Beyond selfish and selfless

Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life, H. Gintis, S. Bowles, R. Boyd, E. Fehr (Eds.). MIT Press, Cambridge, MA (2005). xii + 400 pp., Index, Price: $50.00, ISBN: 0-262-07252-1

An economist offered a woman $10 in exchange for a date. Knowing that she had no weekend engagements, he assessed her opportunity cost at $5, and resolved to offer her $10 (though he valued the date at $100), since he happened to know that she was inclined, or at least indifferent, to a date with him, and therefore, assuming that the woman was properly self-interested, she would happily accept the $5 surplus, and a merry date would undoubtedly result. Like many economists, the man spent the weekend alone.

Economists who take very seriously the view that humans are purely self-interested, and thus invariably respond to incentives, seldom attract dates, they do not have many friends, and their coalitions rarely cooperate to rape and pillage neighboring groups, i.e. they lack humanity. This is the subject of *Moral Sentiments and Material Interests*, a collection of chapters by economists and anthropologists that establishes, beyond any doubt, that people frequently deviate from selfishness and that these deviations are sufficiently systematic and pervasive that they cannot be safely ignored by economic theory. Many of the authors, particularly the economists, instead favor an account of human nature centered around "strong reciprocity," but this proposal is less compelling than the assault on *Homo economicus*.

In the opening chapter, Gintis et al. review evidence that people do not always act to maximize materials gains, arguing that human nature is neither selfish nor unconditionally altruistic. The chapter highlights evidence from experimental games—an important innovation that brings new precision and clarity to age-old debates regarding human nature. Consider the ultimatum game, in which a participant acting as the proposer divides a monetary endowment with another participant, the responder; once the proposer has made a division, the responder accepts or rejects the offer (rejection yields both nothing). The self-interest hypothesis predicts that the proposer will maximize material gain by offering the smallest amount, since a self-interested responder should accept any positive gain. The authors explain that this prediction has been falsified in dozens of experiments worldwide, as proposers typically offer roughly half, and responders frequently reject offers less than 30 percent (reviewed in Camerer, 2003). People are not purely selfish, but are capable of spite – damaging their own material welfare in order to damage others – especially when targets have acted maliciously or unfairly.

Also convincing are the results of experiments with an employer–worker game, a social dilemma in which employers offer a wage and workers subsequently choose an effort level. Self-interest predicts minimum wage offers and effort levels, and here again, the prediction is roundly falsified. Employers offer decent wages, and worker effort increases with wage amount (although only a quarter of workers provided the effort they promised). People are not purely selfish, but are capable of altruism – damaging their own material welfare in order to benefit others – especially when targets have acted benevolently. People are also not purely altruistic, as most workers failed to deliver the effort levels they promised. The chapter similarly reviews evidence of costly altruism and punishment in the public goods game, as well as experiments examining "crowding out" in which incentives have the opposite of the intended effect. An example of the latter is a trust game experiment, in which fines for untrustworthy behavior *decrease* trustworthiness. In sum, the laboratory evidence in this chapter alone makes abundantly clear that people often deviate from selfishness in order to help or harm others.

Part II "The behavioral ecology of cooperation" consists of three chapters by anthropologists addressing the plausibility of a selfish human nature given what is known about human evolution. In chapter two, primatologist Joan Silk writes a concise and enlightening review of the rapidly accumulating data regarding altruism among nonhuman primates—a crucial body of evidence for all scholars of human nature. We learn that primate altruism is real and beautifully principled. From chimp alliances and warfare to vervet alarm calls and howler alloparenting, primates engage in costly actions to benefit others.

Much of primate altruism is illuminated by kin selection (Hamilton, 1964); researchers have found that delivery of benefits is not haphazard, but is often finely tuned to the relatedness of conspecifics. Silk writes, "Primates are discriminating nepotists." Moreover, the determination of relatedness presents primates with special difficulties. Unlike species that can directly detect genetic relatedness (silk gives the example of sea squirts), primates must rely on cues of kinship, e.g. determining maternal kinship by patterns of mother's interaction. Moreover, reliable cues will often differ for maternal and paternal kin, and as social organization and dispersal patterns vary across primate species. What should be absolutely stunning to non-biologists (but a commonplace wonder to biologists) is that primate kin

recognition systems track these subtleties across social ecology. This fact leaves no doubt that many primate species possess sophisticated cognitive machinery that employs cues specific to the species' social ecology to compute relatedness, a computation which in turn shapes the delivery of benefits across conspecific targets. More generally, these facts make clear that altruism falls under the ceaseless gaze of natural selection, creating a market for genes that increase in frequency by specializing in the efficient delivery of benefits to subsets of other organisms, and thereby resulting in countless adaptations, found in innumerable species from bacteria to mole rats, designed to effectively target altruism.

