Towards the Unity of the Human Behavioral Sciences

Herbert Gintis

Politics Philosophy Economics 2004 3: 37
DOI: 10.1177/1470594X04040188

politics, philosophy & economics article

London
Thousand Oaks, CA
and New Delhi

1470-594X
200402 3(1) 37–57

Towards the unity of the human behavioral sciences

Herbert Gintis
Santa Fe Institute, USA

abstract Despite their distinct objects of study, the human behavioral sciences all include models of individual human behavior. Unity in the behavioral sciences requires that there be a common underlying model of individual human behavior, specialized and enriched to meet the particular needs of each discipline. Such unity does not exist, and cannot be easily attained, since the various disciplines have incompatible models and disparate research methodologies. Yet recent theoretical and empirical developments have created the conditions for unity in the behavioral sciences, incorporating core principles from all fields, and based upon theoretical tools (game theory and the rational actor model) and data gathering techniques (experimental games in the laboratory and field) that transcend disciplinary boundaries. This article sketches a set of principles aimed at fostering such a unity.

keywords behavioral science, game theory, experimental economics, rational actor model

1. Introduction

The human behavioral sciences include economics, human biology, anthropology, sociology, behavioral psychology, and political science.[1] We may consider a set of disciplines as *unified* if they are *consistent*, so that in cases where two disciplines deal with the same phenomena, their models are equivalent and *synergic*, each discipline being substantively enriched by the scientific content of the others. The natural sciences achieved unity with the development of quantum mechanics, elementary particle and solid-state physics, and the big bang model of the universe. Such unity is lacking in the human behavioral sciences. Each behavioral discipline models *individual human behavior*, and constructs models

DOI: 10.1177/1470594X04040188
Herbert Gintis, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
[email: hgintis@comcast.net]

of aggregate social behavior compatible with, and often derived from, a model of individual behavior. Unity in the human behavioral sciences requires a common underlying model of individual behavior, which each discipline specializes and enriches for its particular purposes. No current model enjoys such trans-disciplinary status.

Yet, recent developments reveal links across the behavioral sciences sufficiently deep to establish the preconditions for unity. Both sociology[2] and political science,[3] following the pioneering contributions of Coleman, Downs, Olson, Buchanan and Tollison, and others,[4] have begun to adopt the rational actor model, previously espoused virtually exclusively in economics. Game theory, a central element of economic theory, was introduced to biology by Lewontin, Hamilton, and by Maynard Smith and Price,[5] subsequently maturing into an invaluable behavioral tool.[6] In anthropology, the application of experimental game theory to understanding cultural variation is rather new, but quite promising.[7] Conversely, increasing numbers of economists develop behavioral models of social interaction, and draw upon evidence from experimental game theory in modeling behavior. This development is evidenced by the Nobel prize in economics for the year 2002, awarded jointly to two experimentalists: psychologist Daniel Kahneman and economist Vernon Smith.

In this article, I will sketch a set of principles that express my current conception of unity. I will argue the following points. First, game theory provides a trans-disciplinary behavioral lexicon for communication and model-building. For many years, it was widely thought that game theory presupposes methodological individualism and a high level of cognitive functioning on the part of subjects. Were this the case, game theory would be inapplicable to settings where emotional, traditional, and heuristic behaviors are prominent, and where group-level processes and dynamic interactions are common. Contemporary evolutionary and behavioral game theory, however, extend classical game theory to cover such settings. Second, evolutionary biology underlies all behavioral disciplines because *Homo sapiens* is an evolved species whose major characteristics are the product of its particular evolutionary history. Third, evolutionary and behavioral game theory provides the substantive framework for the biology of human behavior. Fourth, the rational actor model, developed in economic theory, is a flexible tool that applies to all the human behavioral disciplines. This model treats actions as instrumental toward satisfying preferences. However, the content of preferences must be empirically determined, and agents may have preferences over actions as well as their outcomes. Moreover, the rational actor model is based on a notion of preference consistency that is not universally satisfied, so its range of applicability must also be empirically determined.

Fifth, progress in modeling human behavior has been hampered by the *underutilization of controlled experiments*, which are common only in behavioral psychology. Game theory and the rational actor model can be used as the basis for formulating, deploying, and analyzing data generated from controlled

experiments in social interaction. Such controlled experiments are replicable across laboratories and foster cumulative knowledge relevant to all behavioral disciplines. Sixth, progress in modeling human behavior has been hampered by the *artificially restricted range of social situations studied by behavioral scientists*. Only anthropology has systematically studied the effects of cultural differences *across societies* on human behavior; only sociology has systematically studied the effects of cultural differences *within societies* on human behavior; and only social psychology has systematically studied the effects of *personality differences* on social interaction. A unified model of human behavior is fostered by taking controlled experiments to the field, and deploying such experiments in a variety of cross-cultural settings across and within societies. Seventh, the demographic success of *Homo sapiens* is due to the ability of humans to sustain a high level of cooperation among non-kin. Whereas biology and economics explain this ability in terms of exchange among self-interested agents, the facts are in line with basic sociology and social psychology: humans often display altruistically prosocial behavior, especially in a form that I call '*strong reciprocity*' — a predisposition to cooperate and to punish non-cooperators at personal cost.[8] Lastly, prosocial behavior in humans can be modeled biologically using the tools of gene-culture coevolution, but the social mechanisms involved must include using the sociological notions of *socialization* and the *internalization of norms*.

2. The language of game theory

Communication across disciplines presupposes a common language. *Game theory* is a universal behavioral lexicon that offers such a common language. In the language of game theory, *players* (or *agents*) are endowed with a set of available *strategies*, and have a range of *information* concerning the rules of the game, the nature of the other players and their available strategies, as well as the structure of payoffs. Lastly, for each combination of strategy choices by the players, the game specifies a distribution of *individual payoffs* to the players. If the game is accurately specified, we can predict the behavior of the players by assuming that they attempt to maximize some preference function involving their personal payoffs, their chosen strategies, the personal payoffs to other agents, and the actions of the other agents.[9] Self-regarding agents maximize their personal payoffs, while other types of agents may care about fairness, the intentions of other agents, the sum of all payoffs, their relative personal payoff, and other aspects of the array of payoffs.

Developments within game theory in recent years have considerably enhanced its value to behavioral disciplines that have traditionally found little use for this analytical tool. First, it is now widely recognized that in many social interactions, agents are not self-regarding, but rather care about the payoffs to and intentions of other players.[10] Second, human actors care not only about material payoffs, but power, self-esteem, and behaving morally — goals that are recognized as central

to many behavioral disciplines.[11] Third, evolutionary and behavioral game theory do not require the extensive cognitive and information-processing capacities of classical game theory, so disciplines in which it is recognized that cognition is a scarce and costly good can make use of game-theoretical models.[12] Thus agents may consider only a restricted subset of strategies,[13] and they may use rule-of-thumb heuristics rather than maximization techniques.[14] Game theory is thus a generalized schema that permits the precise framing of meaningful empirical assertions, but imposes no particular structure on the predicted behavior.

