Portland State University

PDXScholar

Dissertations and Theses

Dissertations and Theses

2004

Fundamental Conditions for the Evolution of Altruism: Towards a Unification of Theories

Jeffrey Alan Fletcher
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Part of the [Other Statistics and Probability Commons](#), and the [Statistical Theory Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Fletcher, Jeffrey Alan, "Fundamental Conditions for the Evolution of Altruism: Towards a Unification of Theories" (2004). *Dissertations and Theses*. Paper 1881.

<https://doi.org/10.15760/etd.1880>

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

FUNDAMENTAL CONDITIONS FOR THE EVOLUTION OF ALTRUISM:
TOWARDS A UNIFICATION OF THEORIES

by

JEFFREY ALAN FLETCHER

A dissertation submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY
in
SYSTEMS SCIENCE

Portland State University
©2004

DISSERTATION APPROVAL

The abstract and dissertation of Jeffrey Alan Fletcher for the Doctor of Philosophy in Systems Science were presented May 6, 2004, and accepted by the dissertation committee and the doctoral program.

COMMITTEE APPROVALS:

Martin Zwick, Chair

Leticia Avilés

Michael T. Murphy

Tom Seppalainen

Thomas Luckett
Representative of the Office of Graduate Studies

DOCTORAL PROGRAM APPROVAL:

George G. Lendaris, Acting Director
Systems Science Ph.D. Program

ABSTRACT

An abstract of the dissertation of Jeffrey Alan Fletcher for the Doctor of Philosophy in Systems Science presented May 6, 2004.

Title: Fundamental Conditions for the Evolution of Altruism:
Towards a Unification of Theories

In evolutionary theory the existence of self-sacrificing cooperative traits poses a problem that has engendered decades of debate. The principal theories of the evolution of altruism are inclusive fitness, reciprocal altruism, and multilevel selection. To provide a framework for the unification of these apparently disparate theories, this dissertation identifies two fundamental conditions required for the evolution of altruism: 1) non-zero-sum fitness benefits for cooperation and 2) positive assortment among altruistic behaviors. I demonstrate the underlying similarities in these three theories in the following two ways.

First, I show that the game-theoretic model of the prisoner's dilemma (PD) is inherent to all three theories. While the PD has been used extensively to model reciprocal altruism, I demonstrate that the n-player PD captures fundamental aspects of multilevel selection and inclusive fitness in that NPD model parameters relate simply to Simpson's paradox, the Price covariance equation, and Hamilton's rule. The tension between hierarchical levels that defines a PD reflects the tension between

Abstract

levels of selection that is explicit in multilevel selection theory, and implicit in the other two theories.

Second, Hamilton's rule from inclusive fitness theory applies to the other theories. As mentioned, I demonstrate that this rule relates to multilevel selection via the NPD. I also show that Queller's generalization of Hamilton's rule applies to the conditional strategies of reciprocal altruism. This challenges the selfish-gene viewpoint by highlighting the fact that it is the phenotypes of others, not their genotypes, that is critical to the evolution of altruism.

I integrate the PD and Hamilton's rule as follows: the evolution of altruism in general involves PD situations in which Hamilton's rule specifies the necessary relationship between 1) the degree of non-zero-sumness within the PD and 2) the degree of positive assortment among altruistic behaviors.

Additional contributions of this research include a demonstration that randomly formed associations can provide the necessary positive assortment for strong altruism to evolve, the development of a new selection decomposition that is symmetrical to the Price equation, and a game-theoretic analysis showing the essential similarity of weak and strong altruism under selection.

Acknowledgments

I wish to express my heartfelt appreciation to all those who have given their comments, advice, criticisms, encouragement, and material support during the time I have spent researching and writing this dissertation. All suggestions were valuable and went into the process of refinement whether or not they are still recognizable in the final mix.

My greatest gratitude goes to my advisor, coauthor, and friend Martin Zwick. He has been extremely generous with his time, intellect, and creativity and nothing has shaped and inspired this dissertation more (or been more rewarding for me) than our lengthy and fruitful conversations. On numerous occasions, I have stopped by Marty's office to relate a new finding or to ask a seemingly discrete question only to have this expand into a two or three hour discussion—and this is often just the beginning of an ongoing exploration. The ideas generated by these conversations often infect us both and many times I have received late night emails from Marty saying he can't sleep until he conveys a few thoughts related to our conversation earlier that day. I wish I knew how to identify the key ingredients of our interactions so as to replicate them with others, but in my experience these brainstorming sessions have been uniquely productive. Certainly, this has been the most stimulating intellectual relationship of my life.

In addition, Marty has endured numerous manuscript drafts in various stages of disrepair and has always managed to take whatever is offered (whether the first version or the tenth) and provide detailed suggestions, questions, and encouragement that markedly improve the final result. Our work has been a true collaboration in the best possible sense in which we often challenge the other's viewpoint, but it is never about being right and always about working together to find the truth of the matter. The results often feel like more than the sum of our individual efforts and for me our interactions are a great example of the potential synergistic benefits of mutual cooperation.

I also wish to thank my other dissertation committee members: Leticia Avilés, Thomas Luckett, Michael Murphy, and Tom Seppalainen. They were always supportive and yet each of them offered challenges from their own perspective and expertise that helped to clarify my own thinking. I want to especially thank Leticia for being an external committee member and traveling to Portland for key meetings. She was also one of the first people to treat me like an "expert" by inviting me to participate in a special workshop on *The Evolution of Sociality* at the *Wissenschaftskolleg zu Berlin* (Berlin Institute for Advanced Study) and by encouraging me to aim high when submitting papers for publication. She has also invited me to give talks to the Ecology and Evolutionary Biology Department at the University of Arizona and the Zoology Department at the University of British Columbia. Thanks also to Tom Luckett for inviting me to talk to his freshman inquiry classes and Michael Murphy for inviting me to give a talk to the Department of

Biology at Portland State University. I am also very grateful to Leticia for financial support for travel to the Berlin workshop and arranging my stay there, as well as a support while we collaborated on a paper in Tucson in the summer of 2002. Finally, thanks also to my comprehensive exam committee Sergio Antoy, Andrew Fraser, Sarah Mocas, and Martin Zwick (as well as contributors Richard Hamlet and George Lendaris) for their time, challenges, and encouragement.

I seldom express things well the first time around and many people (from anonymous reviewers to close friends) have been generous with their time in reviewing (in various stages of completion) the writings that contributed to this dissertation. In addition to my dissertation committee members, I am very grateful for feedback on my written work from Athena Aktipis, Mark Bedau, Cecelia Barry, Rachel Cloues, Niles Lehman, George Lendaris, Sarah Mocas, Peter Mortola, Kim Powers, Andreas Rechtsteiner, Daniel Soose, Eörs Szathmáry, Kim Turnbull, Mark Weber, and Susan Wilmoth, as well as several anonymous reviewers. I am especially grateful to David Wilson for his encouragement and very supportive and detailed review of a manuscript that forms the basis of Chapter 3.

In addition to those who have commented on my written work, I have been very fortunate in having opportunities to discuss this research with many very knowledgeable people. I am especially grateful for the feedback and discussions with fellow presenters at the ALife VII workshop on *ALife's Role in the Group Selection Debate*: Athena Aktipis, Leticia Avilés, Joshua Mitteldorf, John Pepper, and Michael Wade; and with fellow presenters at the Berlin workshop on *The Evolution of*

Sociality: Leticia Avilés, Raghavendra Gadagkar, Peter Hammerstein, Amitabh Joshi, Laurent Keller, John Pepper, and Michal Woyciechowski. I also want to thank my fellow participants at the Santa Fe Institute Complex Systems Summer School in 1999 where I first started to think about the evolution of altruism, especially my project partner Richard Watson and provocative critic Torsten Reil.

Thanks also to Greta Binford, Michael Doebeli, Gene Enneking, Wayne Maddison, Len Nunney, and Tad Shannon for helpful discussions. My research ideas and thoughts on how best to present them have also been shaped by the challenging questions posed by audience members at various conference and seminar talks and I am indebted to them. In addition, I have presented my research in various classrooms. For stimulating discussions, I would like to thank the Philosophy of Biology class at Reed College, the Evolution class at Lewis and Clark College, the 4th grade class at Vose Elementary School, and the Freshman Inquiry, Game Theory, and Artificial Life classes at Portland State University.

Support for this project has come in many forms. After finishing a Master's in Computer Science in 1995, I worked for seven years for MedicaLogic, Inc. (now GE Medical Systems) as a software engineer. Thanks to Mark Leavitt for founding a very egalitarian company that was a rewarding place to work (even if we didn't all get rich). During the first four years, I worked full-time while taking (or sometimes teaching) an occasional class. I want to thank my managers during this time (Blackford Middleton, René Poston, Rylla Riverman, Alan Rosenfeld, and Rick Samco) who were always supportive of my schooling and flexible with my schedule. I

especially thank Mitch Jackson for his support when I wanted to focus more on school and became the only part-time software engineer at this not-yet-profitable startup company. I am also very grateful to Chip Masarie for giving me great flexibility in my schedule when I wanted to spend weeks in Berlin or at the Santa Fe Institute and for always supporting my personal growth, which included me ultimately leaving to concentrate full-time on this dissertation.

I have received helpful support from Portland State University. Since 1996 my tuition has always been covered by an Oregon Laurels Graduate Scholarship and I am very grateful to this program and Berni Pilip for flexibility in letting me use this scholarship over many years in mostly part-time pieces. Thanks also to the Academically-Controlled Auxiliary Activities Committee for support to travel to *The World Congress of the Systems Sciences and ISSS 2000* in Toronto, Canada , to Roberto Santiago and Sangsuree Vasupongayya of the NW Computational Intelligence Lab for letting me use their fast computers to run some of my simulations, and to Dawn Sharafi of the Systems Science Ph.D. program for administrative assistance.

I am grateful to the Santa Fe Institute for financial support to attend their Complex Systems Summer School in 1999 and to the *Wissenschaftskolleg zu Berlin* for hosting me for a couple of weeks in Berlin in May 2001. Also thanks to Angelika Eratz for arranging additional housing in Berlin and to Theo Eratz, Lothar Peter, and Birgit Berndt for hosting me there. Additional thanks to Leticia Avilés and Wayne Maddison for hosting me in Vancouver, B.C.

I am very thankful to my parents, siblings, and large extended family for always encouraging my curiosity and affinity for school—little knowing how many decades of education this would entail! And finally, I am deeply grateful for all the groups—family, academic collaborators, the no-better-offer potluck gang, the men’s group disguised as a book club, the “housing around” home repair coop, the Ross Island solstice/equinox celebration committee, the hootenanny musicians, the Tuesday night basketball players, my kayaking, hiking, and climbing buddies, and the other cooperative groups of friends to which I belong and which greatly enrich my life.

Table of Contents

Acknowledgments	i
List of Tables	xi
List of Figures	xii
List of Symbols and Abbreviations	xiv
Preface	xvi
Chapter 1 Introduction	1
Chapter 2 Background.....	10
THE GROUP SELECTION CONTROVERSY	14
MULTILEVEL SELECTION THEORY	21
<i>Simpson's Paradox</i>	25
<i>The Price Equation</i>	26
INCLUSIVE FITNESS AND THE PROGRESSIVE GENERALIZATION OF HAMILTON'S	
RULE.....	28
<i>Kin Selection</i>	29
<i>Positive Assortment</i>	31
<i>Phenotype vs. Genotype</i>	36
RECIPROCAL ALTRUISM AND EVOLUTIONARY GAME THEORY	38
<i>The Prisoner's Dilemma</i>	40

<i>Iterated Prisoner's Dilemma</i>	41
<i>N-Player Prisoner's Dilemma (The Tragedy of the Commons)</i>	45
Chapter 3 The Evolution of Altruism: Game Theory, Multilevel Selection, and Inclusive Fitness	48
INTRODUCTION.....	48
METHODS	50
<i>The N-Player Prisoner's Dilemma Model</i>	50
<i>Hamilton's Rule</i>	53
RESULTS AND DISCUSSION	54
<i>Simpson's Paradox</i>	54
<i>The Price Equation</i>	58
<i>Alternative Selection Decomposition</i>	60
<i>Selection and Game Theory</i>	63
<i>Altruism via Randomly Formed Groups</i>	68
Reformation Frequency	72
<i>Summary</i>	73
Chapter 4 Strong Altruism Can Evolve in Randomly Formed Groups	76
INTRODUCTION.....	76
CLASSIFICATIONS OF ALTRUISM.....	78
ANALYTIC MODEL.....	79
APPLYING HAMILTON'S RULE	82
ANALYTIC MODEL MODIFICATIONS	86

<i>No Kin Selection</i>	87
<i>No Homogeneous Groups</i>	88
<i>No Population Growth</i>	88
<i>Modification Results</i>	89
SIMULATION MODEL	90
CONCLUSION	98
Chapter 5 Hamilton's Rule Applied to Reciprocal Altruism	103
OBSTACLES TO APPLYING HAMILTON'S RULE TO RECIPROCAL ALTRUISM	105
THE ITERATED PRISONER'S DILEMMA	107
HAMILTON'S RULE	109
APPLYING HAMILTON'S RULE TO RECIPROCAL ALTRUISM	111
HAMILTON'S RULE AND NON-ADDITIVITY	117
CONCLUSION	120
Chapter 6 Summary and Discussion	124
SUMMARY OF RESULTS	124
<i>Fundamentals and Theory Unification</i>	124
The Generality of Hamilton's Rule	125
The Generality of the Prisoner's Dilemma	126
The Generality of Each Theory	127
<i>Other Major Findings</i>	130
DISCUSSION	133
<i>Causal Explanations and Hierarchy</i>	133

Averaging Fallacy	134
Defection Suppression.....	136
Hamilton's Rule and Hierarchy.....	138
Proximate vs. Ultimate Causation	139
Emotions as Proximate Mechanisms.....	141
<i>Are the Conditions that Produce Altruism Rare?</i>	143
<i>Is Nice Necessary?</i>	146
FUTURE STUDIES	149
References	153
Appendices	170
APPENDIX A: HAMILTON'S RULE DERIVATION	170
APPENDIX B: PRICE EQUATION DERIVATIONS.....	172
APPENDIX C: RANDOM GROUP ANALYTIC MODEL	174
APPENDIX D: RANDOM GROUP ANALYTIC MODEL MODIFICATIONS	176
<i>No Benefit to Kin</i>	176
<i>No Homogeneous Groups</i>	177
<i>No Population Growth</i>	178
APPENDIX E: RANDOM GROUPS COMPUTER SIMULATION MODEL.....	179
APPENDIX F: DERIVATION OF G_P' FOR CONDITIONAL STRATEGIES IN	
TFT / ALLD POPULATION	181

List of Tables

Table 2-1. Example of Simpson's Paradox	26
Table 2-2. Historical Advances in Generalizing the r term of Hamilton's Rule.....	37
Table 3-1. Parameter Values for Runs 1-5 used in Figure 3-2 and Figure 3-3.	57
Table 3-2. Percentage of 100 Runs Ending in Cooperator Saturation.	72
Table 5-1. G_A and Predicted P_A , and P_O Values	112
Table 5-2. Hamilton's Rule Minimum b for $\Delta G_P = 0$ in Figure 5-3.	116
Table 5-3. Hamilton's Rule Minimum d for $\Delta G_P = 0$ in Figure 5-4.	120

List of Figures

Figure 1-1. Schematic of Dissertation Core Findings in Chapters 3 to 5.....	7
Figure 2-1. Payoff Matrix for a 2-Player Prisoner's Dilemma.....	40
Figure 3-1. Simple N-Player Prisoner's Dilemma.	51
Figure 3-2. Dynamics in q , r , and n for a Typical NPD Run with 2 Groups.	56
Figure 3-3. Results for 4 Runs with Various Parameters (see Table 3-1).....	57
Figure 3-4. Price Equation and Alternative Decompositions for Run 1.....	62
Figure 3-5. Applying Difference Maximization to NC and PD Games.....	68
Figure 3-6. Dynamic Change in Q for Periodically Randomly-formed Groups.	71
Figure 4-1. Change in Altruist Frequency as a Function of Altruist Benefit.	81
Figure 4-2. Calculated Dynamics in Analytic Model Comparing g and n	85
Figure 4-3. Calculated Altruist Frequency for Analytic Model Modifications.....	89
Figure 4-4. Altruist Frequency for Different Levels of Stochasticity.	93
Figure 4-5. Percentage of Simulation Runs Reaching Altruist Saturation.	95
Figure 4-6. Migration and Percentage of Runs Reaching Altruist Saturation.....	96
Figure 4-7. Dynamics in Q for Individual Simulation Runs with Varying g	98
Figure 5-1. Typical PD Utility (or Fitness) Values for Actor.	107
Figure 5-2. PD Utility (or Fitness) Values Based on Individual Contribution.....	112
Figure 5-3. Change in Fraction of TFT Players for Additive Payoffs.....	116

Figure 5-4. Change in Fraction of TFT Players for Non-Additive Payoffs.	119
---	------------

List of Symbols and Abbreviations

Sub-script	Meaning
<i>i</i>	group index
<i>b</i>	between-group
<i>w</i>	within-group
<i>P</i>	population value
<i>A</i>	actor's value
<i>O</i>	value for others (or opponent)
<i>g</i>	number of generations within groups
<i>a</i>	altruists' values
<i>s</i>	non-altruists' (selfish) value

Super-script	Meaning
<i>o</i>	other-only altruism
<i>w</i>	whole-group altruism
'	primed terms represent values in new generation

Abbreviation	Meaning
PD	prisoner's dilemma game
NPD	n-player prisoner's dilemma game
NC	no-conflict game
TFT	tit-for-tat genotype
ALLD	always-defect genotype
ALLC	always-cooperate genotype
WG	whole-group altruism
OO	other-only altruism
C	cooperate behavior
D	defect behavior

Symbol	Meaning
b	benefit provided by a C behavior
c	cost paid by actor for a C behavior
g	number of generations spent within groups before random redistribution
n_i	number of members in group i
n_g	number of members in a group after g generations
N	number of members in whole population
a_i	number of altruists or cooperators in group i
a_g	number of altruists or cooperators in a group after g generations
A	number of altruists or cooperators in whole population
s_i	number of non-altruists or defectors (selfish) in group i
s_g	number of non-altruists or defectors (selfish) in a group after g generations
S	number of non-altruists or defectors (selfish) in whole population
q_i	fraction of altruists or cooperators in group i
Q	fraction of cooperators in whole population
Q^*	Price idealization about Q
$Q^\#$	alternative decomposition idealization about Q
r	measure of assortment, from Hamilton's rule ("relatedness")
r^o	r value for other-only altruism
r^w	r value for whole-group altruism
r_{dev}	Queller's r value for non-additive benefits of mutual cooperation
w	fitness for an individual (number of offspring)
w_a	fitness for an individual altruist, a
w_s	fitness for an individual non-altruist (selfish), s
w_{av}	average fitness for all group members
w_0	base individual fitness independent of gains or losses due to C or D behaviors
w_i	group i 's fitness as measured by its growth rate, n_i' / n_i
Y	total number of groups in a population
G_P	fraction of <i>genotype</i> in population (here, fraction of TFT)
G_A	fraction of <i>genotype</i> in an actor (here, 1 or 0 for TFT or not)
P_P	fraction of <i>phenotype</i> (behavior) in population (here, fraction of C)
P_A	fraction of <i>phenotype</i> (behavior) by an actor (here, fraction of C)
P_O	fraction of <i>phenotype</i> (behavior) by actor's opponents or others (here, fraction of C)
i	number of iterated games in a pairing
d	deviation from additivity for CC interaction
f_{TT}	fraction of TFT-TFT pairings
f_{TD}	fraction of TFT-ALLD pairings
f_{DD}	fraction of ALLD-ALLD pairings

Preface

Before I became interested in the evolution of altruism, my research focused on a different topic. I was doing investigations in a relatively new discipline called Artificial Life, which uses computer simulations to study life-like processes. Specifically, this research (Fletcher, Zwick, and Bedau 1996; Fletcher, Bedau, and Zwick 1998) focused on developing information-theoretic measures of the match between the complexity of challenges posed by an environment and the repertoire of evolvable responses to these environmental inputs within organisms. I planned to build on this work for a Ph.D. thesis, but then I happened to read a *Scientific American* article by Richard Dawkins (1995) that ultimately set me on a different path.

Dawkins is both a popular science writer and one of the most eloquent and staunchest proponents of reductionism in evolutionary biology. His books, such as *The Selfish Gene* (1976) and *The Extended Phenotype* (1982), are classics that have influenced the thinking of both evolutionary biologists and the general public about what drives evolutionary change. Dawkins argues that much of what we see in the natural world can be explained in terms of the self-interest of genes. In this view the plethora of elaborate mechanisms for accomplishing tasks such as securing nutrition and shelter, and handling competitive and cooperative social interactions are just manifestations of genes endeavoring to increase their representation in future

generations. What I found troublesome about Dawkins' arguments was that there seemed to be a limited role for the influence of processes or properties that emerge at higher levels of organization than that of the gene.

I was convinced that there should be some way to use evolutionary computer models to show that emergent phenomena that affect the outcome of natural selection could not be attributed to their genetic basis alone. Although many other arguments have been made against Dawkins' reductionism (e.g. Wade 1977b; Sober and Lewontin 1982; Sober and Wilson 1998; Lewontin 2000; Gould 2002), I thought these new computer tools for modeling evolutionary processes might be useful in illustrating the limitations of this viewpoint. Fortunately, at around this same time my advisor Martin Zwick heard David Wilson and Leticia Avilés give talks on the role multiple levels of selection play in the evolution of altruism. This led me to read Sober and Wilson's book *Unto Others* (1998) that chronicles the debate in evolutionary theory about altruism and group selection.

It was clear after reading *Unto Others* that the evolution of altruism was a perfect example of higher-level emergent phenomena that could not be explained only in terms of lower-level interactions among competing genes. Additionally, the tension between individual- and group-level selection described in the multilevel selection process for altruistic traits seemed to map naturally onto the tension between individual and collective rationality described by the tragedy of the commons or the prisoner's dilemma (PD) of game theory. Yet while the PD was being used to model

reciprocal altruism, its usefulness in modeling the evolution of altruism in general was not appreciated. Thus this dissertation topic was conceived.

While this has been a very productive research area (and it feels like there is still much more to be done), it has also taken a long time to get to this point. For seven years after completing a Master's in Computer Science in 1995 I worked as a software development engineer for a startup company. During that time, completing course work and doing research was an ever-demanding "hobby" which I mostly did during evenings and weekends. At first my job was more than fulltime, then eventually part-time, and finally for almost two years now I have been able to focus primarily on the research for this dissertation. Over the last four years I have written several papers (all in collaboration with my advisor Martin Zwick) that form the basis of this dissertation. Five papers have been accepted for publication (Fletcher and Zwick 2000a; 2000b; 2001; 2004c; 2004b), four in conference proceedings, one in the *Journal of Theoretical Biology*, and two more manuscripts for journal publication are essentially complete. In addition, I collaborated on a paper with Leticia Avilés and Asher Cutter that has just been accepted to *The American Naturalist* (Avilés, Fletcher, and Cutter 2004), but is not part of this dissertation.

Chapters 3-5 form the heart of the dissertation and below I give some of the history of their development. Chapter 3 is the culmination of work that was ultimately presented in three proceedings papers, as well as a manuscript submitted to *Science* magazine. The first of these papers was presented in July of 2000 at The World Congress of the Systems Sciences and ISSS Conference in Toronto, Canada (Fletcher

and Zwick 2000b). Later that summer I co-coordinated a workshop on “ALife’s Role in the Group Selection Debate” at the ALife VII conference in Portland, Oregon where I also presented a paper (Fletcher and Zwick 2000a). This was a very successful workshop and other presenters and contributors included Athena Aktipis (2000), Leticia Avilés (2000), Joshua Mitteldorf (Mitteldorf and Wilson 2000b), John Pepper (2000a), and Michael Wade (2000) as well as Richard Michod (Michod and Roze 2000a) who contributed a paper but was unable to attend.

In 2001 I was invited to a special 2-day workshop on the “Evolution of Sociality” at the *Wissenschaftskolleg zu Berlin* (Berlin Institute for Advanced Study) in Germany where I presented preliminary findings on the relationship between the n-player PD (NPD) and the Price covariance equation as well as an alternative selection decomposition. This work comprised the last of these proceedings papers which I presented in October of 2001 at the IEEE, Systems, Man, and Cybernetics Conference in Tucson, Arizona and in an invited talk to the Ecology and Evolutionary Biology Department at the University of Arizona. The *Science* submission received a very strong review from David Wilson (who wrote to me directly after being asked to review our paper by the *Science* editors), but was ultimately not selected for publication. I plan to resubmit this paper to another journal soon, supported by new findings detailed below as well as new work done on showing the equivalence between weak and strong altruism, which I first presented at an invited talk to the Evolution, Ecology, and Biodiversity Seminar, Department of Zoology, University of British Columbia in November, 2003.

Chapter 4 grew out of a main criticisms of the *Science* manuscript from one of the other reviewers. This reviewer stated that what we had shown about maintaining altruism through random group formation must only apply to weak altruism. While this assessment was based on a long-standing consensus in the literature that strong altruism could not evolve in randomly formed groups, ultimately I found it was based on *single-generation* models by Hamilton (1975) and independently by Wilson (1975a). It was clear from our NPD model that an *intermediate* number of generations could be most favorable to the evolution of altruism and I knew from our experimental results that this was true of both weak and strong altruism.

In order to address these claims directly, I decided to use the original strong altruism models of Hamilton and Wilson to show that this nearly 30-year-old general consensus was incorrect. This research was first presented at the *Evolution 2003* conference in Chico, California. In a fuller form it has just been published as a paper entitled “Strong Altruism Can Evolve in Randomly Formed Groups” in the *Journal of Theoretical Biology* (Fletcher and Zwick 2004c). This paper forms the basis of Chapter 4. This research also led me to think more about the relationship between strong and weak altruism and to develop a game theoretic argument showing their equivalence when under natural selection (Chapter 3).

Chapter 5 grew out of initial difficulties I encountered in applying Hamilton’s rule to a reciprocal altruism model. While I was convinced that the association between behaviors rather than genotypes was key, my attempts to use phenotype frequencies in Hamilton’s rule had all failed until I came across Queller’s versions

(1985; 1992b; 1992a). Although the example of applying Hamilton's rule to conditional strategies in Chapter 5 is quite simple, as far as I know, it is the first time someone has demonstrated how these equations apply to reciprocal altruism. The paper representing these findings (Fletcher and Zwick 2004b) forms the basis of Chapter 5 and has just been accepted into the proceedings of the IEEE Congress on Evolutionary Computation to be held in Portland, Oregon in June of this year. The major theme of this conference is a celebration of the 20th anniversary of Robert Axelrod's book *The Evolution of Cooperation* (1984). The conference includes an iterated PD competition similar to those used as a basis of this seminal book and Axelrod will be giving the keynote address.

Fittingly, Chapter 5 also returns me to Dawkins' selfish gene theory and other narrow interpretations of Hamilton's rule in which altruistic behaviors are seen as being encoded by genes that help other related copies of themselves. Unlike Hamilton's versions of his equations (on which Dawkins' viewpoint depends), in Queller's more general versions there are no terms for the genotypes of others, only their phenotypic behaviors. These most general interpretations of Hamilton's Rule shows how genetically based altruistic traits can increase via interactions with completely unrelated individuals or even individuals of different species in symbiotic relationships. Phenotype is an emergent property that depends on the interaction of genes and environmental conditions (including the behaviors of others). It is others' phenotypic behaviors that determine the fitness consequences of social interactions to individual's and their genes.

In summary, although the original motivation for the research behind this dissertation began with a negative reaction to Dawkins' reductionist stance (and I was ultimately able to point out some of the limitations in the selfish gene perspective), my research has focused more broadly on the fundamental conditions for altruism to evolve across all theories. The evolution of altruism is a rich and still controversial area of evolutionary theory, which continues to fascinate me and I am sure to continue thinking about for many years to come.

In short, the units of competitive self-interest that make up a genome are utterly interdependent. How did the competitive process of natural selection shape so intricate a mutualism?

Egbert Giles Leigh, Jr, in *Levels of Selection in Evolution* (1999)

Chapter 1

Introduction

Nature is replete with examples of highly cooperative behaviors. From the very effective cooperative hunting and food sharing in human tribes (even among non-relatives) (e.g. Boehm 1993; Hawkes 1993), to worker bees and ants giving up their own reproduction (e.g. Wilson 1975b), to the marvel of the highly specialized and diverse set of individual cells that go into making up an individual organism such as a bee or a person. Yet for every cooperative interaction it is also easy to imagine a way in which one party could take advantage of others to get a little bit more in the cooperative exchange and thereby ultimately make more copies of themselves relative to their more egalitarian cohorts. A hunter who risked a little less danger in the hunt or took a little more than his share of the food would have such an advantage, as would a bee who had her own offspring, or a cell that made more copies of itself at the expense of a well-functioning individual. These less cooperative people, bees, or cells we label *selfish* or *defectors* while the less selfish are *altruistic* or *cooperators*. Here altruism is defined in a strictly evolutionary biology sense in that the fitness of others is enhanced

by being a cooperative hunter, bee, or cell at a fitness cost to the cooperator—that is they could have made more copies of themselves by being selfish.

This question is central to this dissertation: How can such cooperative or altruistic behaviors evolve and be maintained in the face of natural selection which should favor the more selfish? While defections are known to occur in each of these examples (lazy or greedy hunters (Wilson 1998), cheating bees (Oldroyd et al. 1994), or cancerous cells (Nowell 1976; Paulovich, Toczyski, and Hartwell 1997)), they are exceptions to the rule. The three major theories on the evolution of altruism each answer this central question differently and there is thus heated debate as to which theories are correct and which apply in different situations (see Chapter 2). One of the major contributions of this dissertation is to show that the three major theories all share fundamental conditions and rely on the same basic mechanisms to explain the evolution of altruism. Identifying these basic requirements in each theory helps move us towards a more unified understanding of how altruism evolves. Fundamental aspects of the evolution of altruism are embodied in both Hamilton's rule from inclusive fitness theory and the prisoner's dilemma from game theory, and I show how both apply to all three theories. Below I provide a preview of each of the following chapters followed by a summary of this dissertation's contributions to our understanding of how altruism evolves.

Chapter 2 provides a literature review and background information on several of the theories and concepts that are pertinent to this dissertation. I start with some examples of natural phenomena where cooperation or altruistic traits are well

documented and then discuss the major evolutionary explanations of how such behaviors can arise under the process of natural selection. These theories are inclusive fitness theory (which is closely associated with kin selection and selfish gene theories), multilevel selection (an update of group selection) theory, and reciprocal altruism theory. I also discuss several concepts relevant to the evolution of altruism that are related to these theories and used in later chapters. These include Hamilton's rule, the prisoner's dilemma (including iterated and n-player versions), the tragedy of the commons, Simpson's paradox, and the Price covariance equation. Note that because each of the chapters 3-5 is in the form of a standalone paper, some of this background information is repeated in the introductions or discussions of those chapters. In general Chapter 2 covers these topics in more depth.

Chapter 3 is based on an original insight that the tension between within-group and between-group selection described in multilevel selection theory is analogous to the tension between individual and collective rationality described by a tragedy of the commons or n-player prisoner's dilemma (NPD). Although the prisoner's dilemma (PD) has been used to investigate reciprocal altruism, it has not been used explicitly to model multilevel selection. Here I show that an NPD model in multiple groups is a simple yet powerful model of the evolution of altruism. It captures essential factors in model parameters that relate simply to Hamilton's Rule, the Price covariance equation, and a useful alternative selection decomposition I present here. Contrasting this alternative decomposition with the Price equation demonstrates interaction effects

between within- and between-group selection which are not captured in either decomposition alone.

This game theoretic model also allows me to demonstrate essential similarities between strong and weak altruism undergoing natural selection. Lastly, using periodic random groupings of non-conditional strategies I show that, in contrast to conventional wisdom, strong altruism can be selected for and maintained without reciprocity. This model suggests a framework for unifying different approaches to the evolution of altruism. Chapter 3 is based on a manuscript (Fletcher and Zwick 2004a) prepared for submission and draws on three previous proceedings papers (Fletcher and Zwick 2000a; 2000b; 2001).

