhaps less well known, a paper that reflects not only his mind as an organization theorist and scholar, but also his awareness of the importance of some of the fundamental issues in the education of professions and in professional schools (such as business schools). As a person, he was well known for his strong mind and his insistence on going against the centripetal forces of scholarly disciplines, even if it might have been easier for him to stay within one (or two) disciplines. As he said in conversation, 'if you see any discipline dominating you, you join the opposition and fight it for a while'.

The reformation of business schools and management education, symbolized by the work of Simon and his colleagues, had similarities

with the changing of medical schools, which had happened a few decades earlier. Simon's and Flexner's visions for professional education have some similarities although there is little indication that Simon and Flexner overlapped in person. There are lessons from their work that alone – and together – may provide fruitful avenues for future research.

In this chapter, we take as our starting point Simon's organizational analysis of the professional school, and discuss it in the light of some of the changes that happened in other professional schools (medical and engineering), as they went through institutional and intellectual transformations.

Simon directed much of his energy towards creating and reforming much of the intellectual content of one type of professional school (the business school), by creating with colleagues, among other things, the new field of organization studies. Flexner, on the other hand, directed his efforts towards the institutional and societal reform of another type of professional education, namely medical schools.[2] Together, Simon's and Flexner's contributions are powerful not only for understanding some of the tensions facing professional schools as institutions, but they also hold possible implications for how we think of the education of professions in the future. Discussing some of these aspects is the aim of this paper. In particular, the second section discusses the Simon/Flexner vision for professional schools, taking into account the tensions that exist and the concept of professionalism embedded in Simon's and Flexner's visions. Then we discuss how some of the Simon/Flexner insights are embedded in the history of another professional school, the engineering school. We end with some implications and the importance of curiosity in research as emphasized by Simon.

Simon and Flexner on Professional Schools and the Professionalization of Education

> Accurately or not, we perceived American business education at the time as a wasteland of vocationalism that needed to be transformed into science-based professionalism, as medicine and engineering had been transformed a generation or two earlier. (Simon, 1991, p. 138)

> As schools and clinics of medicine multiplied, problems of personnel continued to arise. Where were the men competent to conduct them? They had to be trained or to be brought up to date. Sometimes we could catch a young man and send him abroad to work while the buildings were in process of construction; sometimes we had to

> risk a new but likely candidate. No stereotyped procedure was possible. If we were to create medical schools of the highest standard, the young men who were to form their medical faculties must know not only the best that America could offer but the best that could be obtained anywhere in the world. We readily procured authorization to find men of the requisite caliber and to send them abroad for such periods as might be desirable. We pursued no standardized procedure. (Flexner, 1910, pp. 312–13)

The revolution in business education which Simon (and March, Cyert and others at Carnegie) were on the forefront of is often seen as a 'delayed response' to the Flexner report and what it did to reform medical education. Flexner and Simon shared a belief that the professional school they were concerned about (Simon the business school; Flexner the medical school) required their being rebuilt on the solid ground of several disciplines (making them interdisciplinary), and being developed to be problem driven. The two professional schools also shared historical paths in the sense that, despite their different professional focus, they had been seen as intellectual 'slums'; poor vocational schools which served neither the disciplines nor the professions very well (Flexner, 1910; Gordon and Howell, 1959). Despite the difficulties of intellectual, organizational, and institutional change, and the powerful forces of inertia, both Flexner and Simon managed to transform and lead a kind of revolution in their respective educational institutions.[3]

Why did the ideas about (re)building business schools on the basis of social science disciplines emerge? Did the modern business school that Simon wrote about arise because of demands from businesses or from stimulating empirically inclined deans to facilitate better communication between the academic and business worlds? Did building a science of business follow naturally from existing theories or because increasing growth of, and awareness about, real world business organizations made a distinct education for them necessary? The process of researching such questions might give us an understanding of Simon's insights and the contexts in which they emerged. This understanding, in turn, is critical to realizing and imagining the relevance and impact of the Simon/Flexner vision for professional schools and professional education today.

