BEHAVIORAL AND BRAIN SCIENCES (1989) 12, 683–739
Printed in the United States of America

Selfishness examined: Cooperation in the absence of egoistic incentives

Linnda R. Caporael
Department of Science and Technology Studies, Rensselaer Polytechnic Institute, Troy, NY 12181
Electronic mail: *caporael@mts.rpi.edu*
Robyn M. Dawes
Department of Decision and Social Sciences, Carnegie-Mellon University, Pittsburgh, PA 15213
Electronic mail: *dawes@andrew.cmu.edu*
John M. Orbell
Department of Political Science, University of Oregon, Eugene, OR 97403
Electronic mail: *jorbell@oregon.uoregon.edu*
Alphons J. C. van de Kragt
Department of Political Science, University of Oregon, Eugene, OR 97403

Abstract: Social dilemmas occur when the pursuit of self-interest by individuals in a group leads to less than optimal collective outcomes for everyone in the group. A critical assumption in the human sciences is that people's choices in such dilemmas are individualistic, selfish, and rational. Hence, cooperation in the support of group welfare will only occur if there are selfish incentives that convert the social dilemma into a nondilemma. In recent years, inclusive fitness theories have lent weight to such traditional views of rational selfishness on Darwinian grounds. To show that cooperation is based on selfish incentives, however, one must provide evidence that people do *not* cooperate *without* such incentives. In a series of experimental social dilemmas, subjects were instructed to make single, anonymous choices about whether or not to contribute money for a shared "bonus" that would be provided only if enough other people in the group also contributed their money. Noncontributors cited selfish reasons for their choices; contributors did not. If people are allowed to engage in discussion, they will contribute resources at high rates, frequently on irrational grounds, to promote group welfare. These findings are consistent with previous research on ingroup biasing effects that cannot be explained by "economic man" or "selfish gene" theories. An alternative explanation is that sociality was a primary factor shaping the evolution of *Homo sapiens*. The cognitive and affective mechanisms underlying such choices evolved under selection pressures on small groups for developing and maintaining group membership and for predicting and controlling the behavior of other group members. This sociality hypothesis organizes previously inexplicable and disparate phenomena in a Darwinian framework and makes novel predictions about human choice.

Keywords: altruism; game theory; group behavior; human evolution; prisoner's dilemma; rational choice; selfishness; social cognition; social dilemmas; sociality; sociobiology

Group living has been a central characteristic of the human species throughout human evolution (Foley 1987); there is no archaeological evidence of a transition from isolation to group life. Yet psychologists, anthropologists, and evolutionists have rarely considered what psychological characteristics might have evolved as adaptations to living in small groups. Instead, most theorists assume that human nature is basically selfish and individualistic, an assumption that applies just as well to animals that have evolved in isolation as to those that have evolved in groups. In recent years, "selfish gene" theories (e.g., Dawkins 1976) have provided the ultimate justification for this centuries-old notion of "economic man" (Myers 1983) by arguing – somewhat tautologically – that human behavior evolved to maximize fitness in local environments. A variety of "egoistic incentive" theories have accordingly been proposed to account for the minimal amount of cooperation necessary for group living in terms of the individual selfish payoffs it affords; for example, Axelrod (1984) proposes that cooperation evolves as the result of the payoffs to the cooperating individual when cooperation is reciprocated ("reciprocal altruism"), whereas Alexander (1987) proposes that cooperation evolves as the result of payoffs that accrue to the genes of the cooperating individual when the beneficiaries of the cocperation are genetically related ("kin altruism").

These *egoistic incentive*[1] (EI) notions are really untested metatheories; they seem to be based on cultural beliefs about "human nature" rather than on reasoned argument. In this target article, we present experimental evidence that (1) human cooperation can be controlled by manipulating variables totally unrelated to self-interest and that (2) satisfying self-interest is *not* necessary to

 0140-525X/89 $5.00+.00

elicit this cooperation. These results cannot be explained by EI theories. Instead, we suggest that the cognitive and affective mechanisms underlying the cooperation observed in these studies may have evolved from selection pressures exerted under small-group living conditions for developing and maintaining group membership. According to this "sociality hypothesis," human nature is basically social rather than selfish.[2]

1. Social dilemmas and egoistic incentive theories

In a later section, experiments are described in which people are faced with choices between individual selfishness and social cooperation. These experiments confront people with "social dilemmas" – situations that maximize the conflict between private incentives and group welfare. We use a simple example, the funding of public broadcasting, to introduce the basic concepts and structural features of social dilemmas. (The Appendix includes a glossary for readers unfamiliar with the technical terms of game theory.)

For almost all potential contributors, the success or failure of public broadcasting is determined by what others contribute, usually in private; there are neither tangible social rewards (*payoffs*) for contributing nor sanctions for not contributing (the absence of a sanction is also a payoff). We can "free ride" on the contributions of others by making no contribution and still enjoy the broadcast; if we fear that the contributions of others will not be enough to maintain this public good, we can avoid "throwing away" our own money by withholding it. In fact, no matter how many others contribute, our potential contribution has greater utility for us if it is spent on private consumption rather than on public broadcasting; hence not contributing is a *dominating strategy*. But if all listeners withhold all contributions, the demise of public broadcasting follows, a *deficient equilibrium* for all who enjoy it. The payoff for each individual is higher for not contributing than for contributing, but if each individual follows the dominating strategy, the outcome is one that no individual desires. Choosing the dominating strategy by withholding one's contribution is known as *defection;* accepting a lower individual payoff to sustain public broadcasting is called *cooperation*.

Fossil fuel shortages, overpopulation, "brown-outs," air pollution, and the depletion of fresh water supplies all represent social dilemmas because each individual's payoffs for choosing defection are higher than those for choosing cooperation, no matter what others choose; however, all individuals receive lower payoffs under universal defection than under universal cooperation. Cooperation can be difficult to obtain in social dilemmas because individual decisions typically have a low impact on the collective outcome, especially when the group is large. Also, gains associated with self-interest accrue to the individual whereas the costs – inconveniences, declines in the quality of life, or depletion of resources – are shared by all. Sociobiology's "central problem" (Wilson 1975) can also be framed as a social dilemma: How can cooperation evolve when the genetic payoffs for defection are higher than for cooperation?

Stated in game theoretic terms[3] (see Appendix and Glossary), a social dilemma is a structure in which dominating strategies converge on a deficient equilibrium (Dawes 1980). The outcome of choosing the dominating strategies is "deficient" because some other outcome is preferred by the choosers; it is an "equilibrium" because no group member receives a higher payoff for picking some other option if indeed others choose their dominating strategies. People are assumed to be (1) *rational* decision makers (i.e., they are capable of assessing and acting upon their perceived interests and preferences), and (2) *selfish* maximizers (i.e., only their own costs and benefits enter their decision calculations). Rational choice involves determining and comparing (1) the various outcomes that could result from each alternative course of action and (2) the probability that each outcome will subsequently occur – and then choosing on the basis of personal values (called *utilities*). Selfishness is defined as not taking the consequences for the group into account in this calculation. (Shubik [1964] describes the formal assumptions of "economic man" theories, and Schwartz [1986] describes how they are involved in everyday life.)

According to EI theory, people will always choose the selfish strategy in social dilemmas. The only way to avoid the resulting deficient equilibrium, then, is to embed the dilemma in a larger context involving utilities that make the dominating strategies no longer dominating. These utilities assessed by economic man can be external, such as avoiding sanction or obtaining a direct payment, or internal, such as a positive psychological payoff for "doing the right thing" or "having a clear conscience," irrespective of consequence. Some authors have proposed that these measures of utility can also be more indirect; for example, it would be to the benefit of the current cooperator to have others observe the cooperative act and reciprocate with future acts of their own (Alexander 1987; Campbell 1986).

Social dilemmas are not "trust dilemmas," in which people receive higher individual rewards for choosing a socially beneficial alternative only if others do (as occurs in some business partnerships, for example). The conflict between individual payoffs and group payoffs is maximal in social dilemmas because individuals *always* receive higher rewards for choosing the (dominating) option leading to the deficient equilibrium that no one desires. It should also be noted that social dilemmas are defined in terms of available choice options and payoffs – not in terms of what people actually choose. It is always possible to hypothesize arbitrary egoistic payoffs post hoc whenever an individual is observed to choose what does not appear to be the dominating strategy, or one can simply claim that some such payoffs must exist. It is also possible to maintain that an apparently nonegoistic incentive that motivates behavior must by definition be "selfish," or that choosing a dominated option is impossible, again by definition, because the very fact that it was chosen means the apparently dominating option cannot be dominating. This last argument is known as *revealed preference* (see Glossary) –the term is a double entendre – because the observer claims "revealed" knowledge of an internal cause of behavior (i.e., self-interest) and because behavior "reveals" the actor's utilities (i.e., the matter of self-interest and its subjective value to the actor). (The difficulties of revealed preference are discussed in section 1.4.)

The economic-man prediction for behavior in social dilemmas is that people will neither contribute their own resources nor will they practice restraint in the use of freely available ones. Everyday observation, however, quickly reveals that some people do cooperate in social dilemmas, at least some of the time. EI explanations for cooperative behavior differ only superficially: Theorists who accept the selfish incentive paradigm must transform the social dilemma into a nondilemma. If an inducement or penalty is imposed or developed, the utility of the dominated cooperative behavior is thereby changed so that the consequence of the cooperative choice becomes individually beneficial. Because the cooperative choice is beneficial for the group, these sanctions would explain how selfish behavior is compatible with beneficial group outcomes.

Responses to social dilemmas are ideal for studying EI theories because hypotheses about the necessity of incentive changes are testable. A hypothesized incentive for cooperation would be falsified, for example, if cooperation did not decline when the putative incentive was removed. Similarly, a pattern of findings that was inconsistent with EI theory over a number of experiments would be grounds for doubting its usefulness as a general explanation for human cooperation even if a post hoc EI explanation could be separately generated for each experiment. It is important to test EI theories not only because they are ubiquitous, but because they can have important, and sometimes draconian, implications for social organization (see below). We will now briefly describe incentive changes that have been suggested by prominent EI theorists to account for cooperation.

1.1. Coercion. Thomas Hobbes's (1651/1958) classical justification for absolute government (Leviathan) is based on his analysis of the social dilemma structure of life in its absence. When the self-seeking propensity of mankind is constrained only by others' capacity to defend themselves, where the "covenants without a sword are but words, and of no strength to secure a man at all," the result is the "warre of all against all" (a dominating strategy). This war is the worst outcome for everyone (a deficient equilibrium from which individuals cannot remove themselves). It can (and should) be averted by a government strong enough to make peaceable action the dominant strategy. By coercing people to cooperate, Leviathan changes the incentive structure to make cooperation compatible with egoistic incentives. Hobbes provides not only an evolutionary and ethical justification for strong, coercive state power, but a psychological one as well. People, at least vaguely aware of their own impulses to defect and the resulting individual loss that the collective expression of these impulses would mean, will accept the existence of this power – even when it is used to the selfish advantage of those wielding it.

More recently, Hardin (1968) has rekindled interest in coercion as a strategy for inducing cooperation. He has suggested that exploiting the environment, polluting it, and having children are all dominating strategies, particularly in developing countries. Modern technology and population growth have made these strategies converge on deficient equilibria: pollution, overexploitation of the environment, and overpopulation. Hardin uses the analogy of overgrazing a common pasture: All herdsmen have a dominating strategy to add cattle because their share of the destruction produced by an additional animal costs them less than the profit they accrue from the animal. Yet if all herdsmen choose this strategy, the common resource is destroyed. Hardin's proposal is "mutual coercion mutually agreed upon." Although the process (mutual consent) differs from Leviathan, the result is the same; cooperation becomes compatible with egoistic incentives – provided, of course, that the sanctions enforcing it make the probable payoff of behaving uncooperatively negative.

