Altruistically Inclined?

Economics, Cognition, and Society

This series provides a forum for theoretical and empirical investigations of social phenomena. It promotes works that focus on the interactions among cognitive processes, individual behavior, and social outcomes. It is especially open to interdisciplinary books that are genuinely integrative.

Titles in the Series

Ulrich Witt, Editor. *Explaining Process and Change: Approaches to Evolutionary Economics*

Young Back Choi. *Paradigms and Conventions: Uncertainty, Decision Making, and Entrepreneurship*

Geoffrey M. Hodgson. *Economics and Evolution: Bringing Life Back into Economics*

Richard W. England, Editor. *Evolutionary Concepts in Contemporary Economics*

W. Brian Arthur. *Increasing Returns and Path Dependence in the Economy*

Janet Tai Landa. *Trust, Ethnicity, and Identity: Beyond the New Institutional Economics of Ethnic Trading Networks, Contract Law, and Gift-Exchange*

Mark Irving Lichbach. *The Rebel's Dilemma*

Karl-Dieter Opp, Peter Voss, and Christiane Gern. *Origins of a Spontaneous Revolution: East Germany, 1989*

Mark Irving Lichbach. *The Cooperator's Dilemma*

Richard A. Easterlin. *Growth Triumphant: The Twenty-first Century in Historical Perspective*

Daniel B. Klein, Editor. *Reputation: Studies in the Voluntary Elicitation of Good Conduct*

Eirik G. Furubotn and Rudolf Richter. *Institutions and Economic Theory: The Contribution of the New Institutional Economics*

Lee J. Alston, Gary D. Libecap, and Bernardo Mueller. *Titles, Conflict, and Land Use: The Development of Property Rights and Land Reform on the Brazilian Amazon Frontier*

Rosemary L. Hopcroft. *Regions, Institutions, and Agrarian Change in European History*

E. L. Jones. *Growth Recurring: Economic Change in World History*

Julian L. Simon. *The Great Breakthrough and Its Cause*

David George. *Preference Pollution: How Markets Create the Desires We Dislike*

Alexander J. Field. *Altruistically Inclined? The Behavioral Sciences, Evolutionary Theory, and the Origins of Reciprocity*

Altruistically Inclined?

The Behavioral Sciences, Evolutionary Theory, and the Origins of Reciprocity

Alexander J. Field

Ann Arbor

THE UNIVERSITY OF MICHIGAN PRESS

Published in the United States of America by
The University of Michigan Press
Manufactured in the United States of America
♾ Printed on acid-free paper

2004 2003 2002 2001 4 3 2 1

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

Library of Congress Cataloging-in-Publication Data

Field, Alexander J.
Altruistically inclined? : the behavioral sciences, evolutionary theory, and the origins of reciprocity / Alexander J. Field.
p. cm. — (Economics, cognition, and society)
Includes bibliographical references and index.
ISBN 0-472-11224-4 (cloth : alk. paper)
1. Altruism. 2. Genetic psychology. 3. Social groups. I. Title. II. Series.
BF637.H4 F54 2001
155.7—dc21 2001002648

To my son, Jamie, for his love of history,
and my daughter, Emily, for her love of philosophy

Contents

Preface

Two strangers meet far from the reach of organized society. Each must decide quickly whether to attack, or await the action of the other. Together, they are better off choosing restraint and thus opening up possibilities for mutually beneficial intercourse. Desires for both self-protection and possible wealth enhancement, however, impel each of them toward an initial and immediate aggressive move.

The parable of the good Samaritan reminds us that failure to help can be hurtful. We can easily overlook the symmetrical point: since we are vulnerable to injury from all but the weakest, *failure to harm can be helpful.* In holding in check our ability to damage or destroy, we help our counterparty, both because she has avoided injury at our hands and because she now faces opportunities for gains at our expense that would otherwise have been unavailable. And in forgoing potential gains and exposing ourselves to otherwise avoidable risks, we have harmed ourselves. A surprising but inescapable conclusion: failure to harm can be *altruistic.*

Are we altruistically inclined? Are we, in spite of the counsel of prudence and the temptations of greed, often predisposed, in situations such as that described above, to give up the option of making a first aggressive move? If it is in our nature to be so inclined, how can this possibly be, given what we know of the operation of evolutionary forces?

Discussions of human altruism often have a nebulous and ill-defined quality to them. People commonly question what altruistic behavior is and whether it can truly be distinguished from what is selfish. But in a biological context, altruism has a very precise meaning: behavior by an individual organism that reduces its own reproductive fitness while improving the reproductive fitness of at least one other member of the same species (conspecific). Reproductive fitness affects the relative frequency with which an individual's genes appear in the next generation's gene pool.