Before the Second World War, business schools had emerged largely as vocational schools, often focused on particular crafts or functions of business, or (after the Industrial Revolution) as training grounds for functions such as book keeping. It was only after the establishment of

the field of political economy (starting with Adam Smith) that business found some kind of intellectual base on which it could ground itself more firmly. However, at first, academics weren't very welcoming of business as an intellectual topic (Augier and March, 2011, chapter 2). Helped first by the establishment of certain associations that provided some institutional legitimacy for business topics and specialized functions in business (such as the American Marketing Association established in 1915), more generalized management associations (the American Management Association established in 1923 and the Association to Advance Collegiate Schools of Business established in 1916), and a growing professionalization in the sense of an institutionalization of business functions, all then helped pave the way for seeing the need for upgrading the intellectual content of the schools. The business schools were still mostly serving the business professions; Bossard and Dewhurst noted in 1931 that '[t]he primary aim of the university school of commerce [as business schools were often called back then] is to prepare its students for successful and socially useful careers in business' (p. 55). The lack of a broader (and intellectually deeper) education for business people was acknowledged by deans and faculty across the country, but it was not until the Gordon-Howell report that a good countrywide analysis of the problem was provided.

The forces that helped get the Gordon-Howell report under way included institutional and intellectual developments in places such as the Ford Foundation and the RAND Corporation; for the Gordon-Howell report was as much a symptom as a cause of societal change (Augier and March, 2011, chapters 4–5). In a time and culture of optimistic urgency in the post-war years (with big problems in the world serving as a focal point for attracting different disciplinary minds to think together), intellectual developments that would prove important for the content of business education for Simon and colleagues included operations research and linear programing, game theory, evolutionary economics (which often were pioneered at RAND), and developments in the behavioral and social sciences, which were a priority for the Ford Foundation in the early 1950s (Augier and March, 2011, chapter 5).

Simon was involved in many of these developments, both intellectually (through his contribution to the disciplines), and also by serving on several Ford Foundation committees and advisory groups, focused in particular on trying to establish more fruitful relations between the different disciplines (especially economics and the behavioral social sciences).[4] The Ford Foundation was also aware of the need to do

something about the state of business schools (and had several businessmen on its board and as advisors). The opportunity to blend and integrate work in the disciplines while improving business education became a priority, also linking together different priorities within the Ford Foundation itself. The Foundation launched a number of programs and initiatives, including setting up centers for excellence and grants for particularly promising business schools. It also funded the Gordon-Howell report was providing data, documentation, and an analysis of the fundamental problems of business schools.[5] The Gordon-Howell report, also sometimes considered as a Flexner-like report for business schools, and called for changes such as: a major upgrade of both students and faculty in business schools; rebuilding business education on a solid intellectual foundation; and bringing to bear behavioral social science, mathematics, and statistics in the analysis of business problems. It also said in the opening pages,

> Today it [the business school] is a restless and uncertain giant in the halls of higher education . . . but it is an uncertain giant, gnawed by doubt and harassed by the barbs of unfriendly critics. It seeks to serve several masters and is assured by its critics that it serves none well . . . They search for academic respectability, while most of them continue to engage in unrespectable vocational training. They seek to be professional schools, while expressing doubt themselves that the occupations for which they prepare students can rightfully be called a profession. (Gordon and Howell, 1959, p. 4)

To some extent, their criticism resulted from a tension that exists in all professional education between being relevant for a profession, and pursuing rigorous academic research (Simon, 1967; Bach, 1958; Augier and March, 2007). This is a struggle that most (if not all) professional schools have, with strong internal and external forces working against them and what Simon called a problem of 'mixing oil and water'.

Coinciding with the Ford Foundation efforts, Simon and his colleagues were busy trying to build a business school at Carnegie. Although initially reluctant (Lee Bach and Bill Cooper were the initial ones there to set up shop), he did join Carnegie, as they built the graduate school of industrial administration and created a vision and a program for management education, which became a role model for the Ford Foundation to push business school reform. An important cornerstone was the belief in fundamental academic research: 'I want to stress as strongly as I can my own belief that fundamental research is a major

part of every leading business school, especially those which offer graduate work . . . The function of the university is to be ahead of best practice, not to be trailing a few steps behind the operating business world' (Bach, 1958, pp. 363–4).[6] Simon, Bach, and Cooper were focused on wanting the best students, the best faculty, as they started building the business school.

Simon and his colleagues at Carnegie became a role model for changes in business schools across the country (although schools differed in the extent to which they implemented the changes, or drifted from them) (Augier and March, 2011, chapters 6–7). And, as the Simon quote with which we begin this section indicates, the need for the changes in business schools mirrored the reforms that medical schools had undertaken as a result of the Flexner report decades earlier. As one commentator noted in the wake of the Gordon-Howell report: 'Not since Flexner's 1910 evaluation of medical schools had so much attention been focused on a field of professional education' (Jeuck, 1973, p. 285).