1.2. Reciprocity. Before Hardin's paper Anatol Rapoport (Rapoport & Chammah 1965) conducted experiments on *prisoner's dilemma* games (see Appendix) as analogues of the nuclear arms race. These games have the same structure as social dilemmas except that the game has two players (i.e., defection is dominant, but the strategy leads to an outcome neither player desires). Rapoport considered the arms race to be a prototypical prisoner's dilemma because both sides appear to have a dominating strategy to arm even though both would prefer disarmament and prosperity rather than the resulting escalation of arms production and concentration of resources in the military sector simply to achieve stalemate. Creating prisoner's dilemmas through payoffs in laboratory situations allowed Rapoport and others to investigate how success in avoiding the deficient equilibrium across repeated plays varied with players' overall choice strategy, forms of communication, personality characteristics, and even their gender.

As early as 1967, Amnon Rapoport demonstrated mathematically that the defecting choice in a prisoner's dilemma may no longer dominate when that choice is embedded in a metagame created by repeated (iterated) plays in which reciprocity – or even its possibility – can make cooperation nondominated on particular plays. In iterated dilemmas, players are faced with a very complicated problem of behavioral control in which they try to shape the other players' behavior to yield cooperation, while simultaneously being rewarded or punished themselves. Players may reward cooperative choices by subsequently cooperating and respond to defection with defection (a strategy called "tit-for-tat"). This is analogous to the situation faced by a parent who must decide whether or not to forego a pleasant interaction with a child in the hope of changing future behavior by punishing a past transgression. Like coercion, reciprocity is hypothesized to "work" by making cooperation a selfish choice.

Empirical results suggest that cooperation can result from egoistic incentives because reciprocity exists; individuals making a choice have at least an implicit understanding of this reciprocity and its potential benefits. The strongest and most replicable finding is that subjects who use a tit-for-tat strategy in two-person prisoner's dilemmas attain the best overall outcomes across trials. This strategy consists of cooperating on the first play and then simply matching the opponent's prior choice on successive plays. Thus, cooperation is rewarded with cooperation and defection is punished with defection.[4] Axelrod (1984) replicated these results with two computer tournaments in which various experts in cooperation and conflict submitted overall ("metagame") strategies as computer programs for playing iterated prisoners' dilemmas

against one another. In both tournaments, tit-for-tat garnered the most points.

An interesting elimination of a dilemma through reciprocity was achieved at the height of the nineteenth-century American "cult of domesticity." Catherine Beecher (1847/1972) observed that, "in order that each individual may pursue and secure the highest degree of happiness within his reach, unimpeded by the selfishness of others, a system of laws must be established, which sustain certain relations and dependencies in social and civil life" (pp. 171–72). Beecher defended a system involving the "duties of subordination," which women would undertake. Because the sexes were equal, however, women would undertake their duties only when men's regard for women's best interest demanded it. In compensation for their subordination, women would be treated, by custom and courtesy, as superiors. The assumption that women were better suited to voluntary subordination was founded partially on the belief that women were more capable than men of acting in the absence of egoistic incentives; providence has provided them with a selfless and superior moral nature.[5]

1.3. Conscience. One of the goals of socialization is to take a child "brought into the world with apparently no other thought than its own gratification" (Rushton 1980, p. viii) and teach it concern for others. Campbell (1975) proposes that abstract cultural ideals such as fairness, equity, sharing, and so forth, are instilled in individuals as "conscience": general rules and customs that guide interactions with others. The Freudian version is the superego. Innate selfishness is opposed by internal rewards and punishments that have a higher utility than external ones. In our view, it makes no difference whether a choice payoff is external (such as avoiding a sanction or obtaining rewards through reciprocity in the future) or internal (as in having a clear conscience, heightened self-esteem, or the avoidance of guilt). An individual whose decision can be traced to a positive psychological payoff is acting on the basis of egoistic incentive.

1.4. Inclusive fitness maximization. According to sociobiologists, "altruism" includes not only genetically-based self-sacrifice of genes (the usual biological usage), but also any expenditure of resources, energy, and time that may benefit another organism (Barash 1977). Hence, they try to explain social behavior with the same principles that govern genetic self-sacrifice. According to Alexander (1979), all human behavior, whether categorized as "selfish individualism" or "group altruism," can be explained in terms of striving to maximize the survival by reproduction of one's own genes or copies of one's genes possessed by relatives: That is, the ultimate goal of behavior is to maximize inclusive fitness. (Note that "survival by reproduction" confounds the *survival* of the individual organism and the *reproduction* of the genes, thereby collapsing genetic and individual levels of analyses.) Thus, the solution to sociobiology's "central problem" is to transform the dilemma of altruism into a nondilemma by hypothesizing egoistic incentives for "selfish" genes as an explanation for cooperation. Instances of apparent cooperation among unrelated individuals are best interpreted as the aggregate outcome of individuals pursuing their own (gene-prompted) self-interest (Alexander 1974; Trivers 1971; 1985; Williams 1966).

The sociobiology of choice rests on the revealed preference approach (described in section 1 and Appendix), which also transforms dilemmas into nondilemmas. Even apparently altruistic choices are assumed to reveal underlying utilities to the genes, with people choosing only what maximizes genetic fitness (Chagnon & Irons 1979). The corollary to EI theory, that people are rational, may also be given an explicit genetic justification. Alexander (1979), for example, states that "the 'rational self-interest' of 'utilitarian' explanations of human behavior is really the combination of the genes and chromosomes in perpetuating themselves" (p. 70). This approach is circular, but it does imply the testable claim that humans routinely make rationally self-interested decisions. If choices reveal "underlying" utilities, then they must be consistent, transitive, and independent of the way they are elicited. In reality, however, inconsistencies, intransitivities, and systematic "preference reversals" occur that depend, for example, on whether one's choice is based on comparing commodities directly or on using a system of common currency (see Grether & Plott 1979; Lichtenstein & Slovic 1971).[6] Another example of inconsistency is the following: The finding that people seek sure gains and avoid sure losses leads to predictable preference reversals when the same problem is framed in different ways – a credit card surcharge versus a cash discount, or, more important, lives lost (e.g., 400 of 600) versus lives saved (e.g., 200 of 600; Tversky & Kahneman 1981). As with economic man theories, genetic reductionist accounts can neither accommodate nor dismiss such systematic deviations from rationality.

Although cultural evolutionists also assign a key role to egoistic incentives, they can avoid the shortcomings of revealed preference and the "adaptationist program" by positing intervening selective processes. Boyd and Richerson (1985) explicitly use the social dilemma framework to explore the conflict between self-interest and cooperation. In their view, a theory of cooperation must explain how a tendency to acquire self-sacrificial beliefs could evolve and why altruistic behavior is directed toward some individuals and not others. They propose "conformist transmission" of cultural beliefs and behavior as an explanation. According to Boyd and Richerson, people use a simple rule to judge what behaviors or beliefs to imitate: The more frequent a trait, the greater the value of imitating it. In general, the advantage of conformist transmission is that it provides an easily used "rule of thumb" that increases the probability of acquiring values, goals, beliefs, and behavior favored in the local population. "Excessive cooperation" from the genetic point of view may simply be a byproduct of an otherwise adaptive strategy for judging what behavior and cultural values to imitate.

1.5. Summary. EI theories are variants of a metatheory that assumes, but does not demonstrate, that the only way to "resolve" a social dilemma is to embed it in a larger context in which it is no longer a dilemma. Cooperation is explained as a byproduct of the individualistic pursuit of some goal other than collective success. Goals may be variously identified in the social and behavioral sciences, but as in folk psychology, the organizing principle is that

people try to maximize their chances of obtaining outcomes of purely personal value (Kitcher 1985). Cultural materialism explains the exotic customs of nonindustrial cultures as the "unknowing" cost/benefit calculations of production and reproduction (Harris 1979); psychoanalysis views mental life as being organized around unconscious pleasures and pains (Freud 1900/1965), and political economics posits government for the self-interest of politicians (Downs 1957); psychology stands accused of sanctioning selfishness (Wallach & Wallach 1983), and sociobiology offers the ultimate causal explanation for self-interest – the genes (Alexander 1987). All these theories explain social interaction in terms of incentives that are present in society, but that would be just as highly valued by an isolated nonsocial animal as by a social one.

Experimental studies and computer simulations (such as those described above) have generally supported EI theories by showing that such incentives "work"; however, these variables are manipulated without testing what happens in their absence. Most empirical investigations are of two-person games played repeatedly with the same other player (or presumed player – often in reality a computer program). A cooperative subject may therefore expect to gain more points or money "in the long run" than an uncooperative one, provided that cooperation is reciprocated. (A true dilemma, in contrast, is defined by a dominating strategy of defection, not just one that appears to be dominating because of a short-sighted failure to realize that cooperation yields a higher payoff in a broader context.) Given the existence of reciprocal altruism in most cultures – a fact we do not deny – a player who cooperates in iterated play, especially in a tit-for-tat manner, will garner higher payoffs in the long run than a player who defects; however, the existence of these payoffs does not imply that the factor responsible for such cooperation is the expectation of receiving them. (In our culture, airplane crashes yield monetary payoffs to the survivors and the relatives of those killed with even greater certainty than cooperation yields reciprocity – which implies nothing about the factors responsible for such crashes.) What must be demonstrated is a *lack* of cooperation when the game is not repeated. Moreover, when people are identifiable, their expectation of reciprocal altruism in other contexts may generalize to yield cooperation even when the game is not iterated; consequently, the only two-person games relevant to our concerns are those played once in total anonymity.

Such games have been studied (Rapoport et al. 1976), but other factors have made their interpretation unclear. First, subjects are typically told to make as many points as they can; thus they are *instructed* to seek personal benefit, which at best confronts potentially cooperative players with a choice between cooperating with the other player or with the experimenter. Worse yet, the experimenter may offer a monetary prize to the player or players who "earn" the most points, thereby making the pursuit of joint gain a logical impossibility. Finally, to avoid the "inefficiency" of paying subjects for making a single choice in just one game (Rapoport et al. 1976, p. 90), subjects are often asked to play many different games, though not repeating a particular one. As Rapoport later points out, however, "when the subject's task is to make a long series of choices, and where the structure of the decision problem varies from one [problem] to the next, the subject may adopt an 'across the board' strategy" (1983, p. 377) – thereby making an understanding of the playing of a particular game impossible.

Deutsch's (1961a; 1961b) and Wrightsman's (1966) are the only studies we have been able to find in which subjects make an anonymous single choice without ancillary instructions from the experimenter about what to maximize in order to succeed at the task. The Deutsch studies used imaginary monetary payoffs and the Wrightsman study used real money in some conditions and imaginary money in others. Both these investigators, however, studied the role of personality and attitudinal variables in a single social context rather than the effects of different contexts, which is the emphasis of the experiments reported here.

2. Cooperation in the absence of egoistic incentives

In this section, we report a series of experiments testing the assumption that people will not promote group welfare in the absence of egoistic incentives. All the experiments were monetary public-goods dilemmas in which subjects were given a sum of money and chose anonymously to keep it or contribute it. The general research strategy was to eliminate EI explanations for cooperation by using a "subtractive procedure." Some incentives (e.g., the opportunity to make deals to redistribute the money) were obviated by the experimental design; others were tested directly in the experiments. It is important to note that these laboratory games are not simulations of naturally occurring social dilemmas; they *are* social dilemmas.

