

## 2

---

# The Evolution of Altruism in Humans

The Americans...are fond of explaining almost all the actions of their lives by the principle of self interest rightly understood; they show with complacency how an enlightened regard for themselves constantly prompts them to assist one another and inclines them willingly to sacrifice a portion of their time and property to the welfare of the state. In this respect I think they frequently fail to do themselves justice; in the United States as well as elsewhere people are sometimes seen to give way to those disinterested and spontaneous impulses that are natural to man; but the Americans seldom admit that they yield to emotions of this kind.

Alexis de Tocqueville, *Democracy in America*, 1830, (1945) volume 2, p. 130.

## 2.1 Introduction

Tocqueville's "Americans," a distinguished tradition in biology and the social sciences, has sought to explain cooperative behavior "by the principle of self interest, rightly understood." Richard Dawkins (1989) for instance, states, in the course of the first four pages of *The Selfish Gene*, "a predominant quality to be expected in a successful gene is ruthless selfishness. This gene selfishness will usually give rise to selfishness in individual behavior... Let us try to teach generosity and altruism, because we are born selfish." Similarly, drawing out the philosophical implications of the evolutionary analysis of human behavior, Richard Alexander (1987) writes, "ethics, morality, human conduct, and the human psyche are to be understood only if societies are seen as collections of individuals seeking their own self-interest." (p. 3). From J. B. S. Haldane's quip that he would risk his life to save eight drowning cousins (but not fewer) to the Folk Theorem of modern game theory (§A2), this tradition has clarified the ways that genetic relatedness, repeated play, reputation-building, and other aspects of social interactions among members of a group might confer fitness advantages and other benefits on those engaging in seemingly unselfish behaviors.

Our approach, however, favors Tocqueville, not Tocqueville's "Americans." Following Hamilton we use the term *helping* to describe behaviors that confer benefits on others, reserving the term *altruism* for helping in situations where the helper would benefit in fitness or other material ways

by withholding help (a more complete definition is given in Chapters 3 and 4, and in Appendix A4). Our models and simulations show that these altruistic helping behaviors may proliferate under conditions under which ancestral humans lived, due to the group structure of human populations and the success of groups in which cooperators are common. The unique role of culture in shaping human behavior is centrally involved in this explanation.

## 2.2 Genes, Cultures, Groups, and Institutions

We have treated culture—the ensemble of learned behaviors—as an evolutionary force in its own right rather than simply an effect of the interaction of genes and natural environments. Thus, we have rejected the view that predispositions that are transmitted culturally may constitute the proximate causes of behavior, but they in turn are entirely explained by the interaction of human nature and a people’s environment, so that, for example, the Lamalera whale hunters we discuss in chapter 3 would be said to share valued resources because they have social preferences, but they have social preferences because they live in a place where hunting whales is the best way to make a living, and those who hunt large game do better if they learn how to share.

The parsimony of the approach that treats culture as merely an effect of genes and environments without independent causal force, as well as its ability to discipline theory-building, has appealed to sociobiologists and cultural materialists. Like the principle of self-interest, the hypothesis that the interaction of natural environments and genes affects the evolution of cultures has yielded numerous insights.

But, it is also true that culture affects the natural and social environments in which the relative fitness of genetically transmitted behavioral traits is determined. Luca Cavalli-Sforza and Marcus Feldman (1981), Robert Boyd and Peter Richerson (1985), William Durham (1991), and Richerson and Boyd (2004) and others have provided compelling instances of these cultural effects on genetic evolution. It follows that human cognitive and affective capacities are the product of a dynamic known as *gene-culture coevolution*.

In the pages that follow we advance the view that this coevolutionary process has endowed us with preferences that go beyond the self-regarding concerns emphasized in traditional economic and biological theory, and embrace such other-regarding values as a taste for cooperation, fairness, and

retribution, the capacity to empathize, and the ability to value honesty, hard work, toleration of diversity, and loyalty to one's group.

The genome encodes information that is used both to construct a new organism, to instruct the new organism how to transform sensory inputs into decision outputs (i.e., to endow the new organism with a specific preference structure), and to transmit this coded information virtually intact to the new organism. Since learning about one's environment is costly and error-prone, efficient information transmission will ensure that the genome encode all information relevant to aspects of the organism's environment that are constant, or that change only very slowly through time and space. By contrast, environmental conditions that vary across generations and/or in the course of the organism's life history can be dealt with by providing the organism with the capacity to learn from one's environment, and hence phenotypically adapt to specific conditions.