The background to Flexner's report was the terrible state of medical education in the late nineteenth century, evidenced from increased data collection. For example, the American Medical Association had begun collecting data on medical education since 1901 and established a council on medical education to raise standards. This council wrote a report in 1907 on the need to establish more science in medical education, which was also a central theme for Flexner's 1910 report, *Medical Education in the United States and Canada*.

Based on survey data, analysis, research, and visits to more than 150 medical schools, Flexner recommended that medical education should be rebuilt on underlying scientific disciplines, not just training; knowledge of biology, physics, and chemistry were central for practicing physicians as well as for the future education of medical doctors and educators. As for business schools a few decades later, both institutional, external, and intellectual forces helped shape the need and the possibility for change in professional education. Those factors included:

- Developments in the sciences in the late nineteenth century; in the earlier decades of the nineteenth century there was neither basic scientific training nor much laboratory medicine. These increased after the Civil War and led to the emergence of science teachers.
- Impressive success of German science helped instill a culture of optimism about the prospect of American basic science.
- Organizational forces, including the American Medical Association and the Association of American Medical Colleges (with the AMA in

> particular having suffered political fractions and battles) and the Council of Medical Education asking the Carnegie Foundation for help to study the problem.

Thus Flexner was one voice, albeit an important one, nested in other developments. He became the symbol, and a driving force, for change in medical education, as Simon, Carnegie, and the Gordon-Howell report did for business education.

One of the arguments in business school debates from the time of Simon was, what are the challenges of being a profession and does that entail building management education on a foundation of disciplinary research? This issue was discussed in both Simon (1967) and the Gordon and Howell report.[7] They suggested four criteria for defining a profession. First, a profession should rest on a systematic body of knowledge of intellectual content and on the development of personal skills in the application of that knowledge. Second, it must set up professional standards of conduct and set those above personal gain. Third, it should have an association of members to enforce those standards, and fourth, it should have ways of, and standards for, entering the profession. When Simon and colleagues started out, management did not fit the necessary criteria of a profession. As Lee Bach noted:

> Careful analysis of management and its various facets has given us many insights into what is the gold and what is the dross. But business administration is a new profession. It still operates heavily on rules of thumb and hunches, often unnecessarily so. It is a profession that is growing up rapidly. A crucial part of that growth must be amassing a careful scientific analysis and research to lay bare what is hearsay in management, what is fundamental skill, and what is transient practice. I am personally convinced that careful, fundamental research in the management fields over the next half century can and will vastly improve our present knowledge and skills. (Bach, 1958)

Bach, Simon and colleagues did help rebuild management education on a foundation of behavioral social science, economics, and statistics and developed tools for understanding issues of business practice from those areas. It was a deliberate strategy of making management education more scientific and less vocational. Management was becoming a science-based profession that integrated practical applications and basic ideas through problem driven research.

Engineering Schools – Tensions of Identity

> A social system left to itself gravitates toward a position of equilibrium – of maximum entropy, so to speak. The position of maximum entropy for a professional school is one in which the portion of the faculty that is trained in the profession is absorbed in the culture of the profession, while the portion that is trained in one of the underlying disciplines is absorbed in the culture of that discipline, leaving a deep gulf between them. (Simon, 1967, p. 12)

As seen in the previous section, both Simon and Flexner focused, implictly and explicitly, on the professionalization of schools (business and medicine respectively) by helping to build their respective fields on a sound scholarly and intellectual foundation. They were part of the larger movements that enabled institutional changes. Simon also helped intellectually develop (sub)fields that became important to the development of business schools and management education (in particular, organizational behavior and strategic management). The upgrading of the academic foundations was seen as necessary for the professional schools to benefit from current research and to cultivate professional practitioners who themselves were empirically driven and engaged with lab/field work.

The history of engineering education, in turn, differs somewhat from other professional schools. Unlike business schools that were reluctantly accepted and medical schools that were integrated after the Flexner report, engineering schools have, from their inception, been part of academia. They did not evolve out of the apprenticeship model found in medical and law schools. Instead, engineering education has historically been guided by educators as opposed to practitioners (Grayson, 1977).