The experiments test the basic EI prediction that public goods will be underprovided relative to demand, if they are provided at all. Consider the following illustrative "game" for example: Nine strangers are each given $5. If five or more contribute their money, all nine will receive a $10 bonus whether or not they contributed. If enough members contribute, the net payoff for the contributor would be $10, for a noncontributor $15. If too few contributed, a contributor would lose the contribution, receiving nothing at all, whereas a noncontributor would retain the $5. The game theoretic strategy is to retain the money, but if this strategy is selected by all the members of the group, it leads to a deficient equilibrium: All receive $5 when all could have received $10 by contributing. This game illustrates the kind of decision confronted by people in all of the experiments to be described. Subjects who participate in groups are given money and then required to make a choice between retaining the money or contributing it to a "public good."[7]

The general paradigm was the same for all the experiments, which were conducted in Oregon, Utah, and Arizona over a period of 10 years. More than 1,000 people participated. Subjects were students and townspeople recruited through advertisements in campus and local newspapers.[8] Subjects were scheduled at their convenience to participate in specific groups, which were randomly allocated to experimental conditions. Friends or relatives were prevented from participating in the

same groups (although ultimately the experimenters had to rely on the subjects' honesty in complying with this). When they arrived for the experiment and were seated, subjects were given a promissory note for the same amount of money (usually $4 to $6) that they later had to decide to retain or give to the group. The specific consequences of "giving" or "keeping" the money were explained in some detail, including a "payoff matrix" that specified the monetary consequences for each person who gave or kept as a function of the total number of people in the group who gave. The word "contribute" was avoided; the words used were consistently "give" and "keep." All the experiments involved a single decision to keep or give to the group the full amount of the money in the promissory note. The experimenter read instructions to participants, explained the payoff matrix and answered questions; subjects were subsequently given a short quiz to test their understanding of the instructions and the payoff matrix. (Any failures in this test led to more explanation. The choices were not made until the experimenters were satisfied that all subjects understood the payoff structures.) The experimenter repeatedly emphasized that all decisions were to be made anonymously. If the decision was to "keep," subjects would get the amount of money on the note at the end of the experiment plus the bonus, if the public good was provided. If subjects chose to "give," the payoff matrix specified the payoffs they and the others in the experiment would receive. (The specific consequences, of course, depended on the experiment and condition.)

Subjects were not allowed to talk among themselves except in experiments involving one or more "discussion" conditions. Subjects made their decisions secretly by marking a box on the decision form that could not be seen by other subjects. After the experiment was completed, subjects were released one at a time to the "payroom" where they completed a questionnaire, were told the group outcome, and were paid. Only one subject was ever in the payroom at a time. This procedure was explained to all subjects before they made a decision to give or keep.

To summarize: Subjects were strangers (to the best of the experimenters' knowledge), they made a single decision, their choices were anonymous, and interaction among group members was prevented before and after the session. The prohibition on interaction and the provisions for anonymity were designed to prevent: (a) reciprocity (during the experiment by allowing more than one decision, or after the experiment by splitting the money); (b) coercion (e.g., interpersonal abuse or sanctions); or (c) reputational concerns (which might, for example, lead a subject to hypothesize that giving or keeping would affect others observing this decision to respond in the future – perhaps outside the experiment – to the selfish benefit or detriment of the subject). Thus, the paradigmatic design specifically excluded the "classic" egoistic incentives invoked to explain cooperation.

2.1. Greed, fear, and the pursuit of self-interest. Beginning their research as EI theorists, Dawes et al (1986) designed their initial experiments to distinguish between two selfish motives (Coombs 1973): "Greed" to receive the good without contributing ("free riding") versus "fear" that one may lose money in a futile support of the public good (the "sucker's payoff"). These experimenters conducted a series of "step-level" public-goods games of the type just described; a specified number of contributions ("give" choices, e.g., "5 or more") of $5 supplied to the subjects in the form of a promissory note were required for the experimenters to provide a monetary bonus of $10 to all. That was the "standard" condition in the experiment. In a no-free-ride condition, the opportunity to free-ride was eliminated by truncating the amount paid to subjects at $10 in the event the public good was provided. The experimenters explained this modified structure, complete with the payoff matrix, in the same way they explained the standard structure. If greed is a motive for defecting (not contributing), eliminating the incentive for it should enhance contributing. Contributors might still lose the value of their promissory note, but if the public good was provided, a noncontributor would make no more money than a contributor. In contrast, a money-back guarantee condition eliminated the possibility of futile contributions. The experimenters explained that contributing subjects would keep the $5 they gave to the group if too few subjects overall gave to yield the $10 bonus. Although this condition eliminated the fear of losing their $5 as the result of giving it to the group, subjects could still make $15 rather than $10 by keeping it whenever enough others gave their $5 to yield the $10 bonus. Table 1 illustrates payoffs for the three conditions; it is similar to the payoff matrices presented to the subjects.

Subjects were tested in groups of seven requiring either three or five contributors to obtain the public good. Over three replications, the rate of contributing averaged 51% in the standard condition; 58% in the money-back guarantee condition, and 87% in the no-free-ride condition. (The number of groups in each condition was, respectively, 25, 19, and 25.) The authors concluded that the major motive for not contributing is greed rather than fear of futile contributing.

2.2. The rational pursuit of self-interest. Because step-level games require a specified number of contributors for the public good to be provided, individuals might believe that their contribution was critical. For example, if *exactly* four others have contributed while five or more contributions are necessary, then subjects who contribute their $5 receive the $10 bonus, which is a better outcome than retaining the $5. Contributing is clearly

Table 1. *Payoff matrices for standard dilemma, money-back guarantee, and no-free-ride conditions*

	Less than 5 contributions	5 or more contributions
Standard dilemma		
Contributors	0	10
Noncontributors	5	15
Money-back guarantee		
Contributors	5	10
Noncontributors	5	15
No free ride		
Contributors	0	10
Noncontributors	5	10

rational and consistent with egoistic motivations if such an outcome is likely. It is irrational, however, to believe that one's own contribution is critical if each other subject is assumed to have a constant probability, p, of contributing, because in groups requiring three or five contributors the probability that exactly two of the six others would contribute $[15\,p^2\,(1-p)^4]$ or that exactly four of the six others would do so $[15\,p^4\,(1-p)^2]$ can never be greater than .50 regardless of the value of p. Nevertheless, subjects may be poor estimators of binomial probability distributions. Amnon Rapoport (1967), for example, performed similar experiments and calculated the expected values for contributing in the three conditions based on subjects' *estimates* of three probabilities: (i) the probability that their own contribution would be futile (i.e., that so few others would contribute that one's own contribution would not provide enough to provide the bonus); (ii) the probability that a contribution would be critical, and (iii) the probability that a contribution would be redundant (i.e., that enough others would contribute enough to make their contribution unnecessary). Thus, in a subsequent replication involving the standard, money-back guarantee and no-free-ride conditions in the previous experimental design (a minimum of five contributions of the $5 from seven players required to yield the $10 bonus), the subjects were also asked to indicate on a 100-point scale the likelihood that: (i) fewer than four others would contribute; (ii) exactly four others would contribute; and (iii) five others or more would contribute. Subjects who were maximizing expected value in the standard condition, for example, would contribute only if their estimate of its being critical was greater than .50.

Table 2 shows the average values obtained for the probability estimates. Subjects should cooperate in the three conditions if their estimate of their contribution's being critical, on the average, was large enough so that contributing would have a higher expected value than withholding. (The actual requisite values are given in the table; note that this probability varies by condition; for example, with the money-back guarantee, it is derived by considering only the probability of contributing being critical or redundant, because futile contributions are returned.) Probability estimates did vary significantly by condition, but the proportion of subjects who judged their contribution to be critical did not. In both the no-free-ride and the money-back-guarantee conditions, the perceived probability of a contribution's being redundant is enhanced and of its being futile is diminished by a ratio of 2 : 1. Some directional support for the hypothesis that subjects ascertain and pursue their self-interest in accordance with the expectation model was obtained in the no-free-ride condition only; the non-contributors' estimate of futility was .51 compared to the contributors' estimate of .20. Virtually no contributors, however, believed that their contribution was critical to obtaining the public good with a probability sufficient to justify the "give" decision. In fact, pooling across all conditions, 67% of the contributors believed that so many others would contribute that their own contributions would be redundant.

Table 2. *Predicted and obtained probability estimates of making futile, critical, and redundant contributions*

Experimental condition (expected value) and choice	Subjects' estimated probability that their contribution would be: Futile	Critical	Redundant
By condition			
Standard dilemma (.50)[a]	.47	.21	.32
Money-back guarantee (.37)	.27	.25	.48
No free ride (.27)	.27	.24	.50
By choice and condition			
Standard dilemma			
Contributors (N = 9)	.24	.29	.46
Noncontributors (N = 26)	.55	.18	.26
Money-back guarantee			
Contributors (N = 15)	.25	.29	.46
Noncontributors (N = 20)	.28	.21	.50
No free ride			
Contributors (N = 27)	.20	.23	.57
Noncontributors (N = 8)	.51	.25	.24

[a]Subjects maximizing expected value should cooperate if and only if their probability estimate for their cooperation's being critical is greater than the value indicated in parentheses.

2.3. The reasonable pursuit of self-interest. The foregoing results could still be interpreted according to EI theory. If potential gains are subjectively more valuable than potential losses, a contribution's being critical may not be important and people might be willing to contribute despite their estimates of probable outcomes. After all, even if a contribution is redundant, a contributor will gain $10 rather than $5 if the bonus is provided.

To see whether contributing could be enhanced when subjects had the opportunity to "educate" each other, van de Kragt et al. (1983) tested 12 groups in new conditions in which subjects could communicate with each other. What happened was that every group used discussion to make a decision about *who* would and would not contribute. In 10 groups, exactly the minimal number of contributors was designated (2 groups specified more than the minimum number), everyone so designated contributed, and there was 100% provision of the public good, with overprovision in 3 of the 10 groups (i.e., people who were not supposed to contribute did so anyway). The most common means of designating who was to contribute was a lottery, followed by volunteering; one group attempted to establish which group members were the neediest. Subjects designated as contributors, knowing that only the minimal number was specified, knew that they could not receive the bonus unless they contributed. The fact that all such designated subjects contributed appears to indicate that a contribution's being critical, which had played little role in the expectation analysis, was a powerful motive to contribute.

Subjects' expectations about the number of contributions from others in their groups are consistent with the idea that discussion reduces uncertainty and perceived risk; the pooled variance of expected contributions was 11 times as high in the previous no-discussion conditions as it was in this new discussion condition. Somewhat unex-

pectedly, however, subjects in the no-discussion groups appeared to tailor their expectations to the requirements for providing the bonus; the mean number of expected contributors was always higher when five contributions were required than when three were required. This suggests that subjects either believe that "the group" will tend to its own welfare, or that they are focusing more on incentives confronting the group as a whole than on individual incentives.

As long as one knows that one's own contribution is necessary for obtaining the bonus and one assumes that other designated contributors are equally aware that their contributions too are necessary, self-interest will promote contributing without any help from conscience, normative considerations, or group identity. Again, greed is impossible, as in the earlier no-free-ride conditions. This interpretation, however, is not supported by the explanations the subjects themselves provide for their behavior. A research assistant coded into six categories the explanations subjects gave for their choices in the discussion conditions in which the contributors were designated.[9] Seventy percent of the contributors and 90% of the noncontributors (who were designated not to contribute) attributed their choice to "the group's decision." For comparison purposes, the reasons for contributing in the previous money-back guarantee and standard conditions were collapsed (because they did not produce significant differences). In these conditions, the most common explanations for contributing involved normative constraints, such as appeals to concepts of justice, charity, and equity (40%); selfish considerations such as wanting to make the most money or "looking out for number one" (26%); and, finally, willingness to take a risk (17%). In contrast, comparable explanations in the discussion condition were 2%, 5%, and 0%. Thus, following group discussion, contributors did not explain their behavior in terms of normative constraint or a reasonable expectation of an egoistic payoff. The most common explanation for not contributing in the conditions without discussion was private payoff (80%), followed by unwillingness to take a risk (17%).

