

UNTO

OTHERS

THE

EVOLUTION

AND

PSYCHOLOGY

OF

UNSELFISH

BEHAVIOR

ELLIOTT SOBER

AND

DAVID SLOAN WILSON



Unto Others



*The Evolution and Psychology of
Unselfish Behavior*

Elliott Sober

David Sloan Wilson

Harvard University Press
Cambridge, Massachusetts
London, England
1998

Copyright © 1998 by the President and Fellows of Harvard College

All rights reserved

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Sober, Elliott.

Unto others : the evolution and psychology of unselfish behavior /
Elliott Sober, David Sloan Wilson.

p. cm.

Includes bibliographical references and index.

ISBN 0-674-93046-0 (hardcover)

1. Altruism. 2. Helping behavior. I. Wilson, David Sloan. II. Title.

BF637.H4S65 1998

171'.8—dc21 97-42185

Unto Others

*To altruists everywhere, especially those who are unsure as
to what their motives really are*

Acknowledgments

Maybe it should come as no surprise that we have received so much help in writing a book about altruism. We want to thank the following battalion of people. They come from many different disciplines and helped in many different ways: Richard Alexander, André Ariew, Martin Barrett, Daniel Batson, John Beatty, Len Berkowitz, Howard Bloom, Chris Bohm, Tom Bontly, Deric Bownds, Robert Boyd, Robert Brandon, Felix Breden, Don Campbell, Noël Carroll, Anne Clark, Robert Colwell, Richie Davidson, Jeff Dean, Lee Dugatkin, Nancy Eisenberg, Berent Enç, Murvet Enç, Ted Everett, Robert Frank, Steve Frank, Peter Godfrey-Smith (for the title), Charles Goodnight, Leslie Graves, Jim Griesemer, David Gubernick, William Hamilton, Henry Harpending, Dan Hausman, Jack Hirschleifer, Harmon Holcomb III, Todd Hughes, Robert Jeanne, John Kelly, Bruce Knauft, Hugh LaFollette, Andrew Levine, John Maynard Smith, Rick Michod, Don Moskowitz, Gregory Mougin, Steven Orzack, Stuart Peck, Jane Piliavin, Greg Pollock, Will Provine, William Puka, David Queller, H. Kern Reeve, Peter Richerson, Steve Rissing, Tom Seeley, Larry Shapiro, Alan Sidelle, Barbara Smuts, Dennis Stampe, Chris Stephens, Kim Sterelny, William Talbott, Peter Taylor, Frans de Waal, Mike Wade, Denis Walsh, Doris Williams, George Williams, Edward Wilson, Vero Wynne-Edwards.

We also must extend our thanks to entities that are not individuals; perhaps this also is to be expected in a book about multilevel selection theory. We are grateful for support from Binghamton University, the National Science Foundation, and the University of Wisconsin at Madison. Finally, we want to acknowledge the support and encouragement that have come from two groups that we count as nearest and most dear—our families.

Contents

Introduction: Bentham's Corpse	1
<i>Part I. Evolutionary Altruism</i>	15
1. Altruism as a Biological Concept	17
2. A Unified Evolutionary Theory of Social Behavior	55
3. Adaptation and Multilevel Selection	101
4. Group Selection and Human Behavior	132
5. Human Groups as Adaptive Units	159
<i>Part II. Psychological Altruism</i>	197
6. Motives as Proximate Mechanisms	199
7. Three Theories of Motivation	223
8. Psychological Evidence	251
9. Philosophical Arguments	275
10. The Evolution of Psychological Altruism	296
Conclusion: Pluralism	329
Notes	339
References	363
Index	387

Introduction: Bentham's Corpse

The mummified corpse of Jeremy Bentham occupies a cabinet the size of a telephone booth in University College, London. The head went cheesy some time ago and was replaced by a wax substitute; the original is in a box that sits discreetly between Bentham's feet. Bentham's will decreed that his body was to be preserved and carried into meetings of the University's Board of Trustees (Runes 1959, p. 250). Bentham thought that his stern countenance would inspire future generations to live up to the standards that he and John Stuart Mill advocated in their theory of morality and politics. Bentham and Mill together created the philosophy of *utilitarianism*—the view that people should promote the greatest happiness for the greatest number of individuals. Bentham hoped that a mummy at a meeting would encourage those voting to do the right thing.