3. The rational actor model

The rational actor model assumes that agents have *preferences* reflecting their wants and the tradeoffs among these wants, and that agents maximize their *utility* by choosing from an *action set* that is limited by available information, material resources and time, cognitive capacity, and the agent's physical capacities. Choice is also contingent upon *beliefs* concerning the probabilities of various states of nature, the frequency distribution of types of agents with whom they interact, and the relative effectiveness of different actions. The rational actor model is most highly developed in economics, but it applies to all the disciplines dealing with human behavior.

The rational actor model appears *prima facie* to apply only when extremely stringent conditions are satisfied. However, the model can be shown to apply over any domain in which the agent has *transitive preferences*, in the sense that if he prefers A to B and he prefers B to C, then he prefers A to C, and the agent can *make tradeoffs among outcomes* in the sense that for any finite set of outcomes $A_1, \ldots A_n$, if A_1 is the least preferred and A_n the most preferred outcome, then for any A_i, $1 \leq i \leq n$, there is a probability p_i, $0 \leq p_i \leq 1$ such that the agent is indifferent between A_i and a lottery that pays A_1 with probability p_i and pays A_n with probability $1 - p_i$.[15] Clearly, these assumptions are often extremely plausible. When applicable, the rational actor model's transitivity assumption strongly enhances explanatory power, even in areas that have traditionally abjured the model.[16]

The rational actor model has been underutilized in some behavioral disciplines through several prominent misunderstandings. First, the rational actor model does not require that agents be self-interested. There is no connection between the notion of the transitivity of preferences and the notion that preferences are purely self-regarding. Indeed, one can apply standard choice theory (including the derivation of demand curves, plotting concave indifference curves, and finding price elasticities) for such preferences as charitable giving and punitive retribution.[17] Second, because the rational actor model treats action as instrumental toward achieving rewards, it is often inferred that action itself cannot have reward value. This is an unwarranted inference. For instance, the rational actor model can be used to explain expressive motivation in rational action, including

collective action, that is precluded by the assumption that agents act instrumentally toward satisfying their material needs,[18] since agents may place positive value on the process of acquisition (for instance, 'fighting for one's rights'), and can value punishing those who refuse to join.[19] Third, the areas over which the transitivity postulate holds must be *empirically determined*.

Broadening the rational actor model beyond its traditional form in neoclassical economics runs the risk of developing unverifiable and *post hoc* theories, as our ability to theorize outpaces our ability to test theories. To avoid this, and following the lead of social psychology, we must expand the use of controlled experiments, as suggested above. Often, we find that the appropriate experimental design can generate new data to distinguish among models that are equally powerful in explaining the existing data.[20]

4. Game theory and biology

The analysis of living systems includes one analytical element that does not occur in the non-living world, and is not analytically represented in the natural sciences. This is the notion of a *replicator*, which is a physical system capable of drawing energy from its environment to make relatively accurate copies of itself.[21] The dynamics of replicators are described by the evolutionary notions of replication, mutation, selection, and adaptation.[22]

The most natural setting for replicator dynamics is game theoretical. Replicators endow copies of themselves with a repertoire of strategic responses to environmental conditions, including information concerning the conditions under which each is to be deployed in response to the character and density of competing replicators. Mutations included the replacement of strategies by modified strategies, and the 'survival of the fittest' dynamic (formally called a *replicator dynamic*) ensures that replicators with more successful strategies replace those with less successful ones.[23]

5. Gene-culture coevolution

Genetic replicators transmit information encoded in DNA sequences, through a germ line that is unaffected by environmental conditions. Genetic adaptation to new environments then takes the form of shifts in allele frequencies, and promotion of mutations that better exploit the new environment. In the context of rapidly changing environments, there is a fitness benefit to the transmission of *epigenetic* information concerning the current state of the environment. Such epigenetic information is quite common,[24] but achieves its highest and most flexible form in *cultural transmission* in humans and to a considerably lesser extent in other primates.[25]

There are several basic categories of culture: *conventions* (for example, language use), *techniques and practices* (for example, how to prepare food, how to

make and use tools, and how to treat illnesses), *ethical values* (for example, norms of fairness, reciprocity, and justice), and *transcendental beliefs* (for example, sickness is caused by angering the gods and good deeds are rewarded in the afterlife). A transcendental belief is the assertion of a causal relationship or a state of affairs that has a truth value, but whose truth believers either cannot or choose not to test. There are, of course, other types of beliefs, but these appear to be subsumable under other cultural categories. For instance, one may believe that a certain convention exists, a certain technique is effective, or a certain ethical value is justifiable. To avoid confusion, we treat such beliefs as part of the conventions, techniques and practices, and values that they affirm.

Conforming to *conventions* is adaptive because it is payoff-maximizing to conform when all others are doing so. When an agent must determine the most effective of several alternative *techniques* or *practices*, and if experimentation is costly, it may be payoff-maximizing to copy others rather than incur the costs of experimenting.[26] If everyone else experiments to find the superior technique, it will generally pay simply to follow the majority. By contrast, if everyone else conforms to a single technique in a situation where different techniques are best suited to different environments, then when the environment changes, an individual who experiments may do better than the conformists. Thus, in general, there will be a cultural equilibrium with a positive fraction of both conformists and experimenters. In this sense, the genetic machinery for a predisposition to conform to conventions and to imitate techniques is biologically adaptive.

It is plausible to extend this explanation to transcendental beliefs as well. Such beliefs affirm techniques where the cost of experimentation is extremely high or infinite, and the cost of making errors is high as well. This is, in effect, Blaise Pascal's argument for belief in God and the resolve to follow His precepts. It is supported by Boyer, who models religion as a set of cognitive beliefs that coexist and interact with our other more mundane and testable beliefs.[27] In this view, one conforms to transcendental beliefs because their truth value has been ascertained by others (relatives, ancestors, and prophets), and are as worthy of affirmation as the techniques and practices (such as norms of personal hygiene) that one accepts on faith, without personal verification.

Conventions, techniques, and beliefs are *instrumental* in the sense that they specify how best to achieve certain ends or goals. The remaining cultural category, *ethical norms and values*, is *final* in the sense of specifying what ends or goals to embrace. I discuss the place of cultural values in human behavioral theory below.