Among Silk's most interesting examples is contingent nepotism among Japanese macaques. Females generally intervene on behalf of their sisters in conflicts. But when a female is involved in a rank reversal campaign against her sister, she will intervene on behalf of a nonrelative against her sister. Thus, not only do macaques have specialized machinery for recognizing and helping kin, but they are additionally capable of partnering with nonkin against their sisters, when it is in their interest to do so. Finally, Silk reviews evidence for reciprocal altruism (Trivers, 1971) among several species of primates, describing patterns of grooming, food sharing, and agonistic support that indicate that conditional altruists benefit from favors returned in future interactions.

In chapter three, Kaplan and Gurven discuss the evolution of food sharing among humans. Humans are unique in the extent to which food is shared within and between families. The authors find that much of this activity can be explained by reciprocal altruism since food distribution is controlled by the producer and contingent on the history of transfers with receivers. However, some food sharing, especially large game, involves persistent imbalances and is inconsistent with reciprocal altruism; they argue that tolerated theft, costly signaling, group selection, and/or kin altruism might play a role in explaining these imbalances. The authors further maintain that intergenerational food transfers were required over evolutionary history to sustain the extended development of children, and ultimately human intelligence. If the authors are correct that food sharing played a key role in human evolution, this underscores the possibility that natural selection has been long at work on the cognitive mechanisms underlying food sharing, and that these mechanisms might therefore exhibit intricate design.

In chapter four, Smith and Bird apply costly signaling theory (aka conspicuous consumption) to culturally widespread acts of public altruism such as feasts, funerary rites, and potlatching. Seemingly wasteful behavioral or morphological traits can benefit individuals when these traits reliably signal qualities of the signaler that are otherwise difficult to observe, which occurs when the costs (or benefits) of sending the signal are quality-dependent. Smith and Bird describe a costly signaling model of cooperative actions such as military campaigns or punishment of criminals. However, an important limitation of this model is that costly actions that are neutral or harmful to group welfare are as likely to be favored as group beneficial cooperation. The authors insightfully turn to "broadcast efficiency" to explain why cooperation might be favored over group detrimental signals. People are intrinsically interested in events that increase their own welfare, allowing group beneficial signals to better capture audience attention.

In sum, Part II lucidly demonstrates that the Skinnerian view of human nature embodied by *Homo economicus* is contradicted by what is known about human evolution. Species in all kingdoms of life possess intricate mechanisms designed to deliver benefits and harms to other organisms, even when individually costly. Primates are no exception, and indeed, consistent with their high sociality, primate spite and altruism exhibits dazzling sophistication, which is well illustrated by the diversity of kin recognition schemes. Humans are unique in the complexity of food transfers occurring between and within families, and moreover, food sharing is likely underlain by several distinct mechanisms, each designed to capture different benefits (e.g. future reciprocation, increased status, group advantage, etc.). Finally, widespread public altruism cannot be accounted for by kin selection or reciprocal altruism, again indicating a distinct cognitive system that specializes in altruism designed to signal individual quality. Far from simple and selfish knaves, humans are equipped with an array of devices specialized to deliver benefits and harms to others, even when individually costly.