There is an intermediate case that is not efficiently handled by either genetic encoding or learning from one's environment *de novo* in each generation. When environmental conditions are positively but imperfectly correlated across generations, each generation acquires valuable information through learning that it cannot transmit genetically to the succeeding generation, because such information is not encoded in the germ line. In the context of such environments, there is a fitness benefit to the transmission of information concerning the current state of the environment through some non-genetic information channel. Such information is quite common (Jablonka and Lamb 1995), which is called *epigenetic* by biologists, achieves its highest and most flexible form in *cultural transmission* in humans and to a considerably lesser extent in other primates (Bonner 1984, Richerson and Boyd 1998). Cultural transmission takes the form of vertical (parents to children) horizontal (peer to peer), and oblique (non-parental elder to younger), as in Cavalli-Sforza and Feldman (1981), prestige (higher influencing lower status), as in Henrich and Gil-White (2001), popularity-related as in Newman, Barabasi and Watts (2006), and even random population-dynamic transmission, as in Shennan (1997) and Skibo and Bentley (2003).

The parallel between cultural and biological evolution goes back to Julian Huxley (1955), Karl Popper (1979), and William James (1880). The idea of treating culture as a form of epigenetic transmission was pioneered by Richard Dawkins, who coined the term "meme" in *The Selfish Gene* (1976) to represent an integral unit of information that could be transmitted phenotypically. There quickly followed several major contributions to

a biological approach to culture, all based on the notion that culture, like genes, could evolve through replication (intergenerational transmission), mutation, and selection (Lumsden and Wilson 1981, Cavalli-Sforza and Feldman 1982, Boyd and Richerson 1985).

Dawkins added a second fundamental mechanism of epigenetic information transmission in *The Extended Phenotype* (1982), noting that organisms can directly transmit environmental artifacts to the next generation, in the form of such constructs as beaver dams, bee hives, and even social structures (e.g., mating and hunting practices). The phenomenon of a species creating an important aspect of its environment and stably transmitting this environment across generations, known as *niche construction*, is a widespread form of epigenetic transmission (Odling-Smee, Laland and Feldman 2003). Moreover, niche construction gives rise to what might be called a *gene-environment coevolutionary process*, since a genetically induced environmental regularity becomes the basis for genetic selection, and genetic mutations that give rise to mutant niches will survive if they are fitness enhancing for their constructors.

Our own models of the co-evolution of genetically transmitted individual behaviors and culturally transmitted group-level institutions is another example of the same process. We will see (chapter 7) that the presence of a culturally transmitted convention—resource sharing—is essential to the evolution of a genetically transmitted altruistic trait governed by natural selection. And in chapter 9 we show that the possibility of acquiring advantageous behaviors by social learning could generate the conditions under which a genetically transmitted capacity to internalize norms could evolve. Human cultures, along with the institutional structures they support, are instances of niche construction; i.e., the creation of a particular environment in such a manner that the genetic evolution of the creators is affected thereby (Laland, Olding-Smee and Feldman 2000, Bowles 2000, Laland and Feldman 2004).

Our approach can best be described as gene-culture coevolution in group structured populations. The evolutionary dynamics accounting for human cooperation include the proliferation of new behaviors as the result of learning from one's own experiences and from others as well the transmission and selective replication of genetic and other information from parents to offspring. It also involves selection processes operating at the group level as well as among individuals. Though inspired by biological approaches, especially Cavalli-Sforza and Feldman (1981), Boyd and Richerson (1985),

and Durham (1991), our models do not privilege biological explanation, and are designed to capture the distinctive aspects of human evolution. Our approach may be summarized as follows.

First, while genetic transmission of information plays a central role in our account, the genetics of social behavior is for the most part unknown. Knowledge of the genetic basis of the human cognitive and linguistic capacities that make cooperation on a human scale possible has expanded greatly in recent years, but nothing is known about genes that may be expressed in cooperative behavior, should these exist. No “gene for cooperation” has been discovered. Nor is it likely that one will ever be found, for the idea of a one-to-one mapping between genes and behavior is unlikely given what is now known about gene expression, and is implausible in light of the complexity and cultural variation of cooperative behaviors. Thus, when we introduce genetic transmission in our models (as we do in Chapters 6 to 8), our reasoning operates at the phenotypic level. The ‘A allele’ that accounts for altruistic behavior in Chapter 7 is just a phenotypic character that is transmitted exclusively from parent to child, thus abstracting from diploid reproduction, complex gene interactions, the vagaries of development and other aspects of real human genetic transmission, development and phenotypic expression. Similarly, the four strategies studied in Chapter 6 are just bi-parentally inherited haploid genotypes. In Chapter 9, where we study the evolution of the human capacity to internalize norms, the ‘internalization allele’ is a behavior acquired from parents.

The phenotype-based approach used here is a standard tool for the study of the evolution of social behavior in humans and other animals, and has a cogent justification as a device for abstracting from generally inconsequential complications surrounding the mechanics of genetic inheritance (Grafen 91, Eshel-Feldman 84, Hammerstein 96, Eshel-Feldman-Bergman 98, Frank 98). Moreover, because it uses observable phenotypes rather than unknown genotypes and developmental processes as the basis for analysis, the approach has a firm grasp on reality.