6. The puzzle of prosocial behavior in humans

The success of *Homo sapiens*, as measured by its broad geographical distribution and its considerable share of the Earth's biomass, is based on its unique capacity to use cultural forms to transmit technical knowledge accurately across genera-

tions, and its unique ability to sustain cooperation through space and across time among large numbers of unrelated individuals.[28] How do we explain this cooperation?

Biologists maintain that cooperation can be sustained by *inclusive fitness*, or cooperation among kin, and by individual self-interest in the form of *reciprocal altruism*.[29] Reciprocal altruism occurs when an agent helps another agent, at a fitness cost to herself or himself, with the expectation that the beneficiary will return the favor in a future period. The explanatory power of inclusive fitness theory and reciprocal altruism convinced a generation of biologists that what appears to be altruism (personal sacrifice on behalf of others) is really just long-run self-interest. Economics has developed a similar model of cooperation, based on the notion of long-term, enlightened self-interest, an idea that goes back to Bernard Mandeville's concept of 'private vices, public virtues' and Adam Smith's notion of the 'invisible hand'.[30]

Sociology, by contrast, has used *socialization* to explain cooperation among non-kin. According to Durkheim, the division of labor in society involves assigning individuals to specific *roles*.[31] Individuals are inculcated with *values* and *norms* that induce them to conform to the duties and obligations of the role positions they occupy. This is altruism.

A key tenet of socialization theory is that a society's values are passed from generation to generation through the *internalization of norms*.[32] In the language of optimization theory, internalized norms are accepted not as instruments toward and constraints upon achieving other ends, but rather as *arguments in the preference function that the individual maximizes*. Internalized norms are thus what we termed *ethical values* in our lexicon of cultural forms. In true gene-culture coevolutionary form, a variety of uniquely human prosocial emotions come into play, including prominently *shame*, *guilt*, and *empathy*, directly reinforcing internalized norms.

The programmability of the preference function appears in the form of the human *capacity to internalize norms*, which consists in an older generation instilling the values and objectives of a younger generation through an extended series of personal interactions, relying on a complex interplay of affect and authority. Agents conform to an internalized norm because doing so is an end in itself, and not merely because of the material rewards that follow from norm compliance or punishments that follow from norm violation. For instance, an individual who has internalized the value of 'speaking truthfully' will do so even in some cases where the net payoff to speaking truthfully would otherwise be negative. It follows that where people internalize a norm, the frequency of its occurrence in the population will be higher than if people follow the norm only instrumentally; that is, when they perceive it to be in their narrow material interest to do so. The capacity to internalize is based on a distinctively human psychological predisposition, unrecognized in biology and economics.

7. The internalization of norms

An 'altruistic norm', when acted upon, reduces the bearer's individual fitness or material well-being, but increases the fitness or well-being of other, unrelated, group members. The internalization of altruistic norms appears to be an evolutionary *curiosum* because agents who internalize such norms should be at a fitness disadvantage in comparison with self-interested actors. A closer look at the cultural transmission process, however, offers a resolution to this problem.[33]

Suppose there is an altruistic behavior, A, that imposes fitness cost s on those who embrace it. Suppose also that only a fraction of youth have the genetic capacity to accept ethical norms, and this fraction increases or decreases over time according to the biological fitness of its bearers. Suppose further that altruistic behavior A is transmitted to offspring with this genetic capacity by their parents in an unbiased manner (that is, if neither or both parents embrace A, all of their genetically enabled offspring do the same, and if only one parent embraces A, half of such offspring embrace A). In addition, suppose there is *extraparental transmission* of A, in the form of social pressure (rumor, shunning, and ostracism), rituals (dancing, prayer, marriage, birth, and death), and in modern societies, formalized institutions (schools, churches, and sacred texts). Such extraparental transmission is itself altruistic, since it will generally be individually costly, while the benefits, in the form of a higher frequency of altruism in the group, accrue to others. We handle this, plausibly, we believe, by assuming that the altruistic norm is both to embrace A and to encourage others to embrace it as well, and we include the cost of extraparental transmission in s, the cost of altruism. We measure the strength of extraparental transmission by the parameter γ, such that if the fraction of altruists in the older generation is p_A, then γp_A is the probability that a given non-altruistic child with the genetic capacity to acquire the altruistic norm, will in fact be induced to embrace the altruistic norm.

Suppose, further, that an altruist who meets a non-altruist, which we assume occurs with a probability proportional to the fraction of altruists, switches to the non-altruist's behavior with probability α. Gintis then shows that if α satisfies the inequality:

$$\alpha < \frac{\gamma - s}{1 - \gamma} \tag{1}$$

then the altruistic cultural equilibrium, in which all agents have the genetic capacity to embrace ethical norms, and all actually embrace A, is evolutionarily stable.[34] Note that (1) the larger the fitness cost s of altruism, and (2) the smaller the rate γ of oblique transmission, the lower the maximal rate of 'moral defection' α to the non-altruistic that is compatible with an altruistic cultural equilibrium. Note also that if γ is sufficiently large (specifically, if $\gamma > (1 + s)/2$), then no rate of defection can undermine the altruistic equilibrium, because agents rarely meet non-altruists with whom they can compare their fitness.

The substantive questions, then, are (1) why γ might be positive and large, and (2) why the rate of moral defection might be low. To address (1), note that the rate γ of extraparental transmission depends not only on the willingness of individuals to sacrifice on behalf of the group by engaging in extraparental socialization and by rewarding others who do the same, but also on the structure of social institutions that routinize cultural transmission. There is thus no guarantee that γ will be high, but societies that do effectively organize cultural transmission, and stress ethical norms that are heavily prosocial, will tend to grow and otherwise outcompete societies that do not (this process is referred to above as *weak* group-level selection).

To address (2), we must explain why agents might not defect at a very high rate to fitness-maximizing behavior. The following argument suggests that the psychological constitution of *Homo sapiens* is conducive to a high rate of adherence to moral norms, and hence to the satisfaction of Equation (1). While nearly everyone behaves amorally on some occasions, and some behave amorally much of the time, there is normally a sufficient reserve of moral behavior, including the motivation to punish the moral transgressions of others, to maintain a high level of conformity with group morality.

As we have noted, humans do not maximize fitness, but rather a preference function that is but a rough proxy for fitness under constant environmental conditions. The rapid pace of environmental change and cultural innovation over the past 100,000 years has produced a situation in which the set of needs, desires, drives, pleasures, and pains associated with the human preference function is out of line with the dictates of fitness maximization.[35] Even a random deviation of the human preference function from fitness maximization toward other goals, such as power, esteem, wealth, and pleasure, might be conducive to a relatively slow rate of rejection of moral norms, of which altruistic norms might figure prominently.