Part III "Modeling and testing strong reciprocity" reviews further experimental evidence contradicting the self-interest hypothesis, and additionally describes models of strong reciprocity. A particular emphasis in this section is on strategy heterogeneity among individuals, i.e. some behave selfishly while others act altruistically, reciprocally, etc. In chapter five, Fehr and Fischbacher review further experimental results consistent with data presented in the opening chapter. The authors review results from experiments with the ultimatum game with competition, a double auction experiment, the prisoner's dilemma, the public goods game with punishment, and the gift exchange game. In all cases, the predictions of the self-interest hypothesis are falsified. In chapter six, Falk and Fischbacher attempt to represent strong reciprocity – a tendency to cooperate with cooperators and to punish norm violators – with a

utility function. Their utility function includes a reciprocity parameter, a kindness term, and a reciprocation term. The authors seem to prefer to model strong reciprocity as a preference rather than as a strategy. Given that game theory regards feasible actions, action profiles, outcomes, and preferences over outcomes as primitive elements, it would seem that involving actions directly into utility functions might introduce undesirable circularities. Disappointingly, the authors do not directly address whether strong reciprocity can be more clearly modeled as a strategy or as a preference.

In chapter seven, Boyd et al. describe simulations showing that costly punishment can be a stable strategy against defectors and unconditional contributors, and interestingly, that costly punishment evolves more easily than altruism. However, they conclude that under plausible conditions, costly punishment is unlikely to be produced by genetic group selection; instead, it is more likely to be favored in cultural evolution. In chapter eight, Sethi and Somanathan describe a model of the evolution of reciprocity based on commitment. They show that in groups with changing composition, reciprocity can be stable among selfishness when reciprocators can credibly commit to punishing selfishness, thereby inducing self-interested individuals to contribute to the group. Their claims are supplemented by a mathematical argument in an appendix. The authors additionally review models of the evolution of reciprocity based on parochialism and assortative interaction.

Finally, Part IV "Reciprocity and social policy" examines violations of self-interest in relation to policy. Far from confined to the laboratory, it becomes clear that deviations from selfishness operate in a diverse array of economic settings. In chapter nine, Ostrom argues that increasing the costs of tax evasion can crowd out intrinsic motivation to contribute taxes, removing fees associated with irrigation networks can decrease system maintenance, offering to compensate citizens for accepting a nuclear waste facility in their neighborhood can decrease willingness to do so, and other interesting examples in which incentives operate counterintuitively. Ostrom also reviews analyses of thousands of systems of management created by local users of common resources such as fisheries, irrigation networks, and grazing land (e.g., 200 irrigation systems in Nepal). These systems tend to be enforced more reliably than centralized government enforcement. The success of local enforcement is also due to the huge variety of rules which often include many redundancies that seem to facilitate experimentation and fine tuning of local enforcement systems. Ostrom concludes that naive acceptance of the view that people are selfish has obscured the fact that people do, in fact, engage in costly enforcement, and indeed, these local systems are often superior to centralized government enforcement in producing compliance. In chapter 10, Fong, Bowles, and Gintis review evidence that reciprocity shapes attitudes towards welfare, and thus public support of welfare programs. Particularly, people support giving to those in need when neediness is perceived as resulting from bad luck rather than shirking, and when the recipient is in good moral standing (for a similar argument, see Cosmides and Tooby, 1992, p. 219). In chapter 11, Bewley examines wage rigidity, i.e. why employers seem to favor letting workers go rather than reducing salaries. He considers the possibility that pay cuts might be resented as unfair, diminishing worker morale and potentially inducing good workers to leave.

In chapter 12, Kahan discusses how rewards and punishments can inhibit collective action, focusing on tax compliance, placement of undesirable facilities, and street crime. Two mechanisms in particular are addressed. First, incentives might signal that others are not contributing. For example, increasing penalties for tax evasion can indicate to citizens high levels of noncompliance; this signal can cause reciprocators to reduce their tax payments. Second, incentives can mask voluntary contributions which might otherwise inspire reciprocators to contribute. High penalties might cause reciprocators to infer that others contribute merely to avoid the penalties, thus interrupting any reciprocation of voluntary contributions that might otherwise occur. In the final chapter, Bowles and Gintis repeat the refrain that human nature is neither selfish nor unconditionally altruistic. Instead, they argue that people are best characterized as strong reciprocators, and as a result, what they term "community governance" can be employed to solve problems that can not be solved by individuals, markets, or states. To illustrate the point, the authors review examples such as the efficacy of peer-monitoring in Chicago neighborhoods and risk-pooling in Japanese fishing cooperatives. To explain the advantages of community governance, they point to repeated interactions, information flow, and the effectiveness of informal punishment. They close by noting that economic inequality might impede the capacity of communities to solve problems.