Whereas Simon identified the bringing together of fundamental and applied researchers as one of the primary tensions within business schools, engineering had a somewhat different problem. The identity crisis within engineering schools has traditionally been in being seen as a profession and *differentiating* from scientists. Layton, in his award-winning book *The Revolt of the Engineers*, says, 'Placing the emphasis on the application of professional knowledge, rather than on its creation, distinguishes between the scientist and the engineer' (Layton, 1971, p. 26). Others noted:

> Basically an engineer is not trying to do science; he is doing engineering which is something different. The science he uses is a means to

> that end, and its quality as science is irrelevant so long as it works. A scientist is judged by his publications but an engineer is judged by his large-scale achievements. He can be a great engineer without having published a word... In short engineers are not second hand citizens in the empire of science; they are masters of an empire of their own, of which science is only a part (Johnstone, cited in Emmerson, 1973, p. 299).

Engineering in the United States was initially an all military affair with focus on building naval strength and fortifications. As early as 1778, General Washington called for the establishment of engineering schools, and in 1794 Congress authorized the creation of a Corps of Artilerrists and Engineers in West Point, New York. These visions resulted in the establishment of the West Point Academy, which started granting degrees in 1933.

At the time, engineering was the domain of creative inventors who were creating new and improved modes of agricultural tools and transportation facilities, or individuals who were sent abroad to Europe to develop their technical skills and expertise. In fact, the curriculum at West Point was based on the civil engineering curriculum offered by French institutions (McGivern, 1960, p. 10).

At least in part as a result of increased demand for roads, railroads, canals, and other public utilities, the mid- to late- 1800s saw the growth and diversification of the engineering profession. Starting with a degree in civil engineering offered at Partridge's Academy in 1821, to the Polytechnic College of the State of Pennsylvania offering the first Mechanical Engineering degree in 1854 and a degree in Mining Engineering in 1857, up to the introduction of Electrical Engineering in 1882 and Chemical Engineering in 1888 at the Massachusetts Institute of Technology (Grayson, 1977).

The diversity of the engineering profession meant that there were parts of the profession, such as mechanical and civil engineering, that had emerged out of inventions and the entrepreneurial spirit of people in America, while also being inclusive of chemical and electrical engineering, which were rooted in the sciences. Layton, describing the professionalization of engineering, wrote,

> By asserting that all technology was the work of engineers, they defined their social role. By holding that all technology was applied science, they laid claim to a sophisticated body of knowledge. From these fundamental postulates of esoteric knowledge and social

> service, all of the values of professionalism could be derived. (Layton, 1971, p. 56)

Just like in business and medicine, professional associations played a role in the evolution of engineering education, and the rise of professional societies, such as the American Society of Civil Engineers as early as 1852, laid the foundations of professionalization. However, the growing fragmentation of engineering societies was evidence of the underlying tensions in defining the engineering profession.

> Granted, the qualifier 'civil' included all non-military engineering in the middle of the nineteenth century, but the ASCE's lack of responsiveness to those civil engineers dealing more with dynamic than static structures drove them to form their own societies, the American Society of Mechanical Engineering, in 1880. The establishment of other specialized societies followed and engineering has acquired in the public mind a reputation for being an amorphous mass of specialists with no overarching identity. Today there is an American Society for Engineering Associations, but its very name points to the fact that it is engineering societies that are being served by it rather than the profession of engineering or its practitioners. An individual engineer cannot even join the ASEA. (Petroski, 2001, pp. 2–3)

Among professional schools, engineering schools have probably been the most reflective in understanding the requirements of training engineers both for immediate employment in the workforce and for graduate work. The Society for Promotion of Engineering Education (SPEE) was founded in 1893. In 1946, it became the American Society of Engineering Education (ASEE). Both SPEE and ASEE have undertaken intensive, comprehensive reviews of the state of engineering education. Significant among these are the Wickenden report (1929), which resulted in the creation of accreditation criteria for engineering schools, and the Grinter report (1955), which is the basis of the four-year undergraduate program as it is today. These reports – although Flexner-like in intent – did not have the kind of impact that the Flexner report had on the medical education. One of the reasons cited by Eric Walker, (an IEEE Fellow and former ASEE president) for the lack of sweeping change as a result of these reports has been attributed to the opposite pulls exerted by the need for the definition and training of the professional engineer on the one hand, and on the other hand the wide array of often disparate skill sets that engineering (ranging across civil, mechanical, electrical,

chemical, and more recently environmental and computer engineering) encompasses (Walker, 1971).