However, there is evidence of a more systematic force intervening between biological fitness and human preferences: as we have seen, the human preference function is, to some considerable extent, *programmable*, in the sense that human goals can be altered by socialization. The notion of a programmable preference function is sufficiently unusual that such a mechanism must have arisen as an adaptation, and hence the content of socialization, the actual internalized norms themselves, must be, at least on balance, fitness enhancing. Yet standard sociological theory has not supplied an argument as to why it might be adaptive, and, indeed, has generally ignored evolutionary arguments altogether. We can, however, supply such an argument.

A programmable preference function is the most complex instrument facilitating epigenetic information flows, all of which represent means of transferring information across generations in a manner complementary to, and often more flexible than, genetic transmission.[36] The form that this epigenetic transmission process takes in the case of the internalization of norms is a protracted series of

interactions, controlled by parents and influential elders, undertaken at considerable cost, and reinforced by a complex web of informal sanctions. While cultural learning occurs in many species, programmability of goals is virtually limited to humans because the capacity to be socialized presupposes a high level of cognitive capacity,[37] as well as specialized mental circuitry for valuing interpersonal relationships and making informed social judgments,[38] and specialized emotional capacities that enhance the individual's capacity to attain internalized goals, such as pride, shame, empathy, and remorse.[39] The genetic basis for prosocial emotions is clear from the fact that the inability to experience prosocial emotions, associated with sociopathic personality types, is partially heritable,[40] and is deficient in individuals with damage to specific regions of the brain's frontal lobes.[41]

The capacity to program changes in the preference function *culturally*, indeed, has great adaptive value. By redirecting human goals, and thereby curbing, repressing, and channeling an agent's basic impulses, the agent will have higher fitness than another agent who lacks this capacity. Included among the norms that are commonly internalized are thus norms of personal hygiene, concern for the approval of others, control of temptation, cultivation of a work ethic, and maintaining a long time horizon in decision-making. Such norms are upheld and transmitted in virtually all societies,[42] though a breakdown of cultural transmission in this area occurs in some poorly functioning societies.[43]

For an example of the fitness-enhancing capacity of internalization, note that a sophisticated weapon, such as a sharp knife, may aid an individual in taking revenge upon a transgressor, but the spontaneous impulse to attack an enemy may be fitness-reducing when such weapons are widely available. However, parents can instill in their offspring the norms of 'love thy neighbor' and 'be slow to anger'. Individuals who have acquired this genetic predisposition to internalize norms will pass both this capacity and its content (the conflict-limiting norm itself) to their offspring. For this reason, the internalization of norms may be fitness enhancing.

For a second example, suppose someone invents an aerodynamic spear that is extremely effective in the hunt, but requires daily practice to hone the throwing skills needed to use the spear effectively.[44] Since agents primordially prefer less expenditure of energy to more and have inappropriately short time horizons, they will skimp on daily practice. The hunter who internalizes the norm 'good hunters like to practice' will have an adaptive advantage.

One might object that a non-internalizer could always mimic the behavior of internalizers when it suits his purposes, and do better by violating the norm strategically when it is in his interest to do so. In fact, the non-internalizer *could*, but will not *want* to, emulate internalizers, in the sense that emulating their behavior simply does not maximize his preference function. To pursue the first example above, curbing one's violent tendencies may improve fitness, but the primordial preference function is not geared toward maximizing fitness, but rather toward a set of 'fitness proxies' that entail being violent under earlier

evolutionary, but not contemporary, circumstances. The non-internalizer will, of course, curb his violence for prudential reasons, but not because he in addition values peace, his neighbor's well-being or even his biological fitness. In the second example, the non-internalizer will prefer the larger portion of meat, and the greater prestige that follows from a rigorous practice routine, but, nevertheless, not enough to engage in such a routine.

Once genes for norm internalization are in place, there is nothing preventing altruistic norms from being culturally transmitted, internalized, and acted on in the same manner as personally fitness-enhancing norms. Altruism thus 'hitchhikes' on the personal fitness-enhancing capacity of norm internalization, and hence is an *exaption*, in the sense of Gould and Vrba.[45] It is for this reason that the rate of defection from altruistic norms might be sufficiently low that Equation (1) might hold, as long as fitness costs are not too high and there is some positive level of oblique transmission.[46]

It might be suggested that in a cultural equilibrium with internalized altruistic norms, a mutant family that teaches its children to internalize the personally fitness-enhancing norms, but not the altruistic ones would outcompete families that transmit both personally fitness-enhancing and altruistic norms. However, if part of the ethic of altruism is to punish selfish types, even selfish types will act altruistically, so under plausible conditions, the mutant may have no adaptive advantage. Moreover, using the above notation, if:

$$\alpha < \frac{\gamma - s}{1 + \gamma - s} \tag{2}$$

selfish internalizers are positively disadvantaged with respect to altruistic internalizers.[47]

8. Strong reciprocity and behavioral game theory

Recent experimental research supports the above synthesis.[48] We define *strong reciprocity* as a predisposition to cooperate with others, and to punish those who violate the norms of cooperation, at personal cost, even when it is implausible to expect that these costs will be repaid.

8.1. Evidence from the ultimatum game

In the ultimatum game, under conditions of anonymity, two players are shown a sum of money, say US$10. One of the players, called the 'proposer', is instructed to offer any number of dollars, from US$1 to US$10, to the second player, who is called the 'responder'. The proposer can make only one offer. The responder, again under conditions of anonymity, can either accept or reject this offer. If the responder accepts the offer, the money is shared accordingly. If the responder rejects the offer, both players receive nothing.

Since the game is played only once and the players do not know each other's identity, a self-interested responder will accept any positive amount of money. Knowing this, a self-interested proposer will offer the minimum possible amount, US$1, and this will be accepted. However, when actually played, *the self-interested outcome is never attained and never even approximated*. In fact, as many replications of this experiment have documented, under varying conditions and with varying amounts of money, proposers routinely offer respondents very substantial amounts (50 percent of the total generally being the modal offer), and respondents frequently reject offers below 30 percent.[49]

The ultimatum game has been played around the world, but mostly with university students. We find a great deal of individual variability. For instance, in all of the above experiments a significant fraction of subjects (about a quarter, typically) behave in a self-interested manner. But, among student subjects, average performance is strikingly uniform from country to country.

To expand the diversity of cultural and economic circumstances of experimental subjects, Joseph Henrich, Samuel Bowles, Robert Boyd, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath undertook a large cross-cultural study of behavior in various games, including the ultimatum game.[50] Some experienced field researchers, working in 12 countries on four continents, recruited subjects from 15 small-scale societies exhibiting a wide variety of economic and cultural conditions. These societies consisted of three foraging groups (the Hadza of East Africa, the Au and Gnau of Papua New Guinea, and the Lamalera of Indonesia), six slash-and-burn horticulturists (the Aché, Machiguenga, Quichua, and Achuar of South America, and the Tsimané and Orma of East Africa), four nomadic herding groups (the Turguud, Mongols, and Kazakhs of Central Asia, and the Sangu of East Africa), and two sedentary, small-scale agricultural societies (the Mapuche of South America and Zimbabwe farmers in Africa).