Like Robert Frank's book *Passions within Reason* (1988), this book takes as a starting point the proposition that altruistic behavior is an important empirical category. Like Frank's work, this book explores evolutionary explanations of this phenomenon. But unlike Frank, this book considers the possibility that natural selection—the fundamental motor of

evolutionary dynamics—has operated at the group as well as the individual level. Group selection occurs when selection differentially rewards members of a group as a consequence of the frequency of some trait within it, for example, when groups with higher frequencies of those altruistically predisposed grow more rapidly.

Group selection is not a new idea, but has only slowly been reemerging from an intellectual doghouse. The evolutionary models that most people carry around in their heads start with the premise that natural selection operates exclusively at the level of the individual organism. This poses a fundamental problem for the explanation of altruistic behavior, since by definition, altruism cannot be favored if selection operates only at the individual level.

Much of the history of the social and biological sciences since the 1960s has involved attempts to resolve this apparent contradiction. Considerable progress has been made in understanding altruistic behavior toward kin: for example, the sacrifices that parents make for their children. The theory of inclusive fitness, pioneered by the late William Hamilton, emphasizes that selection occurs ultimately at the level of the gene and, since parents share half their genes with each of their children, sacrifice for offspring may favor genes predisposing to such behavior, even if the sacrifice is not in the material interest of the parent.

The explanation of altruistic behavior toward non-kin is more difficult. The degree of genetic relatedness drops off quickly (second cousins share only 1/32 of their genes). Since altruistic behavior favors the fitness of other conspecific(s) at the expense of the actor, it is hard to see how predispositions to behave altruistically toward non-kin could spread or even survive. Were they to arise through mutation or genetic recombination, such tendencies would seem inevitably to decline in frequency and eventually disappear over time through the operation of natural selection.

If group selection processes are operative, however, it is possible, within a population periodically dividing into smaller groups, for behavioral predispositions to be shrinking in frequency within each individual group, while they are increasing in frequency within the global population. This possibility, admittedly counterintuitive, arises when there is a positive covariance between the frequency of altruists within a group and the rate at which it grows. Thus while altruistic behavior will engender reduced reproductive fitness for the organism exhibiting it within each group, genes predisposing to it can be increasing over time within the global population. The possibility enables us to understand how altruistic tendencies could be favored by evolutionary processes even when they are, by definition, disadvantaged within each individual group.

Most social scientists admit the relevance of altruism in considering

relations among kin. But in relations among non-kin, interest seems to reign supreme, and suggestions that altruistic predispositions have a role to play are, if not rejected, then greeted with considerable skepticism. This presumption persists in the face of considerable evidence, experimental and observational, inconsistent with it. Part of the explanation for this is that we have tended to focus on what sustains or maintains ongoing interaction, as opposed to what allows it to originate.

Altruism may not be necessary to sustain relations of reciprocity. But altruism is necessary for them to originate. The description of a contingently cooperative strategy (Tit-for-Tat, for example) will be formally identical in an environment in which it or similar strategies prevail at low frequency and an environment in which it prevails at high frequency. But whether or not such a strategy is altruistic in an evolutionary sense depends on the frequency of such tendencies and others within the general population. In this respect, inclinations toward such strategies differ from the predispositions of parents to sacrifice for their children, which are altruistic irrespective of the prevalence of such tendencies among others.

My interest in what holds human groups together began with dissertation research more than a quarter of a century ago (Field 1974). That work gave little attention to the possible contribution of evolutionarily determined inclinations in allowing reciprocal, cooperative relations to develop. It seemed pretty obvious that natural selection, by favoring those who helped themselves, meant that Darwin was a problem to be overcome, not part of the solution.

The intervening years led to revisiting the question periodically, exploring and elaborating on the role of institutions and norms in influencing behavior (Field 1981, 1984, 1991). As was true for my dissertation, and in line with conventional social science thinking, none of these articles considers genetically or biologically mediated influences on our abilities to initiate and sustain social and, ultimately, economic intercourse.

A change in perspective was precipitated by a year-long sabbatical at the Social Science History Institute at Stanford University in 1997–98. The break afforded me an opportunity to read broadly and without distraction in a number of areas, some familiar and some entirely new. The process caused a number of inchoate ideas to develop and coalesce in directions not entirely anticipated.