Chapter 4 provides a more thorough treatment of the claim (mentioned above) that strong altruism can *indeed* evolve in randomly formed groups. Currently there is general agreement that altruistic traits involving an absolute cost to altruists (strong altruism) *cannot* evolve when populations are structured with randomly formed groups (Hamilton 1975; Wilson 1975a; Nunney 1985a; Wilson 1990; Nunney 2000; Sober and Wilson 2000). This conclusion implies that the evolution of such traits depends upon special environmental conditions or additional organismic capabilities that enable altruists to interact with each other more than would be expected with random grouping. Here I show, using both analytic and simulation results, that the positive assortment necessary for strong altruism to evolve does not require these additional mechanisms, but merely that randomly formed groups exist for more than one generation. Conditions favoring the selection of altruists, which are absent when

random groups initially form, can naturally arise even after a single generation within groups—and even as the proportion of altruists simultaneously decreases. The gains made by altruists in a second generation within groups can more than compensate for the losses suffered in the first and in this way altruism can ratchet up to high levels. This is true even if altruism is initially rare, migration between groups allowed, homogeneous altruist groups prohibited, population growth restricted, or kin selection precluded.

Until now random group formation models have neglected the significance of multigenerational groups—even though such groups are a central feature of classic “haystack” models (Maynard Smith 1964; Wilson 1987) of the evolution of altruism. I also explore the important role that stochasticity (effectively absent in the original infinite models) plays in the evolution of altruism. The fact that strong altruism can increase when groups are periodically and randomly formed suggests that altruism may evolve more readily and in simpler organisms than is generally appreciated. This chapter is based on a paper recently accepted to the *Journal of Theoretical Biology* (Fletcher and Zwick 2004c).

Chapter 5 shows that very general versions of Hamilton’s inclusive fitness rule (developed by Queller (1985; 1992b; 1992a)) can be applied to traditional reciprocal altruism models such as the iterated PD. While several theorists have shown that multilevel selection and inclusive fitness models are alternative ways of analyzing the same basic mechanism (e.g. Wade 1980; Queller 1985; 1992b; Frank 1998; Sober and Wilson 1998; Fletcher and Zwick 2004a), reciprocal altruism and inclusive fitness are

generally considered *different* mechanisms (e.g. Axelrod and Hamilton 1981; Dawkins 1982) by which cooperative, altruistic traits may evolve. By demonstrating how to apply Hamilton's rule to a reciprocal altruism model with conditional strategies, I challenge the view that these are different mechanisms and show that both inclusive fitness (multilevel selection) models and reciprocal altruism models rely on the same fundamental conditions: 1) sufficiently non-zero-sum helping *behaviors* being 2) sufficiently positively assorted. I also discuss obstacles to this unified view, including phenotype/genotype differences and non-additive fitness (or utility) functions that are typical of reciprocal altruism models. I then demonstrate how Queller's versions of Hamilton's rule remove these obstacles. Finally, the fact that Queller's most general version of Hamilton's rule focuses on the assortment of *phenotypes* (behaviors) rather than the traditional *genotypes* poses a challenge to the selfish gene perspective and shows that inclusive fitness theory can even apply to cooperation across species in symbiotic relationship. This chapter is based on a paper (Fletcher and Zwick 2004b) just accepted to the proceedings of the 2004 *Congress on Evolutionary Computation* to be held in Portland this summer.

Figure 1-1 illustrates areas within evolutionary theory on altruism where this dissertation makes contributions and how these contributions fit into the larger context of current theory. It also indicates which of the core chapters (3-5) addresses each area.

Figure 1-1. Schematic of Dissertation Core Findings in Chapters 3 to 5. Solid lines show established relationships between the three major theories, the NPD, Hamilton's rule and associated phenomena. The dashed lines show relationships newly developed or substantially strengthened by this dissertation along with the associated chapter numbers.

Chapter 6 provides a conclusion and discusses some of the implications of this work for unifying our understanding of how altruism evolves. I also discuss how some of the paradoxes surrounding the evolution of altruism fit into a larger framework of needing to explain causation at appropriate hierarchical levels. In addition, a better appreciation of the role *both* cooperation *and* competition play in the evolution and maintenance of hierarchical biological systems may influence our understanding of human evolution as well as provide a richer basis for societal metaphors drawn from

biology. This final chapter also discusses possible future extensions to the research presented here.

In summary, the major contributions of this dissertation include:

- Identifying the fundamental conditions necessary for the evolution of altruism inherent in all three major theories of how altruism evolves.
- Demonstrating that the n-player prisoner's dilemma (NPD) in multiple groups can be used as a fundamental model of the evolution of altruism in that the paradox of the PD captures the tension between different levels of selection and the parameters of the NPD have simple and intuitive relationships to Hamilton's rule, the Simpson's paradox effect, and the Price covariance equation.
- Providing an integrated framework that can help social science researchers who emphasize game-theoretic models to see their results in the context of Hamilton's rule and multilevel selection, while also enabling biology researchers who focus on relatedness to recognize the inherent game-theoretic character of their models.
- Developing an alternative selection decomposition that is symmetric, but different than the Price covariance equation and helps reveal its underlying assumptions and limitations.
- Showing that the distinction between strong and weak altruism relies on absolute fitness values, whereas selection relies on relative fitness, and

therefore that the dynamics of both types are similar under natural selection.

- Overturning a nearly three decade old consensus among evolutionary biologists that strong altruism could not evolve in randomly formed groups.
- Investigating the role that stochasticity can play in enhancing or dampening selection for altruistic traits.
- Showing how the most general version of Hamilton's rule can be successfully applied to reciprocal altruism models, even when fitness functions are non-additive.
- Demonstrating that in its most general interpretation, the evolution of altruistic traits need not involve selfish genes helping other copies of themselves. This more general interpretation can be used to understand the evolution of cooperation between species—symbiotic relationships for which the selfish gene interpretation has no direct role.

Finally, this dissertation allows me to discuss (in Chapter 6) broader implications of this research including how causal explanations and hierarchy interact in this paradoxical and controversial arena of theories on the evolution of altruism, how both competition and cooperation affect natural selection across different hierarchical levels, how ideas from general systems theory apply to evolving biological systems, and how our understanding of competition and cooperation in biological systems may both influence and be influenced by our understanding of human social systems.

For biologists, I think that the moral is that it pays to be eclectic in our choices of theories. Of course, we have to avoid believing simultaneously in two contradictory theories, or in any one theory that is contradicted by observations that we think are correct. But this leaves plenty of room to be reductionists in one context and holist in another.

John Maynard Smith, in conclusion to *Shaping Life* (1998b)

This possibility of group selection has been regarded as anathema by nearly all evolutionary biologists, although entirely without empirical evidence.

Richard C. Lewontin, in review of *Unto Others* (1998)

Chapter 2

Background

The mechanisms by which self-sacrificing, cooperative, or altruistic behavior may evolve in biological systems have been vigorously debated over the last several decades (e.g. Wynne-Edwards 1962; Maynard Smith 1964; Williams 1966; Wilson 1975a; Dawkins 1976; Wade 1977b; 1985; Queller 1992b; Wilson 1997a; Lewontin 1998; Sober and Wilson 1998; Reeve 1999; Smuts 1999; Katz 2000). Proposed explanations include reciprocal altruism where the self-interest of individuals is served by the exchange of cooperation with others (Trivers 1971; Axelrod and Hamilton 1981; Axelrod 1984), inclusive fitness where the self-interest of genes is served by encoding actions benefiting copies of themselves in other organisms (usually relatives) (Hamilton 1964; 1970; 1972; 1975; 1987), and multilevel selection (often

called group selection) where the self-interest of groups is served by the inclusion of altruistic members and selection at other levels (e.g. gene or individual) is also acknowledged (Wilson 1975a; 1976; 1977; 1990; 1997a; 1997b; Sober and Wilson 1998; Wilson 2004). Although these explanations have mathematically equivalent aspects (Wade 1980; Queller 1985; 1992b; Frank 1998; Sober and Wilson 1998; Fletcher and Zwick 2004b) they clearly differ in their conception of the causal level at which self-interest drives selection for self-sacrifice.

The use of different terms (e.g. *altruism* vs. *cooperation*) has also complicated this debate (Sober and Wilson 1998). Here I use the term *altruism* to describe self-sacrificing behavior where an actor causes others to receive fitness benefits (have more offspring) at an immediate net fitness cost to itself. By *immediate* cost I mean that there was another potential action in a particular social interaction that would have resulted in the actor having more offspring relative to others (this assumes that the actions of others around the actor remained the same for that interaction). In many cooperative interactions there is a potential to take advantage of other cooperators as defection is more advantageous. In such situations cooperation is altruistic and the terms may be used synonymously.

Many different types of such self-sacrificing, cooperative behaviors have been observed. Well defined examples from nature (some of which have been investigated in computer and mathematical models) include reproductive altruism—from female-biased sex ratios (Colwell 1981; Wilson and Colwell 1981; Herre 1985; Nunney 1985a; Avilés 1993) and slower parasite reproduction in the evolution of

virulence (Bull 1994; Lenski and May 1994; Frank 1996; Miralles, Moya, and Elena 1997) to the extreme forms of reproductive sacrifice found in eusocial insects (Rissing and Pollock 1987; Rissing et al. 1989; Bernasconi and Strassmann 1999) and in slime molds (Matapurkar and Watve 1997; Ennis et al. 2000; Strassmann 2000; Strassmann, Zhu, and Queller 2000; Velicer, Kroos, and Lenski 2000). Examples of altruism also include social forms of cooperation such as alarm calling (Sherman 1977; Hoogland 1983; Pepper and Smuts 2000), feeding restraint (Hart, Kohler, and Carlton 1991; Hemptinne and Dixon 1997; Pepper and Smuts 2000; 2002), cooperative colonization of new territory (Mitteldorf and Wilson 2000a; Giraud, Pedersen, and Keller 2002; Rainey and Rainey 2003), and cooperative hunting (Packer and Ruttan 1988; Packer et al. 1991). Genomic coordination and integration is another example of altruism where the normal—and under appreciated—fact of cooperation is revealed when it is undermined by the reemergence of lower-level selfish genetic structures (Cosmides and Tooby 1981; Eberhard 1990; Turner and Perkins 1991; Hickey 1992; Nunney and Miller Baker 1993; Werren and Beukeboom 1993; Albert et al. 1996; Camacho et al. 1997; Smith 1998). Finally, altruistic cooperation is involved in the origin of new levels of organization such as multicellularity (Maynard Smith and Szathmáry 1995; Michod and Roze 2000b; 2001) and in symbiosis (Maynard Smith 1991; Margulis 1993; Frank 1997). There are also of course many examples of selfish behavior, but the paradox for evolutionary biology is: why should there be self-sacrificing behaviors in systems evolved via natural selection—why haven't exclusively selfish strategies won the competition of *survival of the fittest*?

In addition to the examples of cooperation listed above, there are also many such behaviors among humans including well studied examples such as food sharing, risk sharing in peasant farming (Scott 1976), and sharing of limited water resources (Lansing 1994) in less modern societies, as well as more modern examples such as the tons of food quietly donated to the Oregon Food Bank each day or more celebrated sacrifices such as those of firefighters in New York's World Trade Center or soldiers on the battlefield.

In this dissertation I focus on abstract theoretical models of the evolution of altruism that contain no explicit analog of human society or psychology. I am interested in the *most fundamental mechanisms* by which cooperation can arise and be maintained. Thus my findings may be better suited for helping us understand the origins of cooperative interactions in the primordial soup (such as the cooperation among cells in multicellular organisms) than for addressing why a person would volunteer in a soup kitchen. Nevertheless, human evolution has also been subject to these fundamental processes and in the concluding chapter I will touch on how this research may influence our understanding of human behavior as well as how the emphasis on competition in evolutionary theory may provide an incomplete picture for those who justify competitive human behavior in terms of what is “natural.”

The Group Selection Controversy

One explanation for how altruistic behaviors might evolve through the process of natural selection (in both humans and other organisms) is the one given by Charles Darwin: groups of organisms whose members make sacrifices for each other might be more fit as groups and out compete other groups with less cooperative members. As Darwin put it (in somewhat antiquated terms):

It must not be forgotten that although a high standard of morality gives but 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)

Yet it is also true that *within* a group (tribe) individuals would show variation in their levels of competitiveness and the more selfish of these would be at a selective advantage compared to other members within their group. In this quote Darwin is not addressing the emphasis he placed in his earlier work (1859) on individual variation in traits leading to disproportionate individual reproductive success. This opposition of forces contributes to the confusion in analyzing the evolution of altruistic behavior. Natural selection among groups may favor groups with altruistic individuals whereas natural selection within a group favors selfish individuals. What actually evolves is a

net result that depends on the relative strengths of the selective forces occurring at different levels.

There is also considerable debate about what constitutes altruism. David Wilson (1979; 1990) distinguishes “strong” altruism where there is a net absolute cost to an altruist and “weak” altruism where there is just a relative cost compared to other group members. Others have ignored or disagreed with this distinction preferring to reserve the word altruism only for the strong category (Nunney 1985b; Maynard Smith 1998a; Nunney 2000). This debate continues 25 years later (Nunney 2000; Sober and Wilson 2000) and rests on whether one measures fitness in absolute or relative terms. For instance, David Wilson has just published a paper entitled “What’s wrong with absolute individual fitness?” (2004). I weigh in on this debate in chapter 3 using a game theoretic analysis to clarify the issues.

John Pepper (2000b) has made a similar distinction of “whole-group” altruism where altruists contribute to a common good that has some marginal benefit to themselves and “other-only” altruism where altruists give only to others and receive nothing back. This distinction may be more useful than the strong vs. weak distinction because the way in which one calculates assortment (the r term in Hamilton’s rule—discussed further on) is different for each type. Other-only altruism is always strong, but whole-group altruism can be strong or weak depending on whether the altruist’s share of the benefit it provides to the group exceeds its cost. The reason weak altruism may still be considered altruism is that *relative to others* the altruist receives less benefit than it provides others.

Despite the complicated nature of this issue, one might expect that more than 130 years after Darwin's explanation the notion of multilevel selection would be well developed in evolutionary biology. In fact this is far from true. The very existence of group-level selection as a significant natural process is highly controversial, passionately argued, and not generally supported by such prominent evolutionary thinkers as John Maynard Smith (1964; Maynard Smith and Szathmáry 1995; 1998a) and Richard Dawkins. Dawkins goes so far as to compare the appeal of group selection to amateur inventors' fascination with perpetual motion machines (1982).

However, the disfavor with group selection is not due to any definitive proof against it or lack of evidence for it (e.g. see Goodnight and Stevens 1997 for a review of experimental evidence). In *Unto Others* Elliot Sober and David Wilson (1998) chronicle the history of the group selection controversy. As they point out, this history is hardly what the philosopher of science Thomas Kuhn (1970) would call "normal science" in which competing theories and hypotheses are tested and then augmented, rejected, or accepted. Rather group selection theory was discredited without actually being disproved. Sober and Wilson (1998) present a strong case for multilevel selection theory, but the mixed reaction among the reviews speaks to the ongoing passion in this debate (Lewontin 1998; Maynard Smith 1998a; Nunney 1998; Reeve 1999; Smuts 1999; Sansom 2000).

While group selection arguments can be traced back to Darwin (see quote above), the publication of *Animal Dispersion in Relation to Social Behavior* by V.C. Wynne-Edwards (1962) who promoted group selection wholeheartedly kindled the modern

debate. Wynne-Edwards saw limiting the exploitation of resources as beneficial to groups at the expense of individuals that restrain themselves. He then argued that many group behaviors such as the flocking of birds or the vertical movement of plankton were mechanisms by which groups could assess their density and act accordingly. Arguments explaining behaviors in terms of the group or species level were common, but not well formulated. At the same time William Hamilton was developing his formal theory of inclusive fitness focused on selection at the level of genes (1963; 1964). Wynne-Edwards seems to have overstated the case for group selection. In his words: “Evolution at this level can be ascribed, therefore, to what is here termed group-selection—still an intraspecific process, and, for everything concerning population dynamics, much more important than selection at the individual level.” Wynne-Edwards realized that individual selection must also play a role, but claimed that where the two were in conflict “...group selection is bound to win, because the race will suffer and decline, and be supplanted by another in which antisocial advancement of the individual is more rigidly inhibited” (1962 p. 20). In this way he explained many adaptations as being for the good of the species.

This book provoked a strong reaction among evolutionary biologists, none so effective as that of George Williams in *Adaptation and Natural Selection* (1966). Williams’ book addressed the kinds of behaviors that Wynne-Edwards had claimed were group selected. Williams argued that when both individual and group selection scenarios could explain a behavior, individual selection was the more parsimonious and should therefore be preferred. However, Stephen Gould (Gould 2002) has argued

that Williams incorrectly applies the idea of parsimony in logical arguments (Occam's razor) to argue that simpler mechanistic explanations about how nature works should be preferred over more complicated ones. Thus, Gould argues that Williams is making a case for *reductionism* rather than *parsimony*. In addition, Sober and Wilson point out that arguments based on parsimony "certainly do not substitute for a critical test of hypotheses that make different and mutually exclusive predictions" (1998 p. 38). Yet Williams also offered such a verifiable test concerning sex ratios. He was aware of Fisher's (1930) analysis showing even sex ratios are individually selected and concluded that skewed sex ratios would be evidence of a group adaptation. Unfortunately, he lacked adequate data about real examples in nature of biased sex ratios: "I conclude that there is no evidence from data on sex ratios to support the concept of biotic adaptation" (1966 p. 152). (Williams used *biotic adaptation* to mean group adaptation.)

Ironically, only a year later Hamilton published a paper in *Science* entitled "Extraordinary Sex Ratios" (1967) in which he discussed several examples from nature of biased sex ratios. Although he did not emphasize group selection which already had a bad name, in a footnote Hamilton did acknowledge that what actually evolved was a balance between individual selection and group selection.

Unfortunately, most biologists missed the connection between Williams' verifiable test of group selection and Hamilton's paper. Instead the latter was interpreted as furthering Williams' work of explaining apparent group adaptations as truly individual. As Sober and Wilson (1998 p. 42) put it, "At the very time that group

selection theory was entering its dark age, the empirical evidence that should have counted as evidence *for* group selection was accepted as a triumph of individual selection theory!” It wasn’t until 1981 that the connection was made between Williams’ test and the accumulating evidence of biased sex ratios in nature (Colwell 1981; Wilson and Colwell 1981). Even then many biologists “couldn’t believe that an adaptation as well-documented as female-biased sex ratios might count as evidence for a theory as heretical as group selection. Even now the dust has not entirely settled.” (Sober and Wilson 1998 p. 43)

Group selection still carries a significant stigma in evolutionary biology. In *The Major Transitions in Evolution* John Maynard Smith and Eörs Szathmáry (1995 p. 12) state that: “The idea of levels of organization, and hence of levels of selection, is central to this book.” Yet only a few pages earlier they claim (1995 p. 8): “The transitions must be explained in terms of immediate selective advantage to individual replicators: we are committed to the gene-centered approach outlined by Williams (1966), and made still more explicit by Dawkins (1976).” When discussing the origin of life and molecular hypercycles they state, “Molecules that ‘sacrifice’ themselves by producing replicases that serve the good of the hypercycle are ‘altruists’ in a sociobiological sense. To stabilize the hypercycle, one needs conditions in which altruists can spread, or at least coexist with ‘cheaters’” (1995 p. 53). Then on the same page they assume molecules compartmentalized into groups will *automatically* cease to compete among themselves, and further they invoke kin selection in the case where molecules are not compartmentalized. It is very surprising that the kinship of

molecules would be appealed to without offering any plausible mechanism for such an unlikely claim! After all, these molecules do not have genes or heritable genetic traits.

Here are their words:

To put hypercycles in compartments is, in effect, to create individuals with vertical transmission of genetic information, from parent to offspring. Given vertical transmission, the evolution of cooperation between the parts of an individual is to be expected. But if, as we have argued, early chemical reactions occurred on surfaces, we can ask whether cooperation could evolve without compartments, because neighboring molecules would be genetically related. (Maynard Smith and Szathmáry 1995 p. 53)

When discussing aggregates of slime mold cells that form fruiting bodies with stalks, the authors note that individual selection would not favor the formation of non-reproductive parts such as the stalk. As with many such cases in this book, kin selection is appealed to, but Maynard Smith and Szathmáry also add:

But this cannot be the whole story, since cheaters will inevitably arise: when that happens, they are presumably eliminated because aggregates containing cheaters are less efficient in dispersing spores. This explanation is at present speculative, and should be checked by genetic studies on natural populations. (Maynard Smith and Szathmáry 1995 p. 214)

The authors' tentativeness concerning multilevel selection is evident in this last quote, as is their ambiguity about it throughout the book. In a more recent and less technical version of this work they note, "We have to explain how complex entities evolved, despite selection between their components favoring selfish behavior" (Maynard Smith and Szathmáry 1999 p. 20). Yet multilevel selection is conspicuously absent from their list of possible explanations. In fact they state that, "An adequate account requires that we explain the origin of the higher-level entity in terms of

selection acting on the lower-level entities” (Maynard Smith and Szathmáry 1999 p. 19-20). If one of our most prominent evolutionary biologists has trouble being clear about multiple levels of selection in these recent and important works, it is no wonder biologists in general are confused and divided on the subject. Below I discuss each of the three main theories of how such altruistic or cooperative behaviors arise in the face of natural selection.

Multilevel Selection Theory

In its modern form multilevel selection theory was initiated in the 1970’s by David Wilson (1975a; 1976; 1977) and supported by the empirical experiments and analyses of Michael Wade (1977a; 1978; 1979). It differs significantly from the idea of species adaptations that ignited the group-selection controversy (Wynne-Edwards 1962) discussed above in that it is focused on elucidating (with both mathematical and empirical foundations) the more complicated mechanisms of simultaneously acting selective forces. Multilevel selection theory (a generalization of group selection theory) is a straightforward extension of Darwin’s (1859) theory of natural selection. Natural selection of a trait requires three basic ingredients: phenotypic variation in the trait, fitness consequences associated with these variations, and reproduction with heritability of the trait. Multilevel selection theory says that if these three ingredients are present at any level in a hierarchy, then natural selection will take place at that level. If traits experience selective pressure at multiple levels, then what actually

evolves will be the result of the relative strength and directions of these multiple forces.

In some cases these selective pressures may act in the same direction. For example, individuals who are fast runners may be selected for within a group. In addition, groups that contain fast runners may be selected for at the group level. For instance, groups of fast runners might out compete groups of slower runners in hunting ability or warfare. This assumes that the necessary ingredients for natural selection are present at the group level—namely, variation in aggregate running ability among groups, competition between groups where running speed affects group success (fitness consequences), and new groups reflect the running abilities of their parent groups (heritability). Although it may be more common for selective forces at different levels to be complementary, group selection has been most clearly argued in cases where the selective forces are in opposition. Altruistic behavior, by definition, is such a case.

Several researchers have related multilevel selection to ecological considerations and population dynamics. For instance, several studies have shown that altruism can evolve even when group structure is ill-defined. Wilson coined the term “trait-group” (1979) to describe interactions that may be structured with regard to some traits, but not others—and not necessarily due to physically defined groups. For instance even in populations that are well mixed physically, conditional strategies (Trivers 1971; Wilson 1975a; Sober and Wilson 1998) may lead behavioral interactions to be grouped. In uniform environments groups can emerge due to restricted mobility

(viscous populations). While it has been shown that in viscous populations in uniform environments only weak altruism can evolve (Wilson, Pollock, and Dugatkin 1992), more recently, modeling of non-uniform dynamic environments has shown that the strong form of altruism can also evolve in viscous populations due to the ability of groups of altruists to disproportionately expand into areas of environmental disturbance (Harms 2000; Mitteldorf and Wilson 2000a).

In contrast to these *external* influences on population structure (e.g. environmental disturbances) populations can experience their own *internal* dynamics based on the effects of non-zero sum interactions and variation in cooperation. Leticia Avilés (1999) has shown that under harsh conditions populations may depend on the synergistic effects of group living to survive, but cooperation in established social groups may lead to disruptive population dynamics, i.e. boom and bust. Over multiple generations periodic or chaotic dynamics can arise. Whereas the above studies showed the positive effect of environmental disturbances on maintaining altruism, here internal population dynamics, due to cooperation itself, may play a role in providing increased opportunities for the formation of groups and the success of altruism.

Groups of non-similar individuals may also experience multilevel selection. This is especially noteworthy because selection among groups of symbiotic mutualists shows that gene-level selection is not fundamental since symbiots are unrelated. Steven Frank (1997) reviews several models of selection for symbiotic groupings. In addition, selection has been experimentally demonstrated at the level of communities (Goodnight 1990a; 1990b; 2000) and even at the ecosystems level (Swenson, Wilson,

and Elias 2000). Steven Gould has argued that selection may take place even at the species level over geologic time (2002). The eukaryotic cell itself may be the result of selection among groups of non-similar individuals, i.e. groups of organelles that form the eukaryotic cell (Margulis 1970; 1993; Margulis and Sagan 2002).

Finally, multilevel selection when applied to levels at and below that of the individual may be critical to our understanding of the origins of life and the formation of more complex and hierarchical living systems. As previously mentioned it is striking that two works dealing explicitly with these issues (Maynard Smith and Szathmáry 1995; 1999) rely so little on levels of selection arguments, even if multilevel selection is implicit in the mechanisms described. Some investigations of the origins of living cells depend on competition between compartmentalized protocells (Grey, Hutson, and Szathmáry 1995; Szathmáry and Maynard Smith 1997). Richard Michod and Dennis Roze (Michod 1997; Michod and Roze 2000b; 2000a; 2001; Roze and Michod 2001) have modeled the transition from single cell to multicellular life. Applying multilevel selection ideas they have examined methods by which individuals must reduce within-organism variance in order to benefit themselves by discouraging lower level selection among the individual's constituent parts. They argue that such mechanisms are critical to the evolution of higher level entities such as multicellular organisms.

Simpson's Paradox

The evolution of an altruistic trait involves a fundamental paradox that must be dealt with by all theories on the evolution of altruism—*how does a trait that causes others to have more offspring actually end up with more copies of itself?* From the group (or multilevel) selection viewpoint the trait must be decreasing within each group, but increasing overall. This is an example of Simpson's paradox (1951) and Sober and Wilson (1998) have shown it to be key in understanding multilevel selection theory—groups must vary in their composition and those with a higher proportion of altruists must contribute more to the overall population than those with a lower proportion.

A non-biological example helps illustrate the phenomena of Simpson's paradox. At the University of California at Berkeley in the 1970s, the percentage of women graduate school applicants accepted was significantly lower than the percentage of men accepted (Cartwright 1978). Yet, when the University looked at each department they found none were accepting a smaller percentage of women. Again the answer to this paradox lies in the fact that different departments varied in their contribution to the whole. In this case, the cause for this varied contribution was that women were applying in greater numbers to departments that accepted a lower percentage of applicants.

Imagine 70 women and 30 men apply to department A which has 20 positions. If there is no bias with regard to sex, 14 woman and 6 men are accepted. Department B receives 30 applications from women and 70 from men and has 50 positions. Here, if

there is no bias, 15 women and 35 men are accepted. Yet if we aggregate these results, 41 of 100 men are accepted, whereas only 29 of 100 women are accepted. These results are summarized in Table 2-1.

Table 2-1. Example of Simpson's Paradox

	Apply to A	Accepted to A (%)	Apply to B	Accepted to B (%)	Apply to A + B	Accepted to A + B (%)
Women	70	14 (20%)	30	15 (50%)	100	29 (29%)
Men	30	6 (20%)	70	35 (50%)	100	41 (41%)
Total	100	20 (20%)	100	50 (50%)	200	70 (35%)

To return to multilevel selection, groups dominated by cooperators may grow much faster than other groups due to the benefits of mutual cooperation and therefore contribute more to the total, even though *within* these faster growing groups the proportion of cooperators is declining. One of the contributions of this dissertation is to show how Simpson's paradox naturally emerges out of the n-player PD model of multilevel selection that I present in chapter 3.

The Price Equation

George Price, an eclectic and troubled American chemist, game theorist, and science journalist traveled to England in 1967 at the age of 44 to work (without any formal training in biology) on the problem of the evolution of altruism. In a brief period between 1970 and 1973 he made three very significant contributions to evolutionary theory (Frank 1995a). The first was a revolutionary hierarchical model of selection based on covariance that formalized the tension between within-group and between-

group selection (Price 1970; 1995) and helped launch the resurgence of the group selection viewpoint as multilevel selection theory:

$$\Delta Q = \frac{\text{cov}(w_i, q_i)}{E(w_i)} + \frac{E(w_i \Delta q_i)}{E(w_i)}, \text{ or equivalently} \quad (2.1)$$

$$\Delta Q = \Delta Q_b + \Delta Q_w, \quad (2.2)$$

where w_i is a measure of group fitness, here the growth rate of each group (n_i/n_i), E is the expectation, and ΔQ_b and ΔQ_w are the Price between- and within-group components of change in overall cooperator frequency, respectively. (The Price equation is covered in more detail in the next chapter.) This equation (Eq. 2.1) had an important influence on Hamilton that led him to modify his inclusive fitness theory (discussed in the next section). As Hamilton wrote, “Price had not like the rest of us looked up the work of the pioneers when he first became interested in selection; instead he had worked out everything for himself” (1996 p. 172). Hamilton goes on to relate the result of his interactions with Price: “... I now had a far better understanding of group selection and was possessed of a far better tool for all forms of selection acting at one level or at many than I had ever had before” (1996 p. 173). In chapter 3 I explore the relationship between the Price equation and our model of multilevel selection. I also develop an alternative selection decomposition that highlights the underlying assumptions of the Price equation as well as its limitations when there are interaction effects between within-group and between-group selection.

Price’s second major contribution was in launching Evolutionary Game Theory (Maynard Smith and Price 1973). Maynard Smith read Price’s initial paper on the

subject and tried to interest him in further development, but as he put it:

“Unfortunately, Dr. Price is better at having ideas than publishing them” (1972 p. viii).

Maynard Smith went on to develop Evolutionary Game Theory (1982) with little input from Price. Price’s final major contribution was to prove that Fisher’s Fundamental Theorem of Natural Selection (Price 1972; Frank 1995a) applied as broadly as Fisher had claimed—something that most evolutionary theorists had come to doubt. Shortly after this burst of amazing contributions to evolutionary theory, Price drifted away from his academic pursuits. He gave away the bulk of his possessions and spent the last couple of years of his life ministering to the less fortunate on the streets of London until his death by suicide in 1975.

Inclusive Fitness and the Progressive Generalization of Hamilton’s Rule

George Williams, an initial and formidable critic of group selection is very clear in his position that the only reasonable mechanism for the development of *group* adaptations is *group* selection:

...group-related adaptations must be attributed to the natural selection of alternative groups of individuals and that the natural selection of alternative alleles within populations will be opposed to this development. I am in entire agreement with the reasoning behind this conclusion. Only by a theory of between-group selection could we achieve a scientific explanation of group-related adaptations. (Williams 1966 pp. 92-93)

His criticism of group selection was that many traits that looked to some observers like group adaptations were really individual adaptations with advantages to

individuals, i.e. bird's morning songs and zoo plankton's migration to surface waters at night (Wynne-Edwards 1962). Group selection was theoretically possible, but he believed it was "impotent and not an appreciable factor in the production and maintenance of adaptation" (1966 p. 8). Of course altruistic behavior poses an especially difficult challenge to his line of reasoning because by definition altruistic acts are not individually advantageous. He proposed two mechanisms by which altruistic behavior could be explained by individual selection. The first relied on Hamilton's (1963; 1964) idea of inclusive fitness to explain altruism between relatives. The second claim supposed that general mechanisms for parental care were sometimes mistakenly focused on non-relatives.