We can summarize our results as follows.

1. The canonical model of self-interested behavior is not supported in *any* society studied. In the ultimatum game, for example, in all societies either respondents or proposers, or both, behaved in a reciprocal manner.
2. There is considerably more behavioral variability across groups than had been found in previous cross-cultural research. While mean ultimatum game offers in experiments with student subjects are typically between 43 percent and 48 percent, the mean offers from proposers in our sample ranged from 26 percent to 58 percent. While modal ultimatum game offers are consistently 50 percent among university students, sample modes with these data ranged from 15 percent to 50 percent. In some groups, rejections were extremely rare, even in the presence of very low offers, while in others, rejection rates were substantial, including frequent rejections of *hyper-fair* offers (that is, offers above 50 percent). By contrast, the most common behavior for the Machiguenga was to

offer zero. The mean offer was 22 percent. The Aché and Tsimané distributions resemble American distributions, but with very low rejection rates. The Orma and Huinca (non-Mapuche Chileans living among the Mapuche) have modal offers near the center of the distribution, but show secondary peaks at full cooperation.

3. Differences among societies in 'market integration' and 'cooperation in production' explain a substantial portion of the behavioral variation between groups: the higher the degree of market integration and the higher the payoffs to cooperation, the greater the level of cooperation and sharing in experimental games. The societies were rank ordered in five categories: 'market integration' (how often do people buy and sell, or work for a wage?), 'cooperation in production' (is production collective or individual?), plus 'anonymity' (how prevalent are anonymous roles and transactions?), 'privacy' (how easily can people keep their activities secret?), and 'complexity' (how much centralized decision-making occurs above the level of the household?). Using statistical regression analysis, only the first two characteristics, market integration and cooperation in production, were significant, and they together accounted for 66 percent of the variation among societies in mean ultimatum game offers.
4. Individual-level economic and demographic variables did not explain behavior either within or across groups.
5. The nature and degree of cooperation and punishment in the experiments was generally consistent with the economic patterns of everyday life in these societies.

In a number of cases, the parallels between experimental game play and the structure of daily life were quite striking. Nor was this relationship lost on the subjects themselves. Here are some examples.

1. The Orma immediately recognized that the public goods game was similar to the *harambee*, a locally initiated contribution that households make when a community decides to construct a road or school. They dubbed the experiment 'the harambee game' and gave generously (mean 58 percent with 25 percent maximal contributors).
2. Among the Au and Gnau, many proposers offered more than half the pie, and many of these 'hyper-fair' offers were rejected. This reflects the Melanesian culture of status-seeking through gift giving. Making a large gift is a bid for social dominance in everyday life in these societies, and rejecting the gift is a rejection of being subordinate.
3. Among the whale-hunting Lamalera, 63 percent of the proposers in the ultimatum game divided the pie equally, and most of those who did not, offered more than 50 percent (the mean offer was 57 percent). In real life, a large catch, always the product of cooperation among many individual whalers, is meticulously divided into pre-designated parts and carefully distributed among the members of the community.

4. Among the Aché, 79 percent of proposers offered either 40 percent or 50 percent, and 16 percent offered more than 50 percent, with no rejected offers. In daily life, the Aché regularly share meat, which is distributed equally among all households, irrespective of which hunter made the kill.
5. The Hadza, unlike the Aché, made low offers and had high rejection rates in the ultimatum game. This reflects the tendency of these small-scale foragers to share meat, but with a high level of conflict and frequent attempts by hunters to hide their catch from the group.
6. Both the Machiguenga and Tsimané made low ultimatum game offers, and there were virtually no rejections. These groups exhibit little cooperation, exchange, or sharing beyond the family unit. Ethnographically, both show little fear of social sanctions and care little about 'public opinion'.
7. The Mapuche's social relations are characterized by mutual suspicion, envy, and fear of being envied. This pattern is consistent with the Mapuche's post-game interviews in the ultimatum game. Mapuche proposers rarely claimed that their offers were influenced by fairness, but rather by a fear of rejection. Even proposers who made hyper-fair offers claimed that they feared rare spiteful responders, who would be willing to reject even 50-50 offers.

8.2. The public goods game

The *public goods game* has been analyzed in a series of papers by the social psychologist Toshio Yamagishi, by the political scientist Elinor Ostrom and her co-workers, and by economists Ernst Fehr and his co-workers.[51] These researchers uniformly found that *groups exhibit a much higher rate of cooperation than can be expected assuming the standard economic model of the self-regarding actor*, and this is especially the case when subjects are given the option of incurring a cost to themselves in order to punish free riders.

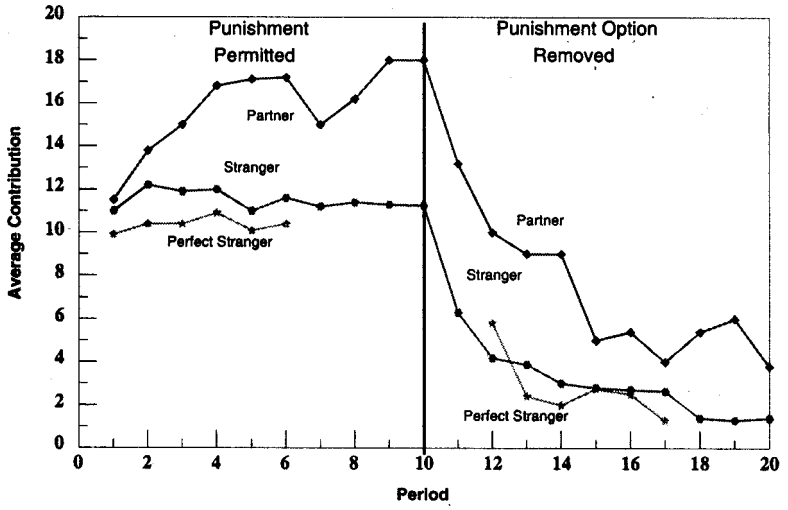
A typical public goods game consists of a number of rounds, say 10. The subjects are told the total number of rounds, as well as all other aspects of the game. The subjects are paid their winnings in real money at the end of the session. In each round, each subject is grouped with several other subjects (say three others) under conditions of strict anonymity. Each subject is then given a certain number of 'points', say 20, redeemable at the end of the experimental session for real money. Each subject then places some fraction of his points in a 'common account', and the remainder in the subject's 'private account'. The experimenter then tells the subjects how many points were contributed to the common account, and adds to the private account of each subject some fraction, say 40 percent, of the total amount in the common account. So if a subject contributes his whole 20 points to the common account, each of the four group members will receive eight points at the end of the round. In effect, by putting the whole endowment into the common account, a player loses 12 points, but the other three group members gain in total 24 (8×3) points. The players keep whatever is in their private account at the end of the round.