2.4. Designating contributors. Two variables were confounded in the foregoing experiment: the decision to designate subjects as contributors and the opportunity to engage in discussion, which has been found to raise the rate of contributing in similar games (e.g., Dawes et al. 1977; Jerdee & Rosen 1974; Rapoport 1974). The reason for such discussion effects is unclear, although the usual assumption is that discussion triggers concern for group welfare or feelings of obligation. Discussion in the experiments just described may simply have served to reinforce how critical the choices of the designated contributors were – both to their own receiving of the bonus, and to others' getting it. If this factor is the basis for the perfect cooperation observed, then *any* method that convinces potential contributors that their contribution is necessary for obtaining the bonus – and assures them that other group members likewise recognize their own contributions to be critical – should yield almost perfect cooperation. Although the explanations of the subjects did not support this interpretation, it needed to be investigated experimentally because subjects do not always have introspective insight into the actual factors influencing their choices (cf., Nisbett & Wilson 1977). Perhaps the subjects were in reality motivated by egoistic incentives, but simply cited nonegoistic factors in explaining their behavior to the experimenters.

Van de Kragt et al. (1986) conducted an experiment with five conditions to test the above possibilities. In all conditions, the general procedures were the same as those described earlier, except that there were nine people in each group, of whom five or more had to contribute if everyone was to receive the $10 bonus. In two conditions a set of contributors was designated by having subjects draw a chip from a bag containing five blue poker chips and four white ones. Blue chip holders were designated as potential contributors who had to make a choice between contributing their $5 promissory notes to the public good (the $10 bonus) or retaining it for themselves. White chip holders had no choice to make; they simply retained their $5 promissory notes and received the additional $10 bonus if all blue chip holders contributed, otherwise not. Thus, if all blue chip holders contributed their note for the $5, a decision made privately and anonymously, a $10 bonus would be distributed to all group members, white chip and blue chip holders alike. In one condition, a discussion followed the drawing; in the other condition, decisions were made immediately after the experimenter was satisfied that all participants understood the game.

There were also two conditions in which there was no designation of which subjects could contribute. One was a replication of the standard conditions in the previous experiments. Another condition had the same structure, but subjects were permitted up to 10 minutes to discuss the problem. This discussion period took place before subjects knew the number of contributions necessary for producing the $10 bonus; they were told that after discussion this number would be from a paper bag with numbered poker chips drawn by a randomly selected subject. The lottery was rigged (an exception to the care usually taken to avoid deception) so that the number five was always drawn. The reason for not making this number known to the subjects prior to discussion was that the previous experiments had indicated that the subjects would designate five among themselves to be contributors had they known that five contributions were necessary. (The experimenter monitored discussion to prevent subjects from using a variant of this procedure.)

In a fifth condition (labeled "super simple") there were five subjects in each group; all were required to contribute to obtain the public good. In this condition, the contributions of everyone in the group were critical for receiving the bonus. Hence, the number of such critical people was the same as in the other conditions, but all members confronted the same problem. All were aware that everyone knew that each was critical; the problem was, accordingly, very simple.

The results are presented in Table 3. As indicated in that table, designating a group of people whose contribution is critical enhances rates of contributing, but not enough to yield the $10 bonus more often than in the standard condition. Unexpectedly, the rate of contributing in the super-simple condition, where self-interest was so clearly contingent on contributing fell far short of the 84% to 100% range anticipated; and in no instance was the public good provided. We will suggest an explanation for

Table 3. *Rates of contributing and providing the public good in designated set experiments*

Condition	Rate of contributing	Rate of providing the public good (provided/total groups)[a]
Designated sets[b]		
Super simple	64%	0% (0/5)
No discussion	84%	54% (6/11)
Discussion	100%	100% (11/11)
No designated sets		
No discussion	47%	60% (6/10)
Discussion	84%	100% (10/10)

[a]Five of nine contributors required for public-good provision.
[b]The percentage of designated contributors (blue chip holders) is reported in these conditions.

this finding later. The finding for the super-simple condition does eliminate the subjects' understanding (of the logic of being critical to obtaining the bonus) as an explanation for the facilitating role of discussion. If this understanding accounted for the effects of discussion, the super-simple game would yield near perfect cooperation, which does not occur.

As indicated in Table 3, the rate of contributing among designated contributors in the discussion condition was 100%; the public good was hence always provided – a replication of the results of the first study where subjects themselves designated contributors. But discussion in and of itself resulted in contributing at the same level as did designating who should contribute without discussion (84%); and discussion led to the provision of the $10 public good in every group.[10] Finally, if it were important that the subjects in the groups themselves – rather than the experiments – designate the contributors, the rate of contributing among such people should not be 100% when they are designated by lot. But it is.

2.5. The consistency between expectations and behavior. In these studies, the experimenters once more investigated subjects' ability to determine their best interests and to pursue them accordingly. Subjects were asked (after their decisions had been made) how many others in their group they expected would contribute. If one expects one's contribution to be critical, then contributing is rational and consistent with the expectation; similarly, if one expects one's contribution to be futile or redundant, withholding is rational and consistent. But "pessimists" (who expect fewer than the required number of people to contribute) or "optimists" (who expect the required number or more to contribute) who contribute anyway are economically irrational. We call the relationship between individuals' expectations about the number of other contributors and their actual contributing behavior *consistency with expectations*.

A widely held intuition is that expectations always precede behavior; people assess (perhaps on an intuitive basis) the likelihood of other group members' contributing and then they behave accordingly. In this view,

Table 4. *Consistency with expectations for contributing and noncontributing*

Condition	Contributors	Noncontributors
Designated sets		
Super simple	67%	89%
No discussion	74%	86%
Discussion	87%	[none]
No designated sets		
No discussion	21%	88%
Discussion	8%	72%

consistency with expectation should be relatively high and independent of the individual's own choice about whether or not to contribute. There is evidence, however, that people's expectations of others are based on how they themselves have behaved: That is, people use their own behavior as a cue in predicting the choices other group members will make or have already made privately (Dawes et al. 1977; Messe' & Sivacek 1979). If so, then contributors would expect higher rates of contributing than would noncontributors – even if this higher rate meant that their contributions would be unnecessary.

Table 4 shows the percentage of subjects whose behavior was consistent with their expectations for contributing and non-contributing subjects in each of the five conditions. When there was no designation of who should contribute, 84% of contributors were *in*consistent. Among these inconsistent contributors, 90% in the no-discussion condition and 96% in the discussion condition believed their contributions would be redundant. In contrast, the behavior of non-contributors was consistent with their expectations in all conditions; 8 of 10 believed that their contributions would be unnecessary. Hence their behavior – but *not* that of contributors – was consistent with their egoistic incentives as they judged them to be.

2.6. Noncontingent contribution. The EI theorist might argue that the high rate of contributing in discussion conditions is still due to the private advantage that is contingent on group welfare, despite the inconsistencies between the behavior and expectations of the contributors. If the link between personal payoff and group payoff were broken, so that an individual's contribution to the group did not benefit the donor, an EI theorist would predict a substantial drop in the rate of contributing. To test this prediction, van de Kragt et al. (1986) designated five people in nine-person groups as contributors by lottery (as in the earlier experiments, by having five blue chips and four white chips drawn from a bag). Subjects were instructed, however, that whether or not any particular member of the group received the bonus depended on the behavior of the *others* in the group. White chip holders received the bonus if at least four blue chip holders contributed, blue chip holders if *all four other* blue chip holders contributed. Thus, each blue chip holder would receive $5 more by not contributing than by contributing no matter what the other blue chip holders did – $15 rather than $10 if they all contributed, $5 rather

than $0 if one or more of these others failed to contribute. Retaining the $5 is the dominating strategy.[11] There were three experimental conditions: one as described, with discussion, one without, and a baseline condition in which each subject received the $10 bonus if four or more of *any* of the remaining eight subjects contributed.

There was no difference in the rate of contributing between the subjects in the baseline condition (30%) and the blue chip holders in the no-discussion condition (28%), indicating that designating the contributing set has no effect if there is no contingency between one's own and others' outcomes. The difference between these rates of contributing and those in the standard dilemma (47%; see Table 3) suggests that some individuals recognize that contributing might serve their personal interest when payoff is contingent on their own choice. In the discussion condition, however, 96% (24 of 25 people) contributed. Thus, even when group interest provides no egoistic incentives, discussion yields essentially the same rate of contributing as it did when a subject received the bonus only if the group did; discussion has this effect even though the rate of contributing in the baseline noncontingent condition is lower than that in the standard condition.

2.7. Conscience. If "a good conscience is our only sure reward," the effect of discussion may be to make the demands of conscience salient. Humans are still selfish, so the argument runs, but internalized social norms provide internal rewards for behaviors that are desirable according to social teaching and punishments for those that are socially undesirable; this thereby provides egoistic incentives to cooperate (cf., "cooperation among genetic competitors" [Campbell 1975; 1982]). If this is so, individuals acting to satisfy the demands of conscience would be as likely to contribute to strangers in another group as to members of their own group. To test this proposition, Orbell et al. (1988) devised a "give-away" dilemma to compare contributions to one's own and other groups under conditions of discussion or no discussion. Groups of 14 subjects were randomly divided (again, by drawing poker chips) into two groups of 7 that met in separate rooms. All subjects were given a promissory note for $6 and told that if they chose to give away their note, 6 "other people in the experiment" would be given $2 each. (Subjects were presented with a payoff matrix as well.) How much – if anything – a contributor received depended on the contributions made by other group members, as did the amount a noncontributor received – which was always $8 more than a contributor received (because contributors neither retained their $6 nor received $2 from their own contribution). The situation is a social dilemma because each subject is $6 better off keeping the money, but if all give theirs away, all receive $12 – a better outcome than the sure $6 everyone can make by keeping the note.

In half of the groups, subjects were initially told that their contributions would benefit members in their own group; in the remaining groups, subjects were told the contributions would go to comparable members in the other group (and that a bonus for them would come from members of the other group as well). This initial belief condition was crossed with final belief at decision-making time. In half the groups of each type, a "switch for experimental purposes" in the identity of the beneficiaries was announced three minutes before the decision. For subjects initially believing that the beneficiaries would be their own group members, the actual benefit would go to the corresponding members in the other group. For those initially believing that the other group would be the beneficiaries, the actual benefit would go to members in their own group (see Table 5). In summary, initial belief (beneficiaries in own group versus other group) and final belief (likewise) were crossed to form four types of groups. Half the groups of each of these four types were allowed up to 10 minutes to discuss the problem, half not. Switching the locus of beneficiaries occurred only *after* discussion was finished. Eight 7-person groups were tested in each of the resulting eight conditions.

Table 5. *Rates of contributing in the beneficiary-switched experiment*

	Initial belief that beneficiaries of contributing are: Own group		Other group	
	Final belief that beneficiaries of contributing are: Own group	Other group	Own group	Other group
No discussion	38%	30%	45%	20%
Discussion	79%	59%	32%	29%

Table 5 shows that, contrary to the clear-conscience hypothesis, discussion does not enhance cooperation when beneficiaries are in the other group. It enhances contributing to the subjects' own group only. Further analysis also showed that the initial belief about the locus of benefit was more important than the actual locus. We interpret the results of the experiment as generally supportive of an *ingroup biasing effect* (Tajfel & Billig 1974), discussed in more detail in the following section.