What it means to be an engineer has always been a contentious issue (and echoes Simon's concerns):

> When words lose precise meaning, or lose the same meaning for everyone, misunderstanding is inevitable and communication chaotic. The word 'engineer' is such a word. It was confounded by a spectrum of connotations even in the nineteenth century, and it has been further confounded by new connotations in the academic world of the twentieth century. (Emmerson, 1973, p. 295)

The Changing Demands on Professional Education

Since engineering has always been an applied field employing problem-based approaches to educate future generations of engineers has been standard practice. The engineering curricula have almost always had to contend with how to combine the theoretical and practical elements of the profession. A profession that demands products are built to meet a specified standard of safety and ethics. As such, learning *how* to build things is a necessary skill to acquire. However, by itself these criteria are incomplete for professionalization. Instead, understanding *why* things are built the way they are adds the necessary theoretical background that is echoed in Flexner's vision of empiricism and laboratory work being made a mandatory aspect of the training of medical professionals.

Until the early 1900s, the education of engineers continued to hold on to the traditional approaches as exemplified in the career of William Burr, a practicing engineer who was an instructor, and later professor, of mechanics at Rensselaer Polytechnic Institute (1872), Harvard (1892), and Columbia (1893), and author of several textbooks on bridge construction and materials:

> More obvious evidence of Burr's priorities could be found in one of his textbooks, *The Elasticity and Resistance of Materials in Engineering* [1883]. He divided it into sections headed 'Rational' and 'Technical', a layout that caused another engineer to accuse Burr of being too theoretical. But Burr made his stance clear in the introduction, arguing that the 'rational' section was important but 'a great number, and perhaps all engineers in active practice . . . [will find it] unnecessary' . . . He wanted students to utilize mathematical analysis as a tool in bridge and structural design, but never doubted that good designers relied as much on experience gained through practice. (Seely, 1999, p. 287)

This approach altered with the inclusion of engineering science, described as 'theoretical analysis regardless of whether it could be applied' (Walker, 1989, p. 107) and a more mathematics-based approach to engineering that became more widely accepted after World War II.[8] However, the pendulum swung too far. Eric Walker commissioned a review of engineering education during his tenure as ASEE president in 1965. This report found that engineering science had, in fact, become the main focus of engineering educators. Recognizing the dangers of this shift, Walker argued that it was dangerous for engineers to become 'enamored by research for its own sake' and that a balance between theory and application was necessary (Walker, 1989, p. 106).

With the changing nature of engineering, shaped by many forces, including the needs of war, economic growth, and population expansion, engineering education has responded and continues to adapt and change to best serve the needs of contemporary society. Because of its inherent empirical focus, a growing awareness of the plight of the impoverished, under-represented, and marginalized sections of society has led the engineering community to, yet again, think about how to re-invent itself.

The case of engineering responding to pressing global and societal issues is but one illustration of how what were previously simple, well-understood, and neat problems metamorphose into problems that have no easy or even right answers. It represents yet another step in the evolution of the role of the engineer and the meaning of the engineering profession. Simon responds to this in his description of the opportunity for research that professional schools offer to researchers in pursuit of basic research. He says, 'he will be confronted with the problems of end use, arising from the environment of business, that he can transform into exciting, non-routine problems of fundamental research.' (Simon, 1967, p. 10).

To work with end users and their life circumstances as the source of a research inquiry requires sensitivity to the ethics of engagement and the nature of implementation of the solution. Flexner was optimistic in thinking that 'Civilized men will resolutely refuse the ill and embrace the good and, unless the world can be governed by the ideals of civilization, nothing can save it from ultimate destruction.' (Flexner, 1915, p. 245). Thinking through the end-use or implications of a solution is not necessarily a shackling of curiosity with utilitarian intent.

One of the most powerful of modern forces that challenges both the intellectual and the institutional structure of professional education is that of globalization; at the intellectual level, it invites us to extend the

Simon/Flexner interdisciplinary vision to include more cross-cultural awareness and research – for the way in which culture influences decision making and behavior is profound, thus influencing, for instance, the way organizations behave in different national settings (Crozier, 1967).