Second, as is conventional in all models of selection, relative payoffs, whether in terms of fitness, material reward, social standing or some other metric, influence the evolution of the population shares of various behavioral types. The resulting dynamics are often clarified using ‘as if’ optimization algorithms, though in doing this we do not attribute conscious optimization to individuals. Nor do we conclude that the resulting outcomes are in any sense optimal. In general they are not. For instance, in the famous *pris-*

*oners' dilemma* game (§A2), the only payoffs compatible with individual optimization are strictly suboptimal for both players. Thus, the aggregation of individually optimal choices is universally suboptimal, except under highly unrealistic conditions.

Individuals with higher payoffs tend to produce more copies of their behaviors in subsequent periods either through the contribution of their greater resources to differential reproductive success or because people disproportionately adopt the behaviors of the more successful members of their group. The latter may occur voluntarily, as when young people copy stars, or coercively as when dominant races, classes, or nations impose their cultures on subjugated peoples.

Third, because positive feedbacks are common in the processes of behavioral and institutional change we study, otherwise identical populations may typically exhibit quite different trajectories, reflecting the multiplicity of equilibria that is typical of models with positive feedbacks. The equilibrium selected need not be that with the higher average payoff. The process of selection among equilibria may be on such a long time scale that two populations described by exactly the same model may exhibit dramatically different distributions of behaviors for thousands of generations. The process of equilibrium selection thus assumes major importance.

Fourth, the emergence, proliferation and extinction of higher-level cultural and biological collections of individuals, such as foraging bands, ethnolinguistic units, and nations, and the consequent evolutionary success and failure of distinct group-level institutions such as systems of property rights, marital practices, and socialization of the young, is an essential, sometimes the preeminent, influence on human evolutionary processes. The maintenance of group boundaries (through hostility towards 'outsiders' for example) and lethal conflict among groups are essential aspects of this multi-level selection process. Within-group non-random pairing of individuals for mating, learning and other activities also plays an important part.

Fifth, chance, in the form of mutation, recombination, developmental accidents, behavioral experimentation, deliberate deviance from social rules, perturbation of the structure of social interactions and its payoffs and other stochastic influences, play an important role. It is often the case that a chance innovation is selectively neutral, not affecting payoffs given the conditions under which it is introduced; in this case it may succeed by chance.

Finally, we measure the empirical plausibility of alternative explanations against the conditions under which early humans lived during the Pleis-

tocene, roughly 1.6 million years before the present until the advent of agriculture beginning about 10,000 years ago, and especially the last 50 or so millennia of this period. We will consider the relevant archeological, climatic, genetic, ethnographic and historical evidence in detail in Chapters 6 to 8 and 10. Here is Christopher Boehm's (2007) summary, based on the common characteristics of the 154 foraging societies (about half of those in the ethnographic record) thought to approximate ancestral "highly mobile...storage-free economic systems":

These highly cooperative nomadic multi-family bands typically contain some unrelated families, and band size, while seasonably variable, seems to be around 20-30 people with families often moving from one band to the other. Band social life is politically egalitarian in that there is always a low tolerance by a group's mature males for one of their number dominating, bossing, or denigrating the others...economic life also tends to be quite egalitarian because of nomadism and a strong sharing ethic which dampens selfish and nepotistic tendencies....regional social networks exist...[and] socially or militarily facilitated group defense of resources is far from infrequent...fueled by ethnocentric tendencies....Drastic resource unpredictability, another likely factor [contributing to group conflict] could have been especially important in the changeable Pleistocene.

Of course, models of the emergence, proliferation and persistence of modern human behaviors must apply to the whole sweep of human history and prehistory as well, including the past 10,000 years.

Having outlined our approach, we now provide an overview of our findings.

### 2.3 Social Preferences

When one is motivated to bear personal costs to help or to hurt others we say that one has *other-regarding preferences*, meaning that affecting the states experienced by someone other than oneself is part of one's motivations. Unlike the conventional self-regarding preferences of *Homo economicus*, social preferences are other-regarding. Generosity towards others, and punishing those who violate norms are commonly motivated by other-regarding preferences, as is hostility to "outsiders." [Sam] In this book, we often use

the term ‘self-regarding’ rather than ‘selfish’ or ‘self-interested’ to describe the conventional assumptions about preferences to avoid the circularity arising from the fact that all uncoerced actions are motivated by preferences and hence might confusingly be termed selfish, leaving only those actions that violate one’s preference ordering to be called unselfish (but would better be called non-rational). Moral behavior—that which follows by ethical commitments—may be motivated by a concern for others but it need not be. We give examples of other-regarding and moral preferences, as well as experimental evidence for their importance, in Chapter 3.