2.8. Discussion. In the experiments described above, allowing subjects a period for discussion reliably elicited contributing. The last experiment indicates that discussion per se is not a completely adequate account for contributing in these social dilemmas. We believe that the best explanation for the effects of discussion is that it engages *ingroup bias*. This tendency to evaluate one's own group positively and to reward it preferentially has been extensively documented in cross-cultural field studies (Brewer & Campbell 1976) and laboratory experiments. Tajfel (1970; 1981; Tajfel & Billig 1974), for example, discovered the remarkable flexibility of the bias in experiments using what is termed a "minimal group situation." Subjects in these studies are randomly assigned group membership ostensibly on the basis of the outcome of a trivial judgment task, such as whether they overestimated or underestimated the number of dots in a display, or on the basis of a preference between two abstract painters. Without ever actually meeting a member of their own group or the other one, participants were

asked to make a choice from a distribution of paired payoffs for a member of each group identified only by a group name and an individual number. Subjects systematically distributed greater sums to those they supposed to be members of their own group than to members of the other group. Further research (reviewed in Brewer 1979) demonstrated that ingroup biasing is a remarkably robust and easily elicited psychological phenomenon, often involving total strangers combined into groups on totally superficial bases.

Tajfel and Turner (1979; Turner 1987; Turner & Giles 1981) argue that group behavior is mediated by a redefinition of the self in terms of shared category membership, or social identity. Social categorizations become internalized as self-images and as regulators of behavior. When social identity is salient, individuals react to themselves and others as exemplars of the group rather than as differentiated individuals. Thus, group identification rather than individual self-interest can be the basis for cooperative behavior. (Note that social identity theory is not an explanation for the ingroup bias; rather, it concerns the elicitation of the bias by cultural categories and social structures.)

Clearly, a group's discussion of the dilemma facing its members is a considerably more substantial basis for establishing group identity than estimating dots on a display or stating a preference for Klee or Kandinsky. Social identity, by reducing the distinctions between one's own welfare and that of others, explains our subjects' willingness to contribute in the face of expectations of their contribution's being redundant. The results of the "give-away" game (section 2.5) are consistent with Brewer's (1979) conclusion that ingroup biasing consists of increased favoritism toward members of one's own group rather than hostility toward members of other groups. This is also consistent with Brewer's later finding (1985; Brewer & Schneider 1987) that being in a small group can, under certain circumstances, mediate contributing to larger collective interests.

Finally, ingroup biasing may explain the dismal rate of contributing in our super-simple game, where self-interest was so clearly compatible with egoistic incentives for cooperation. Brewer (1981), summarizing the results of cross-cultural studies of ethnocentrism, argues that social identification forms the basis for the "depersonalized trust" characteristic of ethnic group identity. There is a universal tendency to evaluate members of one's own group as friendlier, more trustworthy, and more honest. If ingroup identity is the basis for cooperation, then subjects in the super-simple game may have lacked the trust essential for promoting group welfare even when it was so clearly compatible with egoistic motives. Without group discussion or even an identifiable "other group," the public good bonus of $10 was never provided in the super-simple game.

Assuming that a case can be made for the evolution of the ingroup bias, social identity theory illustrates how innate mechanisms as well as evidence of considerable flexibility can be accommodated within a Darwinian framework: The tendency to favor one's own group may be a fixed species characteristic, but the bias is engaged by drawing category boundaries that vary over time, size of social group, context, and the favorableness of the salient distinctions characterizing the group. In the following section, we suggest a general hypothesis about the evolution of human cognition that provides grounds for considering in-group biasing to be an innate tendency.

3. Sociality

Theorists have suggested a variety of conditions leading to sociality: intergroup warfare and cannibalism (Alexander 1971), big game hunting (Washburn & Lancaster 1968), and parental care (Lovejoy 1981), for example. These theorists envision characteristic hominid capacities (e.g., group cohesion, communicating ability, and intelligence) as preconditions for sociality; in contrast, extractive foraging (Parker & Gibson 1979) would not require such capacities. A solitary initial state is often also envisioned. In arguing that tool use evolved out of necessity, for example, Brace (1962, cited in Wolpoff 1987, p. 269) writes, "It would seem a weaponless biped trudging over the savanna with a ripe load of meat would be an exceedingly poor bet for survival." Even with a weapon, however (particularly with anything like the earliest stone tools), a single biped is a poor bet for survival.[12]

Despite the evidence that human ancestors were social, sociobiological models of cooperative behavior (e.g., Alexander 1987; Hamilton 1964; Trivers 1971) also describe ancestors who begin in isolation and evolve to a social state in order to achieve the same egoistic goals they had in their prior isolated state. Yet the catarrhines from which the hominid line derives (35 million years ago) were already social; paleontologists identify the key question to be about the kinds of social groups ancestral catarrhines lived in (Fleagle & Kay 1985). If the so-called first-family collection (Johanson & Edey 1981) can be taken to represent a single contemporaneous group, then bipedal, small-brained Pliocene hominids gathered in mixed sex and age groups at least 3 million years ago. Stone "choppers" were manufactured 1.8 million years ago in hominid campsites showing evidence of food collection, and the same tool technology persisted for about a million years. Our predecessor, *Homo erectus* (1.6 to .5 million years ago), used stone flake technology, penetrated the temperate zone, used fire, hunted cooperatively, and lived in small groups. According to Foley (1987, p. 10), "the high level of sociality represent[s] the starting point in any consideration of human evolution."

The precedence of sociality appears to be difficult to incorporate in evolutionary formulations, however. In spite of his injunction, Foley (1987, p. 223) defines "community" as the species hominids ate, those that ate hominids, and those with which hominids competed: His definition fails to include interactions with conspecifics in social groups as an important, or even identifiable, category. Gowlett's (1984) analysis of the mental abilities discernible from tool making is similarly constrained. There is no hint that the activity must be learned over time from other individuals who themselves mastered the skills under tutelage. Surely the "basic pattern of operation of the human mind" would include this ability as much as the capacity to construct chains of activity that extend in space and time (Gowlett 1984, p. 183).

Hypotheses abut innate mental abilities or tendencies require a basis in evolutionary ecology, a discipline that attempts to identify the problems humans faced when

they were evolving and the solutions to those problems shaped by selection processes (Foley 1984; 1987). Generally speaking, the problems have two sources: the type of animal (e.g., mammal, bipedal, and so forth) and the environment with which it interacts. The following account, which we believe to be consistent with the principles and existing research in evolutionary ecology, could doubtlessly be refined by specialists in the field. Two conditions of early hominid evolution indicate that group living was essential for individual survival. First, isolated nonarboreal individuals would be more vulnerable to predation than individuals in a group. Second, edible resources in the habitat were dispersed in "clumps" (e.g., an antelope carcass or fruit-bearing tree). Frequently, such resources were composed of more material than a single individual could use, and in the case of meat, would involve interspecies competition with other carnivores over carcasses (Potts 1984). Given the morphology and ecology of early hominids, the optimal foraging strategy is to search collectively for resources and share them (Kurland & Beckerman 1985). (Group foraging is optimal when the marginal costs of food sharing are offset by a higher "hit rate" for locating food resources. [See Houston & Macnamara: "A Framework for the Functional Analysis of Behavior" *BBS* 11(1) 1988.] Thus, the interface between the individual and the habitat is a group process: Group living buffers individuals from the risks of predation and starvation. The more effectively a group can forage, the greater the individual benefit of group membership. There are constraints on group size, however. The maximum number of individuals is limited by the carrying capacity of the environment. Similarly, because groups mediate individual survival, a certain minimum number of group members is required. Under such conditions, the small face-to-face group (15 to 30 individuals) becomes a primary locus of selection pressures (cf. Campbell 1983).

The morphology and ecology of hominids are the bases for the sociality hypothesis, which consists of two general propositions:

(1) To the extent that exploiting a habitat may be more efficient as a collective rather than an individual process, not only would more successful groups prevail, *but so also would individuals that were better adapted to group living*. Because a group mediates individual contact with the habitat and the number of niches in groups is limited, fitness should be correlated with the evolution of perceptual, affective, and cognitive mechanisms that support the development and maintenance of group membership (Caporael 1983). Competition within groups should have generated escalating selective pressures for mechanisms supporting increasingly complex social behavior – including the "competition for good will" (Williams 1966) and the detection of within-group "cheaters" (Trivers 1971). These escalating selective pressures may account for the evolution of increased brain size (cf., Trivers 1971), which shows a rapid directional trend typical of specialization for some specific feature of the environment (Eldredge 1985). This trend is puzzling in the case of a generalist species such as *Homo sapiens*.

Rozin (1976) posits hierarchically organized *adaptive specializations* initially "wired" to their specific functions and inaccessible to other systems in the brain. According to him, cognitive evolution is based on the increasing accessibility of older adaptive systems for solving problems in new domains. This notion of accessibility also applies to individual development where independent capacities may become more generally accessible (e.g., children initially understand that a number of marbles may be the same whether the marbles are lumped together or spread out, but they are unable to apply the same logic to identical volumes of water in narrow or wide vessels). Learning involves "bringing to consciousness some of the limited-access programs, the 'cognitive unconscious,' already in the head" (Rozin 1976, p. 246). Rozin's framework is the basis for a second proposition:

(2) Mechanisms evolved for sociality may be accessed by other adaptive specializations (most notably language) and generalized to problems in domains for which they were not adapted – including interactions with nonsocial objects (Humphrey 1976) and large-scale social organizations. For example, children, "primitive" peoples, and even sophisticated scientists may attribute human characteristics to inanimate objects. When prediction or control of a phenomenon is important, and *no* theory exists to explain its behavior, people may "default" to attributing human motives, intentions, and agency until a better explanation can be devised (Caporael 1987; cf. Sellars 1963). In the case of large-scale social organization, group identity is often achieved by words, songs, flags, or other symbols that engage the in-group bias. Achieving organizational goals usually requires subdivision into small groups. Military organization epitomizes both of these features: Participation can be induced by patriotic appeals, but combat is conducted by small face-to-face subgroups that may engage intense ingroup loyalty.

Rozin (1976) developed his theory of cognitive evolution in the context of comparative animal behavior. We believe that the ecological approach recommended by Foley (1984; 1987) – a focus on what were the problems human ancestors had to solve – could add another evolutionary dimension to psychological theorizing about contemporary human behavior (cf., Campbell 1982; 1983; 1986). Two related psychological experiments suggest how such hybrid evolutionary/psychological conjectures could be arrived at. Kramer and Brewer (1986) found that people were more likely to restrain their use of a common resource (taking points worth money from a common pool that was slowly replenished) when they identified with a superordinate group rather than one of its subgroups. They also investigated the effects of how the choice between cooperating (taking fewer points from the pool) and defecting (taking more points) was presented (Brewer & Kramer 1986). A social dilemma can be framed as either a "public-goods dilemma" or the equivalent "commons dilemma" (see Glossary). In a public-goods dilemma, the choice is whether or not to contribute resources in hand to promote the common good; in a commons dilemma, the choice is whether or not to restrain taking resources from the common good to promote individual welfare.[13] Brewer and Kramer found that when subjects identified with the superordinate group and were presented with a problem such as a commons dilemma, they restrained their use of resources regardless of group size. The same dilemma presented as a public-goods problem, however, was found to be sensitive to group size; people contributed significantly more in groups of 8 than in groups of 32.