Closing

> The sometimes explicit premise that utility is the only touchstone of relevance for knowledge in the professional school, and the sometimes implicit premise that inutility is the only touchstone of relevance in the disciplines are mischievous doctrines that have caused untold harm to education in both professions and disciplines. (Simon, 1967, p. 4)

Herb Simon saw the (re)design of business schools and management education as an organizational issue; building an education on the foundation of scholarly disciplines (though in an interdisciplinary way), with some connections to practice involved realizing some fundamental tensions.

Professional schools face pressure towards favoring either disciplinary specialization or practical relevance, not both, and the forces are often strong and with real dangers of falling into competency traps of either direction. March and Sutton articulated it this way: '[Business schools] live in two worlds. The first demands and rewards speculations about how to improve performance. The second demands and rewards adherence to rigorous standards of scholarship' (March and Sutton, 1997, p. 698).

Different schools have approached the tensions differently (Augier and March, 2011, chapters 6–7) and will probably continue to do so, although institutional pressures, such as rankings, provide forces that favor tendencies towards making schools more similar.

Curiosity and the Importance of Questions

Simon refers to himself as 'the cat curiosity couldn't kill'; more than a cute metaphor, that captures how he was led by an unending curiosity about research questions – often empirical in nature – which led to a long and productive career and research program. Good research questions can help on several fronts: from advancing fundamental knowledge on an important issue; to integrating and building on prior research in new ways; and to helping contribute to the evolution of

the sciences. Scholars such as Simon are driven by curious minds and attracted to paradoxes, dilemmas, contradictions in theories, ideas, and in practice – in ways that often lead to new research areas and sometimes entirely new theoretical paradigms. It also, for Simon, represented a way to integrate fundamental knowledge with relevance; or going beyond it, combining knowledge of different disciplines with insights into the dynamics of the real world, useful (perhaps) for an empirical problem.

Simon's own career exemplifies how interest in practical problems allowed his own scholarship to be interdisciplinary, yet disciplined; and (for the most part) empirically valid. An early example is a paper he wrote in 1935 for a graduate class on the management of Milwaukee recreational facilities. Organized around the study of problems in public administration fueled by the growth of municipal recreation facilities, there was a need to study the administration of part and school activities for possible problems of the relations between the city governments and school boards (Simon, 1935, p. 2). Simon often uses this example as the first insight into bounded rationality since he discovered here that neoclassical ideas of utility maximization didn't match the actual budget allocation processes he was studying. The real process, he found, involved issues such as: governance structure; politics in the budget allocation process; and legal issues. Things that a utility maximization perspective on the budget allocation process would not allow. Thus he found: 'My training in economics, evoked in the context of a budget situation, disclosed a contradiction between what theory taught me ought to be happening and what my eyes and ears showed me was actually happening' (Simon, 1991, p. 371).

This illustrates how an empirical curiosity can initiate a research program that becomes quite wide ranging – and it also emphasizes the importance of pursuing the questions even if the existing theories are not equipped to handle them well. Simon's early ideas on bounded rationality weren't well received at first; but eventually became ground breaking for both intellectual development in disciplines, as well as providing an important backbone for interdisciplinary and cross disciplinary sub-fields, such as organization theory and strategic management – both of these central to the development at the institutional level of the emergence of research based business schools which Simon (1967) discussed.

For the importance of being driven by questions – empirical anomalies that facilitate two-way learning between practice and science – wasn't confined to the intellectual developments Simon worked on. He was – as

the introductory quote indicates – very involved in the revolution in management education and changes in business schools.