Moral Sentiments and Material Interests demonstrates substantial and systematic deviations from selfishness in the laboratory and the field, and shows that such deviations can plausibly arise from a number of evolutionary and cultural mechanisms. The self-interest hypothesis is false and potentially damaging. Less persuasive is the particular alternative championed by many (but not all) of the contributors, a view of human nature centered on strong reciprocity.

1. Strong reciprocity?

Why do courted women reliably respond negatively to cash offers? I really do not think the question is as trivial as it sounds. More likely, it sounds trivial precisely because it is not—because the question recruits strong intuitions which are themselves genuinely intriguing. Why do monetary incentives reliably *decrease* the likelihood that a woman will accept an invitation to a date? And why do most of us seem to know this? One might regard the cash-for-date faux pas as simple convention: People just do not do that, and in fact, conventional enticements such as dinner or drinks do seem to increase acceptance. But this is unsatisfying. Ask someone out while standing on your head. People do not do that either, but it is not obvious that being unconventional *per se* is disadvantageous. Offering cash for a date is *offensive*, not merely unexpected. Perhaps the reasons why incentives sometimes backfire with indignation or defiance are genuinely mysterious.

Psychologists have uncovered other effects similar to those discussed in the book under review, for example, the psychology of taboo and outrage (e.g. Fiske and Tetlock, 1997; Tetlock, 2003). Fiske and Tetlock (1997) ask us to imagine offering our mother cash for Thanksgiving dinner, offering our lover hugs in exchange for kisses, or a soldier offering to salute a sergeant on the condition that he stops yelling; in these cases, it is obvious that incentives will not have the desired effect, but why?

There are many possible explanations for deviations from selfishness. For example, organisms that reliably respond to incentives are easily manipulated by others. Frogs reliably respond to wriggling worms, a fact that is exploited by several dozen species of vipers which lure prey by wiggling their tails like worms. It sometimes pays to be skeptical, particularly when incentives arise from other agents. Such skepticism is found in the economic aphorism *there's no such thing as a free lunch*, which derives from the practice of luring saloon patrons with free lunch. Broadly, this possibility suggests that incentives are most likely to backfire when arising from agents. Indeed, experiments have found that when playing against computers or dice, people do behave in a self-interested manner (Gintis et al., chapter one).

Another possibility is that unselfishness directly benefits the genes underlying the associated mechanisms, as with kin altruism, reciprocal altruism, mutualism, and spite. For example, killer X chromosomes produce toxins that spitefully damage themselves in order to inflict greater damage on gametes carrying a Y chromosome, thereby netting a transmission advantage (Burt and Trivers, 2006). Other possibilities include costly signaling (Smith and Bird, chapter four) and commitment (Frank, 1988). Turning to humans in particular, Fiske and Tetlock (1997) have argued that humans exhibit four elementary forms of social relations (communal sharing, reciprocity, authority ranking, market pricing), and that coordinating on one of these mutually exclusive forms might require shunning other forms; for example, offering cash for a date might additionally signal that one regards the relationship as market pricing which is incompatible with the communal sharing expected of romance.

There are many pathways by which unselfishness can arise, and as game theory is yet young, undoubtedly more pathways will be uncovered. Where does strong reciprocity stand among these possibilities? An important limitation of strong reciprocity is that it is not as clearly formulated as other theories such as, for example, reciprocal altruism or kin selection. In the words of contributors Sethi and Somanathan (chapter eight), "The literature on the evolution of strong reciprocity is a patchwork of models, each of which emphasizes a different mechanism under which reciprocators can survive." The problem is that a "patchwork" of models is not as empirically tractable as a single model.

A second drawback of strong reciprocity is that it is probably too ambitious. Its scope is enormous, covering human cooperation in general as well as costly punishment, including both second party retaliation and third party condemnation. Consider a man who spitefully attacks someone for having sex with the man's wife. Is it plausible that this attack is caused by the same cognitive mechanisms that underlie third party intervention in a mugging? (Note here that mate guarding is widespread among species, while third party intervention is extremely rare.) It is likely that strong reciprocity glosses a sizeable set of distinct psychological traits, and that the associated mechanisms are each designed to capture different sets of benefits.