I emerged with a reevaluation of how we can effectively tackle this problem, one whose logic and evidentiary foundation I think important for social and behavioral scientists to seriously consider. It is now far clearer than it was in the 1970s that the natural sciences do more than simply define a problem that the social sciences must resolve. A more nuanced Darwinian approach can enable us to organize and interpret the results of

experimental and other research in ways that facilitate understanding of biological influences both on universal human behavioral propensities and on the structure of our cognitive faculties whereby we acquire knowledge about the world. These in turn can help us understand the emergence of normative structures without which the origins of reciprocity and complex social organization would be impossible.

Evolutionary, biological reasoning has been frequently misused in the past, sometimes in horrific ways, and many readers will approach it with reservation. It is important to enumerate several factors that argue in favor of our being more receptive to it. First, research and, especially, theorizing in this area are, in general, more nuanced and somewhat less prone to overreaching than was the case twenty-five years ago. In particular, there is now more emphasis on understanding genetic influences on human cognitive structure and behavior as reflecting adaptation to the relatively stable ancestral environment of hunter-gatherer existence (a period of at least two million years duration), as well as earlier, and less emphasis on attempts to interpret behavior subsequent to the Neolithic revolution (a period of ten or eleven thousand years at most) as necessarily reflecting adaptation to encountered environments. Second, the understanding of and scientific consensus about the levels at which natural selection can and does operate have been refined as the result of observational, experimental, and theoretical research, as have been assumptions about the interrelationships and balance between "innate" and learned cognitive and behavioral mechanisms. Third, the fruitfulness of inquiry into biological influences on human behavior and cognition has been steadily reinforced by an accretion of observational and experimental data and of new ways of interpreting such data.

Overall progress in these areas over a quarter century is striking in comparison with what one observes in the social and behavioral sciences, and suggests that research along these lines, and perhaps along these lines alone, offers the possibility of transcending the most significant and persisting fault line within them. That divide separates the sociological-anthropological tradition, with its emphasis on culture, norms, institutions, ideology, and emergent properties, from the economic approach, with its assumption of rational choice, and ambivalence toward or outright rejection of all of these concepts.

Research by heterodox scholars has tried to bridge this gap. But many *on both sides of it* remain skeptical that these efforts can lead to a scientifically progressive research agenda. This book is intended for those who, like me, have thought hard about these issues and have often been stymied. Many of us are committed in our work to approaches with explanatory deficiencies that at some level we acknowledge. The argument

and analysis should be of interest to those identifying with either the rational choice or the sociological tradition. But it tries to move beyond the ultimately unproductive opposition of one to the other. Rethinking the implications of evolutionary theory and processes, and in particular relaxing the assumption that natural selection operates only at the level of the individual organism, leads to a rethinking of the strengths and limitations of each. In the context of serious consideration of experimental and observational evidence, it lays groundwork not only for some rapprochement within the social sciences but also, more generally, between the biological and behavioral sciences. This integration, however, entails a different set of implications than those traditionally drawn by advocates of such unification.

Some background in game theory is helpful in understanding the arguments developed here. This is not because game theory, any more than the rational choice approach of which it represents an extension, provides a universal key to understanding human behavior. But in recent years it has become almost impossible to discuss or engage developments in social science, and, increasingly, biological science, without employing its idiom. The main use of game theory in this book is as a means of organizing our thinking about what would be likely outcomes if interacting individuals were strictly self-interested and/or if natural selection operated only at the level of the individual organism.

Since the main focus is on areas where game theory doesn't predict well, it would not be fair to say that the emphasis here is principally on the application of game theory to the social or biological sciences. Those interested in work with more emphasis along these lines, which also treats the experimental literature, should consult Anivash Dixit and Susan Skeath's *Games of Strategy* (1999), Herbert Gintis's *Game Theory Evolving* (2000), or a number of other recent texts.

This book is more wide ranging in scope, more focused on the implications of evolutionary approaches, broadly conceived, for our understanding of essential human predispositions. In exploring the cognitive underpinnings of these tendencies, I also emphasize what has come to be called modularity. Modularity refers to cognitive adaptations, which employ different neurobiological machinery, use different reasoning algorithms, and may lead to different behavioral outcomes depending upon the domain encountered.

As a result of millions of years of selection, humans possess powerful reasoning modules that facilitate foraging and its modern equivalents. These include facility at Bayesian learning—necessary for forming "rational" expectations—as well as competence at, for example, maximizing goals such as caloric yield in allocating time among alternative activities.

The mathematics of constrained maximization, central to economic theory, provides a useful metaphor for modeling the operation of such modules. But in the realm of strategic interaction, as the experimental and observational evidence makes clear, humans possess other algorithms and action inclinations that are equally and in some cases more important in influencing behavior.