Notes

1. Simon preferred to see his legacy and intellectual footprint as integrating, not jumping through, the fields and different disciplines.
2. Simon did also work on larger institutional issues, for instance, being involved in initiatives leading to places such as the Center for Advanced Studies in Behavioral Social Science (CASBSS), in particular through his involvement with the Ford Foundation. At the even larger institutional and science policy level, he was central in the National Academy of Sciences; both in creating room for the social and behavioral sciences there, and on committee work throughout the years.
3. People may disagree about the extent to which the changes they brought about were evolutionary or revolutionary, permanent or temporary.
4. Lee Bach, Simon's dean at Carnegie and partner in setting up the business school at Carnegie, was also involved in some of the Ford Foundation work. So it was no coincidence that Bach himself found that two of the most important 'pillars' intellectually for business schools were economics, and the behavioral social sciences (Bach, 1958).
5. There was in fact a parallel report commissioned by the Carnegie Corporation led by Pierson (Pierson, 1959). Although the report and the structure of the writing were a bit different, the report shared the overall conclusions.
6. As Simon also noted: 'Business school does not stand a chance of recruiting first rate scientists if it insists that all research done within its walls must have direct relevance to business. It will do better to demonstrate its respect for fundamental research by having, and valuing, in its faculty at least some members whose work does not have obvious relevance to business, but does command high respect in its discipline' (Simon, 1967, p. 10).
7. Flexner also had a discussion of what a professon is (1915); interestingly, the criteria he lists do not seem to easily include business as a profession.
8. As noted by Emmerson, this was a remarkable shift from the past when an 1874 presentation to the ASCE argued for the elimination of calculus from the engineering curriculum saying it played no role in the capacities for observation, practical judgment, and the ability to scale, that were the hallmarks of engineering (Emmerson, 1973, p. 262).

References

Augier, M. and J. G. March (2007). The Pursuit of Relevance in Management Education. *California Management Review,* 49 (3), 129–46.

Augier, M. and J. G. March (2011). *The Roots, Rituals, and Rhetorics of Change: North American Business Schools After the Second World War*. Stanford: Stanford University Press.

Bach, L. (1958). Some Observations on the Business School of Tomorrow. *Management Science,* 4, 351–64.

Bossard, J. and J. Dewhurst (1931). *University Education for Business*. Philadelphia: University of Philadelphia Press.
Crozier, M. (2009). *The Bureaucratic Phenomenon* (Vol. 280). New York: Transaction Publishers.
Emmerson, G. S. (1973). *Engineering Education: A Social History*. Newton Abbott: David & Charles
Flexner, A. (1910). *Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching*.
Flexner, A. (1915). *Is Social Work a Profession?* Proceedings of the National Conference of Charities and Corrections.
Gordon, R. and J. Howell (1959). *Higher Education for Business*. New York: Columbia University Press.
Grayson, L. P. (1977). A Brief History of Engineering Education in the United States. *Engineering Education,* 68 (3), 246–64.
Grinter, L. E. (1955). Report on Evaluation of Engineering Education. *Journal of Engineering Education,* 46 (1), 25–63.
Jeuck, J. E. (1973). Business Education: Some Popular Models. *The Library Quarterly*, 43(4): 283–92.
Kuhn, T. S. (2012). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press (first published 1970).
Layton Jr., E. T. (1971). *The Revolt of the Engineers. Social Responsibility and the American Engineering Profession*. Cleveland: Case Western Reserve University Press.
March, J. G. and Sutton, R. I. (1997). Organizational Performance as a Dependent Variable. *Organization Science*, 8 (6), 698–706.
McGivern, J. G. (1960). *First Hundred Years of Engineering Education in the United States (1807–1907)*. Washington, DC: Gonzaga University Press.
Petroski, H. (2001). The Importance of Engineering History. *International Engineering History and Heritage,* 1–7.
Pierson, F. C. (1959). *The Education of American Businessmen: A Study of University-college Programs in Business Administration*. New York: McGraw-Hill.
Seely, B. E. (1999). The Other Re-engineering of Engineering Education, 1900–1965. *Journal of Engineering Education,* 88 (3), 285–94.
Simon, H. A. (1935). Administration of Public Recreational Facilities in Milwaukee. Unpublished manuscript. Herbert Simon Archives.
Simon, H. A. (1967). The Business School: A Problem of Organizational Design. *Journal of Management Studies,* 4 (1), 1–17.
Simon, H. A. (1991). *Models of my Life*. Boston, MA: MIT Press.
Walker, E. A. (1971). The Major Problems Facing Engineering Education. *Proceedings of the IEEE,* 59 (6), 823–8.
Walker, E. A. (1989). *Now It's My Turn: Engineering My Way*. Vantage Press.
Wickenden, W. E. (1929). *Report of the Investigation of Engineering Education, 1923–1929, Volumes I and II*, Society for the Promotion of Engineering Education, Pittsburgh.

Name Index

Notes: bold = extended discussion or term highlighted in text; f = figure, n = endnote or footnote, t = table.

Subject Index

Notes: bold = extended discussion or term highlighted in text; f = figure, n = endnote or footnote, t = table.