A self-regarding player will contribute nothing to the common account. However, only a fraction of subjects, in fact, behave that way. Subjects begin by contributing on average about half of their endowment to the public account. The level of contributions decays over the course of the 10 rounds, until in the final rounds most players are behaving in a self-interested manner.[52] In a meta-study of 12 public goods experiments, Fehr and Schmidt found that in the early rounds, average and median contribution levels ranged from 40 percent to 60 percent of the endowment, but in the final period, 73 percent of all individuals (N = 1042) contributed nothing, and many of the remaining players contributed close to zero.[53] These results are not compatible with the self-regarding actor model, which predicts zero contribution on all rounds, though they might be predicted by a reciprocal altruism model, since the chance to reciprocate declines as the end of the experiment approaches. However, this is not, in fact, the explanation of moderate but deteriorating levels of cooperation in the public goods game.

The explanation of the decay of cooperation offered by subjects when debriefed after the experiment is that cooperative subjects became angry at others who contributed less than themselves, and retaliated against free-riding low contributors in the only way available to them — by lowering their own contributions.[54]

Experimental evidence supports this interpretation. When subjects are allowed to punish non-contributors, they do so at a cost to themselves.[55] For instance, in Ostrom et al. subjects interacted for 25 periods in a public goods game, and by paying a 'fee', subjects could impose costs on other subjects by 'fining' them.[56] Since fining costs the individual who uses it, but the benefits of increased compliance accrue to the group as a whole, the only Nash equilibrium in this game that does not depend on incredible threats is for no player to pay the fee, so no player is ever punished for defecting, and all players defect by contributing nothing to the common pool. However, the authors found a significant level of punishing behavior.

These studies allowed individuals to engage in strategic behavior, since costly punishment of defectors could increase cooperation in future periods, yielding a positive net return for the punisher. Fehr and Gächter set up an experimental situation in which *the possibility of strategic punishment was removed*.[57] They used six- and ten-round public goods games with groups of size four, and with costly punishment allowed at the end of each round, employing three different methods of assigning members to groups. There were sufficient subjects to run between 10 and 18 groups simultaneously. Under the *Partner* treatment, the four subjects remained in the same group for all 10 periods. Under the *Stranger* treatment, the subjects were randomly reassigned after each round. Lastly, under the *Perfect Stranger* treatment, the subjects were randomly reassigned and assured that they would never meet the same subject more than once. Subjects earned an average of about US$35 for an experimental session.



Source: Fehr and Gächter, 'Cooperation and Punishment'.

Figure 1 **Average contributions over time in the Partner, Stranger, and Perfect Stranger treatments when the punishment condition is played first**

Fehr and Gächter performed their experiment for 10 rounds with punishment and 10 rounds without.[58] Their results are illustrated in Figure 1. We see that when costly punishment is permitted, cooperation does not deteriorate, and in the Partner game, despite strict anonymity, cooperation increases almost to full cooperation, even on the final round. When punishment is not permitted, however, the same subjects experience the deterioration of cooperation found in previous public goods games. The contrast in cooperation rates between the Partner and the two Stranger treatments is worth noting, because the strength of punishment is roughly the same across all treatments. This suggests that the credibility of the punishment threat is greater in the Partner treatment because in this treatment the punished subjects are certain that, once they have been punished in previous rounds, the punishing subjects are in their group. The prosociality impact of strong reciprocity on cooperation is thus more strongly manifested, the more coherent and permanent the group in question.

9. Conclusion

Each of the behavioral disciplines contributes strongly to human behavioral science. Taken separately and at face value, however, they offer partial, conflicting, and incompatible models of human behavior. From a scientific point of view, it is scandalous that this situation was tolerated throughout most of the 20th

century. Fortunately, there is currently a strong current of unification based on both mathematical models and common methodological principles for gathering empirical data on human behavior and human nature.

The true power of each discipline's contribution to knowledge will only appear when suitably qualified and deepened by the contribution of the others. For instance, the economist's model of rational choice behavior must be qualified by a biological appreciation that preference consistency is the result of evolutionary forces, and where such forces are absent, consistency will be imperfect and behavior must be augmented by empirical evidence. Moreover, a priori notions that preferences are self-regarding must be abandoned. These are the key tenets of behavioral economics. Second, the sociologist's notion of internalization of norms is generally rejected by the other behavioral disciplines because the ease with which diverse values can be internalized depends on human nature,[59] and the rate at which values are acquired and abandoned depends on their contribution to fitness and well-being.[60] Lastly, there are often swift society-wide value changes that cannot be accounted for by socialization theory.[61] When properly qualified, however, and appropriately related to the general theory of cultural evolution and strategic learning, the socialization theory is considerably strengthened.

Disciplinary boundaries in the behavioral sciences have been determined historically, rather than conforming to some consistent scientific logic. Perhaps for the first time, we are in a position to rectify this situation.

notes

I would like to thank Samuel Bowles and Suresh Naidu for helpful comments and the John D. and Catherine T. MacArthur Foundation for financial support.

1. By 'human biology' I mean the application of biological techniques to modeling human behavior. I use the term 'behavioral psychology' to mean social psychology and psychological decision theory.
2. Michael Hechter and Satoshi Kanazawa, 'Sociological Rational Choice', *Annual Review of Sociology* 23 (1997): 199–214.
3. Kristen Renwick Monroe, *The Economic Approach to Politics* (Reading, MA: Addison Wesley, 1991).
4. See James S. Coleman, *Foundations of Social Theory* (Cambridge, MA: Belknap, 1990); Anthony Downs, *An Economic Theory of Democracy* (Boston, MA: Harper and Row, 1957); Mancur Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge, MA: Harvard University Press, 1965); James M. Buchanan and R.D. Tollison, *The Theory of Public Choice* (Ann Arbor: University of Michigan Press, 1984).
5. Richard C. Lewontin, 'Evolution and the Theory of Games', *Journal of Theoretical Biology* 1 (1961): 382–403; William D. Hamilton, 'Extraordinary Sex Ratios', *Science* 156 (1967): 477–88; John Maynard Smith and G.R. Price, 'The Logic of Animal Conflict', *Nature* 246 (2 November 1973): 15–18.