When a person’s evaluation of states among which he or she may choose is either other-regarding or moral regarding, we say that the individual has *social preferences*. An example exhibiting both other-regarding and moral preferences is the suite of behaviors that we term *strong reciprocity*. Strong reciprocators have a predisposition to cooperate in situations where this is beneficial to others, and they respond to others’ cooperative behavior by continuing or enhancing their level of cooperation, while responding to lack of cooperation by others—and to violations of ethical norms more generally—by punishing the offenders, even at a material cost to themselves, and even when they cannot expect future personal gain from such behavior.

Experimental and other evidence that strong reciprocity and other social preferences are common in most cultures for which we have evidence. The same evidence shows that the fraction of most populations motivated solely by self-regarding preferences is quite modest.

## 2.4 Mutualistic Cooperation

Do social preferences play an important role in the explanation of human cooperation? The answer hinges on whether most forms of cooperation are altruistic or mutualistic. Because mutualistic cooperation will be sustained by individuals with entirely self-regarding preferences, it is readily explained in standard biological and economic models as an expression of self-interest. “Natural selection favors these...behaviors,” wrote Robert Trivers in his “The Evolution of Reciprocal Altruism” (1971), “because in the long run they benefit the organism performing them...two individuals who risk their lives to save each other will be selected over those who face drowning on their own.” (pp. 34–35) Trivers’ explanation initially found favor among biologists and economists because it is consistent with both the common biological reasoning that natural selection will not favor altru-

istic behaviors and with the canonical economic assumption of self-interest. Michael Ghiselin (1974) reflected conventional views in both fields when he concluded: “What passes for cooperation turns out to be a mixture of opportunism and exploitation.” (p. 2)

Trivers identified the conditions under which assisting another would be reciprocated in the future with a likelihood sufficient to make mutual assistance a form of mutualism. These conditions favoring reciprocal altruism included an extended lifetime, mutual dependence and other reasons for limited dispersal so that groups remain together, extended periods of parental care, attenuated dominance hierarchies, and frequent combat with conspecifics and predators. Foraging bands of humans, he pointed out, exhibit all of these conditions. Dawkins (2006) agrees: “In ancestral times we had the opportunity to be altruistic only towards close kin and potential reciprocators.” (p. 221) Michael Taylor (1976) and Robert Axelrod and William Hamilton (1981) subsequently formalized Trivers’ argument using the theory of repeated games (§A2). In economics, analogous reasoning is summarized in the Folk Theorem, which shows that cooperation among self-regarding individuals can be sustained as long as interactions are expected to be repeated with sufficient frequency and individuals are not too impatient (Fudenberg and Maskin 1986).

In subsequent chapters we will give two reasons for doubting the adequacy of this explanation of human cooperation. First is the experimental and other evidence that many (perhaps most) people act in ways inconsistent with the assumption of self-regarding preferences. There are many civic-minded acts that cannot be explained by self-interest, including why people vote, why they give anonymously to charity, and why they sacrifice themselves in battle. Second, in many important human social environments, Trivers’ conditions favoring reciprocal altruism do not hold, yet cooperation among non-kin is commonly observed. These include contributing to common projects when community survival is threatened, and cooperation among very large numbers of people among who do not share common knowledge of one another’s actions. We demonstrate this by showing that the scope of application of the Folk Theorem and similar models are quite restricted, especially in groups of any significant size, once the problem of cooperation is posed in an evolutionary setting and account is taken of “noise” arising from mistaken behaviors and misinformation about the behaviors of others.

## 2.5 The Evolution of Strong Reciprocity

If preferences were entirely self-regarding, the extent of human cooperation would indeed be puzzling. But, if social preferences provide a common proximate explanation of cooperation, the puzzle takes a somewhat different form: how might strong reciprocity and the other altruistic preferences that support cooperation have evolved over the course of human history, given the tendency of both genetic and cultural evolution to favor behavioral traits that on the average are associated with higher levels of material success?

Though as we have seen (section 2.2) there is much more to evolutionary dynamics than payoff-based replication, influence of payoffs, whether in the form of material success or fitness, is rarely absent in evolutionary processes. As a result, the success of altruistic preferences is indeed a puzzle. In chapter 6 therefore we use an evolutionary model driven by relative payoffs to show that individuals behaving as strong reciprocators could proliferate in a population in which they were initially rare, and that their presence in a population could sustain high levels of cooperation among group members.

The intuition behind our model and simulations is the following. In groups with strong reciprocators present, group members whose self-regarding preferences lead them to shirk on contributing to common projects will be punished. Strong reciprocators bear the cost not only of contributing to common projects, but also of punishing the shirking of the self-interested members. If reciprocators are common enough, however, the self-interested members will conform to cooperative norms in order to escape punishment, thereby reducing or even eliminating the fitness differences between the reciprocators and the self-interested members. But this is not enough: human ingenuity or mutation-induced behavioral variation is sure to hit upon the *Cooperator*, those who follows the rule of contributing to common projects so as to avoid punishment, but refrain from punishing. The Cooperators would obviously do better than their fellow group members who were reciprocators as long as some self-interested types shirked some of the time, for the only difference in their payoffs would be the cost of punishing these miscreants, borne by the reciprocators but not the Cooperators.

