EVOlUTION OF THE SOCIAL CONTRACT
Evolution of the Social Contract

BRIAN SKYRMS
University of California
Irvine
For Pauline, Michael, and Gabriel
Two men who pull the oars of a boat, do it by an agreement or convention, tho’ they have never given promises to each other. Nor is the rule concerning the stability of possession the less derived from human conventions, that it arises gradually, and acquires force by a slow progression . . . In like manner are languages establish’d by human conventions without any promise.

David Hume, *A Treatise of Human Nature*
# CONTENTS

**Preface**  
*page ix*

**Acknowledgments**  
*xiii*

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sex and Justice</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Commitment</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>Mutual Aid</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>Correlated Convention</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>The Evolution of Meaning</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Postscript</td>
<td>105</td>
</tr>
</tbody>
</table>

**Notes**  
*111*

**References**  
*129*

**Index**  
*143*
PREFACE

The best-known tradition approaches the social contract in terms of rational decision. It asks what sort of contract rational decision makers would agree to in a preexisting "state of nature." This is the tradition of Thomas Hobbes and – in our own time – of John Harsanyi and John Rawls. There is another tradition – exemplified by David Hume and Jean Jacques Rousseau – which asks different questions. How can the existing implicit social contract have evolved? How may it continue to evolve? This book is intended as a contribution to the second tradition.

Hegel and Marx are, in a way, on the periphery of the second tradition. Lacking any real evolutionary dynamics, they resorted to the fantasy of the dialectical logic of history. It was Darwin who recognized that the natural dynamics of evolution is based on differential reproduction. Something like differential reproduction operates on the level of cultural as well as biological evolution. Successful strategies are communicated and imitated more often than unsuccessful ones. In the apt language of Richard Dawkins, we may say that both cultural and biological evolution are processes driven by differential replication. There is a simple dynamical model of differential replication now commonly called the replicator dynamics. Although this dynamics is surely oversimplified from both biological and cultural perspectives, it provides a tracta-
Preface

A reasonable model that captures the main qualitative features of differential replication. The model can be generalized to take account of mutation and recombination. These biological concepts also have qualitative analogues in the realm of cultural evolution. Mutation corresponds to spontaneous trial of new behaviors. Recombination of complex thoughts and strategies is a source of novelty in culture. Using these tools of evolutionary dynamics, we can now study aspects of the social contract from a fresh perspective.

Some might argue that, in the end, both traditions should reach the same conclusion because natural selection will weed out irrationality. This argument is not quite right, and one way of reading the book is to concentrate on how it is not right. Chapter 1 juxtaposes the biological evolution of the sex ratio with cultural evolution of distributive justice. It shows how evolution imposes a “Darwinian veil of ignorance” that often (but not always) leads to selection of fair division in a simple bargaining game. In contrast, rational decision theory leads to an infinite number of equilibria in informed rational self-interest. Chapter 2 shows that evolution may not eliminate behavior that punishes unfair offers at some cost to the punisher. Such strategies can survive even though they are “weakly dominated” by alternatives that could do better and could not do worse. Chapter 3 widens the gap between rational decision and evolution. If evolutionary game theory is generalized to allow for correlation of encounters between players and like-minded players, then strongly dominated strategies – at variance with both rational decision and game theory – can take over the population. Correlation implements a “Darwinian categorical imperative” that provides a general unifying account of the conditions for the evolution of altruism and mutual aid. Chapter 4 deals in general with situations in which rational choice cannot decide between symmetric optimal options. Evolutionary dynamics can break
Preface

the “curse of symmetry” and lead to the formation of correlated conventions. The genesis of “ownership” behavior and thus the rudiments of the formation of the concept of property are a case in point. Chapter 5 shows how meaning is spontaneously attached to tokens in a signaling game. Here rational choice theory allows “babbling equilibria” where tokens do not acquire meaning, but consideration of the evolutionary dynamics shows that the evolution of meaning is almost inevitable. Throughout a range of problems associated with the social contract, the shift from the perspective of rational choice theory to that of evolutionary dynamics makes a radical difference. In many cases, anomalies are explained and supposed paradoxes disappear.

The two traditions, then, do not come to the same conclusions. There are points of correspondence, but there are also striking differences. In pursuing the tradition of Hume, my aims are explanatory rather than normative. Sometimes, I am happy explaining how something could have evolved. Sometimes I think I can say why something must have evolved, given any plausible evolutionary dynamics. In intermediate cases, we can perhaps say something about the range of initial conditions that would lead to a given result. When I contrast the results of the evolutionary account with those of rational decision theory, I am not criticizing the normative force of the latter. I am just emphasizing the fact that the different questions asked by the two traditions may have different answers.

Although there is real game theory and real dynamics behind the discussions in this book, I have reserved the technical details for scholarly journals. No special background is presupposed. Useful concepts are introduced along the way. I hope and believe that this book should be generally accessible to readers who wish to pursue the fascinating issues of a naturalistic approach to the social contract.

xi
ACKNOWLEDGMENTS

I have many friends to thank for useful comments on some or all of the contents of this book. In particular, I would like to thank Justin d’Arms, Francisco Ayala, Cristina Bicchieri, Ken Binmore, Vincent Crawford, Branden Fitelson, Steven Frank, Alan Gibbard, Clark Glymour, Peter Godfrey-Smith, Bill Harms, Bill Harper, John Harsanyi, Jack Hirshleifer, Richard Jeffrey, Ehud Kalai, Philip Kitcher, Karel Lambert, David Lewis, Barbara Mellers, Jordan Howard Sobel, Patrick Suppes, Peter Vanderschraaf, Bas van Fraassen, and Bill Wimsatt. I owe much to discussions with my late friend and colleague Greg Kavka. The first draft of this book was completed at the Center for Advanced Study in the Behavioral Sciences in 1993–4. I am grateful for financial support provided by the National Science Foundation, the Andrew Mellon Foundation, and the University of California President’s Fellowship in the Humanities. The University of California at Irvine provided a grant of computer time at the San Diego Supercomputer Center, which made some of the larger simulations reported here possible. Substantial portions of “Sex and Justice” and “Darwin Meets ‘The Logic of Decision’ ” are reprinted here with permission of The Journal of Philosophy and Philosophy of Science, respectively.