A final weakness of strong reciprocity is that the evidence presented in *Moral Sentiments and Material Interests* does not seem to support the most popular versions of the theory. Theories of strong reciprocity typically invoke group selection of one kind or another. Claims about biological function are tested by examining whether the associated structure is improbably well designed to perform the proposed function (Williams, 1966). Crucially, adaptations *cannot* be established by evidence of beneficial effects (Williams, 1966). Williams (1966) long ago made clear that a fox might benefit by following its own tracks in the snow, but this does not show whether or not its feet are adaptations designed to produce tracks in the snow. The question is not whether people sometimes benefit the group with cooperation or

costly punishment, but whether the associated mechanisms are improbably well designed to benefit the group. None of the evidence presented in this volume indicates psychological traits well designed to benefit the group.

The main source of evidence allegedly in favor of strong reciprocity is unselfishness in one-shot anonymous games. However, it has been pointed out that human psychology is not designed to deal with one-shot anonymous situations (e.g. Burnham and Johnson, 2005; Hagen and Hammerstein, 2006; Trivers, 2004). The volume responds to this critique (chapters one and five) in a puzzling manner, arguing that experimental evidence of sensitivity to repetition indicates an evolutionary history of one-shot anonymous interactions. However, the argument is not that the probability of repetition did not vary at all—only that it never reached zero. Indeed, sensitivity to the probability of repetition is a key prediction of reciprocal altruism (Trivers, 1971), and in contrast, is not predicted by group-selected strong reciprocity. It does not benefit the group to ignore a crime when no one is watching. Despite claims made in the volume, the acknowledged fact that, "repetition and absence of anonymity dramatically increase the level of cooperation and punishment" (Gintis et al., chapter one) *contradicts* the central prediction of strong reciprocity, which maintains that cooperation and punishment do not depend on levels of anonymity or repetition (for extended critiques of strong reciprocity, see Burnham and Johnson, 2005; Hagen and Hammerstein, 2006; West et al., 2007).

Moral Sentiments and Material Interests is a fascinating volume rich with data from experimental games that brings new precision to fundamental questions about human nature. While strong reciprocity is not much more persuasive than the previous silver bullet, *Homo economicus*, this does not detract from the central message of the book that it is time for economics, like biology, to move beyond quaint depictions of human nature as nice or selfish.

References

Burnham, T., Johnson, D.D.P., 2005. The evolutionary and biological logic of human cooperation. Analyse & Kritik 27, 113–135.
Burt, A., Trivers, R., 2006. Genes in Conflict: The Biology of Selfish Genetic Elements. Harvard University Press, Cambridge.
Camerer, C.F., 2003. Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton.
Cosmides, L., Tooby, J., 1992. Cognitive adaptations for social exchange. In: The Adapted Mind: Evolutionary Psychology and the Generation of Culture. Oxford University Press, New York, pp. 163–228.
Fiske, A.P., Tetlock, P.E., 1997. Taboo trade-offs: reactions to transactions that transgress spheres of justice. Political Psychology 18, 255–297.
Frank, R., 1988. Passions Within Reason: The Strategic Role of the Emotions. W.W. Norton & Company, New York.
Hagen, E.H., Hammerstein, P., 2006. Game theory and human evolution: a critique of some recent interpretations of experimental games. Theoretical Population Biology 69, 339–348.
Hamilton, W., 1964. The genetic evolution of social behaviour. Journal of Theoretical Biology 7, 1–52.
Tetlock, P.E., 2003. Thinking the unthinkable: sacred values and taboo cognitions. Trends in Cognitive Science 7, 320–324.
Trivers, R.L., 1971. The evolution of reciprocal altruism. Quarterly Review of Biology 46, 35–57.
Trivers, R., 2004. Mutual benefits at all levels of life. Science 304, 964–965.
West, S.A., Griffin, A.S., Gardner, A., 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity, and group selection. Journal of Evolutionary Biology 20, 415–432.
Williams, G.C., 1966. Adaptation and Natural Selection. Princeton University Press, Princeton.

Peter DeScioli *
Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA

* Tel.: +1 215 913 8569.
E-mail address: descioli@psych.upenn.edu

7 March 2007
Available online 12 March 2007

10.1016/j.jebo.2007.03.002