[Sam] Our model and simulations of populations who are either Selfish or Cooperators, and are either Punishers or Nonpunishers of shirkers. Thus, there are four types of agents in this model, Cooperator-Punishers, Cooperator-Nonpunishers, Selfish Punishers, and Selfish Nonpunishers. We explore the dynamics of a large population composed of many groups. De-

spite the presence of the Cooperators and Nonpunishers, Punishers do well and attain substantial fractions in the population, leading to a low level of shirking in the long run. The reason revealed by our simulations is that within any group, Nonpunishers do indeed out-compete Punishers, but groups with many Punishers have higher mean fitness than groups with few Punishers, so the cross-group benefits of being a Punisher outweigh the within-group fitness cost of punishing.

An attractive property of this model is that it predicts a heterogeneous population with a considerable fraction of both self-regarding and strong reciprocator types, as is often found in the experimental literature (Chapter 3, Fehr and Gächter 2002).

## **2.6 Multi-level Selection**

Our explanation of why social preferences are common thus hinges on three facts. First, group living is essential to human survival and groups differ in their evolutionary success, some producing large numbers of copies or contributing large numbers of individuals to membership in other groups, while other groups are absorbed into more successful groups or pass out of existence in warfare or during environmental crisis. Second, differential group success plays a central role in the evolution of human behaviors and institutions, less successful groups copying the more successful or being eliminated by them. Well-documented examples of this process include the peopling of much of the world by individuals of European ancestry and the associated spread of European customs and institutions in the past half millennium, and the spread of agriculture from the Middle East to Europe ten millennia ago. Both facts suggest that the structured nature of groups and their interaction strongly affect fitness at the individual and gene levels, a possibility that has long been recognized by biologists (Lewontin 1965, 1970, Wilson 1977, Durham 1991, Dunbar 1993, Laland and Feldman 2004). . Third, cooperative groups tend to prevail in intergroup competition, and a persistence of members with altruistic social preferences supports high levels of cooperation.

But, until recently, most of the formal modeling of evolutionary processes has concluded that group-level effects that would favor the spread of genes contributing to altruistic behavior cannot offset the effects of individual within-group selection operating against altruists, except where special circumstances heighten and sustain genetic differences between groups rel-

ative to differences within the group (Williams 1966, Crow and Kimura 1970, Boorman and Levitt 1973, Maynard Smith 1976). The reason is that the speed of an evolutionary process is proportional to the differences on which it works, so in order for between-group selection to outrun within-group selection, between-group differences must be substantial. But, gene flow due to group exogamy and other sources of migration are thought to preclude this.

Beginning with Darwin, however, a number of evolutionary theorists have suggested that human evolution might provide an exception to this negative assessment of the force of multi-level selection. William Hamilton (1975):331 summarized Darwin's view as follows: "He saw that such traits [as]...courage and self-sacrifice...would naturally be counter-selected within a social group, whereas in competition between groups the groups with the most of such qualities would be the ones best fitted to survive and increase." In *The Causes of Evolution*, J. B. S. Haldane (1932) provided a plausible mechanism for how this might come about. He suggested that in population of small endogamous 'tribes,' an altruistic trait might evolve because the 'tribe splitting' that occurs when successful groups reach a certain size would by chance create a few successor groups with a very high frequency of altruists, reducing within-group variance and increasing between-group variance, a process very similar to that modeled and simulated in Chapters 7 and 8. The small size of typical human groups during most of our evolution thus could play a crucial role in the chance occurrence of one or more groups with a high fraction of altruists, which would then proliferate. Haldane concluded: "evolution in large random-mating populations...is not representative of evolution in general, and perhaps gives a false impression of the events occurring in less numerous species....Our ancestors were mostly rather rare creatures." (p. 213-14) William Hamilton (1975) took up Haldane's suggestion, adding that if the allocation of members to successor groups following 'tribe splitting' was not random but was rather what he called 'associative,' (p. 137) multi-level selection pressures would be further enhanced.

More recent research also suggests that impediments to multi-level selection may be less general than was once thought (Uyenoyama and Feldman 1980, Harpending and Rogers 1987). A number of writers have pointed out that multi-level selection may be of considerably greater importance among humans than among other animals given the advanced level of human cognitive and linguistic capabilities and consequent capacity to maintain group

boundaries and to formulate general rules of behavior for large groups, and the resulting substantial influence of cultural inheritance on human behavior (Alexander 1987, Cavalli-Sforza and Feldman 1973, Boyd and Richerson 1985, 1990, Sober and Wilson 1994, Boehm 1997).