The idea of cognitive and behavioral modularity helps explicate a variety of otherwise anomalous observations. But, like group selection, it is not one that has been widely considered within the social sciences. As a consequence, either is likely to be embraced only after the most careful consideration. This book recognizes the appeal of the familiar, and that we may be drawn to certain explanatory frameworks because of their expressive qualities or their aesthetic appeal, rather than simply their explanatory or predictive power. But it is written under the assumption that evidence and argument ultimately matter, and that the vast majority of scholars in our disciplines are interested in these issues and committed to traditional scientific methods in addressing them.

The arguments here vary in complexity, but many are quite subtle. It is easy to be glib when discussing such matters as essential human predispositions, and when there has been a choice I have erred on the side of providing documentation and seeking clarity of exposition. This makes for a lengthier volume but one I hope will ultimately have more impact. The ideas, models, and analyses explored represent serious attempts at understanding fundamental social phenomena. To analyze them deeply, even if at times critically, is to acknowledge the serious efforts made by scholars to understand these problems.

In striking a balance between being too elliptical and making certain my meaning is understood, some redundancies have crept in. For those readers who get my drift immediately, some tolerance is sought for those less well versed in the technicalities. For those who do not, close reading will perhaps make an argument become clear in a way it had not been before.

Readers approaching this study with a jaundiced view of economics or rational choice theory may question much in the first part of the book as belaboring the obvious. In defense, I can only say that the appeal of the methods associated with this tradition remains very strong in modern behavioral and social science, and that those employing them have a justified sense that they explore the implications of some very powerful human predispositions. Only by carefully delineating the restricted applicability of the underlying models can one hope to make headway in articulating the case for alternative and complementary approaches.

Regardless of one's starting point, it is almost inevitable that some aspects of this book will challenge firmly held, perhaps unexamined,

assumptions. It is helpful to keep in mind that what is obvious and easily accepted by one person may represent a challenge to a deeply entrenched presumption or point of view to another. One of the objectives of this work has been to soften some of the divergences in perspective that define disciplinary divides, and this can only be done by probing fundamentals. The research and writing of this book have been extraordinarily enriching for me and advanced my thinking significantly. I hope readers will share in and benefit from those experiences.

When contemplating a work attempting to cover a broad range of material, one often dips into sections where the knowledge base is strong. If details and analysis seem on target, we read further. The book aims to survive this test from a range of entry points. It covers multiple literatures, and considerable effort has been made to get the details right in each of them. For those whose ideas or arguments are inadvertently misinterpreted, my apologies in advance.

Following a prologue, the first chapter develops the main issues and evidence that occupy the study. Chapter 2 discusses the logic and mathematics of group selection models. Chapters 3 and 4 consider other explorations of altruistic behavior, including work by Robert Trivers, John Maynard-Smith, Robert Axelrod, and Robert Frank. Chapter 5 covers arguments and evidence underlying the concept of modularity. Chapter 6 addresses the heuristics and biases research program and its more limited relevance to the issues addressed here. Chapter 7 considers heterodox utility functions and differences between historical and social scientific explanations, and looks to a more integrated future.

My intent in these chapters is to probe, enlighten, and ultimately persuade. My hope is not that readers will, in light of what is written here, immediately abandon their current lines of inquiry, but rather that they will emerge with a better sense of where efforts fit within a larger scientific enterprise.

I owe debts of gratitude to a wide range of individuals who have read and commented on part or all of this work in manuscript. Particular thanks are due to Paul David; Mark Field; Deborah Garvey; Herbert Gintis; Avner Greif; John Heineke; Jack Hirshleifer; Larry Iannaccone; Terence Kealey; Robert Keohane; Michael Kevane; Daniel Klein; Timur Kuran; Deirdre McCloskey; Ross Miller; Douglass North; Robert Numan; Mel Reder; Tom Russell; Bill Sundstrom; Rick Szostak; Gavin Wright; and participants in seminars at Stanford University, Santa Clara University, the University of California at Berkeley, the University of California at Davis, and the August 2000 Knexus symposium at Stanford.

My wife, Valerie, approaching the work as an intelligent layperson, asked the toughest questions, insisting I define terms I took for common

knowledge, or clarify exposition I thought was clear. If the book is more accessible to a wider audience, she deserves much of the credit. My greatest acknowledgment is to her and my children, for tolerating over an extended period what surely seemed to them abstruse preoccupations. I hope, in the end, they will understand what has concerned me.