Kin Selection

In kin selection Hamilton developed a gene-level explanation of how altruism among relatives could evolve. In its simplest form, a single gene can have allele X which causes an altruistic behavior at cost c to its own fitness and benefit b to recipients of the altruistic act. If the benefit is received by relatives that also contain the allele X , then the fitness of allele X can be calculated in terms of copies of itself in future generations directly descended from X and copies of X due to the benefit it gave relatives. For instance, if at cost c to its own fitness there is a $1/8$ chance that benefit b on average falls to an organism with the X allele, then $b = 8c$ is the balance point. If $b > 8c$ then the altruistic behavior would tend to increase the copies of X in future generations; if $b < 8c$ then the altruistic behavior would tend to decrease copies of X

and be selected against. Here the 1/8 ratio corresponds to a cousin because cousins share 1/8 of their alleles. More generally Hamilton “imagined that an individual might try to maximize a quantity that includes its own fitness plus the fitness of others, weighted by the degree to which they were genetically related. This became the notion of inclusive fitness.” (Sober and Wilson 1998 p. 66) In this case the inclusive fitness of an altruistic trait is given by Hamilton’s Rule (1964) which specifies in deceptively simple terms the conditions for the trait to increase:

$$rb > c, \tag{2.3}$$

where r is the “coefficient of relationship”, b is benefit to recipients, and c is the cost to the altruist.

In this general form, where the inclusive fitness is calculated in terms of the percentage of genes shared (not just the altruistic gene), synergistic and epistatic effects between genes are ignored. For instance, is having 8 times the chance of passing on 1/8 of your genes really equivalent to passing on all of your genes as a whole? In addition, this raises the issue of what it means for genes to be the same—the same *base pairs* (same allele) or the same *phenotypic consequence*? Gradually the meaning of r has moved away from the strict idea of relatedness by descent as inclusive fitness theory has matured.

Positive Assortment

Hamilton's r term is a measure of assortment and there is positive assortment ($r > 0$) when altruistic types (or more accurately behaviors¹) are clumped or grouped. For other-only traits $r = 0$ for a random binomial distribution and can range from $r = -1$ for a uniform distribution (all groups have the same proportional composition as the overall population) to $r = 1$ when all groups are homogeneous with respect to types (or behaviors). For whole-group traits $r = 1/n$ for a random binomial distribution into groups of uniform size n and can range from $r = 0$ for a uniform distribution to $r = 1$ when there is complete grouping (all groups are homogeneous). Because in whole-group altruism some of the behavior of an individual comes back to itself there is always some positive assortment except in a uniform distribution.

As mentioned above, Hamilton's interactions with Price were an important step in the progressive generalization of Hamilton's rule. Using Price's model, Hamilton (1970; 1972; 1975) showed that for selective forces favoring altruism to be stronger than individual selection favoring selfishness, there must be a positive assortment among altruists. In keeping with the focus on the gene level, Hamilton expressed this association at the genetic level, but switched to using the phrase "coefficient of relatedness" to describe his r term—the positive regression between genes regardless of relationship by descent:

¹ This distinction between genetic types and phenotypic behaviors is covered in the next section.

$$r = \frac{\text{cov}(G_A, G_O)}{\text{var}(G_A)}, \quad (2.4)$$

where G_A is the genotype of the potential actor, G_O is the genotype of others the actor interacts with, cov is covariance, and var is variance. Here r is the measure of the positive regression or assortment among *genes* for altruism, and can be interpreted as measuring how likely it is that the benefit offered by altruists fall to other altruists. As Hamilton put it, “the existence of the positive correlation r could be interpreted as implying in this case that there is a chance r that the b units of fitness are definitely given to a fellow altruist, while with chance $(1-r)$ they are given ... to a random member of the population.”¹ (1975 p. 140). The higher this chance of genes for altruism benefiting other genes for altruism (r), the lower the margin of benefit over the sacrifice (b/c) needed in order for altruism to evolve. Groups are implicit in both of Hamilton’s formulations, but the first concerns benefit to relatives in implied *kin groups* (kin selection), whereas here benefit goes to fellow altruists (those with the altruist gene) regardless of lineage.

Richard Dawkins misses the revision that Hamilton (1975) makes to his inclusive-fitness concept and sees this 1975 paper as being supportive in the battle against group selection which he colorfully characterizes thus:

We painfully struggle back, harassed by sniping from a Jesuitically sophisticated and dedicated neo-group-selectionist rearguard, until we finally regained Darwin’s ground, the position that I am characterizing by the label ‘the selfish organism’, the position which, in its modern form, is dominated by the concept of inclusive fitness. (Dawkins 1982 p. 6)

¹ The variable names have been changed in this quote to match the notation used in this dissertation.

Yet, here are Hamilton's own words from his paper:

Because of the way it was first explained, the approach using inclusive fitness has often been identified with 'kin selection' and presented strictly as an alternative to 'group selection' as a way of establishing altruistic social behavior by natural selection (e.g. Maynard Smith 1964; Lewontin 1970). But the foregoing discussion shows that kinship should be considered just one way of getting positive regression of genotype in the recipient, and that it is this positive regression that is vitally necessary for altruism. Thus the inclusive-fitness concept is more general than 'kin selection'. (Hamilton 1975, pp.140-141)

In a later work, Hamilton (1987) was even more direct: "it obviously makes no difference if altruists settle with altruists because they are related... or because they recognize fellow altruists as such, or settle together because of some pleiotropic¹ effect of the gene on habitat preference." In other words, the interactions among members of a population must be clumped (or grouped) with respect to altruistic traits in order for altruism to evolve—they must be disproportionately concentrated within some groups and not others, but grouping of related individuals is just one (if the most common) way to generate differences between these groupings.

This is much different than Hamilton's original notion of inclusive fitness. Originally, inclusive fitness (kin selection) was seen as bringing seemingly group adaptations back under the more parsimonious umbrella of individual selection. In contrast, here kin selection is being brought under the broader umbrella of group effects. It is the positive assortment of altruists within groups that gives altruistic traits their "inclusive fitness." Sober and Wilson (1998 p. 78) point out that in a survey of

¹ *Pleiotropic* means producing many effects or multiple effects from a single gene. So in other words, if an altruistic gene also causes a certain habitat to be preferred this would result in a positive assortment among those with the gene regardless of genealogical kinship.

Science Citation Index during 1994, the original version of Hamilton's theory from two papers in 1964 was cited 115 times, but the updated 1975 version was cited only four times.¹ As they put it:

For much of the evolutionary community, the theory of kin selection was set in stone during the 1960s and thereafter lost its capacity for fundamental change—even at the hands of its own creator! If this is the situation inside the field of evolutionary biology, then it should come as no surprise that scholars from other disciplines who are interested in evolution see group selection primarily as a theory that died many years ago, along with the prospects for genuine altruism in nature. (Sober and Wilson 1998 p. 78-79)

As we have seen, the original arguments developed for inclusive fitness took a gene-level view of selection rather than the more traditional individual level (Darwin 1859). Richard Dawkins (1976; 1982; 1994; 1995) has taken these beginnings to an extreme reductionism in which genes are the only true level of selection. Yet in order to account for the fact that genes cannot go it alone, Dawkins invokes the concept of “vehicles of selection” in which genes must travel together and be selected together. He uses the metaphor of a rowing team in which none of the rowers can reach the finish line ahead of the vehicle in which they all travel. This argument is used to account for why selfish genes would need to cooperate with each other.

If genes were truly free to compete with each other within an individual, we would expect more fecund genes to be more selfishly fit than others. Yet suppression of gene-level selection has been suggested as a necessary step in the evolution of coordinated, cooperative genes within individual organisms (Michod 1996; 1997;

¹ To be fair, Hamilton seems to have not promoted his updated views on group selection very forcefully and this may have been due to its general disfavor among biologists.

Michod and Roze 2000b; 2000a; 2001). The fair process of meiosis itself, in which each allele has an equal chance of making it into a sex cell, is typical of the multicellular colonies we generally call individual organisms and may be one mechanism by which the expected competition between genes is suppressed (Michod 1996). In an article by Leticia Avilés, myself, and Asher Cutter recently accepted to *The American Naturalist*, we support this general idea by showing that a lottery mechanism where helping behaviors are assigned at random can allow much higher levels of cooperation to evolve among non-relatives than in the absence of such a mechanism (Avilés, Fletcher, and Cutter 2004). If vehicles of selection can rein in genes so they restrict their own selfish fitness for the good of the individual, then the selection of larger vehicles composed of individuals, i.e. social groups, could lead to cooperation and self-sacrifice by individuals for the good of the group.

The concept of the vehicle is an admission of multiple levels of selection by reductionism's strongest advocate in evolutionary biology. Dawkins' (1976; 1982; 1994; 1995) insistence on the primacy of genes as replicators is irrefutable, but also not useful for describing higher level phenomena. Explaining, for instance, the cooperation of human parents in raising a child only in terms of gene competition ignores many other emergent levels of interaction and causation. It is like trying to explain the Gulf Stream current only in terms of electrons, protons, and neutrons. Although it is true that all the molecules that make up the water, air, ocean floor, sun, and other factors that influence the current are composed entirely of these elementary particles, like the gene-only replicator view, it is not useful for understanding many

real world phenomena taking place at levels higher than that of the electron. As I show in the next section and also in Chapter 5, one of the biggest challenges to the selfish gene viewpoint comes from advances in inclusive fitness theory itself—the ability to use the positive assortment among emergent *phenotypes* rather than genotypes in Hamilton’s rule.

Phenotype vs. Genotype

Over the years there had been various reports of Hamilton’s rule failing to make accurate predictions (e.g. Cavalli-Sforza and Feldman 1978; Boorman and Levitt 1980; Karlin and Matessi 1983). David Queller noticed that in these models, one of two things was true: 1) there was a difference between phenotype and genotype frequencies or 2) the fitness functions for the benefits of altruism were non-additive. Hamilton’s rule assumes both a one-to-one mapping between phenotype and genotype and additive fitness effects and fails when these assumptions are not met. Queller’s insight was to see that altruism would evolve if there were a positive assortment among *helping behaviors* even in the absence of any positive assortment among *selfish genes*. Queller developed a version of Hamilton’s rule in which the r term is a measure of the assortment between an altruist’s genes and the helping behaviors (phenotypes) of itself and those it interacts with:

$$r = \frac{\text{cov}(G_A, P_O)}{\text{cov}(G_A, P_A)}, \quad (2.5)$$

where G_A is the genotype of the actor, P_O is the average phenotype of others, and P_A is the average phenotype of the actor. This is regardless of how the behaviors are genetically encoded in others (notice no G_O term) and whether or not these phenotypes emerge via environmental interactions (including the behavior of others) or have a one-to-one mapping to their genetic basis. In other words, an altruistic gene can increase in the population if the organism it resides in receives fitness benefits from others more often than average population members—even if these others share *no genes* with the altruist. This allows Hamilton's rule to be used to understand the evolution of conditional strategies (discussed in the next section) as well as cooperation between members of different species in symbiotic relationships. Table 2-2 summarizes the historical changes in the meaning of r in Hamilton's rule. Although several other versions of r have been suggested (see Pepper 2000b for review), these represent the most significant advances in generalizing inclusive fitness theory.

Table 2-2. Historical Advances in Generalizing the r term of Hamilton's Rule.

Contribution Reference	Meaning	Calculated with
Hamilton 1964	Relationship by descent to kin	Genealogical relationship values
Hamilton 1970, 1972, 1975	Positive regression (assortment) between <i>genetic</i> type of actor and <i>genetic</i> type of others	$r = \frac{\text{cov}(G_A, G_O)}{\text{var}(G_A)}$ (Eq. 2.4)
Queller 1985, 1992a, 1992b	Positive regression (assortment) between <i>genetic</i> type of actor and <i>phenotypic</i> type (behaviors) of others	$r = \frac{\text{cov}(G_A, P_O)}{\text{cov}(G_A, P_A)}$ (Eq. 2.5)

Notice that each of the advances in the meaning of r is more general, but does not lose anything when applied under more restrictive conditions. If one takes Eq. 2.5 and assumes phenotypic frequency equals genotypic frequency, then this becomes Eq. 2.4 (note that $\text{cov}(X, X) = \text{var}(X)$). If one further assumes that any positive assortment in genetic frequency is due to relationship by descent then r as the “coefficient of relatedness” in Eq. 2.4 becomes r as the “coefficient of relationship” used in the original Hamilton’s rule, Eq. 2.3. In Chapter 5 I use the most general form of Hamilton’s rule to show the fundamental similarities between reciprocal altruism theory (discussed next) and inclusive fitness theory.

Reciprocal Altruism and Evolutionary Game Theory

While game theory was first established in the field of economics as a framework for studying conflict and cooperation in social systems (von Neumann and Morgenstern 1947), it also provides an important tool for reasoning about the evolutionary consequences of social situations as well as particular interactions between organisms. For instance, Hamilton used it to describe an “unbeatable strategy” in the evolution of female-biased sex ratios (1967; 1971) and Price first used it to reason about ritualized fighting in animals (Frank 1995a). Starting with Price’s work John Maynard Smith developed evolutionary game theory including the notion of an evolutionary stable strategy (ESS) that aids in reasoning about long term evolutionary outcomes when

different behavioral strategies interact (Maynard Smith and Price 1973; Maynard Smith 1982). Robert Trivers (1971) used game-theoretic ideas to explain how seemingly altruistic behaviors among non-relatives could evolve through an exchange of benefits, e.g., “I’ll scratch your back if you scratch mine.” He was well aware of the potential for cheating in such situations and explored mechanisms that might overcome the temptation to defect, including repeated interactions. The political scientist Robert Axelrod and Hamilton expressed game theory’s usefulness in capturing the tensions involved in the evolution of altruism this way:

Many of the benefits sought by living things are disproportionably available to cooperating groups. While there are considerable differences in what is meant by the terms "benefits" and "sought," this statement, insofar as it is true, lays down the fundamental basis for all social life. The problem is that while an individual can benefit from mutual cooperation, each one can also do even better by exploiting the cooperative efforts of others. Game theory in general, and the Prisoner's Dilemma game in particular, allow a formulation of the strategic possibilities inherent in such situations. (Axelrod and Hamilton 1981 p. 1391)

Note that the different levels (groups vs. individuals) are explicit in this quote, yet Axelrod and Hamilton propose that reciprocal altruism (which they then go on to model with the PD) is a distinct *process* from inclusive fitness theory (1981). In Chapter 5 I show that inclusive fitness and reciprocal altruism *are not fundamentally distinct processes*. That is, the fundamentals of the evolution of altruism captured in the most general form of Hamilton’s rule (i.e., the balance among costs, benefits, and the assortment of altruistic behaviors) allow it to be applied to reciprocal altruism as well. In addition, although Axelrod and Hamilton used the PD as a model of reciprocal

altruism, in Chapter 3 I show that the PD (in an n-player version) is also a model of multilevel selection and inclusive fitness.

The Prisoner's Dilemma

The prisoner's dilemma is the most famous and widely used game of game theory. In this imaginary situation two prisoners have been arrested for a serious crime and are being interrogated separately by the police. The prisoners cannot communicate and neither knows what the other will do. If they cooperate with each other and keep quiet there is only enough evidence to convict them of a lesser crime and they will both do minimal jail time. If however one of them turns State's evidence and rats on (defects from) the other who remains silent, then the defector does no jail time, but the cooperator receives a very harsh sentence. In the case where both confess (defect), they both receive an intermediate amount of jail time.

An example of the PD is represented by the payoff matrix in Figure 2-1 below where the numbers represent *a positive measure* of utility. It is simpler to think in terms of positive payoffs rather than the negative payoff of jail time, but the essential features are the same.

		<i>Player 2</i>	
		C	D
<i>Player 1</i>	C	3 / 3	0 / 5
	D	5 / 0	1 / 1

Figure 2-1. Payoff Matrix for a 2-Player Prisoner's Dilemma.

C represents cooperation and D represents defection. The strategies for player 1 are represented as rows and for player 2 as columns. The payoff values for each pair of strategies that meet are listed as *player 1 payoff/player 2 payoff*.

The PD is a non-zero sum game—the sum of the two players' scores varies for different strategy combinations. An essential feature of this game is that there is no opportunity for the players to coordinate their actions and, more critically, there is no higher-level mechanism to enforce any agreements. The players must individually decide which strategy to play without any guarantees about the actions of the other player. To understand the paradox at the heart of the PD, imagine player one trying to decide what to do and using the payoff matrix to reason with as follows: Regardless of what player 2 does, I should defect—I will get 5 instead of 3 if player 2 cooperates, or 1 instead of 0 if player 2 defects (remember these are positive utility values, not jail time). Player 2 is in symmetrical situation and rationally also chooses to defect. So each player follows this *dominant* strategy and gets a payoff of 1, but if they had both cooperated they would have each gotten a payoff of 3. The essential feature of the PD is that the dominant individually-rational strategy for each player leads to a collective sub-optimal or irrational outcome. This outcome (1/1) is “non-Pareto-optimal” because there is another outcome (3/3) to the game that increases the utility of one player without cost to the other. In fact in this case, both players can do better.

Iterated Prisoner's Dilemma

The simulation experiments of Axelrod (Axelrod and Hamilton 1981; Axelrod 1984) addressed the question of whether individual rationality would still favor defection if instead of playing just once or at random, players were forced to play a series of iterated PD games with a probabilistic (rather than fixed) number of iterations. This

work extended and modeled the original idea of reciprocal altruism (Trivers 1971). Axelrod sponsored a tournament in which various strategies implemented in computer programs were played against each other pair-wise in a round robin so that each program played every other program including itself. Each pair-wise iterated game consisted of 200 interactions and the payoff matrix was the one illustrated above in Figure 2-1. Therefore two always-defect (ALLD) strategies playing each other would get a score of 200 each, two always-cooperate (ALLC) strategies playing each other would get a score of 600, and an ALLD playing an ALLC would get respective scores of 1000 and 0. For a summary of subsequent studies of the evolution of cooperation and altruism based on the iterated PD see Dugatkin (1997).

Surprisingly the simplest strategy in the Axelrod tournament turned out to also be the best. Proposed by Anatol Rapoport, it is called Tit-for-Tat (TFT). This strategy cooperates in the first interaction and then always plays the strategy its opponent used in the last encounter. Consistent with the theory of reciprocal altruism, TFT players need the ability to remember previous actions by competitors. The TFT strategy is willing to cooperate, swift to punish a defection, yet forgiving in that it will return to cooperation if its opponent makes the sacrifice of cooperating while TFT is defecting.

Axelrod also explored the iterated PD in an evolutionary context. Here generations of tournaments were held in which the number of players of a particular type (strategy) in the new generation was based proportionately on its tournament scores from the last round. Interestingly, Axelrod wrote to Maynard Smith in England for advice about making evolutionary versions of his iterated PD models. Maynard

Smith informed him that one of the preeminent evolutionary theorists, William Hamilton, was on his own campus at the University of Michigan. Axelrod wrote a seminal paper co-authored with Hamilton entitled “The Evolution of Cooperation” (1981).

Unfortunately, the reason that TFT came out on top in both the tournaments and evolutionary models has been widely misunderstood as being due to its individual fitness or “unbeatability.” As Sober and Wilson (1998), and even Rapoport (1991) have pointed out, TFT can never beat an opponent in any pair-wise iterated interaction because it never defects unless it has already been on the short end of a defection from its opponent. As Rapoport put it, “in every paired encounter, Tit-for-Tat must either draw or lose. It can never win a paired encounter” (Rapoport 1991 p. 93). In this sense TFT is altruistic at the individual level because it often gives more utility than it gets and it never gains more than its opponent does.

The reason TFT won (or was selected for) hinges on the fact that it was often able to play other TFT or similar altruistic strategies where it could run up its accumulated score. Even though in individual competition TFT is inferior, for example to ALLD which is the most fit un-exploitable individual strategy, pairs (groups) of TFT accumulate higher scores than pairs (groups) of ALLD. In the analysis of why TFT was a successful strategy, Axelrod (1984) does note that it tended to score especially well (close to 600) when it played similar strategies, but he does not recognize this *as a group effect*. Maynard Smith comments that, “the programs were ranked according to the total payoff accumulated (not, it should be noted, according to the number of

opponents defeated in the individual matches)” (1982 p. 168). Yet neither he nor Axelrod distinguish the individual and group levels of competition present in this tournament which is obscured by the cumulative method of scoring. Sober and Wilson (1998) emphasize the importance of identifying and separating out the selective forces at different hierarchical levels. They argue that much of the controversy surrounding group selection is due to a failure to do so.

More complicated versions of the iterated PD and their effects on the success of TFT strategies have also been studied. In one investigation the degree of cooperation (not just cooperate or defect) was allowed to evolve (Freen 1996). In another the amount of sacrifice and payoff evolved (Doebeli and Knowlton 1998). Ecological extensions similar to those discussed under multilevel selection have also been applied to iterated PD models. For instance Ferriere and Michod (1995; 1996) have shown the increased fitness of TFT over ALLD in spatial models with player mobility. The effect of population dynamics on the fitness of TFT has revealed that boom and bust dynamics can favor the less individually fit TFT when in direct competition with ALLD (Doebeli, Blarer, and Ackermann 1997). There have also been other studies showing that under different circumstances other (but related) strategies can do better than TFT (see Brembs 1996; Dugatkin 1997 for reviews).

Up until now the notion that TFT (or similar strategies) does well because of high scores received when grouped (paired) with other TFT or cooperating behaviors has not been well formalized. One of the innovations of this dissertation is to demonstrate how to make this argument precise by using Queller’s general versions of Hamilton’s

rule discussed above. Interestingly, Queller (1985) in his original paper developing his equations notes that they should apply to reciprocal altruism, but he says this without offering a demonstration or example. In two later papers (Queller 1992b; 1992a) that expand this original work no mention is made of its relationship to reciprocal altruism.

There is also no indication that Hamilton ever tried to apply his inclusive fitness techniques to Axelrod's models even though, as mentioned, they worked together on a paper (Axelrod and Hamilton 1981) which became a chapter in Axelrod's book (1984). Also in Hamilton's collected papers (Hamilton 1996; 2001), the significance of Queller's generalization is never noted. Another contribution of this dissertation is to draw attention to Queller's fundamental advance in moving from the assortment among *genes* to the assortment among *behaviors* in inclusive fitness theory. This allows Hamilton's rule to be applied much more broadly—specifically to situations where genotype and phenotype differ, as in conditional behaviors. As far as I know, until now no one has actually demonstrated how to apply Queller's equations to reciprocal altruism, as I do in Chapter 5.

N-Player Prisoner's Dilemma

(The Tragedy of the Commons)

The n-player, as opposed to 2-player, PD offers a straightforward way of thinking about the tension between the individual and group levels. In real-world biological and social systems the effects of cooperation or defection are often distributed diffusely to other members of a group, i.e., they do not necessarily arise via pair-wise interactions.

When there is a common and finite resource, each individual benefits by using more than its share of that resource, but when all players apply this individual rationality it can lead to collective irrationality. For example, each country that fishes international waters can increase its utility by taking more of the fish in this common resource, but as more and more countries overfish, the common stock is depleted beyond where it can quickly replenish (Roberts 1997; Tierney 2000). In subsequent years all have less and this leads to decreased utility for both countries that over fish (defectors) and those that don't (cooperators).

Another example is the tension between the common good that we fund through taxation and the individual advantage of paying fewer taxes. The benefit to an individual tax cheat is paid for by either marginal increases in tax rates or marginal decreases in services such as roads, schools, courts, parks, police, etc. There is an incentive to pay less taxes because the gain is more than the loss due to a slight increase in the rate or the marginal lessening of services, but again if the collective consensus is to pay less and less taxes then all may be worse off in the absence of the public services mentioned above.

The fishing example is commonly known as a “tragedy of the commons,” (Hardin 1968) whereas the taxation example is called the “free rider problem” (McMillan 1979; Avilés 2002), yet both are examples of an n-player PD. (Sometimes “tragedy of the commons” is used more generally to refer to both types of situations.)

Interestingly, even though the equivalence (Hardin 1971; Hamburger 1973) between the game-theoretic framework and the less formal notion of a tragedy of the commons

was shown in the early 1970's (not long after Garret Hardin's seminal paper in 1968) it is still not widely appreciated. For instance, a recent entire issue of *Science* magazine was dedicated to discussing the tragedy of the commons "25 years after" Hardin's paper, but nowhere is the prisoner's dilemma mentioned.

George Price was the first to recognize the connection between social dilemmas and levels of selection as demonstrated in the quote at the start of the next chapter. Wilson (1977) in an early graphical illustration of multilevel selection presented a figure and a model satisfying the conditions for an n-player PD, but did not note their game-theoretic interpretation. Surprisingly, despite these hints in the literature, as far as I know the n-player PD has not previously been used explicitly to model multilevel selection as I do in the next chapter.

In the next three chapters I delve more deeply into a few of these issues by presenting a n-player PD model of the evolution of altruism (Chapter 3), demonstrating that altruism (even strongly defined) can evolve in randomly formed groups (Chapter 4), and showing how Hamilton's rule can be applied to reciprocal altruism (Chapter 5).

... the cases discussed where individual selection decreases group fitness are closely and deeply analogous to economic effects recently discussed by Hardin in a paper entitled "The Tragedy of the Commons" (Hardin 1968).

George Price, in 1969 grant proposal (quoted in Frank 1995a)

Chapter 3

The Evolution of Altruism:

Game Theory, Multilevel Selection, and

Inclusive Fitness¹

Introduction

The evolutionary mechanisms by which altruistic behaviors may evolve have been vigorously debated over the last several decades. The most prominent explanations are reciprocal altruism (Trivers 1971; Axelrod and Hamilton 1981; Axelrod 1984), inclusive fitness (Hamilton 1964; 1970; 1975), and multilevel selection (Wilson 1977; Wade 1978; Wilson 1997a; Sober and Wilson 1998). The iterated prisoner's dilemma (PD) naturally lends itself to the study of reciprocal altruism (Axelrod 1984; Dugatkin

¹ This chapter is based on a completed manuscript to be sent for publication and which is coauthored by Martin Zwick, my advisor.

1997), yet real-world biological and social systems often do not involve pair-wise interactions or knowledge of past actions, and consequences of cooperation and defection may be distributed diffusely. The *n*-player prisoner's dilemma (NPD) is a model that captures the diffuse harm to the common good that self-interested behaviors may cause. The NPD encompasses both (Hardin 1971) problems of exploitation of a common resource ("tragedy of the commons" (Hardin 1968)), and problems of inequitable contributions towards a common good ("free-rider problem" (McMillan 1979; Avilés 2002)). Using an NPD model in multiple groups I demonstrate that altruism can evolve, despite the higher within-group fitness of non-altruists, in the absence of reciprocity, cognition, or memory.

The minimal conditions for the evolution of altruism are captured in this model and are: 1) non-zero-sum fitness functions for altruistic behavior, and 2) sufficient non-uniform population structure with respect to altruistic behaviors. Heritability is also assumed in this evolutionary model. The simplicity of this model allows us to connect explicitly the NPD to the other two related theories (Wade 1980; Queller 1985; 1992b; Frank 1998; Sober and Wilson 1998) of altruism evolution: inclusive fitness and multilevel selection, a connection which has not previously been made explicitly, despite hints in the literature. Others (Price in Frank 1995a; Frank 1998; Leigh 1999; Michod 1999; Hauert et al. 2002; Bowles, Choi, and Hopfensitz 2003) have commented on the similarities between social dilemmas and multilevel selection without addressing them explicitly, as I do here.

The NPD model here captures essential factors in the evolution of altruism in model parameters that relate simply to Hamilton's Rule, the Price covariance equation, and a useful alternative selection decomposition I present here. Contrasting this alternative decomposition with the Price equation, I demonstrate interaction effects between within- and between-group selection which are not captured in either decomposition alone. This game theoretic model also allows me to clarify essential similarities between strong and weak altruism undergoing natural selection and suggests a framework for unifying different approaches to the evolution of altruism. Lastly, in contrast to conventional wisdom, I demonstrate that even strong altruism can be selected for and maintained using periodic random groupings of non-conditional strategies.

Methods

The N-Player Prisoner's Dilemma Model

In the simplest form of the model there are only two groups with no dispersal (I relax both these assumptions below). Groups vary initially only in their sizes, n_i , and in the fraction of cooperators, q_i , in each group i . There are no strategies besides always-cooperate (ALLC) and always-defect (ALLD). I follow the frequency of cooperators in each group and across the whole population. Figure 3-1 illustrates a simple NPD with parallel linear fitness functions, w_a and w_s , that gives the fitness *per individual* cooperator (altruist) or defector (selfish) in the vertical axis and where q_i is the

horizontal axis. There are two parameters to this NPD: the slope, b , of the fitness functions and, c , their difference in intercept. (For simplicity I set the cooperator's intercept to 0 so the difference is the defector's intercept value.) The parallel lines used in this model are the simplest fitness curves that satisfy the NPD. The cost of being a cooperator vs. a defector is the intercept difference c . The benefit provided by each cooperator to the group is b . To see this note that the added benefit to *each* group member (including the focal player) in having one additional cooperator in the group (vs. a defector) is b/n_i (the change in q_i is $1/n_i$) and therefore the *total* benefit produced by a cooperator for all group members is $n_i(b/n_i) = b$.

Figure 3-1. Simple N-Player Prisoner's Dilemma.

Fitness functions for individual cooperators (w_a) and defectors (w_s) given as a function of the frequency of cooperators in a group (q_i). The two solid lines have slope b . The dashed line indicates the average fitness (w_{av}), which has a positive slope. The intercept difference of the two functions is given by c .

The defector's fitness line dominates the cooperator's at all q_i and therefore cooperation always involves an altruistic sacrifice relative to defection. The deficient outcome of the NPD here is the fact that the fitness to defectors when all players in a

group defect ($q_i = 0.0$) is *lower* than the fitness to cooperators when all group members cooperate ($q_i = 1.0$), that is, $b > c$ and this is the minimum parameter condition for a NPD. This is also the condition for beneficial non-zero-sumness, i.e., the benefit created by a cooperator exceeds the cost to the cooperator and the average fitness line (w_{av}) has a positive slope. Since defection dominates cooperation, this deficient outcome is an attractor of the dynamics. Thus in a PD situation within-group selection *does not maximize individual fitness*.

At each generation the number of cooperators (and defectors) within each group is increased by the number of individuals utilizing this strategy multiplied by its fitness payoff per individual:

$$a_i' = a_i[1 + w_a(q_i)] \text{ and} \quad (3.1)$$

$$s_i' = s_i[1 + w_s(q_i)], \quad (3.2)$$

where a_i and s_i are the number of cooperators and defectors respectively in group i and primed terms represent values after reproduction. These fitness functions can be interpreted as overlapping generations or as discrete generations where the fitness independent of the altruistic trait is one offspring per individual. Here fitness is fecundity and offspring counts (including fractional components) are determined by the fitness functions from Figure 3-1:

$$w_a(q_i) = bq_i \text{ and} \quad (3.3)$$

$$w_s(q_i) = bq_i + c. \quad (3.4)$$

To aid in comparisons among runs, each generation the total population size is proportionally scaled to its original size, preserving each group's q_i value. For

convenience we define total population variables $A = \sum a_i$, $S = \sum s_i$, $N = A + S$ and $Q = A/N$.

Hamilton's Rule

In this simple model the condition for an increase in the overall frequency of cooperators from one generation to the next, $Q' > Q$, can be derived starting with the NPD Eqs. 3.3 and 3.4. *Appendix A* gives this derivation which results in a form of Hamilton's rule (1964; 1970; 1975):

$$rb > c. \quad (3.5)^1$$

Here we find that the r value in this derivation can be expressed as the between-group over total variance in the cooperate trait,

$$r = \frac{\text{var}_b(q_i)}{\text{var}_t(Q)}, \quad (3.6)$$

where $\text{var}_b(q_i)$ is the weighted between-group variance and $\text{var}_t(Q)$ the total variance among individuals in the cooperator trait. (I will refer to the expression in Eq. 3.6 as the variance ratio.) This is consistent with previous work showing that for altruists that benefit the whole group (Pepper 2000b) as in the model above, r can indeed be expressed in terms of the variance ratio (Breden 1990; Queller 1992b; Frank 1998). For altruism to evolve, the benefit b must not only be greater than the cost c (the minimum NPD condition), but the benefit must be greater than the cost even when the benefit is discounted by the variance ratio. The more structured the population with

¹ Equations will sometimes be repeated for convenience. For instance, Eq. 3.5 is also Eq. 2.1 from Chapter 2.

regard to cooperative interactions (i.e. the closer the variance ratio is to 1), the less non-zero-sum synergy (i.e. $b > c$) is required.