It is a matter of conjecture just how much influence Bentham's corpse has exercised over University College in the years since his death, but it is a certainty that the ideas Bentham defended have profoundly affected the wider culture. Bentham thought that all human activity should aim to maximize pleasure and minimize pain. Underlying this claim about what people *ought* to do was Bentham's picture of how the human mind works *in fact*. According to Bentham (1789), "nature has placed mankind under the governance of two sovereign masters, pain and pleasure." Bentham, here, is endorsing

psychological hedonism—the theory that avoiding pain and attaining pleasure are the only ultimate motives that people have; everything else that we want, we want solely as a means to achieving those twin ends.

Hedonism is a specific version of a more general theory, *psychological egoism*, which claims that every individual's ultimate goal is to benefit him- or herself. Egoism maintains that when we care about what happens to others, we do so only as a means to increasing our own welfare. The view denies that people ever have altruistic ultimate motives. Egoism does not say whether we should rejoice or despair at this feature of the human mind. It claims only to describe how things are in fact.¹ In all our social interactions, we are driven by a single question—“What's in it for me?”

It would be difficult to exaggerate the pervasive influence that hedonism and egoism have had and continue to have on people's thinking. For many, egoism seems obvious, a matter of common sense. People are often unsurprised when others act with ruthless selfishness but find it quite remarkable when others sacrifice themselves for the sake of someone else. If someone says that human beings are by nature selfish, people frequently regard this pronouncement as proceeding from a clear-eyed realism; however, if someone says that human beings are by nature benevolent, people often smile indulgently, thinking that the assertion reflects a propensity to view the world through rose-colored glasses.

Why does psychological egoism have such a grip on our self-conception? Does our everyday experience provide conclusive evidence that it is true? Has the science of psychology demonstrated that egoism is correct? Has philosophy? All these questions must be answered in the negative, or so we will argue. The influence that psychological egoism exerts far outreaches the evidence that has been mustered on its behalf.

Egoism is easy enough to refute when it is given a simplistic formulation. For example, if the egoist claims that the only ultimate goal that people have is to maximize their access to consumer goods, it is not hard to describe behaviors that show that this is false. But if the egoist says that human beings strive for internal, psychological benefits, the proposal is harder to prove wrong. When people sacrifice their own interests to help someone else, the egoist maintains that they

do so in order to feel good about themselves and to avoid feeling guilty. Egoism is a mansion with many rooms. There seems to be room enough in the theory to explain helping behavior and the existence of desires concerning the welfare of others; both are explained as instruments for promoting self-interest.² As a result, the concept of altruism remains an endangered species (Campbell 1994).

Psychological egoism is hard to disprove, but it also is hard to prove. Even if a purely selfish explanation can be imagined for every act of helping, this doesn't mean that egoism is correct. After all, human behavior also is consistent with the contrary hypothesis—that some of our ultimate goals are altruistic. Psychologists have been working on this problem for decades and philosophers for centuries. The result, we believe, is an impasse—the problem of psychological egoism and altruism remains unsolved. A new approach is needed. The novel perspective that we will explore is provided by the theory of evolution.

Bentham died in 1832, more than two decades before the publication in 1859 of Darwin's *Origin of Species*. John Stuart Mill lived well past this watershed event, though he never accepted the theory of evolution by natural selection.³ Darwin's theory gave rise to a fundamental puzzle about the behavior of organisms. The basic idea of natural selection is that characteristics evolve because they help the individuals who possess them to survive and reproduce. A herd of zebra, for example, will gradually increase in running speed because faster zebras do a better job evading predators. Faster zebras are *fitter*—they are more able to survive and tend to have more offspring than slower ones. If offspring resemble their parents, the frequency of fast zebras—the proportion of good runners in the herd—will increase. Notice that a zebra that runs fast benefits itself—not other zebras, not lions, not the whole ecosystem. In this example, natural selection favors those who help themselves. It therefore appears that helping *other* individuals to survive and reproduce at the expense of one's *own* survival and reproduction is the very thing that natural selection will eliminate. In short, natural selection appears to be a process that promotes selfishness and stamps out altruism.