6. John Alcock, *Animal Behavior: An Evolutionary Approach* (Sunderland, MA: Sinauer, 1993); Lee Alan Dugatkin and Hudson Kern Reeve, *Game Theory and Animal Behavior* (Oxford: Oxford University Press, 1998); Herbert Gintis, Eric Alden Smith and Samuel Bowles, 'Costly Signaling and Cooperation', *Journal of Theoretical Biology* 213 (2001): 103–19; Herbert Gintis, 'The Hitchhiker's Guide to Altruism: Genes, Culture, and the Internalization of Norms', *Journal of Theoretical Biology* 220 (2003): 407–18.
7. See Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis and Richard McElreath, 'Cooperation, Reciprocity and Punishment in Fifteen Small-Scale Societies', *American Economic Review* 91 (2001): 73–8; Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr and Herbert Gintis, *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (Oxford: Oxford University Press, 2004).
8. See Herbert Gintis, Samuel Bowles, Robert Boyd and Ernst Fehr, *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life* (Cambridge, MA: MIT Press, 2004).
9. Herbert Gintis, *Game Theory Evolving* (Princeton, NJ: Princeton University Press, 2000).
10. Matthew Rabin, 'Incorporating Fairness into Game Theory and Economics', *American Economic Review* 83 (1993): 1281–302; Theodore C. Bergstrom and Oded Stark, 'How Altruism can Prevail in an Evolutionary Environment', *American Economic Review* 83 (1993): 149–55; James Andreoni and John H. Miller, 'Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism', *Econometrica* 70 (2002): 737–53; Ernst Fehr and Simon Gächter, 'Altruistic Punishment in Humans', *Nature* 415 (10 January 2002): 137–40; Elisabeth Jean Wood, *Insurgent Collective Action and Civil War in El Salvador* (Cambridge: Cambridge University Press, 2003).
11. See Herbert Gintis, 'Solving the Puzzle of Human Prosociality', *Rationality and Society* 15 (2003): 155–87; Samuel Bowles and Herbert Gintis, 'Prosocial Emotions', in *Complex Nonlinear Systems III* (in press); Wood, *Insurgent Collective Action and Civil War in El Salvador*.
12. H. Peyton Young, *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions* (Princeton, NJ: Princeton University Press, 1998); Gintis, *Game Theory Evolving*; Gerd Gigerenzer and Reinhard Selten, *Bounded Rationality* (Cambridge, MA: MIT Press, 2001).
13. Sidney G. Winter, 'Satisficing, Selection and the Innovating Remnant', *Quarterly Journal of Economics* 85 (1971): 237–61; Herbert Simon, 'Theories of Bounded Rationality', in *Decision and Organization*, edited by C.B. McGuire and Roy Radner (New York: American Elsevier, 1972), pp. 161–76.
14. Gigerenzer and Selten, *Bounded Rationality*.
15. David M. Kreps, *A Course in Microeconomic Theory* (Princeton, NJ: Princeton University Press, 1990).
16. Coleman, *Foundations of Social Theory*; Peter Kollock, 'Transforming Social Dilemmas: Group Identity and Cooperation', in *Modeling Rational and Moral Agents*, edited by Peter Danielson (Oxford: Oxford University Press, 1997); Hechter and Kanazawa, 'Sociological Rational Choice'.

17. Andreoni and Miller, 'Giving According to GARP'.
18. Olson, *The Logic of Collective Action*.
19. Barrington Moore, Jr, *Injustice: The Social Bases of Obedience and Revolt* (White Plains, NY: M.E. Sharpe, 1978); Wood, *Insurgent Collective Action and Civil War in El Salvador*.
20. Amos Tversky and Daniel Kahneman, 'Loss Aversion in Riskless Choice: A Reference-Dependent Model', *Quarterly Journal of Economics* 106 (1981): 1039–61; Toko Kiyonari, Shigehito Tanida and Toshio Yamagishi, 'Social Exchange and Reciprocity: Confusion or a Heuristic?', *Evolution and Human Behavior* 21 (2000): 411–27.
21. Edwin Schrödinger, *What is Life?: The Physical Aspect of the Living Cell* (Cambridge: Cambridge University Press, 1958) called this an 'aperiodic crystal'.
22. Richard C. Lewontin, *The Genetic Basis of Evolutionary Change* (New York: Columbia University Press, 1974).
23. P. Taylor and L. Jonker, 'Evolutionarily Stable Strategies and Game Dynamics', *Mathematical Biosciences* 40 (1978): 145–56.
24. Eva Jablonka and Marion J. Lamb, *Epigenetic Inheritance and Evolution: The Lamarckian Case* (Oxford: Oxford University Press, 1995).
25. John Tyler Bonner, *The Evolution of Culture in Animals* (Princeton, NJ: Princeton University Press, 1984); Peter J. Richerson and Robert Boyd, 'The Evolution of Ultrasociality', in *Indoctrinability, Ideology and Warfare*, edited by I. Eibl-Eibesfeldt and F.K. Salter (New York: Berghahn Books, 1998): 71–96.
26. Robert Boyd and Peter J. Richerson, *Culture and the Evolutionary Process* (Chicago, IL: University of Chicago Press, 1985); John Conlisk, 'Optimization Cost', *Journal of Economic Behavior and Organization* 9 (1988): 213–28.
27. Pascal Boyer, *Religion Explained: The Human Instincts that Fashion Gods, Spirits and Ancestors* (London: Heinemann, 2001).
28. Richerson and Boyd, 'The Evolution of Ultrasociality'.
29. For inclusive fitness, see William D. Hamilton, 'The Genetical Evolution of Social Behavior', *Journal of Theoretical Biology* 37 (1964): 1–16, 17–52. For reciprocal altruism, see R.L. Trivers, 'The Evolution of Reciprocal Altruism', *Quarterly Review of Biology* 46 (1971): 35–57.
30. See Kenneth J. Arrow and Gerard Debreu, 'Existence of an Equilibrium for a Competitive Economy', *Econometrica* (1954): 265–90; Robert Axelrod and William D. Hamilton, 'The Evolution of Cooperation', *Science* 211 (1981): 1390–6; Drew Fudenberg and Eric Maskin, 'The Folk Theorem in Repeated Games with Discounting or with Incomplete Information', *Econometrica* 54 (1986): 533–54; Bernard Mandeville, *The Fable of the Bees: Private Vices, Publick Benefits* (1705); Adam Smith, *The Theory of Moral Sentiments* (New York: Prometheus, 2000 [1759]).
31. Emile Durkheim, *Suicide, a Study in Sociology* (New York: Free Press, 1951).
32. Ibid.; Ruth Benedict, *Patterns of Culture* (Boston, MA: Houghton Mifflin, 1934); Margaret Mead, *Sex and Temperament in Three Primitive Societies* (New York: Morrow, 1963); Talcott Parsons, *Sociological Theory and Modern Society* (New York: Free Press, 1967); Joan E. Grusec and Leon Kuczynski, *Parenting and Children's Internalization of Values: A Handbook of Contemporary Theory* (New York: Wiley, 1997).