The meaning of r has changed over the years from a simple measure of relationship via descent (Hamilton 1964) to various statistical measures of similarity (Hamilton 1970; 1975; Wade 1980; Queller 1985; 1992b; Frank 1998). When the benefits provided by an altruist are divided among only *others* in the altruist's group, then r is calculated differently using the covariance between the frequency of the trait in each population member and the average frequency in the *others* each member interacts with in its group (Hamilton 1972; Pepper 2000b). In either case r specifies the assortment between potential actors and their recipients. As we will see in Chapter 5, a more general form of Hamilton's rule shows that it is actually the assortment among altruistic *behaviors* that is most fundamental to the evolution of altruism.

Results and Discussion

Simpson's Paradox

Even though Q can increase, i.e. $Q' > Q$, when Hamilton's rule is satisfied, the frequency of cooperators decreases in every group, i.e. $q_i' < q_i$ for all groups. This is an example of Simpson's paradox (1951), which is key to understanding the role of population structure in the evolution of altruism (Sober and Wilson 1998). Figure 3-2 shows a run (Run 1) in this model where Simpson's paradox is evident. All runs used in Figure 3-2 to Figure 3-4 have a total population of 1,000 divided into two groups

with varying NPD parameters and initial population structures. In Run 1 the *overall* cooperator frequency Q is initially 0.5 and the group sizes are equal, but the *group* cooperator frequencies are $q_1 = 0.1$ and $q_2 = 0.9$. This population structure gives a variance ratio of 0.64, well above the c/b ratio of 0.2 for this run, and therefore Q increases in accord with Hamilton's rule, even though q_1 and q_2 both decrease monotonically. Figure 3-2(a) demonstrates Simpson's paradox in this model by showing the fraction of cooperators in the two groups and in the total population for Run 1. Note that the fraction of cooperators monotonically decreases in both groups while initially increasing overall. This Simpson's paradox effect is due to group 2 (cooperator dominated) rapidly expanding, while group 1 (defector dominated) is shrinking, which is shown in Figure 3-2(c). At the peak of total cooperation in Run 1, group 2 comprises over 95% of the total population.

The Simpson's effect is transient without mechanisms for reestablishing variation among groups, which I demonstrate later. The changes in group size and composition affect the variance ratio (r)—in the case of Run 1, the ratio decreases steadily (Figure 3-2(b)). The generation when the ratio variance drops below c/b is precisely the point when the overall cooperator frequency begins to decline. A vertical dashed lines with arrows indicate this corresponding point for Run 1 in Figure 3-2.

Figure 3-2. Dynamics in q , r , and n for a Typical NPD Run with 2 Groups. Results for Run 1 with a total and constant population of 1,000 and initial groups sizes of 500 in each group. Parameter values are shown in Table 3-1. (a) Frequency of cooperators vs. generation for the total population (Q) and for each group (q_1 and q_2). (b) The between-group over total variance ratio in cooperation frequency (r) vs. generation. The c/b value is also shown. A vertical dashed line with arrows indicates the critical point in Run 1—when r drops below c/b in (b), Q begins to decline in (a). Panel (c) shows how the size of each group changes over the run.

Figure 3-3(a) and (b) are similar to Figures 3-2(a) and (b), but four additional runs with a variety of NPD parameters are compared. The parameters from all five runs are given in Table 3-1. Again, Figure 3-3(a) shows the overall cooperation frequency Q ,

while Figure 3-3(b) gives the variance ratio for each of these runs. The ratio c/b is chosen arbitrarily to be the same for all runs and is shown by a thick unadorned horizontal line in both Figure 3-2(b) and Figure 3-3(b).

Table 3-1. Parameter Values for Runs 1-5 used in Figure 3-2 and Figure 3-3.

Run	a_1	s_1	a_2	s_2	c	b	c/b
1	50	450	450	50	0.20	1.00	0.2
2	50	450	450	50	1.00	5.00	0.2
3	50	450	450	50	0.03	0.15	0.2
4	9	1	1	989	0.10	0.50	0.2
5	50	670	250	30	0.10	0.50	0.2

Figure 3-3. Results for 4 Runs with Various Parameters (see Table 3-1). All runs done with 2 groups and total populations of 1,000. (a) Frequency of cooperators (Q) vs. generation. (b) The between-group over total variance ratio in cooperation frequency (r) vs. generation. The c/b value is also shown. Vertical dashed lines with arrows indicate corresponding points in runs—when r drops below c/b in (b), Q begins to decline in (a).

Runs 2 and 3 show the effects of varying the magnitude of b and c , while keeping c/b and initial population structure (variance ratio) the same as in Run 1. In Run 2 with higher magnitudes the increase and subsequent decrease in Q happens more quickly; in Run 3 with lower magnitudes, the pattern is stretched out over many more generations. Run 4 demonstrates that even with low initial Q values ($Q = 0.01$), a sufficiently high variance ratio can lead to a dramatic increase in cooperators. Run 4 also shows that the variance ratio need not always decrease (Figure 3-3(b)). Changes in group size and composition can in some cases cause the variance ratio to increase transiently without external causes or mixing¹. Finally, Run 5 makes the point that even with an initial high frequency of cooperators, $Q = 0.9$, cooperators will not increase without a sufficient variance ratio. Here the variance ratio is less than c/b and therefore Q decreases monotonically.

The Price Equation

In the runs discussed so far, the transient increase and subsequent decrease in cooperator frequency highlights competing forces—the overall frequency of cooperators, Q , increases while the *between-group* selective force dominates and decreases when the *within-group* force is stronger. Price introduced a covariance equation (1970) which allows us to partition the change in overall cooperator frequency, $\Delta Q = Q' - Q$, into within- and between-group components:

¹ The fact that r can increase from its initial value is critical for strong altruism to evolve in randomly formed groups. This is discussed later in this chapter and more extensively in Chapter 4.

$$\Delta Q = \frac{\text{cov}(w_i, q_i)}{E(w_i)} + \frac{E(w_i \Delta q_i)}{E(w_i)}, \text{ or equivalently} \quad (3.7)$$

$$\Delta Q = \Delta Q_b + \Delta Q_w, \quad (3.8)$$

where w_i is a measure of group fitness, here the growth rate of each group (n_i'/n_i), E is the expectation, and ΔQ_b and ΔQ_w are the Price between- and within-group components of change in overall cooperator frequency, respectively.

As shown in *Appendix B*, these Price equation components of change can be rewritten as:

$$\Delta Q_b = Q^* - Q \text{ and} \quad (3.9)$$

$$\Delta Q_w = Q' - Q^* \text{ where} \quad (3.10)$$

$$Q^* = \frac{\sum q_i n_i'}{N'}. \quad (3.11)$$

The Q^* term in ΔQ_b plays the role of an idealized Q' in which the *before-selection* q_i values are applied to the *after-selection* group sizes, n_i' . The corresponding within-group expression corrects for the ignored changes in cooperator frequency within groups. This Price equation idealization, as also shown in the *Appendix B*, can be expressed in terms that highlight its connection to our linear NPD model,

$$Q^* = \frac{\sum a_i (1 + w_{av} q_i)}{n_i'}, \quad (3.12)$$

where $w_{av} q_i$ gives values for the average fitness line shown in Figure 3-1. The slope of this average fitness line is $b - c$ (see Figure 3-1) or the degree of non-zero-sumness in the NPD model. As a kind of control, if we do a run of this model where *both*

cooperators and defectors receive the average fitness (in this case there is no within-group selection), the change in Q over generations matches the Price between-group component, ΔQ_b , *exactly*. Thus the average slope of the fitness functions (degree of non-zero-sumness) is a measure of the between-group selection force predicted by the Price equation. That is, groups with higher fractions of cooperators get proportionally more of the non-zero-sum advantage and out produce groups that with less cooperators.

Alternative Selection Decomposition

This presentation of the Price between-group idealization suggests an alternative decomposition where the *within-group* component contains the Q' idealization and the *between-group* component is the correction term. This idealization assumes the frequency of cooperators within each group changes, but that the relative size (fitness) of groups does not—I use the *after-selection* q_i' values and the *before-selection* group sizes, n_i . This alternative Q' idealization I denote as:

$$Q^\# = \frac{\sum q_i' n_i}{N} \quad (3.13)$$

and the alternative components of selection can be labeled:

$$alt\Delta Q_w = Q^\# - Q \text{ and} \quad (3.14)$$

$$alt\Delta Q_b = Q' - Q^\#. \quad (3.15)$$

Now we can do a symmetrical control to the one above by setting the slope b to zero. Here there is a difference c between cooperator and defector fitness within groups but no difference between what these types get in different groups (in this case there is no between-group selection), and the resulting actual ΔQ for this run *exactly* matches the $alt\Delta Q_w$ component given by this decomposition. So again the degree of non-zero-sumness (average fitness line) in this model gives the between-group selective force predicted by the Price decomposition, and the cost c of being a cooperator (intercept difference), gives the within-group selective force predicted by the alternative decomposition. The Price within term and the alternative between term act, respectively, as correction terms for these idealizations.

Figure 3-4 shows the *change* from the initial Q value for Run 1 of Figure 3-2 (parameters given in Table 3-1). Also shown are the selection components of Run 1 given by the Price decomposition and the alternative. Notice that the two decompositions give quite different results. In the Price decomposition the equilibrium state ($Q = 0.0$) consists of a balance between a strong between-group force, 0.4, even though group 1 has disappeared (see Figure 3-2(c)), and a strong within-group force, -0.9 , even though cooperators have disappeared. In contrast, the alternative decomposition more intuitively says that the between-group selection force rises as group 2 initially increases over group 1, but that this force goes to zero as the first group disappears. The alternative within-group component steadily decreases to -0.5 to balance the initial Q of 0.5 as Q goes to zero. Which of these two decompositions is more appropriate will depend on the situation being studied. In the runs illustrated

here where the within-group selection force eventually dominates, the alternative decomposition provides more insight.

Figure 3-4. Price Equation and Alternative Decompositions for Run 1. Change in cooperator frequency (ΔQ_t) vs. generation for Run 1 of Figure 3-2 along with the between- and within-group components of selection for this run given by the Price decomposition (ΔQ_b and ΔQ_w respectively) and the alternative decomposition ($alt\Delta Q_b$ and $alt\Delta Q_w$).

Note that neither decomposition is necessarily accurate when there is a mixture of between- and within-group selective forces acting simultaneously. While both decompositions assume the forces can be decoupled, in reality the selective forces affect each other. The Price decomposition posits in its between-group term that all change is due to between-group selection and then assumes the difference between this assumption and the actual change in cooperator frequency is due to the counterbalancing force of within-group selection. The alternative decomposition takes the opposite tack. When the two approaches give roughly the same answer (as in the

first few generations of Figure 3-4), then the forces are roughly decomposable. But as change over longer periods is compared the values given by the two approaches diverge (e.g. compare values at generation 30 in Figure 3-4). This is an indication that over this time period the degree of between-group selection has been affected by changes in group compositions (caused by within-group selection); and that the degree of within-group selection has been affected by group size changes (caused by between-group selection). In addition to offering a symmetric decomposition that may give more intuitive results in some situations, this alternative decomposition is useful in highlighting when the Price equation's assumption of independent effects is inaccurate.

Selection and Game Theory

In addition to the debate over levels of selection and how altruism evolves, there is also disagreement about what behaviors constitute altruism. In a standard game-theoretic analysis rational self-interest is defined by the behavioral choice that produces the highest *absolute* payoff (utility) to a player—regardless of the effect this behavior has on the payoffs of other players (i.e. positive or negative externalities). Individuals maximizing their own utility dictate the expected dynamics in this situation. On the other hand, the engine of natural selection is *differences* in offspring representation in subsequent generations and the dynamics are driven by *relative* fitness (Wilson 2004). Applying the standard game-theoretic (or absolute fitness)

viewpoint to systems under selection leads to contradictory results, as illustrated below.

I stated earlier that the minimum conditions for an NPD in our model require that the average fitness line have a positive slope, i.e. $b > c$, but from the standard game-theoretic viewpoint there is an *additional* constraint: $c > b/n_i$. That is, the cost of being a cooperator must be greater than the cooperator's share of the benefit it creates for the group. This second condition marks the boundary between an NPD and a *No-Conflict* (NC) game, or equivalently the difference between what Wilson (1979; 1990) has termed *strong* and *weak* altruism, respectively. Note that the same behavioral trait (with the same values of b and c) can change between strong and weak altruism depending on changes in group size (Pepper 2000b).

The expected dynamics (from an absolute fitness perspective) in the NC game defined by weak altruism is towards mutual *cooperation* (as cooperation is the dominant strategy), but under selection the dynamics of weak altruism move towards mutual *defection*. Again, this is because under selection *differences* in fitness values drive the dynamics—not just the maximization of absolute fitness amounts (Wilson 2004). Several studies have reported “surprising” results when animal or human subjects choose cooperative or altruistic behaviors that do not maximize absolute fitness or utility, but instead seem more concerned with relative fitness or fairness (e.g. Fehr and Gächter 2002; Boyd et al. 2003; Brosnan and de Waal 2003). Yet these results are not surprising from the perspective of selection. Hamilton in discussing PD situations wrote in 1969:

But natural selection, the process which has made us almost all that we are, seems to give one clear warning about situations of this general kind. When payoffs are connected with fitness, the animal part of our nature is expected to be more concerned with getting “more than the average” than with getting “the maximum possible” (Hamilton 1971 p. 83)

Still some have ignored or disagreed with this distinction preferring to reserve the word “altruism” only for the strong category (Nunney 1985b; Maynard Smith 1998a; Nunney 2000). For instance Nunney has argued that “benevolence” should be used instead of “weak altruism.” He offers the following example:

Suppose that you are offered two financial options. Under the selfish option you receive 10 dollars and keep it all. Under the benevolent option you receive 10 million dollars, but 6 million must be given to a neighbor. Given that neighbors are random samples of a large population, the choice is clearly the benevolent option, a choice based purely on individual greed and not on the general benefit of the neighborhood. Replacing money by fitness, it can be seen that benevolence spreads by individual selection because a net gain of 4 million units of fitness is superior to a net gain of 10 units of fitness. (Nunney 1985b p. 226)

Although it sounds very self-serving to give oneself 4 million dollars, the problem with this example is that it ignores selection among neighbors. If you have a behavioral trait that helps your neighbor have 6 million offspring while it only helps you have 4 million offspring, then only in an infinite world could this trait avoid being weeded out (assuming as Nunney does that only individual selection is operating). In every paired interaction this trait loses ground to its neighbor. On the other hand, such a trait can evolve if the benefits bestowed by altruists fall more to other altruists than to average population members—but this need for positive assortment of altruistic behaviors is the same condition for *strong* altruism to evolve—and in both

cases Hamilton's rule specifies the degree of positive assortment necessary given the fitness costs and benefits.

Returning to our model, in Run 1 (Figure 3-2 and Table 3-1) the first group's size shrinks below $n_i = 5$ between generation 15 and 16, and thus crosses the boundary from strong to weak altruism ($b = 1$, $c = 0.2$, and altruism is weak when $n_i < b/c$ or $n_i < 5$). Under the standard view (focused on absolute fitness), cooperation should now be favored (by individual within-group selection) because this is now an NC game in which cooperation is the dominant strategy (highest absolute payoff). Therefore cooperation should increase when $n_i < 5$ and decrease when $n_i > 5$ for these model parameters. Yet, in reality, cooperation (now weak altruism) continues to be steadily selected against and we observe that the equilibrium for this group is mutual defection. This is because even a weak altruist helps every other group member to have more offspring than itself.

Note that *under selection* in a finite population the essential nature of the PD is retained even for weak altruism—the dynamics lead to mutual *defection* even though collective *cooperation* would result in higher average absolute fitness for group members. In this sense selection converts an NC game into a PD (or equivalently weak altruism to strong) by making the relative amount of externalities (i.e. what others get from an altruist compared to what it gives itself) relevant to the resulting absolute fitness values. These relative fitness effects are ignored in the strong/weak or PD/NC classifications, but are central to natural selection. Note that in the NPD model used here the population is finite as it is proportionally scaled to its original size each

generation, but even in an infinite population the *percentage* of altruists (whether strong or weak) decreases within groups every generation.

To illustrate the conversion from weak to strong altruism under selection, I start with a 2-player NC game with a dominant strategy to cooperate. I then convert it by asking what the payoff values would be in the case that *differences* in absolute values represented the *actual* utilities. This idea is similar to “social motives” in game theory (Davis 1989) where the given utility values do not represent the actual utilities of the players who may have other motives than absolute payoff maximization, such as “fairness” or in this case “difference maximization.” I illustrate this conversion in Figure 3-5. When two strategies combinations result in the same difference calculation I add an amount δ to the one that has the higher absolute fitness. This implements the idea that when there is no difference between opponents it is preferable to get the higher absolute amount of utility (fitness).

Note that when we start with a NC (No-Conflict) game (first row Figure 3-5), difference maximization (selection) *converts the game to a PD*. On the other hand, when we start with a PD game (second row Figure 3-5), difference maximization (selection) still results in a PD. So strong altruism remains strong under selection, but altruism designated as weak with regard to absolute fitness values is converted to strong altruism under selection. Whereas cooperation is the dominant strategy when only absolute fitness values are considered in a NC game, selection converts this situation into a PD where defection is the dominant strategy—and to cooperate is to *make an altruistic sacrifice*.

NC		<i>Player 2</i>		taking $\longrightarrow$ utility differences	PD		<i>Player 2</i>	
							C	D
<i>Player 1</i>	C	5 / 5	1 / 3		<i>Player 1</i>	C	$0 + \delta /$ $0 + \delta$	$-2 / 2$
	D	3 / 1	0 / 0			D	2 / -2	0 / 0

PD		<i>Player 2</i>		taking $\longrightarrow$ utility differences	PD		<i>Player 2</i>	
							C	D
<i>Player 1</i>	C	3 / 3	0 / 5		<i>Player 1</i>	C	$0 + \delta /$ $0 + \delta$	$-5 / 5$
	D	5 / 0	1 / 1			D	5 / -5	0 / 0

Figure 3-5. Applying Difference Maximization to NC and PD Games. Illustrates effect of converting a game using the *differences* in original utility values (left-hand games) to calculate converted utility values (right-hand games). When two different strategy combinations result in equal differences, a value δ is added to the one with the highest original (absolute) utility. The game-theoretic classification of each game is indicated in the upper left corner of each payoff matrix—Prisoner's Dilemma (PD) or No-Conflict (NC). The strategies for player 1 are represented as rows and for player 2 as columns. The payoff values for each pair of strategies that meet are listed as *player 1 payoff/player 2 payoff*.

Altruism via Randomly Formed Groups

So far I have explored the simplest case of multilevel selection where the within-group level is represented by only two alternative strategies and the between-group level is represented by only two distinct groups. As we have seen, in this case the increase in cooperators is transient because once one group dominates there is no longer a between-group selection force (the variance ratio goes to zero) and the within-group more fit defect strategy takes over this single group. In order to illustrate that mechanisms which reestablish variance among groups can maintain altruism

indefinitely, I have modified our model to include periodic random redistributions of the population among multiple groups. Between redistribution events reproduction takes place and group sizes vary with cooperator and defector fitness as previously explained. This modification shares features of Maynard Smith's haystack model (1964; Wilson 1987) including that groups exist for multiple generations and there is a global mixing phase followed by random distribution into many groups.

Interestingly, even among those that debate whether weak altruism should be considered altruism, there is consensus that strong altruism cannot evolve in randomly formed groups (Hamilton 1975; Wilson 1975a; Nunney 1985a; Wilson 1990; Maynard Smith 1998a; Nunney 2000; Sober and Wilson 2000). This conclusion is based on single-generation models. In Chapter 4 I address this conclusion directly using the original models upon which this conclusion is based (Fletcher and Zwick 2004c). Here I will compare weak and strong altruism in randomly formed groups that exist for multiple generations in our NPD model. To guarantee that only strong altruism is operating (regardless of changes in group size) I use other-only altruism in which altruists give nothing to themselves and their benefit is divided evenly among the *others* in the group. In the other-only case the x -axis in Figure 3-1 would be the fraction of others cooperating in a group and the q_i value for calculating a_i' in Eq. 3.3 is $(a_i - 1)/(n_i - 1)$ and for calculating s_i' in Eq. 3.4 is $a_i/(n_i - 1)$; that is, the fraction of cooperators in *others*. By definition, the other-only runs are guaranteed to involve strong altruism whereas the whole-group runs begin with weak altruism conditions for comparison.

I contrast this other-only (strong) altruism with the whole-group altruism we have used thus far. As mentioned earlier, the value of r from Hamilton's rule differs for these two types (Pepper 2000b). For other-only altruism the expected value of r when groups of size n_i are formed at random is 0 (Pepper 2000b) so on average no positive values of b and c will satisfy Hamilton's rule ($rb > c$). For whole-group altruism, randomly formed groups produce an average r value of $1/n_i$ (Pepper 2000b) so the condition becomes $b/n_i > c$ which is the definition of weak altruism. Thus the consensus that strong altruism cannot evolve via randomly formed groups. Below I show this conclusion is not correct when groups exist for multiple generations (as they do in haystack models (Maynard Smith 1964; Wilson 1987)). I first contrast the dynamics of other-only (strong) and whole-group (weak) altruism for individual runs in Figure 3-6 and then give aggregate results in Table 3-2.

Figure 3-6 shows that the Q values for both other-only and whole-group altruism follow the same familiar hump-shaped pattern seen in Figure 3-2 and Figure 3-3 except that in the other-only runs Q decreases initially. The similar general dynamics of the other-only (strong altruism) and whole-group (weak altruism) runs is another indication of the similarities between strong and weak altruism under selection—effectively they are both PDs. The initial decrease in Q for other-only runs is expected because on average $r = 0$ after groups are formed and therefore Hamilton's rule cannot be satisfied. Surprisingly however, within a couple of generations Q begins to increase and then follows the familiar hump-shaped pattern until another group reformation event. As I will show in Chapter 4, this is due to the r value increasing from its initial

value of 0 over multiple generations within groups. Note the corresponding points for whole-group and other-only runs where groups are reformed every 40 generations (indicated with a vertical dashed line with arrows in Figure 3-6). In runs where groups are reformed close to when Q is peaking (e.g. every 20 generations runs), cooperation can ratchet up to saturation, whereas in runs where groups are reformed long after the peak in Q (e.g. every 40 generations runs), cooperation tends to be eliminated. This is true for both types of runs in each pair.

Figure 3-6. Dynamic Change in Q for Periodically Randomly-formed Groups. Frequency of cooperators (Q) vs. generation for 6 runs where an initial population of 500 cooperators and 500 defectors is randomly assigned to 100 groups and periodically randomly redistributed into groups. Three runs each are done for whole-group (WG) and other-only (OO) altruistic benefit for redistribution frequencies of 20, 30, and 40 generations. For all runs $b = 20$ and $c = 1.0$. A vertical dashed line with arrows indicates corresponding points at generation 40 for the whole-group and other-only runs.

In Figure 3-6 there are three pairs of runs each consisting of an other-only (OO) run and a whole-group (WG) run. For each run a total population of 500 cooperators and 500 defectors is distributed randomly into 100 groups and then randomly redistributed periodically. Parameters are the same across runs except the number of generations between group reformation events varies among pairs. In this analytic model we allow fractional counts of cooperators and defectors from generation to generation, but do not use counts less than one in group reformation events. This tends to weed out residual fractions and because there is no mutation, the extremes of $Q = 1.0$ and $Q = 0.0$ act as attractors and intermediate values do not persist indefinitely.

The results in Figure 3-6 show typical individual runs done with the same initial seed for comparison. I also did 100 runs at each reformation frequency with the same parameters but different random number seeds. In all cases runs were done until either cooperator or defector saturation was reached. Table 3-2 shows the percentage of runs reaching cooperator saturation for these parameter conditions as well as shorter and longer periods between group reformations. These results support the trends discussed above.

Table 3-2. Percentage of 100 Runs Ending in Cooperator Saturation.
The same model parameters as for Figure 3-6 were used.

	Reformation Frequency					
	1	10	20	30	40	50
Whole-group	100%	100%	100%	66%	12%	12%
Other-only	0%	100%	65%	10%	8%	7%

Again, the results shown here for strong altruism are in contrast to the commonly held belief that strong altruism cannot evolve in randomly formed groups (Hamilton 1975; Wilson 1975a; Nunney 1985a; Wilson 1990; Maynard Smith 1998a; Nunney 2000; Sober and Wilson 2000). In the next chapter (Fletcher and Zwick 2004c) I explore the issue of strong altruism evolving in randomly-formed groups in more detail with an individual agent-based model and show that this general result does not depend on kin selection and even holds under fairly high levels of migration between groups each generation. The evolution of altruism via random associations may be one method by which new integrated units of selection (Margulis 1993; Maynard Smith and Szathmáry 1995; Michod 1999) at increased hierarchical levels originate.

Summary

By embodying non-zero-sumness, population structure (assortment), and heredity in their most basic forms, this NPD model offers a simple framework for understanding the paradoxical nature of the evolution of altruism, integrating such central concepts as Simpson's paradox, Hamilton's rule, and the Price covariance equation. It also suggests an alternative selection decomposition which is more intuitive in some situations and helps emphasize the coupled nature of within- and between-group selection acting over multiple generations. I show that a game-theoretic framework is also useful in understanding the similarities between weak and strong altruism undergoing selection. Finally, I contrast other-only (strong) and whole-group (weak) versions of the NPD model to highlight both their initial differences immediately after

random group formation and their overall dynamical similarities. This also illustrates, in contrast to conventional wisdom, that strong altruism can evolve in periodically randomly formed groups that are multigenerational. In the next chapter I illustrate more fully the reasons that this conventional wisdom does not hold by starting with the original analytical models on which it is based (rather than the NPD model used here) as well as developing modifications to these models that explore, among other factors, the role of kinship, stochasticity, and migration in the evolution of altruism.

Recently, game-theoretic models have been demonstrated where cooperation increases even without reciprocity. In these cases social interactions are clumped by various mechanisms including the presence of non-players (Hauert et al. 2002), the need for sufficiently similar arbitrary tags (Riolo, Cohen, and Axelrod 2001), and social institutions for conformity within groups (Bowles, Choi, and Hopfensitz 2003). In the context of the model presented here, we would expect these results with their various cost, benefit, and population structure parameters also to conform to Hamilton's rule, although this kind of analysis is not usually undertaken in such papers ((Bowles, Choi, and Hopfensitz 2003) is an exception although the whole-group version of r is used for a model which is other-only). Recent models showing the evolution of cooperation among non-kin (Avilés 2002; Avilés, Abbot, and Cutter 2002) are also understandable in the framework presented here. Although this model could be interpreted as containing clonal cooperators and defectors, it is common in game-theoretic models of the evolution of cooperation (Maynard Smith 1982; Axelrod 1984) to interpret these two strategies as representing social behaviors—which may be

based on non kin-related, but heritable, genotypes. The evolution of altruism does not actually require either reciprocity or kinship. What is essential is: 1) sufficiently non-zero-sum benefits for altruistic behaviors (where behaviors are heritable) and 2) sufficiently non-uniform interactions among these behaviors. As demonstrated here, the necessary combination of population structure and non-zero-sumness is specified by Hamilton's rule.

This overall framework can help social science researchers who emphasize game-theoretic models to see their results in the context of Hamilton's rule and multilevel selection, while also enabling biology researchers who focus on relatedness to recognize the inherent game-theoretic character of their models.

If costs and benefits combine additively, and groups are formed randomly, then altruism cannot evolve. But if altruists tend to associate with altruists, and non-altruists with non-altruists, then altruism can evolve. This conclusion is agreed.

John Maynard Smith, in review of *Unto Others* (1998a)

Chapter 4

Strong Altruism Can Evolve in Randomly Formed Groups¹

Introduction

Nearly three decades ago Hamilton (1975) and Wilson (1975a) independently developed models which were interpreted as showing that strong altruism (involving an absolute cost to altruists) cannot evolve in randomly formed groups. This conclusion is still generally accepted even among those who debate how best to define altruism and the mechanisms by which it evolves (Hamilton 1975; Wilson 1975a; Nunney 1985b; Wilson 1990; Maynard Smith 1998a; Nunney 2000; Sober and Wilson 2000). Here I challenge this conclusion by exploring what happens when groups exist for more than one generation. Multigenerational groups are a central feature of

¹ This chapter is based on a paper published in the *Journal of Theoretical Biology* (Fletcher and Zwick 2004c).

Maynard Smith's classic "haystack" model (1964; Wilson 1987), but the role of multiple generations within groups was not explored in Hamilton's (1975) and Wilson's (1975a) models. Although the initial conditions after random group formation favor non-altruists over altruists, paradoxically these conditions can switch to favor altruists after even one generation of selection. Thus even though the overall proportion of altruists decreases after one generation, it can increase *even more* after a second generation spent within groups.

Besides single-generation groups, these original analytic models rely on other simplifying assumptions such as an infinite population and no migration between groups. I begin by showing how strong altruism can evolve under the assumptions of the original models, with the only modification being delayed reformation of random groups. Multigenerational groups introduce additional issues such as interactions among related offspring, persistent homogeneous groups of altruists, and exponential growth of population size. I explore model modifications—preventing altruists from benefiting kin, precluding homogeneous groups from forming, and adding a population-level carrying capacity—that mitigate each of these factors. I find that under *all* these modifications (imposed both separately and concurrently) strong altruism, although dampened, can still evolve in randomly formed multigenerational groups. I then transform the basic analytic model into an evolutionary simulation in which population size is finite and stable. In the simulation model random group formation from a finite population and migration between groups both reduce positive assortment and therefore dampen selection for altruism. On the other hand, increased

stochasticity in benefit distribution and culling due to carrying capacity can enhance the likelihood that altruism will evolve compared to processes with minimum stochasticity.

Classifications of Altruism

Both the analytic and simulation models discussed here involve what Pepper (2000b) has termed an *other-only* altruistic trait because none of the altruist's benefits come back to itself, as opposed to *whole-group* traits (also called group-beneficial traits) where the benefit is divided among *all* group members including the altruist. Wilson (1979; 1990) previously classified altruistic traits in a related but different way as either *strong* (involving an absolute cost to altruists) or *weak* (involving only a relative cost to altruists). Other-only altruistic traits are always strong while whole-group traits are strong if the cost to an altruist is greater than its share of the benefit it provides. Note that the same whole-group behavior involving the same sacrifice and provided benefit may be strong or weak depending on group size (Pepper 2000b).

In contrast to strong altruism, Wilson showed (1979; 1990) that weakly altruistic traits *can* increase when groups are randomly formed every generation. That is, for an infinite population where a binomial trait is randomly redistributed every generation, the resulting between-group component of total variance can be enough for weak, but not for strong, altruism to evolve. Nevertheless in finite populations where fitness is relative the distinction between strong and weak altruism may be less important as

both types are selected against within groups and require selection (or differential productivity) among groups in order to increase (Wilson 1979; 1990). In this paper I focus on other-only, strong altruism (the most restrictive situation) to address the random group models of Hamilton and Wilson directly, but the consequences of multigenerational groups and stochasticity also apply to weak, whole-group traits and therefore these traits can even more readily increase via randomly formed groups than was previously shown (Wilson 1979; 1990). A side-by-side comparison of the dynamics of weak and strong altruism was given in Chapter 3.