Among the consequences of these distinctive human capacities are the suppression of within-group phenotypic differences through egalitarianism, coinsurance, consensus decision making, conformist cultural transmission leading to large between-group differences, forms of social differentiation supporting high levels of assortative interactions both within and between groups, and the frequency of between-group conflict. Other animals do some of these things, but none does all of them on a human scale. All of these aspects of human social life enhance the force of group-level selection relative to individual level selection.

Our joint work with Boyd and Richerson (Boyd, Gintis, Bowles and Richerson 2003) shows, through agent-based simulations, that for some cooperative behaviors—notably punishing those who violate cooperative norms—multi-level selection on culturally transmitted traits can be decisive even for very large groups with substantial rates of intergroup migration. The reason for this surprising result is that if most members of a group are adhering to the norm, the costs incurred by those who are predisposed to punish violators are very small for the simple reason that violations are infrequent. Thus while within-group selection against the cooperative behavior exists, it is very weak in the neighborhood of the cooperative equilibrium. This supports the persistence over long periods of substantial between-group differences in group composition, some in which virtually all individuals are predisposed to cooperate and to punish those who do not, and other groups composed of virtually all self-regarding individuals. Additional between group variance is provided by inter-group conflicts following which the winning groups absorb the losers and then divide. Even a random group division process produces group difference by sampling errors, especially if group size is limited, exactly as Haldane had anticipated.

In Chapter 7 we explore whether similar processes promoting a genetic predisposition to behave altruistically could have been at work under conditions likely to have obtained during the Late Pleistocene. This is essentially an empirical question, requiring estimates of the frequency and consequences of conflict and environmental crises among human ancestral groups and the degree of genetic differences between groups. Using historical, ethnographic, paleoclimatic and archeological data to assess the former

and genetic material collected from recent foraging populations to investigate the latter we came to a surprising conclusion. It appears quite likely that our ancestors lived under exactly those extraordinary conditions that would allow a genetically transmitted altruistic behavior to proliferate by means of group selection.

## **2.7 The Coevolution of Institutions and Behaviors**

Because multi-level selection is part of the explanation of the evolutionary success of cooperative individual behaviors, it is likely that group-level characteristics—such as the institutions that suppress within group competition, as well as relatively small group size, limited migration, or frequent inter-group conflicts—that enhance multi-level selection pressures co-evolved with cooperative behaviors. Thus group-level characteristics and individual behaviors may have synergistic effects. This being the case, cooperation is based in part on the distinctive capacities of humans to construct institutional environments that limit within-group competition and reduce phenotypic variation within groups, while rendering between-group competition both frequent and lethal. The result of these characteristically human institutions is to heighten the relative importance of between-group competition, and hence to allow individually-costly but ingroup-beneficial behaviors to coevolve with these supporting environments through a process of multi-level selection.

The idea that the suppression of within-group competition may be a strong influence on evolutionary dynamics has been widely recognized in eusocial insects and other species (Ratnieks 1988, Ratnieks and Visscher 1989, Frank 1995, Reeve and Keller 1997, Seeley 1997). Alexander (1979), Boehm (1982) and Eibl-Eibesfeldt (1982) first applied this reasoning to human evolution, exploring the role of culturally transmitted practices that reduce behavioral variation within groups. Examples of such practices are leveling institutions, such as resource sharing among non-kin, namely those which reduce within-group differences in reproductive success or material well-being. These practices are leveling to the extent that they result in less pronounced within-group differences in material well-being or fitness than would have obtained in their absence.

By reducing within-group differences in individual success, such practices may have attenuated within-group genetic or cultural selection operating against individually-costly but group-beneficial practices. As a result,

the groups adopting them sustain higher frequencies of individuals willing to sacrifice their interests to benefit others. Such groups clearly enjoy advantages in intergroup contests. Hence, the evolutionary success of social institutions that reduce phenotypic variation within groups may be explained by the fact that they retard selection pressures working against in-group-beneficial individual traits and the fact that high frequencies of bearers of these traits reduces the likelihood of group extinctions.

Drawing on joint work with Jung-Kyoo Choi and Astrid Hopfensitz, we model an evolutionary dynamic along these lines. The novel feature of this approach is that genetically and culturally transmitted individual behaviors as well as culturally transmitted group-level institutional characteristics are subject to selection, with intergroup contests playing a decisive role in group-level selection. We show that intergroup conflicts may explain the evolutionary success of both. First, altruistic forms of human sociality towards non-kin. Second, group-level institutional structures such as resource sharing that have emerged and diffused repeatedly in a wide variety of ecologies during the course of human history. Ingroup-beneficial behaviors may evolve if they inflict sufficient costs on outgroup individuals when group conflicts occur, and if group-level institutions limit the individual costs of these behaviors and thereby attenuate within-group selection against these behaviors.