33. Gintis, 'The Hitchhiker's Guide to Altruism'.
34. Ibid.
35. Richerson and Boyd, 'The Evolution of Ultrasociality'.
36. See Bonner, *The Evolution of Culture in Animals*; Boyd and Richerson, *Culture and the Evolutionary Process*; Jablonka and Lamb, *Epigenetic Inheritance and Evolution* and their 'Epigenetic Inheritance in Evolution', *Journal of Evolutionary Biology* 11 (1998): 159–83.
37. Michael Tomasello, *The Cultural Origins of Human Cognition* (Cambridge, MA: Harvard University Press, 1999).
38. Antonio R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: Avon Books, 1994).
39. Bowles and Gintis, 'Prosocial Emotions'.
40. Linda Mealey, 'The Sociobiology of Sociopathy', *Behavioral and Brain Sciences* 18 (1995): 523–41.
41. Damasio, *Descartes' Error*.
42. Donald E. Brown, *Human Universals* (New York: McGraw-Hill, 1991).
43. Robert B. Edgerton, *Sick Societies: Challenging the Myth of Primitive Harmony* (New York: Free Press, 1992).
44. William H. Calvin, 'A Stone's Throw and its Launch Window: Timing Precision and its Implications for Language and Hominid Brains', *Journal of Theoretical Biology* 104 (1983): 121–35.
45. Stephen J. Gould and Elizabeth Vrba, 'Exaption: A Missing Term in the Science of Form', *Paleobiology* 8 (1981): 4–15.
46. By the same token, even antisocial norms can hitchhike on the internalization capacity, and it not infrequently does just that. See Edgerton, *Sick Societies*. Gintis, 'The Hitchhiker's Guide to Altruism' shows that the tendency of higher fitness groups to outcompete lower fitness groups provides a strong tendency toward the circumscription of antisocial norms.
47. Gintis, 'The Hitchhiker's Guide to Altruism'.
48. This material is developed in Herbert Gintis, Samuel Bowles, Robert Boyd and Ernst Fehr, 'Explaining Altruistic Behavior in Humans', *Evolution & Human Behavior* 24 (2003): 153–72.
49. Colin Camerer and Richard Thaler, 'Ultimatums, Dictators, and Manners', *Journal of Economic Perspectives* 9 (1995): 209–19; Werner Güth and Reinhard Tietz, 'Ultimatum Bargaining Behavior: A Survey and Comparison of Experimental Results', *Journal of Economic Psychology* 11 (1990): 417–49; Alvin E. Roth, Vesna Prasnikar, Masahiro Okuno-Fujiwara and Shmuel Zamir, 'Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study', *American Economic Review* 81 (1991): 1068–95.
50. Henrich et al., 'Cooperation, Reciprocity and Punishment in Fifteen Small-Scale Societies'.
51. See Toshio Yamagishi, 'The Provision of a Sanctioning System as a Public Good', *Journal of Personality and Social Psychology* 51 (1986): 110–16; 'The Provision of a Sanctioning System in the United States and Japan', *Social Psychology Quarterly* 51 (1988): 265–71; 'Seriousness of Social Dilemmas and the Provision of a Sanctioning System', *Social Psychology Quarterly* 51 (1988): 32–42; Elinor Ostrom, James Walker and Roy Gardner, 'Covenants With and Without a Sword:

Self-Governance Is Possible', *American Political Science Review* 86 (1992): 404–17; Simon Gächter and Ernst Fehr, 'Collective Action as a Social Exchange', *Journal of Economic Behavior and Organization* 39 (1999): 341–69; Ernst Fehr and Simon Gächter, 'Cooperation and Punishment', *American Economic Review* 90 (2000): 980–94; 'Altruistic Punishment in Humans'.

52. Robyn M. Dawes and Richard Thaler, 'Cooperation', *Journal of Economic Perspectives* 2 (1988): 187–97; J.O. Ledyard, 'Public Goods: A Survey of Experimental Research', in *The Handbook of Experimental Economics*, edited by J.H. Kagel and A.E. Roth (Princeton, NJ: Princeton University Press, 1995), pp. 111–94.
53. Ernst Fehr and Klaus M. Schmidt, 'A Theory of Fairness, Competition, and Cooperation', *Quarterly Journal of Economics* 114 (1999): 817–68.
54. James Andreoni, 'Cooperation in Public Goods Experiments: Kindness or Confusion', *American Economic Review* 85 (1995): 891–904.
55. John M. Orbell, Robyn M. Dawes and J.C. Van de Kragt, 'Organizing Groups for Collective Action', *American Political Science Review* 80 (1986): 1171–85; Kaori Sato, 'Distribution and the Cost of Maintaining Common Property Resources', *Journal of Experimental Social Psychology* 23 (1987): 19–31; Toshio Yamagishi, 'The Provision of a Sanctioning System in the United States and Japan'; 'Seriousness of Social Dilemmas and the Provision of a Sanctioning System'; 'Group Size and the Provision of a Sanctioning System in a Social Dilemma', in *Social Dilemmas: Theoretical Issues and Research Findings*, edited by W.B.G. Liebrand, David M. Messick and H.A.M. Wilke (Oxford: Pergamon, 1992), pp. 267–87.
56. Ostrom et al., 'Covenants With and Without a Sword: Self-Governance Is Possible'.
57. Fehr and Gächter, 'Cooperation and Punishment'.
58. Ibid. For additional experimental results and analysis, see Samuel Bowles and Herbert Gintis, 'Homo Reciprocans', *Nature* 415 (10 January 2002): 125–8.
59. John Tooby and Leda Cosmides, 'The Psychological Foundations of Culture', in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, edited by Jerome H. Barkow, Leda Cosmides and John Tooby (New York: Oxford University Press, 1992), pp. 19–136; Steven Pinker, *The Blank Slate: The Modern Denial of Human Nature* (New York: Viking, 2002).
60. Gintis, 'Solving the Puzzle of Human Prosociality'; 'The Hitchhiker's Guide to Altruism'.
61. Dennis H. Wrong, 'The Oversocialized Conception of Man in Modern Sociology', *American Sociological Review* 26 (1961): 183–93; Herbert Gintis, 'Welfare Economics and Individual Development: A Reply to Talcott Parsons', *Quarterly Journal of Economics* 89 (1975): 291–302.