Darwin was aware that organisms in nature sometime behave in ways that appear altruistic. For example, a honeybee sacrifices its life

for the colony when it uses its barbed stinger to attack intruders to the nest. And many of the most praiseworthy human qualities—honesty, charity, trust, and heroism—appear to benefit others at expense to self. Darwin explained these characteristics by saying that natural selection sometimes acts on *groups*, just as it acts at other times on *individuals*. An altruist may have fewer offspring than a nonaltruist within its own group, but groups of altruists will have more offspring than groups of nonaltruists. In a famous passage from *The Descent of Man*, Darwin used the principle of group selection to explain the evolution of human morality:

It must not be forgotten that although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe, yet that an increase in the number of well-endowed men and advancement in the standard of morality will certainly give an immense advantage to one tribe over another. There can be no doubt that a tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to aid one another, and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection. At all times throughout the world tribes have supplanted other tribes; and as morality is one important element in their success, the standard of morality and the number of well-endowed men will thus everywhere tend to rise and increase. (Darwin 1871, p. 166)

Although Darwin never discussed how important a role group selection played in the history of life, his practice was to appeal to this process only rarely. Darwin's successors were less abstemious, invoking the process widely and often uncritically. According to Allee (1951), dominance hierarchies exist to minimize within-group conflict, so that the entire group can be more productive. According to Wynne-Edwards (1962), individual organisms restrain themselves from consuming food and from reproducing, so that the population can avoid crashing to extinction. And according to Dobzhansky (1937), whole species maintain genetic diversity to cope with new environmental challenges; like savvy investors, they diversify their portfolios because the future is uncertain. Many biologists happily invoked these and other group-level explanations while at the same

time explaining the evolution of camouflage, disease resistance, and other traits as individual-level adaptations. Biologists often simply chose the level of explanation that they found more intuitive, appealing to individual adaptation on Mondays, Wednesdays, and Fridays, and to group adaptation on Tuesdays, Thursdays, and Saturdays. Theorizing was unconstrained, and adaptationist explanation was similarly unrestrained.

All this changed in the 1960s, when group selection was attacked by a number of biologists. The most thorough and devastating critique was G. C. Williams's 1966 book, *Adaptation and Natural Selection*. Williams touched a nerve, and his vigorous rejection of adaptations that exist for the good of the group spread quickly through the community of evolutionary biologists. For the next decade, group selection theory was widely regarded as not just false but as off-limits, as far as serious evolutionary thought was concerned. At best, group adaptation was regarded as a theoretical possibility but as so enormously unlikely that alternative explanations should be preferred whenever possible. The following passage (Ghiselin 1974, p. 247) illustrates the fervor with which altruistic, group-level explanations were rejected in favor of accounts that appeal to selfishness:

The economy of nature is competitive from beginning to end . . . The impulses that lead one animal to sacrifice himself for another turn out to have their ultimate rationale in gaining advantage over a third . . . Where it is in his own interest, every organism may reasonably be expected to aid his fellows . . . Yet given a full chance to act in his own interest, nothing but expediency will restrain him from brutalizing, from maiming, from murdering—his brother, his mate, his parent, or his child. Scratch an “altruist,” and watch a “hypocrite” bleed.

The interpretation of human behavior was similarly transformed. In *The Biology of Moral Systems*, Alexander (1987, p. 3) shows the extent to which biologists abandoned the idea that genuinely self-sacrificial behaviors are part of our evolutionary legacy:

I suspect that nearly all humans believe it is a normal part of the functioning of every human individual now and then to assist someone else in the realization of that person's own interests to the actual

net expense of those of the altruist. What this greatest intellectual revolution of the century [i.e., the individualistic perspective in evolutionary biology] tells us is that, despite our intuitions, there is not a shred of evidence to support this view of beneficence, and a great deal of convincing theory suggests that any such view will eventually be judged false.

The concept of selfishness that became established in evolutionary biology, like the concept of egoism in psychology, is a mansion with many rooms. It claims to explain such apparently altruistic traits as the bee's barbed stinger and human morality. These and other characteristics are said to be only apparently altruistic because individuals who help others receive benefits in return or promote their "genetic self-interest" by helping copies of their own genes that are found in the bodies of others. Genuinely altruistic traits that evolve by group selection became an endangered species in evolutionary biology during the 1960s and 70s, just as genuine psychological altruism has long been an endangered species in the social sciences.