These findings suggest the following hybrid evolution-

ary/psychological conjecture: In the habitats in which hominids evolved, the distribution of resources may have posed something like a public-goods problem rather than a commons-dilemma problem. The more frequent problem in the evolutionary context was to distribute available resources rather than to practice restraint in the use of resources. If food sources were dispersed but profitable, optimal group size depended on the probability of encountering a given resource clump and on the return per individual (Kurland & Beckerman 1985). Thus, lacking "evolutionary experience," people may be indifferent to group size (all else being equal) in commons-dilemma problems, but in public-goods problems group size may serve as a cue about whether or not to cooperate.[14]

The evolutionists' research on resource distribution and foraging could help explain psychological findings. Psychological research could also be useful for theorizing about human evolution. For example, the ease with which in-group biasing effects can be elicited might indicate that intergroup conflict may not have existed at the high level that is popularly assumed (Lewin 1987). The bias is more consistent with the view that there were cycles of macroband formation and dispersion as subgroups. The two kinds of groups would have conflicting advantages and disadvantages related to information flow, maintaining organizational flexibility, and merely surviving (Jarvenpa & Brumbach 1988). This explanation is less exotic than positing cannibalism or warfare as "prime movers" in human evolution, but could be closer to the truth.

It is surprising that group living might be overlooked or even rejected as a significant factor in the evolution of characteristics that we recognize as human. Alexander (1979), for example, briefly considers the hypothesis that the advantages of group living were so powerful that humans developed the tendency to live in groups. He rejects this hypothesis on two grounds: (1) Group living is so deleterious to reproduction that it can be accounted for only by external threats (most notably, other humans), and (2) the hypothesis is distasteful. Alexander believes that positing a tendency to live in groups would mean that humans were inflexible "captives of [their] genetic history" (p. 222), compulsively living in groups despite the drawbacks. Sociobiologists minimize the importance of archaeological research and fossil evidence (Washburn & McCown 1978) because in their view, contemporary behavior and social institutions can be explained in terms of people's attempts to maximize their inclusive fitness.

We feel that human sociobiological theory may be unduly influenced by the "representativeness heuristic" (Kahneman et al. 1982), according to which effects resemble their originating causes. Assuming that there was an initial solitary human state, labeling human sociality as a "problem," analyzing human behavior by examining self-interested motives, and focusing on nepotistic and reciprocal transactions (Alexander 1979) – these all reflect features of population geneticists' models: single genes, the genetic problem of self-sacrificial altruism, individual selection, kin selection, and selection for reciprocity. Human action is taken to resemble gene action. The heuristic error is to mistake evolutionary forces for the psychological mechanisms that result from them. Because phenotypic "selfishness" is representative of genetic "selfishness" (the replacement of alternative alleles in a population by an unmotivated, mechanistic process), the former is insulated from the effects of natural selection, precluding the evolution of "cooperative" cognitive biases that ultimately benefit the organism's genes.

Our pessimism about the value of explaining contemporary human behavior in terms of fitness maximizing derives not only from the above considerations, but also from the sociality hypothesis. If cognitive evolution involves the increasing accessibility of adaptive specializations, and the hallmark of intelligence is increasing access to such specializations in solving new kinds of problems (Rozin 1976), then fitness may explain the *origins* of mechanisms underlying behavior, but not their role in domains for which they were not originally specialized. Cheney et al. (1986) draw similar conclusions about nonhuman primate intelligence. Primates typically perform poorly in laboratory settings with nonsocial problems (e.g., transitive inference with paired stimuli) that are logically similar to problems they regularly solve in social relationships (e.g., inferring dominance hierarchy from paired interactions). The poorer performance may be explained by the need to extend sophisticated cognitive mechanisms adapted for manipulating social objects to a different domain: the manipulation of nonsocial objects. Our own evolutionary analysis and the primate work suggest that *social* cognition may be the "general case" of human cognitive activity, and nonsocial "cold" cognition (e.g., logical problem solving and statistical reasoning) a derivative ability (Ostrom 1984; Zajonc 1980a, 1980b).

Part of EI's allure is its heuristic convenience. It organizes a variety of behaviors under a single explanatory umbrella. As a heuristic device, it fails less because it makes incorrect predictions (in fact, deviations from its predictions, as we have shown, are informative) than because it precludes alternative views of human nature (Abelson 1976). EI theory creates the illusion that the ultimate *ends* of human behavior – the satisfaction of self-interest – are known and only the *means* remain to be discovered. There is no apparent reason to test the validity of EI assumptions; and possible disconfirmations are "explained away" with "just-so" stories or claims about revealed preference. EI theory is hence "preemptive" in that there is no apparent need to consider alternatives. We have tried to show that an equally useful alternative can be developed that is consistent with evolutionary principles.

4. Summary and conclusions

All social theories make psychological assumptions, even those that reject the significance of psychology (Jahoda 1982). The assumptions are frequently drawn from folk psychological beliefs that ascribe considerable explanatory power to people's selfishness and intuitive rationality. These EI explanations may be so entrenched that economically costly and irrational social practices must be explained away as satisfying selfish ends "unconsciously." Sahlins (1976) criticizes such explanations on the grounds that "economic man" is an idiom of Western culture, not a fact of human nature.

In sociobiology, the folk concept of selfishness is supplemented with a genetic account to explain moral systems. Alexander (1987, p. 109) proposes that such sys-

tems are built on the basis of five "giving rules" that humans are evolved to learn: (1) give to a genetic relative, (2) give when the recipient is likely to give you back more than was received, (3) give if sanctions for the failure to give are likely, (4) give if observers of the giving are likely to provide you with benefits in the future, and (5) "*in all other situations do not give*" (emphasis ours). Contrary to Alexander's rule (5), subjects in our investigations gave away their money.

The experiments we report provide evidence against such EI theories. The studies eliminated widely accepted explanations for cooperation: reciprocity, friendship, kinship, custom, and conscience. Cooperation was enhanced by the elimination of free-riding, the assignment of social roles, and, of course, egoistic incentives. Egoistic incentives are neither necessary nor sufficient for individuals to support group welfare. In the absence of discussion, both choice and people's explanations for their choice are consistent with egoistic incentives – either monetary rewards or internal rewards for "doing the right thing." But group discussion was found to increase the rate of cooperative choices dramatically and led both contributors and non-contributors to expect overprovision of the public good, resulting in a normatively irrational choice for those who nevertheless decided to contribute.

A proponent of EI theory might still argue that discussion serves as a cue for reciprocity: In the long run, responding to such cues benefits the individual even if there is no reciprocation for any specific social interaction. Because most people's prior prosocial behavior is not anonymous and occurs in the context of discussion, subjects might more or less "automatically" contribute in a similar context – the way they automatically use turn signals without distinguishing empty roads and populated ones. This explanation has shortcomings. For example, Dawes et al. (1977) found that the discussion of an irrelevant topic (the proportion of people at various income levels in Eugene, Oregon) led subjects to contribute in a similar dilemma game at the same level as no discussion, and rates of contributing in these conditions were significantly lower than when subjects discussed their social dilemma. It seems that if discussion is to serve as a cue, it must be specifically targeted toward the problem at hand. In addition, other manipulations of identity – in the absence of discussion, or even without meeting other group members – are as effective in promoting group welfare as discussing the dilemma. In Brewer's (1985; Kramer & Brewer 1986) depleting-resource dilemma experiments, people were told that a computer lottery would determine the value of accumulated points. Subjects contributed more points in a "common fate" condition in which a single lottery determined the value of points for all group members than in a differentiated fate condition in which the lottery determined the value of points for members in a subgroup. Subjects who were led to believe (by false feedback on a computer screen) that members of their own subgroup were most responsible for the overuse of the resource compensated by restraining their own use. Those who believed that the overuse was due to members of other subgroups increased their use. Group discussion appears to involve psychological mechanisms that can also be engaged by other manipulations.

The sociality hypothesis was introduced as an alternative to EI theory and as an explanation for the ingroup bias. According to this hypothesis, the cognitive/affective mechanisms underlying certain human behaviors evolved in a social setting involving small face-to-face groups. Once evolved, these mechanisms could then be applied in domains for which they were not specifically adapted. Separating the history of mechanisms from their current utility has several advantages. For example, we could predict (and experimentally test) "mismatches" between once useful mechanisms and their role in the modern world (e.g., heuristic rules-of-thumb that lead to error under some contemporary problem-solving conditions; cf., Cosmides 1985). Evolutionary ecologists and psychologists could join forces with specialists in other disciplines in seeking a fuller understanding of innate mechanisms and their origins. This "bi-level" approach might also help to forge a more satisfying compromise between "innatists," who argue that the primary determinants of social behavior are biological, and "environmentalists," who favor individual learning and culture as primary causal factors (cf., Caplan 1978). If we distinguish past functioning and present functioning, the considerable flexibility of human behavior need not be inconsistent with innate mechanisms.

The last objection to the sociality hypothesis we will address here is likely to come from biologists. We have argued that evolved cognitive and affective mechanisms are a factor in the experimental effects we report. It was adaptive for ancestral humans to identify automatically with an ingroup and to accept its goals as their own. Such automatic commitment to group goals is *not* – in Maynard Smith's (1984) terms – an evolutionarily stable strategy (ESS). A strategy of maintaining group membership by appearing to pursue group goals while *successfully* defecting whenever possible to individual advantage could "destabilize" a group cooperative strategy (i.e., hypocrites would replace cooperators). As Logue (1984, p. 111) points out, however, "an animal that ordinarily optimizes well may be confronted with a complex problem involving many alternatives that is beyond the animal's present mental capacity." The consistent choice of alternatives that successfully fake social commitment while secretly defecting from this commitment is a problem beyond our present capacity. There is a trade-off between devoting energy to seeking out such alternatives and obtaining the results of successful group activity; both involve mental "competition." As Simon (1980) and many others have noted, our cognitive capacities are limited. Frank (1988) argues that the easiest way to be accepted by others when appearing to be a cooperator committed to the group is to actually be a cooperator, thereby preserving energy for pursuits other than dissembling. Many of these pursuits are likely to be more advantageous than secretly defecting. Moreover, the negative payoffs for a single mistaken choice can be devastating. Campbell (1975) and Axelrod (1984), among others, have pointed out that the sanctions for uncovered hypocrisy in most societies are severe – occasionally extending to total ostracism from the group. Such an alternative should accordingly be chosen only rarely.

Despite our objections to EI theory, we do not believe that the effort expended over centuries to devise egoistic incentives for cooperation is wasted. Many of the social

problems that must be solved today involve large groups of people unknown to each other – a novel departure from the conditions of human evolution. Our work does call for considerably broadening the scope of research on social dilemmas, however, as well as a serious reexamination of the "selfish human nature" assumptions that dominate the human sciences. Self-interest is only one "invisible hand" described by Adam Smith. The other was a universal "fellow feeling," one that even "the greatest ruffian, the most hardened violator of the laws of society, is not altogether without" (Smith 1759/1976, p. 7).

APPENDIX

A useful prototype for social dilemmas is the prisoner's dilemma, which is usually illustrated with the following scenario. Two prisoners are charged with a crime. The police want a confession from at least one prisoner. The prisoners are separated and each is offered the same deal: "If you confess and your partner doesn't, you'll get a light sentence and your partner will be imprisoned for a long time; if neither of you confess, you'll both serve a few years in prison; if both of you confess, you'll both serve many years in prison." The choices confronted by the prisoners ("players") can be represented as a game matrix that shows the outcomes ("payoffs") for all combinations of confessing and not confessing. In the matrix below, these are represented as years in prison.

		Prisoner B	
		Confess	Not confess
Prisoner A	Confess	7, 7	1, 10
	Not confess	10, 1	2, 2

By convention, the first number in each cell entry is the payoff for Prisoner A and the second is the payoff for Prisoner B. Each player prefers to minimize the number of years in prison. Prisoner A observes that if Prisoner B confesses, then A's own confession will lead to 7 years in prison rather than 10. If Prisoner B does not confess, then A's confession will lead to 1 year rather than 2. Therefore, regardless of what Prisoner B chooses, Prisoner A should choose confessing. But Prisoner B is confronted with the same choice, and therein lies the dilemma: If both confess, then both are worse off than if both had not confessed. In this example, confessing is a dominating strategy (because it leads to the fewest years in prison), but it results in a deficient equilibrium (because both prisoners could have spent fewer years in prison if they had chosen the alternative, nondominating strategy).