Our simulations show that if group-level institutions implementing resource sharing or within-group non-random pairing among group members are permitted to evolve, group-beneficial individual traits co-evolve along with these institutions, even where the latter impose significant costs on the groups adopting them. The simulations also show that cooperative individual behaviors and within-group variance-reducing social institutions could proliferate when they are initially rare in the population. In the absence of these group-level institutions, however, group-beneficial traits evolve only when intergroup conflicts are very frequent, groups are small, and migration rates are low. Thus the evolutionary success of cooperative behaviors in the relevant environments during the first 120,000 years of anatomically modern human existence may have been a consequence of distinctive human capacities in social institution-building.

## 2.8 War and Parochial Exclusion

No social practice is more emblematic of humans than warfare. Chimpanzees and fire ants engage in lethal group conflicts, but no animal matches our ancestors for the scale and frequency with which they killed members of neighboring groups. As we have seen, there is good reason to think that these conflicts were among the extraordinary conditions allowing the emergence and proliferation of altruism among humans. But, we are then faced with another puzzle. Those engaging in hostilities against their neighbors exhibit a special kind of altruism, one that links hostility toward outsiders with a willingness to confer benefits on group members. How could this suite of behaviors—parochial altruism—have evolved, given both the mortal risks involved and the fact that hostility towards outsiders would bear additional cost, such as forgoing a large mating pool and opportunities for exchange and coinsurance with outsiders? In Chapter 8 we take up this challenge, modeling and then simulating the evolution a population in which individuals may be either parochial or tolerant of outsiders and either altruistic or not in their behaviors towards fellow group members. Under parameter values likely to reflect late Pleistocene conditions, our hypothetical populations evolve to one of two well defined states: a prevalence of parochial altruists sustained by frequent lethal conflict or a prevalence of tolerant non-altruists who engage in mutually beneficial relations among groups. Thus, it seems likely that within-group cooperation and hostility towards “outsiders” co-evolved.

## 2.9 The Internalization of Norms

Social institutions promote multi-level selection in part by reducing within-group differences, so that even relatively small differences in between-group average fitness may play a predominant role in evolution by determining which groups survive and grow, and which do not. But institutions are not alone in this regard. Indeed, the human psyche itself fosters within-group cultural uniformity by internalizing social norms and actively punishing norm violators.

An *internalized norm* is a pattern of behavior that is desired its own sake, in addition to, or despite, any effect the behavior has on personal fitness or material well-being. Internalized norms are reinforced by the social emotions guilt and shame (see chapter 10), which induce conformity to basic values even when individuals are tempted to sacrifice normative goals on the

alter of current convenience. The ability to internalize norms is extremely widespread if not ubiquitous among humans. While widely studied in the sociology and social psychology, it has been virtually ignored outside these fields (but see Caporael et al., 1989 and Simon, 1990).

All successful cultures foster norms that enhance personal fitness, such as future-orientation, good personal hygiene, positive work habits, and control of emotions. Cultures also promote altruistic norms that subordinate the individual to group welfare, fostering such behaviors as bravery, honesty, fairness, willingness to cooperate, and empathy with the distress of others (Brown 1991).

Given that most cultures promote cooperative behaviors, and if we accept the sociological notion that individuals internalize the norms that are passed to them by parents and other influential elders, then we have a plausible proximate explanation of cooperation. If even only a fraction of society internalized the norms of cooperation and punish free riders and other norm violators, a high degree of cooperation might be maintained in the long run. But, this does not explain cooperation. The puzzles are two: why do we internalize norms, and why do cultures promote cooperative behaviors?

In Chapter 9, we provide an evolutionary model in which the capacity to internalize norms develops because this capacity enhances individual fitness in a world in which social behavior has become too complex and multifaceted to be fruitfully evaluated piecemeal through individual case-by-case assessment. Internalization transforms norms from *constraints* that one can treat instrumentally towards maximizing well-being, into *goals* that are then valued as ends rather than means. It is not difficult to show that if an internal norm is fitness enhancing, then for plausible patterns of socialization, an allele for internalization of norms, should one exist, would be evolutionarily stable.

We use this framework to model Herbert Simon's (1990) and Linnda Caporael's et al.'s (1989) explanation of altruism. They suggested that altruistic norms could 'hitchhike' on the general tendency of internal norms to be fitness-enhancing. However, they provided no formal model of this process and their promising ideas have been widely ignored. This chapter shows that their insight can be analytically modeled and is valid under plausible conditions. A straightforward multi-level selection argument then explains why individually fitness-reducing internalized norms are likely to be prosocial as opposed to socially harmful: groups with social internal norms will outcompete groups with antisocial, or socially neutral, internal norms.