Analytic Model

I focus on Hamilton's model (1975) because he developed a formal proof that altruism cannot evolve in single-generation randomly formed groups (Wilson's model (1975a) is similar in all important aspects). In this model a haploid infinite population is randomly subdivided into groups of equal size, n . Group members interact for one generation, affecting each other's fitness (offspring count), before the population is pooled and then again randomly assigned to new groups. In every generation each altruist behaves in a way that costs itself c offspring and provides a *total* benefit of b offspring divided evenly among the other $n-1$ group members. Each non-altruist receives its share of benefits, but does not provide any benefit to others. Therefore, within every group non-altruists have more offspring than altruists, but groups with more altruists have more offspring per capita than groups with less. This is an example

of multilevel selection where here selection within groups opposes selection between groups.

Hamilton (1975) using Price's covariance equation (1970) showed that under his model's assumptions, between-group selection (due to the variance between groups in altruist frequency, q) must always be weaker than average within-group selection (due to the expected variance in the altruistic trait within groups) and therefore the overall frequency of altruists, Q , must decrease in every generation. (Capital letters indicate whole population values; small letters indicate group values.) To illustrate this, I calculate $\Delta_1 Q$ for an infinite binomial distribution, where Δ_1 indicates that the change occurs over one generation within groups, $g = 1$. The variable g is the number of generations spent within groups before each reformation event. (See Appendix C for model details.) Figure 4-1(a) shows how $\Delta_1 Q$ depends on the level of benefit, b , provided by altruists for different starting Q values. (For convenience, all results reported in this paper use $c = 1$ such that benefit b is also the benefit to cost ratio.) The results shown in Figure 4-1(a) are the same for any group size n . Note that as benefit increases, zero is an upper limit on $\Delta_1 Q$ —hence the conclusion that strong altruism cannot increase under the assumptions of this model for all values of Q and n (Hamilton 1975; Wilson 1975a).

Figure 4-1. Change in Altruist Frequency as a Function of Altruist Benefit. (a) and (b) compare the effect of different starting Q values on $\Delta_g Q$ after one ($g = 1$) and two ($g = 2$) generations spent within groups, respectively, where founding group size $n = 4$ (although in (a) the results are the same for all n). (c) compares the effect of different n for multigenerational groups ($g = 2$) when altruism is rare—here initial $Q = 0.001$. The cost $c = 1$ in all calculations.

Yet quite different results are obtained if groups persist for even one additional generation ($g = 2$) before random mixing and the formation of new groups. Figure 4-1(b) shows how the change in Q after two generations within groups, $\Delta_2 Q$, depends

on benefit values for different starting Q values. The only difference between Figure 4-1(a) and (b) is that the latter measures the change in altruist frequency after an additional generation spent within groups where the fitness functions are deterministically and recursively applied. In this case strong altruism can clearly increase ($\Delta_2 Q > 0$) for sufficient values of benefit. Figure 4-1(c) shows that smaller groups give a larger increase in altruist frequency which is consistent with previous findings on the relationship between group size and the evolution of altruistic traits (Boyd and Richerson 1988; Avilés 1993). Additionally Figure 4-1(c), for which $Q = 0.001$, shows that strong altruism can increase due to multigenerational groups even when the altruistic trait is rare, although higher benefit levels are needed for $\Delta_2 Q > 0$ when Q is low.

Applying Hamilton's Rule

We can also understand these results in terms of Hamilton's rule (1964) which states that the condition for an altruistic trait to increase in the next generation is:

$$rb > c, \quad (4.1)$$

where r is the "coefficient of relatedness" or more generally the regression coefficient between the frequency of the trait in recipients and actors (Hamilton 1972). Thus r is a measure of positive assortment—the degree to which the benefits of altruists fall to other altruists. The value of r differs for other-only and whole-group traits because in the latter case, but not the former, altruists are recipients of their own actions (Pepper

2000b). I use superscripts w and o respectively for whole-group and other-only measures of r . For whole-group traits r^w is the between-group variance in q over the total variance in the altruistic trait (Breden 1990; Frank 1995a). For an infinite binomial population of randomly formed groups of size n , the variance ratio $r^w = 1 / n$ (Pepper 2000b). Thus according to Hamilton's rule (Eq. 4.1) the trait increases after one generation if $b / n > c$, but for whole-group traits this means that an altruist's share of its benefit must be greater than its cost—this is the definition of weak altruism so as Wilson (1979; 1990) noted only weak traits can increase after one generation.

For groups of uniform size the r values are related by the following expression (Pepper 2000b):

$$r^o = \frac{nr^w - 1}{n - 1}. \quad (4.2)$$

Therefore $r^o = 0$ for an initial random distribution where $r^w = 1 / n$. Obviously there are no positive values of b and c that can satisfy Hamilton's rule (Eq. 4.1) for an other-only (strong) altruistic trait when $r = 0$ and such traits must decrease. Note however, that any modifications to the model that make $r^o > 0$ can yield an increase in Q , given a sufficient value of b . Hamilton noted that any assortment of altruists beyond that produced at random could allow altruism to increase (1975). Surprisingly, for many parameter settings r^o increases above zero after one generation of selection—even as the proportion of altruists decreases. That is, this transient one-generation-long “population viscosity” of the original models is enough (without any other mechanisms for creating positive assortment) to create conditions that favor altruism

in the following generation. If groups are randomly reformed after this single generation then this gain in positive assortment is destroyed before being used by selection; r^o returns to zero and Q subsequently declines. On the other hand, additional generations within groups can take advantage of this increased positive assortment so that strong altruism increases, as shown in Figure 4-1. (Whether altruism actually increases or not depends on parameters including Q , b , and n .)

Note that although Hamilton emphasized a ratio of variances in his proof, in this other-only model the regression coefficient between actors and recipients, r^o , is an easier to interpret measure of the changing conditions affecting altruism as selection occurs. For instance, after one generation of selection (starting with randomly formed groups), the between over total variance, r^w , can decrease while r^o increases. It is the increase in r^o that accurately reflects whether altruism can increase in the next generation. For most parameter settings both measures increase after one generation, but the range of parameters where r^w decreases is greater than the range in which r^o decreases. In general the r values can decrease when there is a combination of low Q , low b , and high n .

According to Hamilton's rule whether altruism increases in the second generation within groups depends on whether r^o after the first generation is greater than c / b . Of course it is not enough for altruism to increase in the second generation for altruism to increase overall. The increases in subsequent generations within groups must make up for any losses in the initial generation(s). Figure 4-2(a) shows the expected dynamic change in Q values over successive generations when groups persist for one and two

generations before random reformation and Figure 4-2(b) shows the concurrent changes in r^o .

Figure 4-2. Calculated Dynamics in Analytic Model Comparing g and n . Results for one and two generations spent within groups ($g = 1, g = 2$) for groups of two different sizes ($n = 3; n = 10$). (a) shows the dynamics in overall altruist frequency Q given different starting Q values and different sized groups. (b) shows the concurrent change in the regression coefficient between actors and recipients, r^o . The critical r^o value of c/b is also shown with a solid horizontal line. Here $Q = 0.1$, $b = 10$, and $c = 1$. Both r^o and Q are calculated at the end of the indicated generation and after group reformation if it occurs. (a) and (b) use the same legend.

Altruist frequency Q decreases monotonically when groups are reformed every generation and $r^o = 0$ before each round of selection. On the other hand when groups

exist for two generations, Q oscillates (and can ratchet upward). The every-other-generation saw-toothed peaks in Q correspond to similar (but offset) oscillating peaks in r^o (Figure 4-2(b)). Here r^o increases after a generation within groups and the critical c / b value is indicated with a solid horizontal line. Troughs on the other hand correspond to global mixing, new group formation, a decrease of r^o back to zero and a subsequent decrease in Q . In Figure 4-2 I also show a case with the same parameters except bigger group size ($g = 2; n = 10$). Here, although Q can increase during the second generations within groups, it is not enough to make up for losses in the first generations. Note that when peaks in r^o fail to reach the c / b value (after generation 21 in Figure 4-2(b) for $g = 2; n = 10$), as predicted by Hamilton's rule, Q can no longer increase and instead falls during both generations within groups (Figure 4-2(a)).

Analytic Model Modifications

Multiple generations within groups complicate the simple single-generation model in several ways: 1) kin interactions within groups become possible; 2) the contribution of homogeneous groups of altruists increases—these groups uniquely retain their initial (maximal) level of altruism; 3) the additive frequency-dependent fitness functions can now lead to exponential growth of the population. Yet as I demonstrate below, while not inconsequential, none of these factors are essential to explain why strong altruism increases in randomly formed multigenerational groups—especially when altruism is initially rare. The following three paragraphs elaborate on each issue

and describe modifications to the basic model to address them. I follow this with a summary of the results produced by each modification.

No Kin Selection

In the original model groups are formed by randomly selecting individuals from an infinite population and therefore groups contain unrelated individuals. In a second generation within groups, when the benefits provided by an altruist are divided among other group members, some of this benefit (in the form of additional offspring) will fall to those with the same parent as the altruist. In general the proportion of benefit falling to relatives (defined by common ancestry) in subsequent generations will depend on parameters n , b , and Q , but this proportion is bounded by $1 / n$ (Appendix D). This limit is approached for high b and Q , but when altruism is initially rare this proportion is much less significant. For example, for groups of size 4, the maximum possible proportion of benefit that could fall to relatives is $1 / n = 25\%$, but for $Q = 0.1$ and $b = 10$ the observed proportion of benefit given to relatives is actually 0.61% during the second generation within groups. For the same Q and b the proportion decreases with larger group size and more generations within groups. To eliminate kin selection I modify the model so that altruists only divide their benefit among non-relatives (Appendix D).

No Homogeneous Groups

In the infinite population of this model, homogeneous groups of altruists will be randomly created whenever $Q > 0$. These groups are unique in being the only group composition for which q cannot decrease with successive generations within groups. They are also the fastest growing groups as they contain no free-riding non-altruists. One might suspect that such homogeneous groups account for altruism being able to increase after multiple generations within groups. To check this I modify our model (Appendix D) such that immediately after group formation all homogeneous groups of altruists have one altruist switched to a non-altruist. Note that this artificially decreases Q , making it even harder for altruism to evolve.

No Population Growth

Even with additive (linear) fitness functions, multiple generations within groups can cause a population to grow exponentially (Wilson 1987). To study the effect of stable population size I implement a global carrying capacity by scaling the offspring count of all population members each generation by the inverse of the expected overall growth rate during that generation (Appendix D). This holds the population size constant (albeit infinite) at every generation, but allows groups with more altruists to have relatively more offspring each generation than groups with less.

Modification Results

Figure 4-3 compares the results for each of these modifications with the unmodified model for two generations within groups, $g = 2$. I also include results for the original model where $g = 1$. For each of the three modifications $\Delta_2 Q$ is dampened, but still positive given sufficient benefit. This is true even when all of the modifications are imposed simultaneously. That is, in a model where no benefit is given to kin, homogeneous groups are always corrupted, and population size is held constant, strong altruism can still increase after two generations within groups.

Figure 4-3. Calculated Altruist Frequency for Analytic Model Modifications. Shows altruist frequency ($\Delta_g Q$) after two ($g = 2$) generations within groups as a function of altruist benefit, b , for several modifications of the original binomial model including preventing homogeneous groups from forming, scaling the population size to its original size each generation, and distributing altruist benefit only to the non-relatives of an altruist. The results for the original model after one ($g = 1$) and two ($g = 2$) generations are also shown for comparison. The original $Q = 0.25$, $n = 4$, and $c = 1$ in all calculations.

In contrast to the *unmodified* dynamic model shown in Figure 4-2 where altruism tends to evolve to $Q = 1.0$ or $Q = 0.0$ given enough generations, it does not *necessarily* evolve to saturation under all these modifications. Corrupting homogeneous groups for example necessarily keeps $Q < 1.0$. In the case of a population-level carrying capacity, for $n = 4$, $Q = 0.1$, $g = 2$, and $b = 15$, a stable limit cycle is reached in which Q oscillates every other generation between 0.616 and 0.636. (Yet as shown in the next section, when stochasticity is introduced populations tend to evolve to one extreme or the other in these models.)

Simulation Model

So far, like Hamilton, we have used the assumption of an infinite population in order to calculate the expected distribution of group compositions when individuals are randomly distributed. But infinity here has two special consequences. First it converts a seemingly stochastic process (random group formation) to a deterministic one—the *expected value* of r^o is produced by *every* group reformation event. For any *finite* population, group reformation events will produce r^o values that fluctuate both above and below the average value. Second, the expected value of r^o is lower (i.e. less than zero) for groups formed randomly from a finite population (compared to an infinite one). This is because in the finite case where “sampling without replacement” is used we have a hypergeometric (rather than binomial) distribution. Once an individual of a certain type is assigned to a random group, the probability of assigning another

individual of that type is lower than in the first assignment. This results in the formation of less homogeneous groups and more heterogeneous groups (than in the infinite binomial distribution) which decreases the overall assortment—thus on average r^o is negative. Based on empirical calculations it appears that the expected negative value of r^o depends only on population size (not n or Q) and is $r^o = -1 / (N-1)$ for equal-sized groups randomly formed from a finite population of size N . (I have not yet found a proof of this.) Particular random group formation events will result in r^o values which fluctuate around this lower value.

In addition, a finite simulation model can incorporate the consequences of other stochastic processes including the way altruistic benefit is distributed and the way individuals subject to a carrying capacity are eliminated. I now transform the analytic model above into a computer simulation of a finite evolving population and the following features and options (see Appendix E for further details):

1. When reforming groups each individual is assigned at random to an unfilled group (rather than by using the expected distribution).
2. The benefit value used by each group, each generation and the distribution of this benefit to other group members is done in one of two ways:
 - i) Low stochasticity: the benefit level is the same for all groups during a simulation run and is divided as evenly as possible (in units of whole offspring) with only any *remainder* distributed randomly among other group members.

- ii) High stochasticity: the benefit level used in a group is drawn from a Poisson distribution in whole units where the given benefit value is the mean and *each unit* of benefit is then distributed at random to other group members.
3. Population size is held constant each generation by a global carrying capacity in one of two ways:
- i) Low stochasticity: group sizes are proportionally scaled back (as in the no-growth analytic model), but only whole organisms are removed proportionally with any remainder removed randomly (as below).
 - ii) High stochasticity: excess population offspring are removed at random (without regard to the altruistic trait or group membership).

The first options in 2. and 3. above minimize stochasticity while still preserving the simulation model's integral organisms, whereas the second options introduce more stochasticity. Figure 4-4 compares the change in altruist frequency after 2 generations within groups, $\Delta_2 Q$, for the no-population-growth run of the analytic model (from Figure 4-3), the simulation model with stochastic group formation but minimal other stochasticity, and each of the more stochastic choices introduced separately, and then simultaneously. Each data point represents the average of 1,000 runs done with different random number seeds. Note that strong altruism evolves less easily in the finite simulation with stochastic group formation (than in the comparable infinite analytic model). From this base, adding more stochasticity in benefit distribution has a

slight positive effect whereas adding more stochasticity in elimination due to carrying capacity has a strong positive effect. The latter case results in an even bigger increase in altruism than in the no population growth analytic model. Using both options simultaneously does even better.

Figure 4-4. Altruist Frequency for Different Levels of Stochasticity. Shows change in altruist frequency for after two generations within groups ($\Delta_2 Q$) as a function of altruist benefit, b , for several variations of the finite simulation model with no population growth. The base case is stochasticity in random group formation, but otherwise minimal noise. To this case I add high stochasticity in benefit distribution, high stochasticity in implementing carrying capacity, and both options simultaneously. Each data point is the average of 1,000 runs done with population size $N = 1,000$. The other parameters for all runs are $Q = 0.25$, $n = 4$, and $c = 1$. The results for the infinite analytic model with no population growth using the same parameters are also shown for comparison.

These methods of adding stochasticity are somewhat *ad hoc* and I do not imply that additional stochasticity will necessarily increase $\Delta_2 Q$. In fact, the stochasticity in random group formation appears to dampen selection for altruism— $\Delta_2 Q$ is less than

would be expected in an *average* hypergeometric distribution. The effect of a particular method of introducing random noise will depend on its relative impact on within- and between-group selection. I do however show that it is possible for altruism to evolve even more easily in stochastic finite populations than it does in deterministic infinite models.

We now explore the long-term behavior of this simulation. Because here there is no mutation, empirically we observe that $Q = 1.0$ and $Q = 0.0$ act as stable equilibrium points and intermediate values do not persist indefinitely. All runs were done until one of these equilibrium points was reached and I use the percentage of 1,000 runs reaching altruist saturation, $Q = 1.0$, as a measure of how readily altruism evolves under the given conditions. For all runs and figures described in the rest of this paper we use the high stochasticity option in implementing carrying capacity, but minimum stochasticity in benefit distribution. If the results shown in Figure 4-4 apply generally, then these settings are more favorable to the evolution of altruism than the comparable analytic model, but less favorable than if we had used high stochasticity in both processes simultaneously. As was the case for the analytic model, Figure 4-5(a) shows that both higher starting Q and higher benefit values favor selection for altruism and Figure 4-5(b) shows that less altruistic benefit is required to evolve altruism for smaller group sizes.

Figure 4-5. Percentage of Simulation Runs Reaching Altruist Saturation. Shown as a function of altruist benefit, b . (a) compares the effect of different starting Q values, where the number of generations spent within groups $g = 2$ and initial group size $n = 4$. (b) compares the effect of different n where $Q = 0.1$ and $g = 2$. The total population size $N = 1,000$ and cost $c = 1$ in all runs. 1,000 runs were done for each unique set of parameters with different random number seeds. All runs were done until $Q = 1.0$ or $Q = 0.0$. Here the high stochasticity carrying capacity option and the low stochasticity benefit distribution are used.

We now investigate the effect of migration in our simulation model where the migration rate, m , specifies the probability that an individual will leave its group during each generation, moving to a randomly selected group (weighted proportionately by group size). The idea here is that larger, thriving groups are proportionately more attractive to migrants, but similar results obtain when migrants join groups at random, independent of group size. Figure 4-6 shows how the interaction between the number of generations spent within groups and the migration

rate influence selection for strong altruism. Predictably, migration lessens selection for altruism (Figure 4-6) by working to dampen the positive assortment, r^o , between actors and recipients each generation, but for intermediate numbers of generations spent within groups, even at relatively high migration rates (i.e. 30%), strong altruism evolves to saturation in some runs.

Figure 4-6. Migration and Percentage of Runs Reaching Altruist Saturation. Shows percentage of simulation runs reaching altruist saturation as a function of the number of generations spent within groups g . Compares the effect of different migration rates, m , where for all runs $Q = 0.1$, $g = 2$, $b = 10$, $n = 4$, $N = 1,000$, and $c = 1$. 1,000 runs were done for each unique set of parameters with different random number seeds. All runs were done until $Q = 1.0$ or $Q = 0$. For $m = 0.4$ all runs resulted in $Q = 0.0$ (data not shown). High stochasticity carrying capacity and low stochasticity benefit distribution are used.

Figure 4-6 also shows that even without migration ($m = 0.0$) intermediate numbers of generations within groups are most favorable to the evolution of altruism. The advantage of an intermediate number of generations is consistent with similar findings in haystack models (Wilson 1987) and models of biased sex ratios (Wilson and Colwell 1981). Note that in the simulations of Figure 4-6 it is initially unlikely that any homogeneous groups of altruists will form. With initial $Q = 0.1$ and $n = 4$ the probability of forming homogeneous altruist groups is one in 10,000 and only 250

groups are formed ($N = 1,000$, $n = 4$) at each group reformation. Yet, in the absence of homogeneous groups strong altruism can still initially increase overall even as q declines in every group. This is because groups with a higher frequency of altruists grow faster—population Q increasing while every group q decreases is an example of Simpson's paradox (Simpson 1951; Sober and Wilson 1998). With more generations within groups Q must eventually decrease as the altruists are eliminated from every group. Altruism evolves most readily when the number of generations spent within groups takes full advantage of the increase in Q due to Simpson's paradox, but avoids the inevitable decline in Q .

Figure 4-7 illustrates this tension. Here representative individual runs are shown for 2, 4, and 10 generations within groups using the same parameters as Figure 4-6 without migration ($m = 0.0$). To aid in comparison the same random number seed (same initial group distribution) is used in all three runs. For 10 generations within groups ($g = 10$), reformation clearly takes place well after peak Q values are reached and altruism eventually goes extinct. For $g = 4$, reformation takes place near peak Q values and altruism rapidly evolves towards saturation—even though altruism always decreases in the first generation after reformation. (This is true for $g = 10$ also, but harder to see as the rate of decline after reformation matches the rate before.) On the other hand when groups are reformed every other generation ($g = 2$), the potential additional increase in Q that would result from staying within groups longer is lost and altruism increases more gradually. Note that the initial increase in Q in these three runs takes place in the absence of homogeneous groups. For $g = 2$ and $g = 4$ no such

groups are formed until Q reaches about 0.3 (initial $Q = 0.1$) and in the case of $g = 10$ homogeneous groups never formed. In contrast, in the infinite analytic model homogeneous groups are always initially present and more generations within groups can allow these fastest-growing groups to become more and more dominant, even if initially rare.

Figure 4-7. Dynamics in Q for Individual Simulation Runs with Varying g . Shows altruist frequency Q for individual simulation runs of $g = 2$, $g = 4$, and $g = 10$ generations spent within groups. For all runs initial $Q = 0.1$, $b = 10$, $n = 4$, $N = 1,000$, and $c = 1$. To aid in comparison, all three runs were initiated with the same random number seed (same initial distribution into groups). High stochasticity carrying capacity and low stochasticity benefit distribution are used.

Conclusion

The main purpose of this paper is to demonstrate that strong altruism can evolve in randomly formed groups and thereby challenge a presumed theoretic limitation on the evolution of altruism. Although allowing groups to last more than one generation introduces new complications, I have demonstrated that kin selection, homogeneous groups, and population expansion are not essential to account for this phenomenon.

The fundamental explanation is that, for many initial conditions, after even just one generation of selection in randomly formed groups, the assortment between altruists and their potential recipients increases (above the expected initial value for randomly formed groups) as measured by the regression coefficient, r^o . The groups that are by chance initially dominated by altruists grow larger compared to other groups and even though the fraction of altruists declines in these groups, the absolute number of altruists poised to benefit other altruists in a subsequent generation increases. On the other hand, the groups that are by chance dominated by non-altruists do not grow as large *and* the relatively few altruists in these groups are eliminated or greatly diminished after one to several generations within groups. This also increases positive assortment as these non-altruists are stuck with each other and will receive less benefit from altruists than they did in the first generation. Of course the few non-altruists lucky enough to end up in altruist-dominated groups are the fittest individuals, but overall the conditions that favored non-altruists in the initial random distribution can switch to favor altruists in subsequent generations.

I emphasize again that even when groups are multigenerational, the vast majority of the benefit provided by altruists will fall to non-relatives—especially when altruism is initially rare (Appendix D). Altruism evolves due to the positive assortment among heritable helping behaviors regardless of whether there is a positive assortment among relatives by descent. The regression coefficient used here, r^o , measures the assortment, not genealogical relationship. This positive assortment can be viewed equivalently (Wade 1980; Queller 1985; 1992b; Frank 1998; Sober and Wilson 1998) as causing

selection on the altruistic trait (allele) via inclusive fitness or as causing selection among groups that vary in their trait composition. While interactions among kin in nature no doubt often contribute to the positive assortment of altruistic traits, kin interactions are not in themselves a requirement for altruism to evolve.

Whether strong altruism evolves in nature via mechanisms similar to those illustrated here will depend on the degree to which the assumptions of these models are representative of natural conditions. For instance, in both the analytic and simulation models we saw that strong altruism can evolve even when population size is held constant by a global carrying capacity. In nature, in addition to population-level limits on growth there are often limits on group size. While not explored here, group-level limits will dampen between-group selection for altruism, so further investigation is needed to elucidate the relative import of global vs. local levels of population control in the evolution of altruism. A lack of mutation is also unrealistic. I experimented with mutation in our models (data not shown), but in the simple binary genetics used here a mutation that switches behavioral types exerts pressure towards $Q = 0.5$ and thus favors altruism when Q is initially low. This is because the more common type experiences more mutations. Even if this bias could be compensated for, low mutation rates are unlikely to alter our basic results, which are robust under fairly high levels of migration among groups and even when homogeneous groups of altruists are “mutated” to contain at least one non-altruist.

This model started with the original assumptions of the random group models (Hamilton 1975; Wilson 1975a) and added the idea of multigenerational groups from

haystack models (Maynard Smith 1964; Wilson 1987). Just as Wilson (1987) created a simulation model to study a more realistic version of Maynard Smith's (1964) original haystack model, I have created a simulation model that adds finite population size, stochasticity, and multigenerational groups to the original analytic random group models. Whereas Wilson's (1987) haystack simulation corrected the "worst case" assumption made by Maynard Smith (1964) that groups would persist until altruism was eliminated in all mixed groups; here we correct an opposite "worst case" assumption made in random group models that groups only exist for a single generation. As demonstrated here and in the haystack simulations (Wilson 1987), an intermediate number of generations within groups is most favorable to the evolution of altruism.

Maynard Smith (1998a) in discussing different views on the evolution of altruism recently echoed the original findings of Hamilton (1975) and Wilson (1975a) and the current consensus opinion in the quote given at the beginning of this chapter. Many mechanisms which result in a positive assortment among self-sacrificing behaviors have been proposed including passive methods such as foraging in non-uniform resource distributions which can be depleted (Pepper and Smuts 2002), continuous population viscosity with periodic environmental disturbances (Mitteldorf and Wilson 2000a), the coevolution of group joining and cooperative behaviors (Avilés 2002), and the presence of non-participants (Hauert et al. 2002), as well as active methods such as kin recognition (Gamboa, Reeve, and Holmes 1991), conditional strategies based on past actions (Trivers 1971; Axelrod and Hamilton 1981) or reputation (Nowak and

Sigmund 1998; Panchanathan and Boyd 2003), policing (Frank 1995b; 2003), punishment of non-altruists (Boyd and Richerson 1992; Fehr and Gächter 2002; Boyd et al. 2003), the coevolution of cultural institutions that constrain individual behaviors (Bowles, Choi, and Hopfensitz 2003), and even recognition of arbitrary tags (Riolo, Cohen, and Axelrod 2001). Here I have shown in both deterministic and stochastic models that when groups exist for more than one generation such specific or more complex mechanisms for creating positive assortment, although certainly important if present, are not needed—the positive assortment that develops between randomly created multigenerational groups can suffice for between-group selection to dominate within-group selection and thus for strong altruism to evolve.

The reciprocal altruism strategy can be selected for the same reason as altruism towards kin: covariance between the performance of the behavior and the recipients frequency of the altruism allele.

David C. Queller, in *Nature* (1985)

Chapter 5

Hamilton's Rule Applied to Reciprocal Altruism¹

Over two decades ago Axelrod and Hamilton (Axelrod and Hamilton 1981; Axelrod 1984) suggested two alternative mechanisms for the evolution of cooperative (altruistic) traits: 1) when the benefits of altruism fall to relatives, inclusive fitness (Hamilton 1964; 1970; 1972) is the likely mechanism; and 2) when benefits fall to non-relatives, reciprocal altruism (Trivers 1971) is the likely mechanism. Around this same time researchers began to show that multilevel (group) selection theory (Wilson 1975a; Wade 1978), a third alternative not emphasized by Axelrod and Hamilton, is mathematically equivalent (Wade 1980; Queller 1985; 1992b; 1992a; Frank 1998; Sober and Wilson 1998) to Hamilton's inclusive fitness theory. This theory shows that Hamilton's coefficient of relatedness can be treated as a measure of assortment

¹ This chapter is based on a paper accepted into the proceedings of the Congress on Evolutionary Computation and will be presented there in Portland, Oregon June 2004 (Fletcher and Zwick 2004b). It will also be submitted in an expanded form to a biology journal.

among altruistic traits (or behaviors) rather than strictly as a measure of kinship (Hamilton 1972) and that the differential productivity of groups constitutes a group-level selection process (Price 1970). Other researchers have gone further in claiming that reciprocal altruism also can be unified with multilevel selection and inclusive fitness theories. Queller (1985) originally suggested that a generalization he developed of Hamilton's rule could be applied to reciprocal altruism and more recently Sober and Wilson (1998) have claimed such a unification is possible, yet there has not been a clear demonstration of *how* to apply Queller's equations to traditional reciprocal altruism models. In fact in a subsequent expansion of his original result Queller (1992b; 1992a) does not mention its application to reciprocal altruism.

Here I take up this thread of inquiry and demonstrate that Queller's equations do indeed provide a foundation for the unification of reciprocal altruism, inclusive fitness, and multilevel selection theories. I do this by using Hamilton's inclusive fitness rule to predict whether a reciprocal altruism strategy will increase or not within a classic iterated prisoner's dilemma (PD) model. This model is similar to those used by Axelrod and Hamilton in their original work. In this way I support the idea that, rather than relying on alternative mechanisms, these theories offer different ways of understanding a common principle by which self-sacrificing traits can increase. This principle is that altruism can evolve if there exists: 1) sufficient non-zero-sum benefits of cooperation such that on average the benefit provided to others by an altruist exceed the cost to the altruist, 2) sufficient positive assortment among heritable altruistic behaviors such that on average altruistic genotypes benefit more from these behaviors

than average population members. The “sufficient” positive assortment and benefit to cost values necessary for altruism to increase are given by Queller’s generalization of Hamilton’s rule. This rule applies whether the source of positive assortment among altruistic behaviors is due to conditional strategies, higher interaction levels among relatives, or other causes of population grouping (even among non-relatives).

Obstacles to Applying Hamilton’s Rule to Reciprocal Altruism

There are two major obstacles in unifying reciprocal altruism with inclusive fitness and multilevel selection theories—one stemming from the difference between genotype and phenotype and the other from non-additive fitness functions. The first obstacle is that by definition there is not a one-to-one relationship between behavior and genotype in reciprocal altruism models. In contrast, inclusive fitness and multilevel selection models most often assume that there is a cooperation or altruism “gene” (or some additive combination of genetic determinants) and that the frequency of the genetic trait is proportional to the frequency of the cooperative or altruistic behavior. With reciprocal altruism models, however, behaviors are conditioned on environmental circumstances (e.g. the behaviors of others) and are not generally proportional to genetic frequencies. For example, if there is a trait such that cooperation is conditioned on what others did in previous encounters, then the frequency of this trait does not directly predict the frequency with which an individual

displays cooperative behaviors. For this one must know the previous behaviors of those with whom the individual interacts.

Because the round-robin tournament structure used by Axelrod and Hamilton ensures no positive assortment among player types (ignoring the play against ones “shadow” self (Axelrod 1984)), it seems that altruism cannot evolve via either inclusive fitness or multilevel selection. Both Hamilton (1975) and Wilson (1975a) (working independently) have shown that strong altruism does not evolve when groups are formed randomly because there is not enough assortment in the altruistic trait. Recently, however, we showed that altruism *can* evolve in randomly formed groups if these groups are multigenerational (Fletcher and Zwick 2004c, Chapter 4). Here I will demonstrate that altruism can also evolve in single generation groups (pairings) when *behaviors* are positively assorted—even in the absence of assortment between the interacting *genetic types*.

The second obstacle to unification is that the fitness (or utility) values used in reciprocal altruism models are not necessarily additive and therefore are not decomposed into the familiar benefits and costs commonly employed in inclusive fitness and multilevel selection models. This non-additivity is true of the most common payoff matrices used in iterated PD experiments (Figure 5-1), including those of Axelrod and Hamilton. I explain this further in the section on non-additivity.

Both of these obstacles, as I demonstrate below, are effectively handled by Queller's generalizations of Hamilton's rule. I begin by briefly reviewing the iterated PD experiments of Axelrod and Hamilton as well as Hamilton's rule and Queller's

contributions. I then demonstrate how Hamilton's inclusive fitness rule can be applied to an additive PD experiment involving a conditional strategy, and then finally do the same when fitness payoffs are non-additive.