The concepts of psychological egoism and altruism concern the motives that people have for acting as they do. The act of helping others does not count as (psychologically) altruistic unless the actor thinks of the welfare of others as an ultimate goal. In contrast, the evolutionary concepts concern the effects of behavior on survival and reproduction. Individuals who increase the fitness of others at the expense of their own fitness are (evolutionary) altruists, regardless of how, or even whether, they think or feel about the action. Many researchers are careful to draw this distinction between the psychological and evolutionary concepts. Nonetheless, the concepts of selfishness in biology and the social sciences are often thought to be compatible and to reinforce each other. If evolutionary altruism is absent in nature, why should psychological altruism be present in human nature?

In this book, we will thoroughly explore the concepts of altruism and selfishness in evolutionary biology, psychology, and philosophy. In contrast with the views just outlined, our argument builds a strong case for both evolutionary and psychological altruism. However, the relationship between these two concepts is not simple. The case for evolutionary altruism requires showing that group selection has been

an important force in evolution. The case for psychological altruism requires showing that an ultimate concern for the welfare of others is among the psychological mechanisms that evolved to motivate adaptive behavior. Both arguments are evolutionary, but they are sufficiently different that we have divided the book in two.

We mentioned before that group selection was once regarded as both thoroughly confused and thoroughly refuted. Nevertheless, it would be a mistake for the reader to regard us as two heretics crying out in the wilderness. During the 1970s, a robust theory of group selection emerged that could withstand the earlier criticisms. Readers who think they are familiar with the subject may be surprised to learn that even G. C. Williams, the icon of the individual selection movement, has accepted the evidence for group selection as the best explanation of important biological adaptations such as female-biased sex ratios and reduced virulence in disease organisms. In short, rather than defending a heretical new theory, we will be reporting and extending a transition in evolutionary thought that is already in full swing.

If group selection has become respectable again, the reader may well wonder why the news is not generally known. One reason is that ten or twenty years is not a long time for certain kinds of scientific change, especially when the subject is as emotionally loaded as altruism and selfishness. The rejection of group selection during the 1960s was based on an evaluation of the theories and evidence available at the time. Unfortunately, the verdict has been transmitted more faithfully through the years than the reasons behind it. Many evolutionary biologists learned just one thing about group selection during their graduate training—"Don't do it!" They avoid the hypothesis partly because it seems scandalous, and partly because they sometimes feel unqualified to evaluate the arguments. As a result, the modern theory of group selection has developed in partial isolation, even within the field of evolutionary biology. Articles that treat group selection as uncontroversial appear in the most respected journals alongside other articles that continue to treat it as a bogeyman. One of the purposes of our book is to present the arguments for and against group selection in enough detail so that readers—biologists and nonbiologists alike—can judge for themselves.

Although the modern theory of group selection is already well developed and empirically supported, the psychological question

about altruistic ultimate motives remains open. Some psychologists think that experimental evidence now exists to decide the question, but many of their colleagues disagree. Indeed, some have suggested that psychological experiments are incapable of distinguishing between altruistic and selfish ultimate motives. It is our ambition, in this book, to outline an evolutionary theory of psychological motives that can solve this problem. Since this is a relatively new enterprise, our case for psychological altruism is more provisional than our case for evolutionary altruism.

The idea that human behavior is governed entirely by self-interest and that altruistic ultimate motives don't exist has never been supported by either a coherent theory or a crisp and decisive set of observations. The entire debate has been characterized by an intellectual pecking order in which an egoistic explanation for a given behavior, no matter how contrived, is favored over an altruistic explanation, even in the absence of empirical evidence that discriminates between the two approaches. It is interesting that a similar pecking order existed during the 1960s for the subject of evolutionary altruism, which made the case against group selection appear much stronger than it actually was. Intellectual pecking orders are sometimes justified—for example, when one of the approaches appears very weak on theoretical grounds—but the group selection debate moved beyond this stage in the 1970s and now is conducted on an even playing field. Alternative theories have equal status and generate different predictions that can be tested empirically. The debate over psychological altruism will never make real progress until it undergoes the same transition. We think that our analysis of psychological altruism will help move the debate onto the same type of even playing field. If psychological mechanisms are partially designed by natural selection to motivate adaptive behaviors, there is good reason to expect these psychological mechanisms not to funnel all behavior through the narrow tube of egoistic ultimate motives.