## 2.10 Social Emotions

Among the consequences of the internalization of norms is the important role of the social emotions in sustaining cooperation. *Social emotions* are physiological and psychological reactions, including some, like shame, guilt, empathy, and sensitivity to social sanction, that induce individuals to undertake cooperative social interactions. By contrast, other social emotions, such as the desire to punish norm violators, reduce free-riding when the social emotions fail to induce sufficiently cooperative behavior in some fraction of members of the social group (Frank 1987, Hirshleifer 1987).

In Chapter 10 we consider a public goods game (§4.4,A2) where subjects maximize a utility function that captures five distinct motives: personal material payoffs, one's valuation of the payoffs to others which depends both on one's degree of reciprocity and one's sense of guilt or shame when failing to contribute one's fair share to the collective effort of the group. We have evidence of shame if players who are punished by others respond by behaving more cooperatively than is optimal for a material payoff-maximizing individual. We present indirect empirical evidence suggesting that such emotions play a role in the public goods game.

Using this utility function we show that reciprocity, shame, and guilt increase the level of cooperation in the group. Reciprocity motivates the punishment of those who do not contribute, and shame enhances the effectiveness of this punishment. The presence of reciprocity motives means that individual behaviors depend on their beliefs about the others' behaviors, so that groups may get locked into an outcome in which each individual's low level of cooperation sustains the ill will of the other members, who respond by contributing little.

Finally, we seek to explain why human behaviors are so often driven by emotions. Like cooperation itself, this is an evolutionary puzzle: why would an animal with extraordinary cognitive capacities develop ways of short-circuiting the brain? The most likely answer is threefold. First, emotions economize on costly cognitive capacities, providing a low-budget guide to behavior that is far from infallible, but on balance contributes to individual and group success. Second, the immediate gratification supplied by meeting emotional needs counterbalances the excessively short time horizon typical of human decision-makers, which by themselves would underweigh the long-term personal benefits of actions on behalf of others in the present. Finally, and partly for this reason, groups in which social emotions support-

ing cooperation are common will tend to proliferate their members and to survive environmental and military challenges.

## **2.11 Conclusion**

Our goal is to explain something that actually happened: the evolution of a uniquely cooperative species. There are now literally dozens of models of the evolution of cooperation, and we do not propose any novel causal mechanisms underlying the evolution of human behavior. Rather the models we introduce serve to clarify causal mechanisms already proposed, sometimes only verbally, and to identify the empirical conditions under which the relevant models might work. We then ask if these models, singly or in combination, can explain the distinctive aspects of human cooperation and how these might have evolved under the ecological and social conditions experienced by early humans.

But, considered literally as a history, our work is surely lacking, for while we have mined the archeological, genetic, linguistic, historical and ethnographic records with care, empirical knowledge of the conditions under which our ancestors lived tens of millennia ago is sparse, and inferences made from contemporary foragers subject to error. By default, we have produced hypothetical histories: self-consistent, dynamic trajectories that would account for the cooperative nature of humans, and that in light of what we do know about the distant past, might have occurred.

Our reasoning, while necessarily speculative, has been disciplined in three ways. First, the forms of cooperation we seek to explain and the causal processes supporting them are confirmed by historical, genetic, ethnographic and archeological experimental data. Second, our account is based on a plausible evolutionary dynamic involving gene- culture coevolution in group-structured populations, the consistency of which can be demonstrated through both verbal argument and mathematical modeling. Third, our agent-based simulations generate literally hundreds of thousands of artificial histories based on various plausible assumptions about human ecology and social interactions in the past. These show that our explanations can account for the evolution of cooperation under parameter values consistent with what can be reasonably inferred about the environments in which humans have evolved.

As the overviews of our detailed arguments presented above make clear, human cooperation, whether in the laboratory or in natural settings, takes

many forms, and these have not evolved by a single mechanism but by many. Because most of the human past approximates the experience of foraging bands more closely than other contemporary or recent societies we give special attention to cooperation among hunters and gatherers. In this respect, the assessment by Polly Wiessner (in a personal communication) based on her extensive field work in Southern Africa and New Guinea captures our understanding of the complexity and diversity of cooperation.

In all the societies where I have worked...kin selection, reciprocal altruism and strong reciprocity...operate in different realms. Kin selection can account for much generalized assistance such as everyday sharing, reciprocal altruism lies at the heart of many long term economic relations that appear altruistic, and strong reciprocity maintains norms that make group living possible. One could have a family group cooperating on the basis of kin selection and many individual partnerships...that are based on reciprocal altruism, but without strong reciprocity to create a common matrix of norms, it seems impossible to have large cooperative communities. Kin selection, reciprocal altruism and strong reciprocity account for different aspects of cooperation—together they make cooperative communities possible.

In the pages that follow we emphasize strong reciprocity and other forms of altruism as proximate motives for cooperation because of the central role that they play in sustaining cooperation, and the limited attention given to altruistic cooperation in the biological and social sciences to date. We begin by documenting the importance of these altruistic behaviors and the motives sustaining them.