The Iterated Prisoner's Dilemma

The prisoner's dilemma (PD) game captures a fundamental problem of social life—individually rational behavior may lead to a collectively irrational and deficient outcome. Figure 5-1 illustrates a typical 2-player PD payoff matrix (and the one used by Axelrod and Hamilton (Axelrod and Hamilton 1981; Axelrod 1984)). The game is symmetric and each player has a dominant strategy to defect (D). This mutual defection results in each player receiving a payoff of 1, but if they had both cooperated (C), they could have each received a payoff of 3. In this game it is presumed that players choose their strategy simultaneously and there is no knowledge or guarantees about what the other player will do. The dilemma is that cooperation makes a player vulnerable to exploitation—in this case the exploiting defector gets 5 while the lone cooperator gets 0.

		opponent's behavior	
		C	D
actor's behavior	C	3	0
	D	5	1

Figure 5-1. Typical PD Utility (or Fitness) Values for Actor. Shows actor's payoff given its own and opponents behaviors. Behaviors are either cooperate (C) or defect (D).

An n-player version of the PD in which cooperators contribute to the common good (also called *whole-group* altruism (Pepper 2000b)) models familiar social dilemmas such as the “tragedy of the commons” (Hardin 1968) and the “free-rider problem” (McMillan 1979; Avilés 2002). Elsewhere I have demonstrated that an n-player PD with fixed strategies played across multiple groups captures fundamental features of the evolution of altruism via multilevel selection (or equivalently inclusive fitness) (Fletcher and Zwick 2000a; 2000b; 2001, Chapter 3). Here I will show that a 2-player game with conditional strategies also fits into this broader framework.

Although in a PD situation it is individually rational to defect in single plays of the game, Axelrod and Hamilton (Axelrod and Hamilton 1981; Axelrod 1984) showed that conditional cooperative strategies can do well overall when games are iterated. This combination of iterated games and conditional play can create positive assortment among *behaviors* even when there is none among *player types*. In their experiments many different types of players (submitted by researchers from many fields) competed in a round-robin tournament in which each submission played all others an average of 200 times. The most successful strategy in Axelrod and Hamilton's experiments was also one of the simplest. Submitted by social scientist Anatol Rapoport, it was called Tit-For-Tat (TFT). TFT always cooperates with an opponent on the first move and then in all subsequent moves simply plays whatever the opponent did in the last game. Even though TFT never came out ahead in any single match of 200 iterated games (Rapoport 1991; Sober and Wilson 1998), its overall score was superior as it was able to minimize exploitation by defecting

strategies such as Always-Defect (ALLD), while taking advantage of mutual cooperation when it met other “nice” strategies. When viewed as an evolutionary process where utility payoffs of the PD are used to determine the number of each player type in the next generation, Axelrod and Hamilton showed that both ALLD and TFT are Evolutionarily Stable Strategies (ESS). That is, they cannot be invaded by individual mutations or migrants exhibiting a different strategy. Since these original experiments more than 20 years ago, much research has been done on the iterated PD (e.g. see Dugatkin 1997 for review).

Hamilton's Rule

Hamilton's rule gives the condition necessary for an altruistic trait to increase in a subsequent generation and is deceptively simple (1964):

$$rb > c \quad (5.1)$$

where b is the average benefit provided by an altruist and c is the average cost to the altruist. The complications lie in the meaning of the r term which has been progressively generalized over the years. Originally thought of as a simple measure of relatedness via descent (Hamilton 1964), Hamilton (after interacting with Price (1970)) broadened the meaning of r to be a measure of the interaction of genetic types regardless of relatedness:

$$r = \frac{\text{cov}(G_A, G_O)}{\text{var}(G_A)} \quad (5.2)$$

where G_A is the genotype or breeding value with respect to the altruistic trait for each potential actor and G_O is the average genotype or breeding value of those that interact with the actor. After several examples of Hamilton's rule failing were reported in the literature (e.g. Karlin and Matessi 1983), Queller showed that these were due to phenotype/genotype differences and that it was the altruistic *behaviors* of others and itself that determined whether the actor's genotype increased, not the *genotype* of others. Queller generalized Hamilton's r term as follows (1985; 1992b; 1992a):

$$r = \frac{\text{cov}(G_A, P_O)}{\text{cov}(G_A, P_A)} \quad (5.3)$$

where P_O is the average phenotypic value of others interacting with the actor and P_A is the average phenotypic value of the actor.

Note that Eq. 5.3 is more general than Eq. 5.2. In the case of no genotype/phenotype difference ($P_O = G_O$ and $G_A = P_A$) Eq. 5.3 reduces to Eq. 5.2. The absence of any G_O component in this most general interpretation of Hamilton's r term by Eq. 5.3 is especially noteworthy because it challenges the selfish gene (Dawkins 1976; 1982) view of inclusive fitness in which genes encode altruistic behaviors only because these behaviors help other copies of the genes residing in those receiving the altruism.

Based on the definition of covariance, Eq. 5.3 can also be written as:

$$r = \frac{\sum (G_A - G_P)(P_O - P_P)}{\sum (G_A - G_P)(P_A - P_P)} \quad (5.4)$$

where the summation is over each individual in the population (each potential actor) and P_P is the average population phenotypic value. I use this equation to calculate r below. Note that here there is no mutation and the genetics are haploid, but Frank (1998) has expanded Queller's equations to accommodate different degrees of fidelity in transmission from one generation to the next.

Applying Hamilton's Rule to Reciprocal Altruism

Here I provide a simple example of applying Hamilton's rule to reciprocal altruism using a population consisting of two classic evolutionarily stable (ESS) types, TFT and ALLD. Because one of the types (TFT) uses conditional behaviors we must measure genotypes and phenotypes separately. Figure 5-2 illustrates a general view of the PD in which an altruist provides a benefit value of b to its opponent at a cost c to itself. w_0 is the base fitness value unrelated to the altruistic trait. The d term is a deviation from additivity and is discussed in the next section; in this section $d = 0$. Behaviors are either cooperate (C) which has a phenotypic (P) value of 1 or defect (D) which has a phenotypic value of 0. Note that the fitness values in Figure 5-2 depend only on the phenotype of the actor (P_A) and the opponent (P_O), not on their genotypes. Since we will keep track of the fraction of TFT players in the population, G_P , we give the TFT type a genotypic (breeding) value of 1 and ALLD a value of 0.

actor's behavior	opponent's behavior	
	C ($P_O = 1$) contributes b	D ($P_O = 0$) contributes 0
C ($P_A = 1$) sacrifices c	$w_0 + b - c$ (+ d)	$w_0 - c$
D ($P_A = 0$) sacrifices 0	$w_0 + b$	w_0

Figure 5-2. PD Utility (or Fitness) Values Based on Individual Contribution. Shows PD utility (or fitness) values *for the actor* given its own and opponents behavior. The fitness (or utility) values for the actor are represented as the sum of additive contributions from the opponent and its own sacrifice. The w_0 term is an additive base fitness value uncorrelated with C and D behaviors. The d term is the deviation from additivity (addressed in next section).

In this population of two types there will be three possible pairings each with set values for G_A , P_O , and P_A . Table 5-1 gives the values for each of the player types when it is the potential actor (A) and its opponent (O) is either the same type or different. To calculate r using Eq. 5.4 the overall *average* behavior of the actor (P_A) and opponents (P_O) must be calculated based on the frequency of each situation in Table 5-1.

Table 5-1. G_A and Predicted P_A , and P_O Values for Individual TFT and ALLD Players.

Actor (A)	Opponent (O)	G_A	P_A	P_O
TFT	TFT	1	1	1
TFT	ALLD	1	$1/i$	0
ALLD	TFT	0	0	$1/i$
ALLD	ALLD	0	0	0

The value $1/i$ in Table 5-1 represents the fact that when playing an ALLD player, TFT will cooperate once on the first play out of i total plays in this pairing. To use Eq. 5.4, we need additionally the values of i , G_P , and P_P . We take i and G_P to be

parameters of the model, but the overall fraction of cooperate behaviors, P_P , can be calculated from the frequencies (f) with which different pairings occur:

$$P_P = \frac{f_{TT} 2i + f_{TD}}{2i}, \quad (5.5)$$

where the numerator of Eq. 5.5 represents $2i$ cooperate behaviors in all TFT-TFT pairings plus one cooperate behavior in all TFT-ALLD pairings.

I first provide a simple example of using Eq. 5.4, Table 5-1, and Eq. 5.5 to calculate Hamilton's r term for conditional strategies in an iterated PD, and then provide a more general case for an infinite population. For simplicity, first imagine a population with just 2 TFT players and 2 ALLD players where each player plays each of the other players four times ($i = 4$) in an iterated PD game defined in terms of b and c as in Figure 5-2 (with no deviation from additivity, $d = 0$). For these 4 players there are 6 unique pairings and $f_{TT} = 1/6$, $f_{TD} = 4/6$, and $f_{DD} = 1/6$. Substituting into Eq. 5.5 gives an overall expected fraction of cooperate behaviors of $P_P = 1/4$. $G_P = 1/2$ for this situation (50% TFT type), $G_A = 1$ for TFT, and $G_A = 0$ for ALLD. Lastly we must calculate the average P_A and P_O values based on the situation for each type of actor. As an actor a TFT player is paired once with the other TFT and once with each of the two ALLDs. So using Table 5-1, for TFT the averages across the three pairings are $P_A = (1 + 1/i + 1/i) / 3 = 1/2$ and $P_O = (1 + 0 + 0) / 3 = 1/3$. The averages for ALLD are $P_A = (0 + 0 + 0) / 3 = 0$ and $P_O = (0 + 1/i + 1/i) / 3 = 1/6$. We are now ready to substitute into Eq. 5.4 where the summations are over each player as the actor and the values are

the same for each of the two TFT players (listed first) and each of the ALLD players (listed second):

$$r = \frac{\sum (G_A - G_P)(P_O - P_P)}{\sum (G_A - G_P)(P_A - P_P)} \quad (5.4 \text{ repeated})$$

$$r = \frac{2(1-1/2)(1/3-1/4) + 2(0-1/2)(1/6-1/4)}{2(1-1/2)(1/2-1/4) + 2(0-1/2)(0-1/4)} = \frac{1}{3} \quad (5.6)$$

Thus, with this degree of assortment among the *behaviors* of others and actors in this 4-player population, according to Queller's version of Hamilton's rule we would expect the TFT type to increase (gain more utility than the ALLD type) when $b/c > 3$. I confirm this below, but first consider a more general case.

The expected frequency of pairing types in general can be calculated from G_P and population size for various ways of forming pairs (e.g. a binomial or hypergeometric distribution). The round-robin tournaments used by Axelrod and Hamilton are equivalent to the expectation for a hypergeometric distribution (sampling without replacement) if one ignores the *ad hoc* modifications they used of having each player play a "shadow" self as well as an opponent playing randomly. For a round-robin tournament (hypergeometric distribution) of large population size, the frequencies of pair types approaches those of a binomial distribution. Here for simplicity I will assume random pairings from an infinite population and therefore use a binomial distribution. In this case $f_{TT} = (G_P)^2$, $f_{TD} = 2G_P(1 - G_P)$, and $f_{DD} = (1 - G_P)^2$. We thus have $P_P(i)$ and also from Table 5-1, G_A , $P_A(i)$, and $P_O(i)$. Now using only G_P and i as

parameters to this simple model, I can calculate r from Eq. 5.4 and use it in Hamilton's rule (Eq. 5.1) to predict whether TFT will increase or not.

In order to check the predictions of Hamilton's rule, I also calculate the fraction of TFT in a subsequent generation, G_P' , using the frequencies (f) of pairings and the PD payoffs (utilities) for each player from Figure 5-2. This assumes that the number of each player type in the new generation is proportional to the fitness (or utility) values gained in the previous generation. A generation is defined as one round of pairings each with i games between paired players. G_P' is then given by (derivation in Appendix F):

$$G_P' = \frac{f_{TT} 2i(w_0 + b - c + d) + f_{TD}(iw_0 - c)}{f_{TT} 2i(w_0 + b - c + d) + f_{TD}(2iw_0 + b - c) + f_{DD} 2iw_0} \quad (5.7)$$

where for now $d = 0$.

For the simple 4-player population example above where $r = 1/3$, we can confirm that a b/c ratio of 3 is the balance point. Using $b = 3$, $c = 1$, $w_0 = 1$, $d = 0$, and f_{TT} , f_{TD} , and f_{DD} values of $1/6$, $4/6$, and $1/6$ respectively (as above), $G_P' = G_P = 0.5$. If $b > 3$ while all other parameters are held constant (including c), then TFT will increase as predicted by Hamilton's rule; if $b < 3$ TFT will decrease.

Figure 5-3 shows the change in the fraction of TFT in the infinite binomial population after one generation (or tournament), $\Delta G_P = G_P' - G_P$, as a function of benefit level where cost and base fitness are held constant at $c = 1$ and $w_0 = 1$. Calculation results are shown for a few different starting G_P and i values. Note the general trends— ΔG_P is positive for higher initial G_P and more game iterations.

Figure 5-3. Change in Fraction of TFT Players for Additive Payoffs. Shows change in fraction of TFT players in population after one generation, ΔG_P , as a function of benefit level, b . Data is shown for indicated initial G_P and i . Cost c and base fitness w_0 are both held constant at 1.

For our purposes, the main point is to notice that the precise value of b needed to get an increase in TFT, $\Delta G_P > 0$, is again given by Queller's version of Hamilton's rule. Rearranging Eq. 5.1 gives the condition for TFT to increase of $b > c/r$. Table 5-2 gives the c/r values for the five example set of parameters used in Figure 5-3 (where cost is held at $c = 1$). Each c/r value in Table 5-2 precisely predicts the boundary condition for the benefit level needed for TFT to increase, i.e. $\Delta G_P > 0$.

Table 5-2. Hamilton's Rule Minimum b for $\Delta G_P = 0$ in Figure 5-3.

G_P	i	Minimum $b = c/r$
0.1	7	2.6667
0.1	5	3.5000
0.3	2	4.3333
0.1	3	6.0000
0.1	2	11.0000

These results from Figure 5-3 and Table 5-2 confirm that Queller's version of Hamilton's rule which utilizes the *phenotypes of others* can be used to predict whether a conditional reciprocal altruism strategy (e.g. TFT) will increase, whereas Hamilton's version of his rule which utilizes the *genotypes of others* does not give the correct prediction in this situation. Queller's more general version predicts the necessary balance of levels of positive assortment in helping *behaviors* and the ratio of benefits to costs necessary for a conditional trait to increase.

Hamilton's Rule and Non-Additivity

The version of Hamilton's rule given in Eq. 5.1 relies on the assumption that fitness (utility) benefits and costs are additive and due to the independent effects of individual behaviors. But note that it is not possible to pick positive values for w_0 , b , and c in Figure 5-2 that will sum to the values found in Axelrod and Hamilton's classic version of the PD in Figure 5-1. This latter PD (Figure 5-1) is non-additive. One way to view non-additivity is as the presence of some positive or negative synergy for mutual action (where defection is considered non-action). As an example of positive synergy, imagine a cooperative trait that involves hunting for one's paired partner—when both hunt simultaneously the catch may exceed two times the individual result. Negative synergy can be seen in terms of diminishing returns. In a different ecological context two hunters might interfere with each other or be exploiting the same limited resource—their total might then be less than the sum of individual efforts. The d term

in Figure 5-2 gives the deviation from additivity. Now we can use the values $b = 4$, $c = 1$, $w_0 = 1$, and $d = -1$ to arrive at the familiar PD in Figure 5-1¹.

Again we can rely on Queller (1985; 1992b; 1992a) who also developed a version of Hamilton's rule that accommodates a non-linear term. In this case the condition for an altruistic trait to increase is (1985; 1992b; 1992a):

$$rb + r_{dev}d > c \quad (5.8)$$

where r is still given by Eq. 5.4 and r_{dev} is given by:

$$r_{dev} = \frac{\text{cov}(G_A, P_A P_O)}{\text{cov}(G_A, P_A)} = \frac{\sum (G_A - G_P)(P_A P_O - \overline{P_A P_O})}{\sum (G_A - G_P)(P_A - P_P)} \quad (5.9)$$

Notice that the difference between r_{dev} and r is that P_O is replaced by the interaction term $P_A P_O$ in the numerator. This r_{dev} term scales the amount of deviation from additivity (d in Eq. 5.8) for mutual cooperation (C-C interactions) by the proportion of an actor's behaviors that take place in such mutually cooperative interactions.

Again, to verify Hamilton's rule in this reciprocal altruism model we will need to calculate G_P' using the deviation in the payoff for mutual cooperation, which happens only in TFT-TFT pairings (see Eq. 5.7). Figure 5-4 shows the effect of non-additive deviations on ΔG_P where parameters are set to resemble those in Axelrod and Hamilton's iterated PD experiments, $b = 4$, $c = 1$, $w_0 = 1$, and $i = 200$. Again, when $d = -1$ this corresponds to the PD used in their experiments (Figure 5-1). There are again general trends such that ΔG_P is easier to make positive for higher initial G_P and higher

¹ Note that a negative d value here implies that the classic PD (Figure 5-1) penalizes mutual cooperation—making it more difficult for cooperation to evolve than in an additive situation.

values of d . Note also that TFT can gain in the population even when relatively rare, given the high number of iterated interactions, $i = 200$. The more linear appearance of the curves in Figure 5-4 (compared to Figure 5-3) is due to scale differences—here I start with much smaller G_P values and plot much smaller changes.

Figure 5-4. Change in Fraction of TFT Players for Non-Additive Payoffs. Shows the change in the fraction of TFT players in a population after one generation, ΔG_P , as a function of the deviation from additivity for mutual cooperation, d . Data is shown for indicated initial G_P . Cost c and base fitness w_0 are both held constant at 1, $b = 4$, and the number of iterated games $i = 200$.

To check the application of Hamilton's rule under these non-additive conditions we rearrange Eq. 5.8 to get the condition:

$$d > (c - rb) / r_{dev}. \quad (5.9)$$

Table 5-3 shows the value of $(c - rb) / r_{dev}$ for the parameters used in Figure 5-4. For each set of parameters, these values precisely predict the threshold value of d where ΔG_P becomes positive, thus, demonstrating that this version of Hamilton's rule can be successfully applied to reciprocal altruism models, even when fitness values are non-additive.

Table 5-3. Hamilton's Rule Minimum d for $\Delta G_P = 0$ in Figure 5-4.

G_P	Minimum $d = (c - rb) / r_{dev}$
0.005	-1.985
0.004	-1.735
0.003	-1.318
0.002	-0.485
0.001	2.015

Conclusion

I have demonstrated how reciprocal altruism can be analyzed using Hamilton's inclusive fitness rule. The key is to use general versions of Hamilton's rule appropriate to the situation under study. In reciprocal altruism with its conditional strategies and repeated interactions there can be a positive assortment *among behaviors*, even when there is no positive assortment *among genetic types*. Therefore these differences must be accommodated in the analysis. Similarly, when the fitness consequences of individual behaviors are non-additive, this deviation must be accounted for. Queller (1985; 1992b; 1992a) has provided versions of Hamilton's rule that handle both of these situations.

This example has been minimal with only two different types, a fixed number of interactions, and predictable fractions of interacting pair types from an infinite binomial distribution. But these techniques can be used for more complicated situations if one uses the actual frequency of phenotypes and genotypes observed in all interactions.

This demonstration of an underlying unity in mechanisms for the evolution of altruism broadens the meaning of both reciprocal altruism and inclusive fitness. Fundamentally, the evolution of altruism depends on the positive assortment of social behaviors such that “nice” types help each other more than they help average population members. In this sense, altruism requires reciprocity in order to evolve whether based on conditional strategies or some other source of positive assortment. The notion of inclusive fitness must also be broadened beyond the selfish gene viewpoint (Dawkins 1976; 1982) when phenotype differs from genotype. For instance, if one imagines a single Always-Cooperate (ALLC) individual interacting in a population where all remaining individuals are TFT, the fitness of the ALLC player would be equal to the fitness of a TFT player—not because it selfishly helps copies of its own genetic type (ALLC), but because it successfully exchanges cooperate behaviors with the TFT type. Axelrod and Hamilton emphasized that symbiotic interactions were an example of cooperation between different species that could not be based on relatedness. In the above example one could view the ALLC and TFT types as different species. Cooperation across species is a clear demonstration that the “selfish-gene” of kin-selection theory is only a special case of a much more general

phenomenon. Although it sounds strange to say, inclusive fitness as captured by Hamilton's rule can involve different species as well as non-relatives of the same species.

Reciprocal altruism is not fundamentally different from inclusive fitness or multilevel selection. As I have shown using Queller's equations (1985; 1992b; 1992a), inclusive fitness can be generalized to encompass reciprocal altruism as a special case. As already noted, inclusive fitness and multilevel selection have been shown to be mathematically equivalent. It thus follows that all three theories offer different accounts of the same phenomenon. In fact there are many specific mechanisms by which altruism can evolve, including passive ones such as foraging in non-uniform resource distributions (Pepper and Smuts 2002), continuous population viscosity with periodic environmental disturbances (Mitteldorf and Wilson 2000a), the coevolution of group joining and cooperative behaviors (Avilés 2002), the presence of non-participants (Hauert et al. 2002), and multigenerational randomly-formed groups (Fletcher and Zwick 2004c), as well as active methods such as kin recognition (Gamboa, Reeve, and Holmes 1991), conditional behavior based on the reputations (Nowak and Sigmund 1998; Panchanathan and Boyd 2003) or past behaviors (Trivers 1971; Axelrod and Hamilton 1981; Axelrod 1984; Dugatkin 1997) of others, policing (Frank 1995b; 2003), punishment of non-altruists (Boyd and Richerson 1992; Fehr and Gächter 2002; Boyd et al. 2003), the coevolution of cultural institutions that constrain individual behaviors (Bowles, Choi, and Hopfensitz 2003), and even recognition of arbitrary tags (Riolo, Cohen, and Axelrod 2001). What all these

mechanisms have in common is: 1) sufficient non-zero-sum benefits of cooperation, and 2) sufficient positive assortment among heritable altruistic behaviors.

Be warned that if you wish, as I do, to build a society in which individuals cooperate generously and unselfishly towards a common good, you can expect little help from biological nature.

Richard Dawkins, in the introduction to *The Selfish Gene* (1976)

The stupid way to be selfish is ... seeking happiness for ourselves alone and in the process becoming more and more miserable. The intelligent way to be selfish is to work for the welfare of others.

The Dalai Lama, in *The Way to Freedom* (1994)

Chapter 6

Summary and Discussion

In this final chapter I summarize the major findings of this dissertation and discuss my results in the larger context of how the evolution of altruism is understood in both biological and human terms.

Summary of Results

Fundamentals and Theory Unification

The major goal of this dissertation is to help clarify the debate surrounding the evolution of altruistic traits. Identifying the most fundamental elements of the process allows us to see more clearly the similarities among the major theories on how altruism evolves. These theories include inclusive fitness with its emphasis on selfish

genes and familial relationship, multilevel selection theory with its emphasis on between- and within-group selection, and reciprocal altruism with its emphasis on conditional strategies and memory. The fundamental requirements for altruism to evolve are:

1. Non-zero-sum fitness benefits for cooperation
2. Positive assortment among heritable cooperative behaviors

In the preceding chapters I demonstrate how each of these theories embodies these fundamental requirements in two general ways: by showing how Hamilton's rule applies to each theory and by demonstrating that the PD is inherent in models of each theory. I also highlight the way multilevel selection theory emphasizes the need for positive assortment in altruistic behaviors and the tension between levels of selection. While these aspects are implicit in the other theories, the emphasis in inclusive fitness theory on relatedness *per se* (rather than groupings possibly due to relatedness) and reciprocal altruism theory's lack of distinction between the assortment among genetic types vs. assortment among helping behaviors, both obscure what multilevel selection theory makes explicit.

The Generality of Hamilton's Rule

The first demonstration of unification is in the general applicability of Hamilton's rule. In Chapters 3 and 5 I show that Hamilton's rule ($rb > c$) from inclusive fitness theory can be successfully applied to models from the other two theories. The reason for the success of Hamilton's rule across these seemingly different explanations is that it

embodies in its variables the two basic common requirements for selection of altruistic traits listed above:

1. The $b > c$ condition (even when r is maximal at 1) expresses the need for cooperative behaviors to produce non-zero-sum fitness advantages to the system as a whole.
2. Its r term captures the degree of assortment among actors' genes and the helping *behaviors* of others.

Especially noteworthy is the demonstration in Chapter 5 that a general version of Hamilton's rule developed by Queller applies to reciprocal altruism with the phenotype/genotype differences of its conditional strategies. This demonstrates it is the *phenotypic* behaviors of others that determine the fitness of an altruistic trait, not their *genotypes*. This thus challenges the selfish-gene notion that altruistic behaviors between individuals evolve via help given to other copies of themselves in others. This broader view allows the same fundamental requirements for altruism to be applied to cooperation among non-relatives with different genes and even to cooperation across species in symbiotic mutualistic relationships.

The Generality of the Prisoner's Dilemma

The second demonstration of unification is the prisoner's dilemma as a general model of the conditions that must be overcome for altruism to evolve. In the context of altruism, requirement #1 (non-zero-sumness) implies a PD situation and the NPD evolutionary model with multiple groups can encompass requirement #2 (positive

assortment) as well. Thus this game-theoretic model captures these fundamental features and the conditions for altruism to increase in our model are equivalent to Hamilton's rule. This model enables me to demonstrate simple relationships between Hamilton's rule, Simpson's paradox, and the Price covariance equation.

In addition it highlights the fundamental similarities among these theories from a game-theoretic vantage point. While the PD has long been known as a model of reciprocal altruism, here I show for the first time that it is inherent in fitness functions for the other two theories as well. The PD defines the problem of altruism and Hamilton's rule gives the necessary balance between the degree of non-zero-sumness in the PD ($b > c$) and the degree of assortment in behavioral interactions among PD players (r). Hamilton's rule can thus be seen as defining the conditions for altruism to increase, the conditions for a Simpson's paradox, or in game-theoretic terms, the conditions for cooperators to gain more utility than defectors in a NPD situation—that is, the conditions for overcoming a tragedy of the commons or free-rider problem.

The Generality of Each Theory

Not only can the most general form of Hamilton's rule from inclusive fitness theory be applied to the other theories, but each of the other theories in turn has universal aspects when considered in *its* most general form. For example, reciprocity is universally required for altruism to evolve in that the positive assortment requirement (#2 above) can be interpreted as a need for altruists to exchange more fitness benefits

with each other than with average population members (whether or not this is mediated by conditional strategies or memory of past actions).

Hierarchical structure in behavioral interactions is also universally required in that this positive assortment requirement #2 can additionally be seen in terms of multilevel (or group) selection where (in the case of whole-group altruism) r is a measure of what proportion of the variance in the trait is due to differences among groups (whether or not these groups are well-defined physical groups). For instance, we saw in the last chapter that conditional strategies can group behaviors even when there is no assortment in pairings or in the number of interactions. Multilevel selection theory in its most general form includes both physically defined groups as well as grouped behavioral interactions. In addition, fitness at the group level can be measured in terms of groups that bud new daughter groups or in terms of the disproportionate contribution to a global population. In either case the more fit (faster growing) groups are those with a higher proportion of altruists due to the non-zero-sum advantage ($b > c$) of altruistic behavior #1. Yet at the same time non-altruists are more fit within every group. This tension between selection at different levels that is explicit in multilevel selection theory is also implicit in the others—that is, inclusive fitness theory and reciprocal altruism theory specify, respectively, the need for *sufficient* levels of inclusive fitness or reciprocity among altruists (measurable with r) in order to overcome direct individual selection for selfish behavior.

The most general versions of these theories may diverge from their original, and more narrowly conceived, assumptions. For example, well-defined physical groups in

group selection theory or memory of past actions in reciprocal altruism theory are not included in the more general form of these theories. Again, in the most general form of inclusive fitness theory (Hamilton's rule) developed by Queller (1985; 1992b; 1992a), the original assumptions of kinship and selfish genes that help other copies of themselves are both absent. These more narrow assumptions therefore are *not fundamentally necessary* for altruism to evolve.

Of course when conditions do match the narrower assumptions, the original insights hold: kinship in viscous populations *is* important in producing positive assortment in behaviors and when there is a one-to-one mapping of genotype to phenotype genes may encode behaviors that help copies of themselves in others. Likewise when conditional strategies are present that allow altruists to minimize their exploitation by non-altruists, this is an important factor in creating positive assortment among behaviors. And well-defined groups in direct competition for existence that vary in their proportion of altruists clearly *do* experience selection at the group level. But if the goal is to understand what is most fundamental, it is the most general forms of these theories that help us see what is universally necessary for altruism to evolve. *Factors that are not fundamentally necessary* include: interactions among kin, selfish genes benefiting other copies of themselves, physically distinct groups with physically distinct daughter groups, and cognition and memory about past interactions or reputations.

In the taxonomy of species there is a long-standing tension between “splitters” and “lumpers” and this may be analogous to the tension between reductionists and

holists in evolutionary theory. I imagine the “splitters” or reductionists would be uncomfortable with the paragraphs above. They might complain: What is the use of group selection without well-defined groups or of inclusive fitness when it no longer involves copies of genes in others (kin)? It is indeed unfortunate when advances in our theoretical understanding are hampered by historical terminology. The multilevel selection movement of the last three decades has never shed its group selection heritage; and instead of having “Queller’s Rule” overthrowing the selfish gene viewpoint we have “Queller’s generalization of Hamilton’s inclusive fitness rule”—regrettable and yet understandable.

I have tried to describe the most fundamental features of selection for altruism with phrases not necessarily associated with any one of the current theories, e.g. *non-zero-sumness* and *positive assortment*, while still illustrating how these fundamental features are embodied in current theories. Hopefully, in this way researchers most comfortable in a particular theory will have a better understanding of how their preferred theory relates to the others.

Other Major Findings

The game theoretic analysis of Chapter 3 is also useful in teasing out the differences and similarities between strong and weak altruism. Namely, under the process of natural selection (which relies on relative fitness values) this distinction (which relies on absolute fitness values) becomes less meaningful—the difference is a matter of degree rather than a fundamental distinction about what is achievable by within- and

between-group selection. Weakly altruistic traits are selected against within groups as are strongly altruistic traits, and Hamilton's rule describes the necessary conditions for each to increase. While *a priori absolute* fitness values for weak altruism do not define a PD, under selection in a finite environment the *actual* resulting fitness values *do* define a PD—or equivalently strong altruism (i.e. if the population is scaled or if fitness success is considered in terms of fractional representation in the whole population). That is, if fitness (utility) differences rather than absolute values are used, the game being played changes from a no-conflict (NC) game in which cooperation would be the individual dominant strategy to a PD in which defection is the individual dominant strategy. Thus this game theoretic analysis unifies weak vs. strong altruism in a more general framework that incorporates selection.

Historically, one place where the weak vs. strong distinction has been emphasized is in models involving randomly formed groups. Weak altruism is always whole-group and because in this case some of the altruist's benefits fall to itself, weak altruism initially increases more easily than strong. In fact when same-sized groups are randomly formed from an infinite population the initial amount of assortment produced is never enough for strong altruism to increase, but can be enough for weak altruism to evolve. In chapter 4 I show that although for almost three decades this finding was thought to hold in general, it actually only applies to single-generation groups because the amount of assortment can increase over subsequent generations. I demonstrate that strong altruism can indeed increase in multigenerational randomly formed groups even under various restrictions including most notably the absence of

kin selection. I also show that stochasticity in group formation, benefit distribution, and culling due to carrying capacity can influence whether altruism will increase overall.

Finally, in Chapter 3 I provide a unique analysis of the Price covariance equation which both simplifies it and reveals its underlying assumptions. This allows me to develop an alternative selection decomposition that makes distinct and symmetric assumptions about the components of selection. While many biologists assume the Price equation accurately partitions selection into within and between-group components, this symmetric decomposition makes clear that the Price equation is just one possible decomposition and that it is based primarily on an idealization of the effect of *between-group* selection. The alternative decomposition is based primarily on an idealization of the effect of *within-group* selection. The differences between the values given by the two approaches highlights the interaction effects between selection at different levels, which are not accounted for in either decomposition alone. In Chapter 3 I also show how these two decompositions relate to the basic NPD model—the average slope of the fitness functions ($b - c$) is a measure of the between group component of selection given by the Price equation and the intercept difference (c) is a measure of the within-group component given by the alternative decomposition. The other components are correction terms for the assumptions made in each idealized decomposition.