At the risk of sounding defensive, we feel we should address a criticism that is often leveled at advocates of altruism in psychology and of group selection in biology. It is frequently said that people endorse such hypotheses because they *want* the world to be a friendly and hospitable place. The defenders of egoism and individualism who

advance this criticism thereby pay themselves a compliment; they pat themselves on the back for staring reality square in the face. Egoists and individualists are objective, they suggest, whereas proponents of altruism and group selection are trapped by a comforting illusion.

This criticism is made so often that it is tempting to reply in kind, with conjectures about the psychological benefits that defenders of egoism and individualism extract from believing their pet theories. However, speculations about the motives that prompt someone to defend a theory are irrelevant. They are *ad hominem*. The point is to discover which theories are *true*. What is needed is a focused attention on the evidence for *theories*, not on the psychological quirks of *theorists*.

In any event, it is worth saying here that our goal in this book is not to paint a rosy picture of universal benevolence. Group selection does provide a setting in which helping behavior directed at members of one's own group can evolve; however, it equally provides a context in which hurting individuals in other groups can be selectively advantageous. Group selection favors within-group niceness *and* between-group nastiness. Group selection theory does not abandon the idea of competition that forms the core of the theory of natural selection; rather, it provides an additional setting in which competition can occur. Not only do individuals compete with other individuals in the same group; in addition, groups compete with other groups.⁴

Similar remarks apply to the story we will tell about psychological altruism. We will not suggest that everyone has a thoroughgoing and saintly dedication to helping others—that people always treat the well-being of others as an end in itself and never think of their own welfare. Rather, our objective will be to show that concern for others is *one* of the ultimate motives that people *sometimes* have. Even if we are right, our view leaves plenty of room for the hypothesis that individuals spend a good deal of time looking out for number one.

This book draws on four disciplines—evolutionary biology, social psychology, anthropology, and philosophy. In discussing material from each of these fields, we have tried to begin at the beginning. Our goal is not to address the handful of people who already are conversant with all four areas, but to reach people who know something about only one, or even about none of them at all. Beginning at the

beginning also has the virtue, we feel, of forcing one to rethink fundamentals. This has benefited our own thinking about altruism; we think it will benefit our readers as well.

The significance of a book is the result of an interaction between its contents and the diverse conceptual frameworks of its readers. We anticipate that our readers will come from at least three very different conceptual backgrounds; these can be labeled *individual-level functionalism*, *group-level functionalism*, and *anti-functionalism*. To avoid needless controversy, we want to describe how our argument will relate to these three points of view.

We have already described individual-level functionalism; it is the view that individuals are the primary functional units. Group behavior is “just” the product of interactions among individuals, and groups are not functionally organized in their own right. As G. C. Williams (1966) put the point, a fleet herd of deer is just a herd of fleet deer—the group runs fast not because this benefits the group but because it benefits each individual. The individualistic tradition in evolution and methodological individualism in the human sciences are examples of individual-level functionalism. Against this background, the primary message of our book is that groups, too, can be functional units and that individuals sometimes behave more like organs than like organisms.

Although individual-level functionalism is the dominant tradition that we are opposing, group-level functionalism represents a long-standing point of view; it embodies the opposite belief that groups are the primary functional units. Herds of deer run fast, and have other characteristics, because those traits benefit the herd. Outside of science and cross-culturally, the idea that individuals exist to benefit their society may be more common than the idea that society is merely a collection of selfish individuals. Group-level functionalism also was common among the founding fathers of sociology, anthropology, and social psychology, who often treated culture and society as organic wholes that obey their own higher-order laws. Though less common today, it still exists as a minority view and is even the dominant tradition in some subdisciplines of biology and the human sciences. Our book is a mixture of good and bad news for group-level functionalism. The good news is that we can offer the first robust theory of group-level functionalism. The bad news is that it is not nearly as