Again by convention, not confessing is termed "cooperating" and confessing is termed "defecting"; people choosing these strategies are termed "cooperators" and "defectors," respectively. In this article, we have tended to use the terms "contributors" and "noncontributors" because they are more descriptive of our subjects' behavior than is the conventional usage.

The scenario above and the following glossary should help the reader unfamiliar with game theoretic terminology to follow the discussion in the target article.

GLOSSARY

Commons dilemma: A social dilemma in which the cooperative strategy requires refraining from behavior that benefits oneself at a cost shared by other group members. Overgrazing a common pasture destroys the commons.

Cooperation: Choosing a dominated (*q.v.*) alternative (or "strategy") that yields higher payoffs to all *other* players than the alternatives dominating it no matter what the other players choose (e.g., not confessing in a prisoner's dilemma). In other words, choosing a strategy that benefits others.

Defection: A choice or strategy that yields a higher payoff to the player choosing it (no matter what combination of strategies other players choose) than does an available cooperative strategy.

Dominating strategy: This is a strategy that (1) yields at least as high a payoff for the individual choosing it as any other alternative strategy no matter what others choose, *and* (2) leads to a higher payoff than any other strategy for at least one combination with others' choices. If dominating strategies exist for all players, then the outcome resulting from their mutual choice is the sole equilibrium in the game; if the outcome is deficient, the game is a social dilemma.

Equilibrium: An outcome resulting from each player's choice of a particular strategy is an equilibrium outcome if no player can obtain a higher outcome by unilaterally choosing another strategy. An equilibrium is *deficient* if some other possible outcome yields higher individual payoffs to every player. A game may have more than one equilibrium; for example, consider a two-person game in which each player chooses a strategy, A or B, yielding $1 to both if both choose A, $2 to both if booth choose B, and $0 to both if one chooses A and one chooses B. Mutual choices of A or of B both result in equilibrium outcomes; the equilibrium resulting from a joint choice of A is deficient.

Evolutionarily stable strategy (ESS): A dominating strategy or, alternatively, a "mixed" strategy where choosing between available strategies on a probabilistic basis guarantees that if all players adopt the mixed strategy no other strategy or combination of strategies can yield a higher expected payoff.

Public-goods dilemma: A social dilemma in which the cooperative strategy requires a contribution to a good that is shared by all, whether or not they contribute. Public broadcasting is a public good.

Revealed preference: Outcome preferences inferred from actual choices without reference to any criterion of preference independent of the observed choice. If a player does not choose an apparently dominating strategy, then a revealed preference analysis requires that some other strategy be identified as "dominating.": This is because *by definition* at least one payoff (of any possible sort, including inferred "psychic rewards") that could result from the strategy chosen must be preferred to the payoff resulting from choosing the apparently dominating strategy.

Utilities: Personal values of outcomes. In a revealed preference analysis, actual choices can be used as *measures* of utilities provided such choices do not violate any of the axioms of rational choice (e.g., if A is preferred to B and B is preferred to C, then A must be preferred to C too; see von Neumann & Morgenstern 1947). Many researchers infer utilities from choices made in artificially simple situations to determine whether actual choice is consistent with rational choice axioms. Thus, rather than assuming that actual choice by definition yields utilities, their concern is whether a particular choice maximizes expected utility.

ACKNOWLEDGMENTS

We are grateful to Marilynn Brewer, David Buss, Donald Campbell, Glen Culbertson, Peter Richerson, David Sloan Wilson, the editor of *BBS*, and the reviewers of this article for their valuable comments and suggestions (not all followed) on earlier versions of this paper. The empirical research was supported by National science Foundation Grants SES-8605284 to Dawes, Orbell, and van de Kragt and SES-8308610 to these three authors and Randy Simons. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation or those of the sponsoring universities. We regret the untimely death of Alphons J. C. van de Kragt on October 16, 1989.

NOTES

1. We use "egoistic incentive theory" to refer to the very broad category of explanations of human behavior (used even in ancient times, e.g., by Thucydides) based on the premise that human nature is "fundamentally selfish" (Richards 1987). The historical continuity and wide currency of these explanations justify identifying them as "the selfish paradigm" (Schwartz 1986). "Economic man theories" are versions of the paradigm supposedly legitimized by science (e.g., reinforcement theory, economics, sociobiology). Economic man theories may be loosely woven accounts of human nature or quantitative formalizations of the conditions of choice behavior or decision making. The common thread is the premise of basic selfishness.

Nevertheless, the term "selfishness" does not work well as a descriptive category in science. Not only is it enormously value-laden, it is associated with different causes (e.g., genes or original sin) depending on one's initial assumptions. Without a shared framework, discourse using an identical term can be confusing. For example, distinctions between individual and group interests can be obliterated as in Campbell's (1982) "clique selfishness." This refers to choosing to cooperate for the benefit of face-to-face subgroups (e.g., disciplinary departments) at the expense of the larger collective group (e.g., a university). We resort to the neutral, albeit awkward, term "egoistic incentive" because it highlights two important assumptions in the selfishness paradigm. (1) People's behavior can be explained as responses to incentives, and (2) incentives are valued by the individual, not a group entity. This does not mean we have eliminated the term "selfishness" from our vocabulary – it is still useful in contexts where it connotes a continuity between scientific work and cultural milieu.

2. To claim that humans are basically social is not to claim they are basically altruistic (Trivers 1985) or that they have an innate moral sense (Richards 1987). Obviously, human behavior can be characterized as selfish as well as cooperative and altruistic. Our concern is with the species-typical "cognitive machinery" (i.e., mental structures and process) that underlies certain kinds of behavior. We suggest that this machinery evolved because it enhanced fitness under the historical circumstances of human evolution. This machinery cannot be reduced to a single motivation. The reader who believes that EI theories are true by definition may defend the belief on grounds that whatever leads to evolutionary success is *ipso facto* "selfish." This explanatory framework has two flaws. First, it conflates the cognitive machinery with the genes "building" the machinery. Second, it leads to a tautological explanation of prosocial behavior: Because humans are basically selfish, one can assume that their behavior reveals egoistic preferences even if an unobservable incentive must be posited (e.g., to "feel good"). The assumption that EI theories are true "by definition" eliminates the possibility of disconfirming the assumption by empirical research.

3. Game theory is the formal description of decision making under conditions of conflict. The characteristics of a game include assumptions about the players (self-interested rationality); rules for specifying how resources (points, money, and so forth) may be used; strategies specifying what should be done under various contingencies; and payoffs, which depend on the strategies used by the players. A game may be a descriptive device (e.g., Maynard Smith 1984), it may be modeled by a computer program (Axelrod 1984), or it may be played by people – subjects participating in experimental games (as reported in this target article). In a game, each player selects a strategy and receives a payoff, the value of which is contingent upon the strategy selected by the player and the strategies of the other players. Games are distinguished (and frequently named) by variations in the rules, strategies, and payoffs. (For details, see Appendix.)

4. A tit-for-tat strategy in symmetric games (as Axelrod's [1984] were) can be reformulated in a logically equivalent form making no reference to reciprocity: Defect if and only if the sum of points received is less than the sum of points the other player has received. This equity-based framing results in cooperation on the first play, defection after the other player's first defection (if it occurs), and cooperation after the other player's first switch from defection to cooperation (if it occurs) – as in tit-for-tat. A player or computer program based on achieving absolute equity would thus respond identically to one based on absolute reciprocity; moreover, it would be just as effective (or ineffective, if it results in alternation) with another player sensitive to absolute equity as with one sensitive to absolute reciprocity. This reformulation illustrates the potential ambiguity of results consistent with a single model of cooperation, or strategy for "achieving" it. In contrast, throughout this target article we have attempted to devise experimental tasks that yield differential predictions from contrasting models, or that at least compare the presence of an explanatory factor with its absence.

5. In defense of Beecher (1847/1972), we must point out that her arguments are representative of the ones women frequently use both to explain a complex network of powers over which they have little control and to try to force that network to adjust to their own interests (see Friedlander et al. 1986). Suffragettes successfully collected "compensation" for their subordination and touted their Providentially superior moral nature to obtain the vote.

6. In the Grether and Plott (1979) study, subjects preferred a gamble in which they could win $4 with a probability of 35/36 and lose $1 with a probability of 1/36 to a gamble that paid $16 with a probability of 11/36 and a loss of $1.50 with a probability of 25/36. Nevertheless, these same subjects were willing to pay more for the opportunity to play the *second* gamble.

7. As this game is presented, it is not strictly true that subjects *always* receive more money when holding back their contribution – i.e., that not contributing the $5 is strictly a dominating strategy. If *exactly* four other subjects contribute, a subject receives a final payoff of the $10 bonus for contributing while retaining only $5 as a result of not contributing. This slight deviation will be addressed later in the paper when more complicated experiments designed to eliminate it are discussed. The latter are variations of the experiment described, which was the first in the series investigated and is discussed here because of its simplicity.

8. The advertisements offered respondents $4 for showing up for a scheduled appointment and "between $0 and $__ depending on your decision and the decisions of others in the experiment." The second amount varied between $18 and $38 depending on the experiment. The ads also emphasized that no prior experience was necessary. On average, about two-thirds of the subjects were students, the rest were townspeople, often unemployed.

In discussions with other researchers, we found that objections to the subject recruitment and population were of two kinds: Some objected that the small amount of money made cooperation especially easy. Others suggested that cooperation in the experiments could be traced to subject characteristics such as moral beliefs or religiousness; to suspicion that the experimenters were duplicitous and noncooperation would be exposed; or to failure to understand experimental instructions. Concerning the first objection: The money was very real to these subjects, who can be assumed to have needed it; there was also substantial variance in the amounts available, depending on the decisions made. As to the second objection, we acknowledge all these factors – plus the large number of other possible subject characteristics, beliefs, attitudes, fears, and so forth – influencing why people cooperate. In one study, for example, we found a slight positive correlation between cooperation and frequency of church attendance. We assume, however, that such variables are randomly distributed across conditions by the procedures and that the emphasis on anonymity of choice and the lack of deception convinces most subjects.

9. We are frequently asked what our subjects are "thinking" during these experiments. Their explanations of their behavior tell us less about the underlying psychological factors than about how people perceive their behavior. Both behavior and perceptions of it are influenced by cultural context, personal values, and the sorts of explanations people prefer. In other words, verbal reports do not explain behavior; rather, they too are data requiring explanation by recourse to cognitive and social factors.

10. This 84% figure is somewhat misleading when compared to the 100% contributing rate in a designated set game with discussion. In the former case, every group member is potentially a contributor; hence the average percentage contributing is calculated over all group members. In the latter case, the average percentage contributing is based only on the designated contributors because others were "designated" noncontributors.

11. If four of the designated five contributed, then the defector would get the bonus, but the other four would not because there would not be enough others contributing for them to do so. The bonus is not a "pure" public good because it admits exclusion rather than equal provision to all group members.

12. Landau (1984) has analyzed the narrative structure of evolutionary scenarios. She found the structure to be identical to that used for folk tales and hero myths. It seems likely that the tacit assumption of an initial solitary state is an unwitting artifact of this narrative structure rather than an intended theoretical proposition.