Discussion

In this section I discuss how the findings of this dissertation fit into the larger context of the debate on the evolution of altruism. This includes unifying several points of confusion or contention under a common theme of causal explanations being offered at inappropriate hierarchical levels. I also touch on the question of how common or rare altruism is and whether cooperation is as natural as competition when selection takes place in hierarchical biological (and social) systems. Finally, although this dissertation does not focus on human altruism, these discussions allow me to mention some ways in which these issues relate to cooperation among human beings.

Causal Explanations and Hierarchy

One of the things that makes the evolution of altruism fertile for contentious debate (and an interesting dissertation topic) is that it embodies several paradoxical features. We have already seen how the paradox of the PD and Simpson's paradox are intimately associated with selection for altruistic traits. Here I expand this theme by reviewing other confusing issues concerning the evolution of altruism and discuss them in terms of common problems in the study of hierarchical systems. These additional issues include the sub-optimization problem, proximate vs. ultimate causes, the averaging fallacy, the PD, and the structure/function distinction. These various ways in which our understanding of the evolution of altruism is confused can be at least partially attributed to a common theme—explanations of causation being offered at inappropriate hierarchical levels.

While we can understand the temptation among scientists to reduce explanations to lower levels—after all, reductionist science has been phenomenally successful during the last century—more recently there has been an increased appreciation of phenomena that emerge in systems at higher levels (e.g. Kauffman 1993; Casti 1994; Zwick 2004). If a process in a system depends on emergent properties, then explanations of this process using only system elements at a level lower than the emergent property will never be satisfactory. For example, the selection for alleles which cause altruistic behaviors among individual organisms will never make sense *only* at the level of genes because this increase depends on the non-zero-sum benefits of mutual cooperation and positive assortment of phenotypic behaviors that *do not exist at the gene level*. Ignoring the combination of non-zero-sum benefits and population group structure necessary for cooperation to evolve (by overpowering individual within-group selection) is equivalent to ignoring the *dilemma* in the prisoner's dilemma, the *problem* in the free-rider problem, or the *tragedy* in the tragedy of the commons.

Averaging Fallacy

In the case of selection for altruism, by averaging across groups, the existence of groups are ignored even though clumped interactions account for selection of the trait. As Sober and Wilson put it, this kind of averaging has the effect of “defining group selection out of existence” (1998 p. 32). They refer to this as the “averaging fallacy” (1998 p. 31), but as Lewontin points out, it is not just the error of assuming the

probability of an average is the same as the average of individual probabilities—it is a “metaphysical one about causal reality” (1998).

For instance, Dawkins (1982) claims that group selection is not needed to understand the evolution of multicellularity because individual cells are better off being part of an aggregate than they would be if they were on their own—therefore individual selection (at the cell level) is enough to explain multicellularity. Maynard Smith and Szathmáry make a similar claim in the quote used earlier in chapter 2: “...the evolution of cooperation between the parts of an individual is to be expected” (1995 p. 53). These assertions rely on a common misconception that individual (within-group) selection maximizes individual fitness. When the evolution of altruism is understood in the context of the PD, it is clear that this is not true.

One of the contributions of this dissertation is to describe the evolution of altruism in terms of social dilemmas with which we all have direct experience. The preceding arguments are similar to one that states: “Of course citizens willingly pay their taxes because they would be worse off in the anarchy of no government and no public services,” or “Of course countries do not over fish international waters because the consequences of over fishing affect them.” While it may well be true that conditions under anarchy or unregulated fishing would be worse for all, it is also true that each person is better off individually to not pay taxes (while hopefully everybody else does), or each country is better off over fishing (while hopefully all other countries show restraint). In the absence of a higher level of organization or control

(e.g. tax collection enforcement or enforceable fishing treaties) this does in fact lead to mutual defection or a tragedy of the commons.

The same holds in the case of multicellularity—each cell is more fit by not contributing to the collective good, while other cells do. Only with a higher level of selection where groups of cells that cooperate effectively with each other out-compete cell groups locked in an NPD, can the *expected mutual defection* be overcome and these cells end up in a situation with higher individual cell fitness. Minimally this selection may be driven by random differences between groups of cells, but may also be enhanced by higher-level mechanisms that penalize defection and thus change the payoffs (fitness) values at lower levels. Michod has rightly recognized this tension between levels and noted the need of higher levels to evolve mechanisms that suppress lower-level competition—he has especially emphasized this in the origins of multicellularity (Michod 1996; 1997; Michod and Roze 2000b; 2001). Only by overcoming the drive towards mutual defection (through this kind of suppression) can higher levels stabilize.

Defection Suppression

This tension can also be seen as a *sub-optimization* problem familiar in all hierarchical systems. Sub-optimization refers both to the fact that optimization of each subsystem results in sub-optimization at the whole-system level and conversely that optimization at the higher level results in sub-optimization at the subsystem level (Zwick 2004).

There is a natural tradeoff between optimization (selection) at the different hierarchical levels.

The development of suppression mechanisms and their subsequent refinement allows selection at higher levels to be more effective (Frank 1995b; 2003). For instance, the process of fairly allotting alleles (via crossover and meiosis) to sex cells is so central to reproduction in higher animals and plants that it is often taken for granted, yet if genes were actually free to behave selfishly we would not expect this process to be so egalitarian—genes that caused more copies of themselves in new cells would be more individually fit. The fact that suppression of this lower level competition is so integral to sexual reproduction is an example of what Ludwig von Bertalanffy has called “progressive systematization” in which each successive level of hierarchy requires mechanisms for stabilizing lower levels (von Bertalanffy 1968). This is also related to Martin Zwick’s (1978) idea of a limited number of hierarchical levels “after which the hierarchy often becomes consolidated in a stable and coherent whole (which may become a base unit for still higher levels).” Sober and Wilson offer a similar idea in the context of multilevel selection where initially between-group selection may be in strong opposition to within-group selection, but over successive generations this tension may become less pronounced as within-group selection favors less individually costly ways to accomplish the group-level functionality, while between-group selection favors mechanisms that more effectively suppress within-group selection (Sober and Wilson 1998). In the same vein, Robert Wright (2000) has proposed the idea that the development of hierarchy itself is driven by the non-zero-

sum advantages available for mutual cooperation at lower levels, but he does not emphasize the need to overcome the tragedy of the commons (of mutual defection) and in fact does not support the idea of group-level selection (see Wilson 2000 for review).

When lower level competition is suppressed very efficiently it does not draw our attention. For instance, returning to multicellularity, we think of different cell lines (skin, liver, heart, etc.) as naturally cooperating to make a well functioning animal, but when suppression mechanisms fail the ever-present selection pressures at lower levels reveal themselves. This is the case with cancer in which fast growing individual cell lines are individually more fit at the expense of the common good for all cells in being part of a healthy body.

Hamilton's Rule and Hierarchy

One of the problems with the simplicity of Hamilton's rule and the inclusive fitness point of view is that it obscures the tension and interdependence between these *whole-system* vs. *subsystem* levels of causation. That is, the r term captures the population structure without emphasizing the required grouped interactions in a whole-system context. Again, partly this is historical in that the r term started as a measure of kinship, yet even when thought of strictly in terms of kin selection, r works because the grouped interactions of relatives (subsystem) are contrasted against the background of a population of non-relatives (whole-system) (Hamilton 1964). If you have instead a single unstructured population (one well-mixed group) that is highly

related (say 99.9% clonal cooperators), cooperation will steadily decline to extinction as the 0.1% of defectors steadily increase to saturation. The r term is zero in a well-mixed population (even composed of relatives) because it measures how clumped or grouped interactions are *compared to the average*, or how much altruists benefit other altruists *compared to what they give average population members*. If behavioral interactions are not grouped (no assortment) then altruism will not evolve even where relatedness is very high.

Although the r term of Hamilton's rule captures the assortment or group structure of interactions, the interpretation of inclusive fitness (caused by differences *between groups*) as being equivalent to within-group individual fitness (caused by differences *within groups*) obscures this hierarchical tension. Hamilton's rule asks a question: Is there enough group structure for altruism to evolve given a certain benefit to cost ratio? But it does not emphasize that the answer involves opposing vectors of causation. Multilevel selection theory on the other hand makes this tension explicit by emphasizing each component of selection, as in the Price covariance equation or the alternative decomposition presented in Chapter 3.

Proximate vs. Ultimate Causation

This dissertation uses a strict evolutionary biology definition of altruism—behaviors that cause an organism to have relatively less fitness (offspring) while causing others they interact with to have relatively more are altruistic. Yet, this definition does not match our common use of the term which involves physiological or emotional factors.

For instance, if it brings some one emotional pleasure to cooperate with another (even if there are fitness costs) would we call this behavior altruistic? In their book *Unto Others* Sober and Wilson (1998) separate phenotypic traits and their fitness consequences from the proximate mechanisms within an individual that may lead to the behavior. This is similar, respectively, to the Aristotelian distinction between *final cause* and *efficient cause*.

Presumably natural selection maximizes fitness (final cause) by constructing proximate mechanisms (efficient causes) for every selected behavioral trait, but whether it is initiated by nerve reflex, hormonal cascade, or a memory of pleasure is not directly relevant to the fitness consequences (assuming the proximate mechanisms are equally efficient in eliciting the behavior). If the behavioral trait with fitness consequences (final cause) is to forage for food when biochemical energy stores are low, the exact mechanism that triggers this behavior (an uncomfortable sensation of hunger or fond memories of food) is of no consequence. In the same way, if an altruistic behavior lowers individual within-group fitness while raising the fitness of other group members, whether the proximate mechanism is a pleasurable emotion or nerve mediated reflex, the fitness consequence for the behavior is the same.

This distinction is also related to the function vs. structure distinction that is common to hierarchical systems (Zwick 2004). Each system can be seen as a whole with its own properties that interface with the external environment. For instance, an organism interacting with its environment has a particular fitness. A system can also be viewed as a collection of interacting subsystems. The subsystems and their

interactions account for the way the system as a whole interacts with its environment, but it is not usually necessary to understand these lower level interactions to describe the higher level ones. For instance, organisms may contain subsystems such as nervous, circulatory, pulmonary, hormonal, and genetic, but Darwin (1859) was able to describe the process of natural selection in which individual organisms interact with their larger environment (including other organisms) without relying on the workings of these subsystems to describe this process. Focusing on structural detail often obfuscates the prominence of function when examining behavioral traits. This is related to the genotype/phenotype distinction. Selection among individuals can only act on functional phenotypes and where the trait in question is altruism towards others, only the external consequences of this altruism are relevant to fitness measures, not their structural (genetic) underpinnings. Again, this is what Queller's version of Hamilton's rule illustrates—altruistic traits gain from the *phenotypic behaviors* of those they interact with, not their underlying *genotypic* basis.

Emotions as Proximate Mechanisms

One the other hand, the existence of proximate mechanisms in organisms leads us to ask: what ultimate (final) behaviors are being selected for? For instance, recognizing the proximate mechanism of hunger in animals (including ourselves) helps us appreciate the historic selective pressures for behaviors such as food foraging and consumption of high caloric foods. Similarly, there is a growing body of evidence for proximate mechanisms in humans and other primates that support the idea that traits

for self-sacrificing, altruistic behaviors have been selected for. In a recent study capuchin monkeys were trained to perform a simple task for which they were rewarded with a piece of cucumber. This all worked fine (perform task—get cucumber) until a test subject was placed next to another capuchin monkey receiving a more desirable grape for performing the same task. In this situation many test subjects threw down their cucumber and had a temper tantrum (Brosnan and de Waal 2003).

Similarly, in studies of human cooperation involving small groups playing “public goods” games (equivalent to an NPD) subjects reported experiencing the proximate mechanism of emotional anger towards defectors in the game to such an extent that it motivated them to give up some of their own financial rewards in order to punish the defectors—even though these defectors were anonymous and unlikely to be encountered by the punisher in the future (Fehr and Gächter 2002) (also see Fehr and Fischbacher 2003 for review). These studies suggest both that we (and other animals) may have evolved to be more concerned with relative than absolute fitness (why give up money in your pocket or a perfectly good cucumber?) and that we are willing to make altruistic sacrifices to help discourage cheating in others.

Finally there is also new physiological evidence for proximate emotional mechanisms in humans. Areas of the brain associated with pleasure have been shown to “light up” (increased blood flow) on functional CAT scans when subjects act (or reason about acting) altruistically in games involving cooperation and defection and brain areas associated with anger light up more when subjects are subjected to the selfish behaviors of others (Greene et al. 2001; Sanfey et al. 2003).

Are the Conditions that Produce Altruism Rare?

The textbook viewpoint of the last few decades is that group selection (although theoretically possible) requires special circumstances that are rarely seen in nature and therefore altruism (except among close relatives) must also be rare (Maynard Smith 1964; Williams 1966; Nunney 1985b; Maynard Smith 1998a). This dissertation supports a reassessment of this traditional view in several ways. As discussed above, understanding the underlying PD nature of the evolution of altruism (illustrated in Chapter 3) helps to avoid misidentifying situations that require self-sacrificing traits as due to individual within-group selection. In addition, we have seen how the common view that altruism must be defined in absolute terms does not make sense under selection (Chapter 3) and that even when defined this way, strong altruism can evolve via randomly generated associations, rather than requiring special circumstances or additional organismic capabilities as previously believed (Chapters 3 and 4).

The answer to this question of altruism rareness of course also depends on how realistically these models and parameter values map onto natural situations. Yet Hamilton's rule helps us see that there is a possible tradeoff. If the positive assortment of interactions is less than we have assumed, then this can be compensated for with a greater benefit to cost ratio. And while it may seem excessive to have benefit values several times higher than costs, it is actually easy to imagine such situations if one considers the benefits involved in, for instance, division of labor and economy of scale. For example, the chances that a prairie dog sentinel (Sherman 1977) giving a

warning call is actually captured by a predator may be very small, but the act of warning others may reduce their chances of being caught many fold. Or as another example, altruistic *Pseudomonas* bacteria (Rainey and Rainey 2003) that exude a sticky polymer reproduce *slightly less* often than bacteria that do not make this substance, but the benefit to other bacteria in being held close to the air/liquid interface by the polymer allows them to reproduce *much faster* than they would otherwise (stuck in the anoxic liquid). Finally, the slime mold cells (Strassmann, Zhu, and Queller 2000) that make up a spore stalk may not survive anyway as environmental conditions deteriorate and so give up relatively little fitness, but they may increase the chances of survival (and subsequent offspring production) of other cells in the spore body from almost zero to substantial.

Another important point about benefit levels is that none of the examples used in this dissertation embody the non-linear benefits we normally associate with the synergy of collective cooperation and these non-linearities can contribute to the more common occurrence of altruism. The only model that deviates from additivity is the traditional PD payoff values used in Chapter 5, which actually have a *negative* deviation for mutual cooperation. The $b > c$ non-zero-sum requirement implies a certain kind of synergy—more benefit is produced than what it costs, but these additive models assume that each additional cooperator contributes an equal amount to the fitness of others. So synergy in the different sense of non-linearity is not invoked.

For example, imagine a food-sharing trait where an altruistic hunter consumes energy and risks injury that on average amounts to one rabbit's worth of calories in

order to catch 4 rabbits. The hunter then gives these rabbits to its group or tribe where they are divided evenly among all tribe members. Here the b/c ratio is 4. The additive models presented here assume that if the tribe contains 3 such altruistic hunters that they would collectively produce 12 rabbits at a cost of 3—still a b/c ratio of 4. But what if 4 hunters expending their efforts collectively are enough to bring home a buffalo! Now the ratio of benefit to cost jumps dramatically and makes this altruistic hunting trait more likely to evolve, but still only via the positive assortment of altruists—now it just requires less positive assortment (r) because the b/c value is higher. Note that if we assume the buffalo is shared equally among group members and we put ourselves in the position the fourth hunter, we might mistakenly believe this hunter joins the hunt for purely selfish reasons—a slab of buffalo is better than its share of the 12 rabbits. But remembering that fitness is a relative process we see that the other, say, 10 members of the tribe that stay in camp and do not hunt are even more fit—they get their share of the buffalo without paying the one rabbit's worth of cost. This is similar to the example involving neighbors with millions of dollars discussed in Chapter 3.

The point here is that the models used in this dissertation which show how altruism can evolve have done so without using a known and strong mechanism that promotes cooperation. It is easy to imagine in nature (and society) many *non-linear synergies* for collective action—in acquiring food, shelter, protection from predators, raising of young, etc. This kind of non-linear benefit while not modeled here and not

typical of other evolutionary models, may be very important in determining the actual level of self-sacrificing cooperation in hierarchical natural (and social) systems.

Is Nice Necessary?

The goal here is not to propose a Pollyannaish view of nature in which cooperation dominates, but rather to emphasize that competition is only half the picture. One would expect that selfish behaviors would increase when there is less higher-level competition (and less selection for suppression of local competition discussed earlier), whereas increased levels of altruistic behavior would require more strife at higher levels. For a biological system to effectively compete in natural selection it must engender some level of cooperation and sub-optimization among its parts (subsystems), while at the same time selective processes at lower levels will select for more local optimization. Those systems with more cooperative, self-sacrificing parts are likely to be more fit than systems with more selfish parts.

At the level of an individual cell cooperation among cell organelles including genes will be selected for; at the level of individual organisms, those with cooperating parts (e.g. cells) have a selective advantage; and at the level of groups of individuals in competition, cooperation and altruism among individuals within groups is selected for. Steven Gould has even suggested that as species either go extinct or fill new biological niches, there may be species-level selection (Gould 2002). In this sense natural selection is a kind of *fractal process* that can be taking place *simultaneously* on multiple hierarchical levels. Paradoxically, selection for selfish individual fitness at

any particular level may be opposed by selection for altruism by the selective pressures at a higher level. In this since altruism and selfishness are linked in a kind of *ying-yang* relationship of mutual dependence that allows natural selection to bootstrap cooperation up the hierarchical structure of life via competition at yet a higher level.

In addition to increasing our understanding of how selection and hierarchy interact in biological systems for the sake of scientific understanding, establishing a more balanced view of the role both cooperation and selfishness play in biological evolution may provide more balanced metaphors borrowed from biology about the natural tensions between self-interest and the common good. Darwin's view of natural selection that emphasized competition between organisms has been used to justify everything from genocide and eugenics to *laissez faire* capitalism. This dark view is echoed by Dawkins' quote at the beginning of this chapter, but may be based on an incomplete picture of nature—and therefore support an oversimplified view of society. For example, political rhetoric both in support of smaller or bigger government rarely acknowledges the natural tensions between market forces encompassing individual choices and government's role in protecting the common good. As the environmentalist David Orr has said: "The market is how we say 'I'; government is how we say 'we'" (Orr 2004). While both forces are important in society, it is the same averaging fallacy (discussed above) to believe that competition in free markets *alone* will maximize individual utility or lead towards the common good. One of the roles of government is in fact to help solve the PD situations that market or market-like individual actions cannot solve.

Conversely, it may also be true that social and environmental conditions have influenced our scientific understanding towards one that emphasizes the selfish half of the equation. As Gould put it: “We all have a tendency to spin universal theories from a limited domain of surrounding circumstances” (1988 p. 16). For instance, Daniel Todes (1987) has argued that both the abundance of the tropics in which Darwin did much of his research, and the more crowded conditions of London and England that influenced Malthus (1798) (from whom Darwin drew influence), help account for the emphasis on direct competition, whereas the under populated and harsh environmental conditions of Russia led many there, most famously Petr Kropotkin (Kropotkin 1902), to question both Malthus and Darwin’s emphasis on competition. They instead emphasized the need for—and the ubiquity of—cooperation or “mutual aid” in confronting the “struggle for existence” especially where the more formidable opponent was bleak environmental conditions rather than other organisms. In game-theoretic terms this latter idea is referred to as a “game against nature,” which is the subject of “decision theory,” when distinguished from game theory (Luce and Raiffa 1985).

Studies of mutual cooperation in humans have emphasized this mutual aid side rather than direct competition in situations where mutual defection may spell disaster for all. For example, James Scott’s study of subsistence peasant farmers found that mutual help in minimizing risks from crop failure were more prominent than maximizing individual benefit through direct competition (1976). Similarly cooperative systems have been documented in the sharing of a limited and vital

resource such as fresh water on the island of Bali (Lansing 1994). Interestingly when mutual defection brings mutual elimination, this may be thought of in game-theoretic terms as an n-player game of chicken (Oosterhout 1996), rather than a PD. In two-player chicken it is individually more fit to cooperate if the other player defects and there is no dominant strategy. In the n-player case it can be fitter to cooperate when *most* players are defecting. When levels of cooperation rise to levels sufficient enough to avoid disaster, the game again becomes an NPD. This produces a stable equilibrium point that is intermediate between total defection and total cooperation (even in the absence of group-level selection). For instance, Gretchen Oosterhout (1996) has shown that a fishery can be modeled as a NPD (tragedy of the commons) or n-player Chicken depending on the severity of the consequences of over-fishing. While an NPD captures altruistic cooperation, depending on the circumstances cooperation may be self-serving. Only by distinguishing between these situations (i.e. determining if the fitness functions describe a PD or game of chicken) can their underlying selection forces be assessed. Witnessing seemingly cooperative behaviors by themselves is not enough to decide without knowing the fitness consequences.

Future Studies

Although I have many research ideas that I have not had time to pursue, here I just mention a few direct extensions to topics covered in this dissertation.

- The discussion of chicken above highlights that there are other games besides the PD in the study of cooperation. While the PD captures altruism, other game-theoretic frameworks such as decision theory and coalition theory could be used to assess the boundaries beyond the strict definition of altruistic behavior. This may be important in investigating absence/persistence boundaries under harsh conditions as touched on above.
- In Chapter 5 I used Queller's version of Hamilton's rule to analyze a reciprocal altruism situation, but only two alternative types (TFT and ALLD) were used. One goal is to develop a way to apply Hamilton's rule to three or more types involving conditional strategies.
- While I have argued that Queller's equations challenge the selfish-gene viewpoint and could apply to symbiotic relationships across species, I would like to develop a model explicitly involving different species and show that this version of Hamilton's rule can indeed predict when symbiotic cooperation will increase.
- In the analysis of the Price equation and development of an alternative decomposition in Chapter 3, I noted that neither decomposition gives an accurate picture of the interaction effects between within-group and between-group selection. It seems as if a more accurate decomposition lies between the two and it may be possible to find some general way of averaging them to arrive at a better decomposition.

- All of the models used here could be made more realistic both ecologically (e.g. more flexible population structures and environmental factors such as food resources) and evolutionarily (e.g. mutation, sexual reproduction, polygenetic traits).
- One of the problems of talking about non-linear synergies for cooperation as I did in this chapter is that by definition they are not decomposable to independent effects of individual participants. Yet, only using linear fitness functions is very unrealistic and does not capture much of what we mean by the benefits of mutual cooperation. It would be nice to develop models that encompass non-linear synergies in order to investigate how such fitness relationships influence the evolution of cooperative behaviors.
- Although I only touched on it in this concluding chapter, I am very interested in the way metaphors derived from our understanding of nature influence our understanding of society. It would be interesting to think about the best way of presenting this more complex picture of coupled competition and cooperation in nature to a general audience.

Part of the appeal in pursuing the research behind this dissertation lies in the fact that the evolution of altruism is so controversial. This debate of course has historical and social roots, but I believe it is also fueled by a lack of common understanding as to what is in dispute, as well as at times rather emotional support of particular viewpoints. It is my hope that this dissertation can provide a helpful framework for understanding the most fundamental conditions necessary for altruism to evolve and

that this will increase the level of useful discussion among those that come from different traditions or find a particular theory more appealing. Only by conducting this debate from a common foundation can we hope to form testable hypothesis that are capable of influencing the thinking of all, and thereby advance our common understanding of how evolutionary processes shape our natural world. I look forward to being part of this discussion in the future.

References

- Aktipis, C. A. (2000). "The Effect of Behavioral Assortment on Selection Dynamics: Externalities, Information Processing and Subjective Commitment." *Artificial Life VII Workshop Proceedings, Portland, Oregon, 2000*, pp. 90-93.
- Albert, B., Godelle, B., Atlan, A., De Paepe, R., and Bouyon, P. H. (1996). Dynamics of Plant Mitochondrial Genome: Model of a Three-Level Selection Process. *Genetics* 144:369-382.
- Avilés, L. (1993). Interdemic Selection and the Sex Ratio: A Social Spider Perspective. *The American Naturalist* 142:320-345.
- . (1999). Cooperation and Non-Linear Dynamics: An Ecological Perspective on the Evolution of Sociality. *Evolutionary Ecology Research* 1:459-477.
- . (2000). "Modeling the Evolution of Sex Ratios and Other Traits Subject to Selection at Multiple Levels." *Artificial Life VII Workshop Proceedings, Portland, Oregon, 2000*, pp. 94-95.
- . (2002). Solving the Freeloaders Paradox: Genetic Associations and Frequency-Dependent Selection in the Evolution of Cooperation among Nonrelatives. *Proceedings of the National Academy of Science, USA* 99:14268-14273.
- Avilés, L., Abbot, P., and Cutter, A. (2002). Population Ecology, Nonlinear Dynamics, and Social Evolution I: Associations among Non Relatives. *The American Naturalist* 159:115-127.
- Avilés, L., Fletcher, J. A., and Cutter, A. (2004). The Kin Composition of Groups: Trading Group Size for Degree of Altruism. *The American Naturalist*:In Press.
- Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books, Inc.

- Axelrod, R., and Hamilton, W. D. (1981). The Evolution of Cooperation. *Science* 211:1390-1396.
- Bernasconi, G., and Strassmann, J. E. (1999). Cooperation among Unrelated Individuals: The Ant Foundress Case. *Trends in Ecology and Evolution* 14:477-482.
- Boehm, C. (1993). Egalitarian Society and Reverse Dominance Hierarchy. *Current Anthropology* 34:227-254.
- Boorman, S. A., and Levitt, P. R. (1980). *The Genetics of Altruism*. New York: Academic Press.
- Bowles, S., Choi, J.-K., and Hopfensitz, A. (2003). The Co-Evolution of Individual Behaviors and Social Institutions. *Journal of Theoretical Biology* 223:135-147.
- Boyd, R., Gintis, H., Bowles, S., and Richerson, P. J. (2003). The Evolution of Altruistic Punishment. *Proceedings of the National Academy of Science, USA* 100:3531-3535.
- Boyd, R., and Richerson, P. J. (1988). The Evolution of Reciprocity in Sizeable Groups. *Journal of Theoretical Biology* 132:337-356.
- . (1992). Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups. *Ethology and Sociobiology* 13:171-195.
- Breden, F. (1990). Partitioning of Covariance as a Method for Studying Kin Selection. *Trends in Ecology and Evolution* 5:224-228.
- Brembs, B. (1996). Chaos, Cheating and Cooperations: Potential Solutions to the Prisoner's Dilemma. *Oikos* 76:14-23.
- Brosnan, S. F., and de Waal, F. B. M. (2003). Monkeys Reject Unequal Pay. *Nature* 425:297 - 299.
- Bull, J. J. (1994). Virulence. *Evolution* 48:1423-1437.
- Camacho, J. P. M., Shaw, M. W., López-León, M. D., Pardo, M. C., and Cabrero, J. (1997). Population Dynamics of a Selfish B Chromosome Neutralized by the

- Standard Genome in the Grasshopper *Eyprepocnemis plorans*. *The American Naturalist*:1030-1050.
- Cartwright, N. (1978). Causal Laws and Effective Strategies. *Nous* 13:419-437.
- Casti, J. (1994). *Complexification*. New York: Harper-Collins.
- Cavalli-Sforza, L. L., and Feldman, M. W. (1978). Darwinian Selection and "Altruism". *Journal of Theoretical Biology* 14:263-281.
- Colwell, R. K. (1981). Group Selection Is Implicated in the Evolution of Female-Biased Sex Ratios. *Nature* 190:401-404.
- Cosmides, L. J., and Tooby, J. (1981). Cytoplasmic Inheritance and Intragenomic Conflict. *Journal of Theoretical Biology* 89:83-129.
- Dalai Lama XIV. (1994). *The Way to Freedom*. San Francisco: Harper.
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. London: J. Murray.
- . (1871). *The Descent of Man and Selection in Relation to Sex*. London: J. Murray.
- Davis, R. G. (1989). Organizational Inducements and Social Motives: A Game Theoretic Analysis. Ph.D., Portland State University.
- Dawkins, R. (1976). *The Selfish Gene*. New York: Oxford University Press.
- . (1982). *The Extended Phenotype: The Gene as the Unit of Selection*. Oxford: W.H. Freeman and Co.
- . (1994). Burying the Vehicle. *Behavioral and Brain Sciences* 17:661-617.
- . (1995). God's Utility Function. *Scientific American*:80-84.
- Doebeli, M., Blarer, A., and Ackermann, M. (1997). Population Dynamics, Demographic Stochasticity, and the Evolution of Cooperation. *Proceedings of the National Academy of Science, USA* 94:5167-5171.

- Doebeli, M., and Knowlton, N. (1998). The Evolution of Interspecific Mutualisms. *Proceedings of the National Academy of Science, USA* 95:8676-8680.
- Dugatkin, L. A. (1997). *Cooperation among Animals, an Evolutionary Perspective*. New York: Oxford University Press.
- Eberhard, W. G. (1990). Evolution in Bacterial Plasmids and Levels of Selection. *The Quarterly Review of Biology* 65:3-22.
- Ennis, H. L., Dao, D. N., Pukatzki, S. U., and Kessin, R. H. (2000). Dictyostelium Amoebae Lacking an F-Box Protein Form Spores Rather Than Stalk in Chimeras with Wild Type. *Proceedings of the National Academy of Science, USA* 97:3292-3297.
- Fehr, E., and Fischbacher, U. (2003). The Nature of Human Altruism. *Nature* 425:785-791.
- Fehr, E., and Gächter, S. (2002). Altruistic Punishment in Humans. *Nature* 415:137-140.
- Ferriere, R., and Michod, R. E. (1995). Invading Wave of Cooperation in a Spatial Iterated Prisoner's Dilemma. *Proceedings of the Royal Society London B* 259:77-83.
- . (1996). The Evolution of Cooperation in Spatially Heterogeneous Populations. *The American Naturalist* 147:692-717.
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. New York: Dover.
- Fletcher, J. A., Bedau, M. A., and Zwick, M. (1998). "Effect of Environmental Structure on Evolutionary Adaptation." *ALife VI, Cambridge, MA, 1998*, pp. 189-198.
- Fletcher, J. A., and Zwick, M. (2000a). "N-Player Prisoner's Dilemma in Multiple Groups: A Model of Multilevel Selection." *Artificial Life VII Workshop Proceedings, Portland, Oregon, 2000a*, pp. 86-89.
- . (2000b). "Simpson's Paradox Can Emerge from the N-Player Prisoner's Dilemma: Implications for the Evolution of Altruistic Behavior." *Proceedings*

of The World Congress of the Systems Sciences and ISSS 2000, Toronto, Canada, 2000b.

- . (2001). "Altruism, the Prisoner's Dilemma, and the Components of Selection." *IEEE Systems, Man, and Cybernetics Conference, Tucson, AZ, 2001*.
- . (2004a). The Evolution of Altruism: Game Theory, Multilevel Selection and Inclusive Fitness. *In Prep*.
- . (2004b). "Hamilton's Rule Applied to Reciprocal Altruism." *Congress on Evolutionary Computation, Portland, Oregon, 2004b*, pp. In Press.
- . (2004c). Strong Altruism Can Evolve in Randomly Formed Groups. *Journal of Theoretical Biology* 228:303-313.
- Fletcher, J. A., Zwick, M., and Bedau, M. A. (1996). Dependence of Adaptability on Environmental Structure in a Simple Evolutionary Model. *Adaptive Behavior (Special Issue on Environment Structure and Behavior)* 3/4:283-315.
- Frank, S. A. (1995a). George Price's Contributions to Evolutionary Genetics. *Journal of Theoretical Biology* 175:373-388.
- . (1995b). Mutual Policing and Repression of Competition in the Evolution of Cooperative Groups. *Nature* 377:520-522.
- . (1996). Models of Parasite Virulence. *Quarterly Review of Biology* 71:37-78.
- . (1997). Models of Symbiosis. *The American Naturalist* 150:S80-S99.
- . (1998). *Foundations of Social Evolution*. Princeton: Princeton University Press.
- . (2003). Perspective: Repression of Competition and the Evolution of Cooperation. *Evolution* 57:693-705.
- Frean, M. (1996). The Evolution of Degrees of Cooperation. *Journal of Theoretical Biology* 182:549-559.

- Gamboa, G. J., Reeve, H. K., and Holmes, W. G. (1991). Conceptual Issues and Methodology in Kin-Recognition Research: A Critical Discussion. *Ethology* 88:109-127.
- Giraud, T., Pedersen, J. S., and Keller, L. (2002). Evolution of Supercolonies: The Argentine Ants of Southern Europe. *Proceedings of the National Academy of Science, USA* 99:708 - 712.
- Goodnight, C. J. (1990a). Experimental Studies of Community Evolution I: The Response to Selection at the Community Level. *Evolution* 44:1614-1624.
- . (1990b). Experimental Studies of Community Evolution II: The Ecological Basis of the Response to Community Selection. *Evolution* 44:1625-1636.
- . (2000). Heritability at the Ecosystem Level. *Proceedings of the National Academy of Science, USA* 97:9365-9366.
- Goodnight, C. J., and Stevens, L. (1997). Experimental Studies of Group Selection: What Do They Tell Us About Group Selection in Nature? *The American Naturalist* 150:S59-S79.
- Gould, S. J. (1988). Kropotkin Was No Crackpot. *Natural History* 97:12-19.
- . (2002). *The Structure of Evolutionary Theory*. Cambridge: Belknap Press.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science* 293:2105-2108.
- Grey, D., Hutson, V., and Szathmáry, E. (1995). A Re-Examination of the Stochastic Corrector Model. *Proceedings of the Royal Society London B* 262:29-35.
- Hamburger, H. (1973). N-Person Prisoner's Dilemma. *Journal of Mathematical Sociology* 3:27-48.
- Hamilton, W. D. (1963). The Evolution of Altruistic Behavior. *The American Naturalist* 97:354-356.

- . (1964). The Genetical Evolution of Social Behavior I and II. *Journal of Theoretical Biology* 7:1-52.
- . (1967). Extraordinary Sex Ratios. *Science* 156:477-488.
- . (1970). Selfish and Spiteful Behavior in an Evolutionary Model. *Nature* 228:1218-1220.
- . (1971). "Selection of Selfish and Altruistic Behavior in Some Extreme Models," in *Man and Beast: Comparative Social Behavior*, pp. 59-91. Washington, D.C.: Smithsonian Institution Press.
- . (1972). Altruism and Related Phenomena, Mainly in Social Insects. *Annual Review of Ecology and Systematics* 3:193-232.
- . (1975). "Innate Social Aptitudes of Man: An Approach from Evolutionary Genetics," in *Biosocial Anthropology*. Edited by R. Fox, pp. 133-155. New York: John Wiley and Sons.
- . (1987). "Discriminating Nepotism: Expectable, Common, Overlooked," in *Kin Recognition in Animals*. Edited by D. J. C. Fletcher and C. D. Michener, pp. 417-437. New York: John Wiley and Sons.
- . (1996). *Narrow Roads of Gene Land*. Vol. 1 Evolution of Social Behaviour. Oxford: W. H. Freeman/Spektrum.
- . (2001). *Narrow Roads of Gene Land*. Vol. 2 The Evolution of Sex. Oxford: Oxford University Press.
- Hardin, G. (1968). The Tragedy of the Commons. *Science* 162:1243-1248.
- Hardin, R. (1971). Collective Action as an Agreeable N-Prisoner's Dilemma. *Behavioral Science* 16:472-481.
- Harms, W. (2000). Biological Altruism in Hostile Environments. *Complexity* 5:23-28.
- Hart, D. D., Kohler, S. L., and Carlton, R. G. (1991). Harvesting of Benthic Algae by Territorial Grazers: The Potential for Prudent Predation. *Oikos* 60:329-335.

- Hauert, C., De Monte, S., Hofbauer, J., and Sigmund, K. (2002). Volunteering as Red Queen Mechanism for Cooperators in Public Goods Games. *Science* 296:1129-1132.
- Hawkes, K. (1993). Why Hunter-Gatherers Work: An Ancient Version of the Problem of Public Goods. *Current Anthropology* 34:341-361.
- Hemptinne, J. L., and Dixon, A. G. (1997). Are Aphidophagous Ladybirds (Coccinellidae) Prudent Predators? *Biological Agriculture and Horticulture* 15:151-159.
- Herre, E. A. (1985). Sex Ratio Adjusted in Fig Wasps. *Science* 228:896-898.
- Hickey, D. A. (1992). Evolutionary Dynamics of Transposable Elements in Prokaryotes and Eukaryotes. *Genetica* 86:269-274.
- Hoogland, J. L. (1983). Nepotism and Alarm Calling in the Black-Tailed Prairie Dog, *Cynomys Ludovicianus*. *Animal Behavior* 31:472-479.
- Karlin, S., and Matessi, C. (1983). The Eleventh R. A. Fisher Memorial Lecture: Kin Selection and Altruism. *Proceedings of the Royal Society London B* 219:327-353.
- Katz, L. D. Editor. (2000). *Evolutionary Origins of Morality: Cross Disciplinary Perspectives*. Thorverton, UK: Imprint Academic.
- Kauffman, S. A. (1993). "Self-Organization and Adaptation in Complex Systems," in *The Origins of Order: Self-Organization and Selection in Evolution*, pp. 173-235. New York: Oxford University Press.
- Kropotkin, P. (1902). *Mutual Aid: A Factor in Evolution*, 1989 reprint edition. Montréal: Black Rose Books.
- Kuhn, T. (1970). *The Structure of Scientific Revolutions*, 2nd edition. Chicago: University of Chicago Press.
- Lansing, J. S. (1994). *The Balinese*: Harcourt Brace.

- Leigh, E. G. (1999). "Levels of Selection, Potential Conflicts, and Their Resolution: The Role of the "Common Good"," in *Levels of Selection in Evolution*. Edited by L. Keller, pp. 15-30. Princeton: Princeton University Press.
- Lenski, R. E., and May, R. M. (1994). The Evolution of Virulence in Parasites and Pathogens: Reconciliation between Two Competing Hypotheses. *Journal of Theoretical Biology* 169:253-265.
- Lewontin, R. C. (1970). The Units of Selection. *Annual Review of Ecology and Systematics* 1:1-18.
- . (1998). Survival of the Nicest? *The New York Review of Books*:59-63.
- . (2000). *The Triple Helix : Gene, Organism, and Environment*. Cambridge: Harvard University Press.
- Luce, R. D., and Raiffa, H. (1985). *Games and Decisions Introduction and Critical Survey*. New York: Dover Publications, Inc.
- Malthus, T. (1798). *An Essay on the Principle of Population*. London: J. Johnson.
- Margulis, L. (1970). *Origin of Eukaryotic Cells*. New Haven: Yale University Press.
- . (1993). *Symbiosis in Cell Evolution: Microbial Communities in the Archean and Proterozoic Eons*, 2nd edition. New York: W.H. Freeman and Co.
- Margulis, L., and Sagan, D. (2002). *Acquiring Genomes : A Theory of the Origins of Species*. New York, NY: Basic Books.
- Matapurkar, A. K., and Watve, M. G. (1997). Altruist Cheater Dynamics in Dictyostelium: Aggregated Distribution Gives Stable Oscillations. *The American Naturalist* 150:790-797.
- Maynard Smith, J. (1964). Groups Selection and Kin Selection. *Nature* 201:1145-1147.
- . (1972). *On Evolution*. Edinburgh: Edinburgh University Press.

- . (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
- . (1991). "A Darwinian View of Symbiosis," in *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis*. Edited by L. Margulis and R. Fester, pp. 26-39. Cambridge, MA: MIT Press.
- . (1998a). The Origin of Altruism. *Nature* 393:639-640.
- . (1998b). *Shaping Life*. New Haven: Yale University Press.
- Maynard Smith, J., and Price, G. R. (1973). The Logic of Animal Conflict. *Nature* 246:15-18.
- Maynard Smith, J., and Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford, England: W.H. Freeman and Co.
- . (1999). *The Origins of Life: From the Birth of Life to the Origin of Language*. Oxford: Oxford University Press.
- McMillan, J. (1979). The Free-Rider Problem: A Survey. *Economic Record* 55:95-107.
- Michod, R. E. (1996). Cooperation and Conflict in the Evolution of Individuality. II. Conflict Mediation. *Proceedings of the Royal Society B* 263:813-822. *Proceedings of the Royal Society London B* 263:813-822.
- . (1997). Evolution of the Individual. *The American Naturalist* 150:S5-S21.
- . (1999). "Individuality, Immortality, and Sex," in *Levels of Selection in Evolution*. Edited by L. Keller, pp. 35-74. Princeton: Princeton University Press.
- Michod, R. E., and Roze, D. (2000a). "A Multi-Level Selection Theory of Evolutionary Transitions in Individuality." *Artificial Life VII Workshop Proceedings, Portland, Oregon, 2000a*, pp. 82-85.
- . (2000b). Some Aspects of Reproductive Mode and the Origin of Multicellularity. *Selection* 1:97-109.

- . (2001). Cooperation and Conflict in the Evolution of Multicellularity. *Heredity* 86:1-7.
- Miralles, R., Moya, A., and Elena, S. F. (1997). Is Group Selection a Factor in Modulating the Virulence of RNA Viruses? *Genetical Research* 69:165-172.
- Mitteldorf, J., and Wilson, D. S. (2000a). Population Viscosity and the Evolution of Altruism. *Journal of Theoretical Biology* 204:481-496.
- . (2000b). "Population Viscosity and the Evolution of Altruism." *Artificial Life VII Workshop Proceedings, Portland, Oregon, 2000b*, pp. 96-99.
- Nowak, M. A., and Sigmund, K. (1998). Evolution of Indirect Reciprocity by Image Scoring. *Nature* 393:573-577.
- Nowell, P. C. (1976). The Clonal Evolution of Tumor Cell Populations. *Science* 194:23-28.
- Nunney, L. (1985a). Female-Biased Sex Ratios: Individual or Group Selection. *Evolution* 39:349-361.
- . (1985b). Group Selection, Altruism, and Structured-Deme Models. *The American Naturalist* 126:212-230.
- . (1998). Are We Selfish, Are We Nice, or Are We Nice Because We Are Selfish? *Science* 281:1619-1620.
- . (2000). Altruism, Benevolence and Culture: Commentary Discussion of Sober and Wilson's 'Unto Others'. *Journal of Consciousness Studies* 7:231-236.
- Nunney, L., and Miller Baker, A. E. (1993). The Role of Deme Size, Reproductive Patterns, and Dispersal in the Dynamics of T-Lethal Haplotypes. *Evolution* 47:1342-1359.
- Oldroyd, B. P., Smolenski, A. J., Cornuet, J. M., and Crozier, R. H. (1994). Anarchy in the Beehive. *Nature* 371:749.
- Oosterhout, G. (1996). An Evolutionary Simulation of the Tragedy of the Commons. Ph.D., Portland State University.

- Orr, D. (2004). "The Nature of Design." ISEPP Science, Technology and Society Lecture Series. Portland, Oregon, March 12, 2004.
- Packer, C., Gilbert, D. A., Pusey, A. E., and O'Brien, S. J. (1991). A Molecular Genetic Analysis of Kinship and Cooperation in African Lions. *Nature* 351:562-565.
- Packer, C., and Rutan, L. (1988). The Evolution of Cooperative Hunting. *The American Naturalist* 132:159-198.
- Panchanathan, K., and Boyd, R. (2003). A Tale of Two Defectors: The Importance of Standing for Evolution of Indirect Reciprocity. *Journal of Theoretical Biology* 224:115-126.
- Paulovich, A. G., Toczyski, D. P., and Hartwell, L. H. (1997). When Checkpoints Fail. *Cell* 88:315-321.
- Pepper, J. W. (2000a). "An Agent-Based Model of Group Selection." *Artificial Life VII Workshop Proceedings, Portland, Oregon, 2000a*, pp. 100-103.
- . (2000b). Relatedness in Trait Group Models of Social Evolution. *Journal of Theoretical Biology* 206:355-368.
- Pepper, J. W., and Smuts, B. B. (2000). "The Evolution of Cooperation in an Ecological Context: An Agent-Based Model," in *Dynamics in Human and Primate Societies: Agent-Based Modeling of Social and Spatial Processes*, pp. In Press. Oxford: Oxford University Press.
- . (2002). Assortment through Environmental Feedback. *The American Naturalist* 160:205-213.
- Price, G. R. (1970). Selection and Covariance. *Nature* 277:520-521.
- . (1972). Fisher's 'Fundamental Theorem' Made Clear. *Annals of Human Genetics* 36:129-140.
- . (1995). The Nature of Selection. *Journal of Theoretical Biology*:389-396.

- Queller, D. C. (1985). Kinship, Reciprocity and Synergism in the Evolution of Social Behavior. *Nature* 318:366-367.
- . (1992a). A General Model for Kin Selection. *Evolution* 46:376-380.
- . (1992b). Quantitative Genetics, Inclusive Fitness, and Group Selection. *The American Naturalist* 139:540-558.
- Rainey, P. B., and Rainey, K. (2003). Evolution of Cooperation and Conflict in Experimental Bacterial Populations. *Nature* 425:72-74.
- Rapoport, A. (1991). Ideological Commitments and Evolutionary Theory. *Journal of Social Issues* 47:83-99.
- Reeve, H. K. (1999). Multi-Level Selection and Human Cooperation. *Evolution and Human Behavior* 21:65-72.
- Riolo, R. L., Cohen, M. D., and Axelrod, R. (2001). Evolution of Cooperation without Reciprocity. *Nature* 414:441-443.
- Rissing, S. W., and Pollock, G. B. (1987). Queen Aggression, Pleometrotic Advantage and Brood Raiding in the Ant Veromessor Pergandei (Hymenoptera, Formicidae). *Animal Behavior* 35:975-981.
- Rissing, S. W., Pollock, G. B., Higgins, M. R., Hagen, R. H., and Smith, D. R. (1989). Foraging Specialization without Relatedness or Dominance among Co-Founding Ant Queens. *Nature* 338:420-422.
- Roberts, C. M. (1997). Ecological Advice for the Global Fisheries Crisis. *Trends in Ecology and Evolution* 12:35-38.
- Roze, D., and Michod, R. E. (2001). Mutation, Multilevel Selection, and the Evolution of Propagule Size During the Origin of Multicellularity. *The American Naturalist* 158:683-654.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The Neural Basis of Economic Decision-Making in the Ultimatum Game. *Science* 300:1755-1758.

- Sansom, R. (2000). A Definitive Case for Group Selection. *Complexity*:33-35.
- Scott, J. C. (1976). *The Moral Economy of the Peasant: Rebellion and Subsistence in Southeast Asia*. New Haven: Yale University Press.
- Sherman, P. W. (1977). Nepotism and the Evolution of Alarm Calling. *Science* 197:1246-1253.
- Simpson, E. H. (1951). The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society B* 13:238-241.
- Smith, N. G. C. (1998). The Dynamics of Maternal-Effect Selfish Genetic Elements. *Journal of Theoretical Biology* 191:173-180.
- Smuts, B. B. (1999). Multilevel Selection, Cooperation, and Altruism: Reflections on Unto Others: The Evolution and Psychology of Unselfish Behavior. *Human Nature* 10:311-327.
- Sober, E., and Lewontin, R. C. (1982). Artifact, Cause and Genic Selection. *Philosophy of Science* 49:157-180.
- Sober, E., and Wilson, D. S. (1998). *Unto Others, the Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.
- . (2000). Morality and 'Unto Others': Response to Commentary Discussions. *Journal of Consciousness Studies* 7:257-268.
- Strassmann, J. E. (2000). Bacterial Cheaters. *Nature* 404:555-556.
- Strassmann, J. E., Zhu, Y., and Queller, D. C. (2000). Altruism and Social Cheating in the Social Amoeba *Dictyostelium Discoideum*. *Nature* 408:965-967.
- Swenson, W., Wilson, D. S., and Elias, R. (2000). Artificial Ecosystem Selection. *Proceedings of the National Academy of Science, USA* 97:9110-9114.
- Szathmáry, E., and Maynard Smith, J. (1997). From Replicators to Reproducers: The First Major Transitions Leading to Life. *Journal of Theoretical Biology*:555-571.

- Tierney, J. (2000). "A Tale of Two Fisheries," in *New York Times Magazine*.
- Todes, D. P. (1987). Darwin's Malthusian Metaphor and Russian Evolutionary Thought, 1859-1917. *Isis* 78:537-551.
- Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *Quarterly Review of Biology* 46:35-57.
- Turner, B. C., and Perkins, D. D. (1991). Meiotic Drive in *Neurospora* and Other Fungi. *The American Naturalist* 137:416-429.
- Velicer, G. J., Kroos, L., and Lenski, R. E. (2000). Developmental Cheating in the Social Bacterium *Myxococcus Xanthus*. *Nature* 404:598-601.
- von Bertalanffy, L. (1968). *General Systems Theory: Foundations, Development, Applications*. New York: Braziller.
- von Neumann, J., and Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
- Wade, M. J. (1977a). An Experimental Study of Group Selection. *Evolution* 31:134-153.
- . (1977b). The Selfish Gene. *Evolution* 32:220-221.
- . (1978). A Critical Review of the Models of Group Selection. *The Quarterly Review of Biology* 53:101-114.
- . (1979). The Primary Characteristics of *Tribolium* Populations Group Selected for Increased and Decreased Population Size. *Evolution* 33:749-764.
- . (1980). Kin Selection: Its Components. *Science* 210:665-667.
- . (1985). Soft Selection, Hard Selection, Kin Selection, and Group Selection. *The American Naturalist* 125:61-73.
- . (2000). "Interactions in Metapopulations: Effects on Adaptation and Diversity." *Artificial Life VII Workshop Proceedings, Portland, Oregon, 2000*, pp. 79-81.

- Werren, J. H., and Beukeboom, L. W. (1993). Population Genetics of a Parasitic Chromosome: Theoretical Analysis of *Psr* in Subdivided Populations. *The American Naturalist* 142:224-241.
- Williams, G. C. (1966). *Adaptation and Natural Selection*. New Jersey: Princeton University Press.
- Wilson, D. S. (1975a). A Theory of Group Selection. *Proceedings of the National Academy of Science, USA* 72:143-146.
- . (1976). Evolution on the Level of Communities. *Science* 192.
- . (1977). Structured Demes and the Evolution of Group-Advantageous Traits. *The American Naturalist* 111:157-185.
- . (1979). Structured Demes and Trait-Group Variation. *The American Naturalist* 113:606-610.
- . (1987). Altruism in Mendelian Populations Derived from Sibling Groups: The Haystack Model Revisited. *Evolution* 41:1059-1070.
- . (1990). Weak Altruism, Strong Group Selection. *Oikos* 59:135-140.
- . (1997a). Altruism and Organism: Disentangling the Themes of Multilevel Selection Theory. *The American Naturalist* 150:S122-S134.
- . (1997b). Introduction: Multilevel Selection Theory Comes of Age. *The American Naturalist* 150:S1-S4.
- . (1998). Hunting, Sharing, and Multilevel Selection: The Tolerated-Theft Model Revisited. *Current Anthropology* 39:73-97.
- . (2000). Nonzero & Nonsense: Group Selection, Nonzerosumness, and the Human Gaia Hypothesis. *Skeptic* 8:84-89.
- . (2004). What Is Wrong with Absolute Individual Fitness? *Trends in Ecology and Evolution*:In Press.

- Wilson, D. S., and Colwell, R. K. (1981). Evolution of Sex Ratio in Structured Demes. *Evolution* 35:882-897.
- Wilson, D. S., Pollock, G. B., and Dugatkin, L. A. (1992). Can Altruism Evolve in Purely Viscous Populations. *Evolutionary Ecology* 6:331-341.
- Wilson, E. O. (1975b). *Sociobiology*. Cambridge, MA: Belknap Press.
- Wright, R. (2000). *Non-Zero, the Logic of Human Destiny*. New York: Pantheon Books.
- Wynne-Edwards, V. C. (1962). *Animal Dispersion in Relation to Social Behavior*. Edinburgh: Oliver and Boyd.
- Zwicky, M. (1978). "Some Analogies of Hierarchical Order in Biology and Linguistics," in *Applied General Systems Research: Recent Developments & Trends*. Edited by G. Klir, pp. 521-529. New York: Plenum Press.
- . (2004). "Elements and Relations: Aspects of a Scientific Metaphysics," Manuscript In Preparation.

Appendices

Appendix A: Hamilton's Rule Derivation

Here I derive Hamilton's rule starting with the basic NPD fitness functions and the condition that the fraction of cooperators in the population increases.

Starting with $Q' > Q$ we get:

$$\frac{A'}{N'} > \frac{A}{N}.$$

By the fitness equations 3.3 and 3.4:

$$\frac{A + \sum a_i q_i b}{N + \sum a_i q_i b + \sum s_i (q_i b + c)} > \frac{A}{N}.$$

Cross multiplying and isolating c/b on the right side yields:

$$\frac{N(\sum a_i q_i - QA)}{NQS} > \frac{c}{b}.$$

We cancel N and expand terms to give:

$$\frac{\sum a_i q_i - 2AQ - AQ}{A(1 - Q)} > \frac{c}{b},$$

and then:

$$\frac{\sum n_i q_i^2 - 2Q \sum n_i q_i - Q^2 \sum n_i}{A - 2AQ + AQ} > \frac{c}{b}.$$

This we rewrite as:

$$\frac{\sum n_i (q_i^2 - 2Qq_i + Q^2)}{A - 2AQ + NQ^2} > \frac{c}{b},$$

which gives:

$$\frac{\frac{\sum n_i (q_i - Q)^2}{N}}{\frac{A(1 - Q)^2 + S(0 - Q)^2}{N}} > \frac{c}{b},$$

or:

$$\frac{\text{var}_b(q_i)}{\text{var}_i(Q)} > \frac{c}{b}.$$

This is Hamilton's rule for a whole-group trait where r is the between-group over total variance:

$$rb > c.$$

Appendix B: Price Equation Derivations

Here I show how the Price Covariance equation can be interpreted in terms of an idealized Q' and how this relates to our NPD fitness functions.

Starting with the Price between-group term,

$$\Delta Q_b = \frac{\text{cov}(w_i, q)}{E(w_i)},$$

where w_i is the growth rate of a group, n_i'/n_i . By definition we get:

$$\Delta Q_b = \frac{E(w_i q_i)}{E(w_i)} - \frac{E(w_i)E(q_i)}{E(w_i)}.$$

This can be written as

$$\Delta Q_b = \frac{\frac{1}{N} \sum n_i' w_i q_i}{\frac{N'}{N}} - Q.$$

Canceling $1/N$ and using the definition of w_i we get:

$$\Delta Q_b = \frac{\sum n_i' q_i}{N'} - Q$$

which gives

$$\Delta Q_b = Q^* - Q$$

where

$$Q^* = \frac{\sum n_i' q_i}{N'}.$$

Now given $\Delta Q = Q' - Q$ and $\Delta Q = \Delta Q_b + \Delta Q_w$, it follows that $\Delta Q_w = Q' - Q^*$.

Starting with

$$Q^* = \frac{\sum n_i' q_i}{N'}$$

from above we expand n_i' to get

$$Q^* = \frac{\sum [q_i n_i (1 + w_{av} q_i)]}{N'}$$

or

$$Q^* = \frac{\sum [a_i (1 + w_{av} q_i)]}{N'} = \frac{A^*}{N'}$$

where A^* is the idealized A' given by the Price between-group component of selection.

That is, the Price equation's idealization about the new number of altruists in the population is given by adding the existing number in each group, a_i , to the amount each receives based on the average fitness line in the NPD model (Figure 3-1).

Appendix C: Random Group Analytic Model

Here I describe Hamilton's original model with recursion added to accommodate multiple generations within groups. If a_g , s_g , and n_g are respectively the number of altruists, non-altruists (selfish individuals), and total individuals in a group after g generations spent within groups, then:

$$a_g = a_{g-1} \left(1 + b \frac{a_{g-1} - 1}{n_{g-1} - 1} - c \right), \quad (\text{C.1})$$

$$s_g = s_{g-1} \left(1 + b \frac{a_{g-1}}{n_{g-1} - 1} \right), \text{ and} \quad (\text{C.2})$$

$$n_g = n_{g-1} \left(1 + (b - c) \frac{a_{g-1}}{n_{g-1}} \right). \quad (\text{C.3})$$

In Hamilton's model g is always one, but in our model I vary g by using these equations recursively—inputting the results from one generation into the calculations for the next. Note that when first formed all groups are size n , but after reproduction group sizes vary. (Terms without g subscripts indicate initial values, i.e. n is n_0 .)

The overall number of altruists, A_g , and individuals, N_g , in the population after g generation within groups is then the number contributed (after g generations) by groups of every possible original composition ($a = 0$ to n) times the number of such groups expected in a random binomial distribution. If Y is the total number of groups, then the expected count of groups with a initial altruists out of n group members is:

$$h(a) = Y \binom{n}{a} Q^a (1-Q)^{(n-a)}. \quad (\text{C.4})$$

The total population values after g generations spent within groups are then given by:

$$A_g = \sum_{i=0}^n h(i) a_g(i) \quad \text{and} \quad (\text{C.5})$$

$$N_g = \sum_{i=0}^n h(i) n_g(i), \quad (\text{C.6})$$

where $a_g(i)$ is the a_g value from Eq. C.1 given the particular starting a value specified by the index i and similarly for $n_g(i)$. Although Y is infinite, it cancels in the calculation of $Q_g = A_g / N_g$ and $\Delta_g Q = Q_g - Q$.

Appendix D: Random Group Analytic Model

Modifications

Here I describe modifications to the analytic model such that altruists do not give benefit to kin, homogeneous groups are disallowed, and the population size is kept constant.

No Benefit to Kin

The modified fitness functions for when altruists only give to non-relatives are implemented by substituting the size of altruist kin groups, k , for the minus-one term in Eqs. C.1 and C.2. The minus-one term subtracted the altruist from the number of its beneficiaries; here we subtract the altruist's kin (those having a common ancestor) as well. A preceding superscript k is used to designate fitness calculations that subtract k instead of one from a group's altruist count and group size:

$${}^k a_g = {}^k a_{g-1} \left(1 + b \frac{{}^k a_{g-1} - k_{g-1}}{n_{g-1} - k_{g-1}} - c \right), \quad (\text{D.1})$$

$${}^k s_g = {}^k s_{g-1} \left(1 + b \frac{{}^k a_{g-1}}{n_{g-1} - k_{g-1}} \right), \quad (\text{D.2})$$

where the size of a kin group of altruists in generation g is given by:

$$k_g = k_{g-1} \frac{a_g}{a_{g-1}}. \quad (\text{D.3})$$

This is the size of the kin group in the last generation times an altruist's clutch size for this generation. The initial k value $k_0 = 1$ (altruists are only related to themselves). Shifting benefit from kin to non-kin in this way does not affect the total group size and Eq. C.3 works for calculating n_g .

Note that in the unmodified model the average proportion of a group that is related to an altruist, k_g / n_g , can never be above $1 / n$ and therefore the proportion of an altruist's benefit that falls to kin $(k_g - 1) / (n_g - 1)$ is also bounded by $1 / n$. To see this note that k_g / n_g will be largest within homogeneous groups of altruists compared to mixed groups. In such groups (given our convention that $c = 1$) k_g is multiplied by b each generation and total group size also increases with b . Therefore the proportion k_g / n_g remains at its original value of $1 / n$. In all other groups this proportion falls with successive generations. Only when Q is high (so that homogeneous altruist groups are common) or when b is high (so homogeneous altruist groups grow proportionally bigger than other groups) is this limit approached.

No Homogeneous Groups

In an infinite population there will always be some homogeneous groups whenever $Q > 0$. For this modification, each time groups are randomly formed we eliminate homogeneous groups by simply converting all groups where the altruist number $a = n$ to groups where $a = n - 1$. In this way all homogeneous groups of altruists have one member switched to a non-altruist. Because non-altruists always increase faster than altruists, groups become more non-homogeneous with each successive generation.

Note that this modification causes Q to decrease, but this effect is small when Q is small.

No Population Growth

The global carrying capacity is implemented by scaling back all offspring numbers each generation by N_{g-1} / N_g where N_g is first calculated without scaling. We use a preceding asterisk to denote values calculated with scaling. For instance, the number of altruists in a group after g generations with scaling is,

$$^*a_g = a_g \frac{N_{g-1}}{N_g} \quad (\text{D.4})$$

and similarly for group size *n_g , where scaling is imposed at each recursion (generation). Whole population values with scaling *A_g and *N_g then sum over *a_g and *n_g instead of a_g and n_g respectively in Eqs. C.5 and C.6 and $^*Q_g = ^*A_g / ^*N_g$.

Appendix E: Random Groups Computer Simulation

Model

For each run of the model, individuals ($N = 1,000$) are initially randomly distributed into groups of size n using a random number generator to assign individuals to unfilled groups. The proportion of altruists and non-altruists is determined by the starting Q value. The sequential steps of the simulation are then:

1. In each group the new number of altruists and non-altruists (to the closest whole individual) are determined (using either the low or high stochasticity method described in the text).
2. Individuals are eliminated (using either the low or high stochasticity method described in the text) until the original population size N is reached.
3. If g generations have passed within groups since the last group reformation, all individuals are randomly assigned to new groups of size n ; otherwise if the migration rate, m , is greater than zero, mN individuals are chosen at random from the whole population and moved to new random locations in the population array which is ordered by groups—consequently larger groups are proportionately more likely to send out and receive migrants.

These steps are repeated until an equilibrium at $Q = 0.0$ or $Q = 1.0$ is reached. For runs where $n = 6$, N was 1,002 instead of 1,000 and initial $Q = 0.0998$ instead of 0.1000 to allow an even distribution into groups.

Appendix F: Derivation of G_P' for Conditional Strategies in TFT / ALLD Population

Here I derive a formula for the frequency of TFT players in the next generation (G_P') based on the fitness (utility) values gained in the previous generation. These values are calculated using the PD payoff values of Figure 5-2, the frequencies of each type of pairing, and the number of iterations played. Figure 5-2 is repeated here for reference.

actor's behavior	opponent's behavior	
	C ($P_O = 1$) contributes b	D ($P_O = 0$) contributes 0
C ($P_A = 1$) sacrifices c	$w_0 + b - c$ (+ d)	$w_0 - c$
D ($P_A = 0$) sacrifices 0	$w_0 + b$	w_0

The idea is to calculate a numerator that represents the total fitness (offspring) for TFT players. This is their fitness in TFT-TFT (TT) and TFT-ALLD (TD) pairings scaled by the frequency of each type of pairings. In the TD case the TFT plays C on the first move and gets the sucker's payoff, but for the next $(i - 1)$ moves TFT plays D as does its opponent. The denominator is the total of all payoffs for each pairing (TT , TD , DD scaled by f_{TT} , f_{TD} , and f_{DD} respectively) over i iterations.

$$G_P' = \frac{f_{TT} 2i(w_0 + b - c + d) + f_{TD}(w_0 - c + (i - 1)w_0)}{f_{TT} 2i(w_0 + b - c + d) + f_{TD}(w_0 - c + 2(i - 1)w_0 + w_0 + b) + f_{DD} 2iw_0}$$

This can be reduced to:

$$G_P' = \frac{f_{TT} 2i(w_0 + b - c + d) + f_{TD}(iw_0 - c)}{f_{TT} 2i(w_0 + b - c + d) + f_{TD}(2iw_0 + b - c) + f_{DD} 2iw_0}$$

which is Eq. 5.7.