13. In the seven-person "give some" games described earlier, the choice between giving $12 to other group members versus keeping $6 is equivalent to a "take some" (commons dilemma) choice between retaining $12 with no effect on other group members versus taking an additional $6 with a $12 fine assessed against the other group members – $2 each – for this $6 personal gain.

14. Clearly, to pursue this line of thought would require a better understanding of the human-habitat relationships that were important to human adaptation. The ecology of a relatively mobile hominid group that travels from resource to resource regularly presents choices that differ from those of a more sedentary group that stores and conserves food resources. The timing of the emergence of different practices would suggest different cognitive predictions. In addition, a better understanding is needed of how features in the environment are selected as relevant for making choices. Liebrand (1984), for example, does not find reductions in contributing when group size expands. Brewer (personal communication) explains the effect of how the problem is presented as a function of the salience of individual versus collective identity inherent in the task. Public-goods problems focus attention on one's own behavior, whereas collective interests are the focus for commons-dilemma problems.

Open Peer Commentary

Commentaries submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Ambivalent sociality: The human condition

Marilynn B. Brewer

Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90024-1563

In exposing and challenging the pervasiveness of individual self-interest assumptions in the behavioral sciences, Caporael et al. have taken a major step toward freeing our science from the constraints imposed by that view of human nature. The study of prosocial behavior has long been harnessed to the implicit assumption that the occurrence of cooperative or altruistic behavior is *problematic*, calling for extraordinary adjustments of our basic models of behavioral choice in order to account for apparent deviations from purely self-interested motivation.

In the social-dilemma literature, this assumption is manifest in the idea that collective dilemmas can be resolved only if individuals are in some sense "tricked" into violating hedonistic rationality – either by changing the rules of the game or by obscuring the basic conflict between individual and collective welfare. The impressive program of research undertaken by Dawes and his colleagues effectively challenges this view of collective decision making. Using the subtractive method, they have demonstrated the presence of self-sacrificing cooperative behavior in the absence of any apparent egoistic incentives. More important, the near 100% rate of cooperative choice obtained following group discussion demonstrates convincingly that, under the appropriate conditions, ingroup cooperation is just as "natural" as self-gratification as a rule for individual decision making.

If Caporael's sociality hypothesis is accepted, it prompts social and behavioral scientists to address the interesting questions posed by the task of identifying the "perceptual, affective, and cognitive mechanisms that support the development and maintenance of group membership" (sect. 3, para. 6). In the interests of furthering that research agenda, I would like to correct two general misconceptions prevalent in evolutionary psychology that might interfere with our ability to explicate those mechanisms.

The error of confusing form and function. The first misconception is the assumption that form necessarily reflects function. In sociobiology this is represented by the search for the "altruistic gene" (and its psychological correlate, the altruistic drive) as the basis of prosocial behavior. It is very possible, however, that the mechanisms that support group cooperation are not directly represented phenomenologically as "self-sacrifice" or "collective welfare," but rather as cognitive extensions of self-interest.

Psychological theories of self, dominated by individualistic conceptualizations of human nature, have seldom acknowledged the blurring of distinctions between self and others in the formation of self-concepts. Yet we know that social isolation has devastating effects on human functioning (both cognitive and affective), and that many behaviors that are irrational in terms of individual self-interest are explicable when one takes into account the individual's identification with significant social groups. Much of human behavior is sustained because it is symbolic of or helps preserve an individual's self-conception as a member of particular social groups.

This aspect of self-concept is explicitly recognized in Social Identity Theory (Turner 1987; Tajfel & Turner 1985), which posits that identification with significant social groups entails "a shift towards the perception of self as an interchangeable exemplar of some social category and away from the perception of self as a unique person" (Turner 1987, p. 50). Our own experimental research on social dilemmas (e.g., Brewer & Kramer 1986; Kramer & Brewer 1984) has demonstrated that the intensity of an individual's identification with a group has a profound effect on a subject's willingness to restrict individual gain to preserve a collective resource. Our work also supports the contention that there are inherent limits to the *size* of social groups that can engage such cooperative mechanisms. If group identification is a symbolic extension of an individual's self-concept, then that identification must provide for distinctiveness and differentiation, as well as for a sense of belonging. If distinctive social identities are essential to human functioning, then what Campbell (1982) calls "clique selfishness" may well be a more powerful and intractable form of selfishness than even the most extreme individual self-interest.

The error of unidimensional thinking. The second misconception characteristic of evolutionary psychology is the representation of "selfishness" and "altruism" as endpoints of a single bipolar dimension, with the idea that "human nature" falls somewhere along this continuum. In contrast to this view, I have posited elsewhere (Brewer 1976) that self-gratification and collective identity are independent, opposing processes that are reflected in ambivalence and variability of responding in the face of conflict between individual and collective welfare.

The potential variability of choice behavior in social dilemmas is exhibited most strongly under conditions of collective crisis. In our resource-dilemma experiments, subjects tend to behave relatively homogeneously when the availability of the shared resource is not threatened. However, when the resource pool is being depleted, individual choices are polarized. When a collective social identity has been made salient, group members respond to a crisis by reducing dramatically their own resource use. However, when a collective social identification has not been provided, individuals respond to the depletion of the resource pool by increasing immediate resource use at the cost of long-term resource availability. Because subjects are randomly assigned to group conditions in these experiments, we can assume that such variability is not simply a function of individual differences in social motives. The results suggest instead that both response modes are present in all individuals, subject to selective engagement in different social contexts.

Sociality: Costs, benefits, and mechanisms

Thomas Caraco

Behavioral Ecology Group Biological Sciences, State University of New York, Albany, NY 12222

I found the target article's arguments disappointing. Caporael et al.'s conclusion that selection favored attributes maintaining group membership in early humans is almost certainly correct, but hardly novel.

Natural selection, by definition a process influencing a population's gene frequencies, can ultimately only imply "selfishness." That is, evolutionary self-interest must rule at the ultimate level when natural selection occurs. The neo-Darwinian concept of inclusive fitness follows logically, if only approximately, from the population-genetic definition of natural selection. Primacy of this ultimate-level "selfishness" need not obtain for all forms of genetic evolution, but the target article equates evolution with natural selection.

The mechanisms favored as a consequence of natural selection, however, defy such a simple categorization, particularly when we consider the ecological complexity of a social group. Ultimate self-interest (defined by the dynamics of gene frequencies) can favor selfish, cooperative, altruistic, and mutualistic responses by one individual toward other individuals. Caporael et al. recognize this distinction and still misinterpret the difference in its verbal analysis of an n-person prisoner's dilemma.

Overall, the target article lacks rigor and logical consistency. Natural selection is marshalled in, critiqued, and later recalled as the organizing concept. Initially, the authors appear to dismiss natural selection (or is it just mechanisms inhibiting cooperation?) as a "cultural belief" and not a "reasoned argument." But in section 3, paragraph 4, they adopt a "Darwinian framework" to suggest an explanation of "adaptations" for group living. I was put off by a series of apparent contradictions; most occur because Caporael et al. so often confuse ultimate and proximate phenomena. An increased frequency of a "selfish" gene modulating a behavior need not imply that selection has favored a greater propensity to defect in a social interaction. The authors do not appear to appreciate that evolution by natural selection in solitary and social organisms is ultimately a "selfish" process in both cases, because it is natural selection. The outcomes of selection in the two cases, however, can vary from extreme intraspecific aggression to highly cooperative or even altruistic social interactions. The lack of clarity concerning ultimate and immediate causation will not generate much sympathy among evolutionary biologists. Genes always compete when exposed to natural selection; the dynamics of the process allow only competitive "selfishness." The behavioral mechanisms through which they compete may include cooperation, which surely has been the case during primate evolution. In their second footnote, Caporael et al. attempt to discriminate between "cognitive machinery" (i.e., proximate mechanisms) and the influence of natural selection (ultimate phenomena), but the text suffers from a semantic confusion. I hope other readers found subtleties I missed.

The experiments seem to indicate that humans cooperate more often than the currently popular tit-for-tat scheme (Axelrod & Hamilton 1981) might predict. Perhaps this behavior represents an epiphenomenon of an evolutionary heritage of social life in groups organized largely by bonds of kinship. But Caporael et al.'s sociality hypothesis suggests a plausible scenario to me. Many primates require continued membership in the same social group for survival and reproduction. Presumably, this was true for early humans. Natural selection would indeed have continued to favor individuals whose attributes promoted their position within the group, provided the basis of those attributes was directly or indirectly genetic and heritable. Cooperation with group members, even when defection in a particular interaction might yield a greater momentary benefit, could have been an important element in the maintenance of group cohesion. The mechanistic propensity to cooperate, especially coupled with in-group biasing (increasing the probability of reciprocity) seems a logical outcome of individual selection.

Caporael et al. are not the first to assume that in human bands an individual's survival and the survival of the individual's offspring were enhanced by group membership. Lack of an egoistic incentive (a mechanism inhibiting cooperation) might have imposed a cost of sociality at times, but the benefit of group membership advanced individual fitness sufficiently to overcome the cost. If I have paraphrased Caporael et al.'s argument correctly, the sociality hypothesis reduces to the standard cost-benefit paradigm of behavioral ecology. Considering cooperation as a cognitive mechanism, rather than as a phenotype with a fitness, does not suggest a novel perspective. Again, I may have missed some subtle point, but I can't recommend this target article to students.

Selfishness reexamined

R. I. M. Dunbar

Department of Anthropology, University College London, London WC1E 6BT, England

The conclusions developed in this target article are, I believe, essentially correct, but the arguments on which they are based perpetuate three important (and depressingly common) errors concerning sociobiological explanations of behaviour. It is essential to be absolutely clear about just what predictions sociobiological theory would make about human behaviour lest we run the risk of throwing the baby out with the proverbial bathwater.

The first misunderstanding is the suggestion that the selfishness of genes implies the selfishness of individuals. Although in principle this is one possible interpretation, it is neither essential to, nor a necessary consequence of, any sociobiological theory. The equation of genetic and individual selfishness conflates ultimate and proximate levels of explanation. It is for this reason that it is possible for genuine altruism (mediated by motivations of altruism) to evolve through the selfishness of genes. The issue hinges on a proper reading of the term "selfish" in this context: The phrase "selfish gene," coined by Dawkins (1976), is intended as a metaphor – a reminder to us that when we determine the costs and benefits of behaviour, we must do so in terms of the number of copies of a given gene gained and lost. It carries no motivational connotations.

The equation of genetic and individual selfishness also fails to note the importance of "mutualism" as a driving force in the evolution of sociality (see Dunbar 1988; 1989; Maynard Smith 1982; Wrangham 1982). This is especially so for mammals in general and for primates in particular. To ignore the advantages of cooperation that arise through mutualism is to ignore the very basis of mammalian sociality. In this context, the error is perhaps forgivable on Caporael et al.'s part: Those few sociobiologists who have engaged, with notable eclat, in popularising human sociobiology have tended to perpetuate the same error. The reason, I suspect, lies in the fact that, without exception, their backgrounds lie in the study of invertebrates or nonmammalian vertebrates where the more direct control of genes over behaviour renders a mechanistic view a very reasonable approximation. But it is extremely difficult to find a sociobiologist who works on mammals (and, a fortiori, on primates) who espouses such a simplistic view (see, for example, Cheney et al. 1986; Dunbar 1984; 1988; Smuts 1985). Quite why both those who have dabbled in human sociobiology and their critics have been so desperate to class our own species with the bees and ants is an interesting question for the sociologists of science.

The second source of misunderstanding concerns the historical origins of human sociality. Caporael et al. make a great play

of the fact that sociobiologists assume human behaviour as we see it now arises out of an asocial precursor state, whereas, as they rightly point out, both the archaeological record and the evidence from living primates indicate that hominids have always been social. Unfortunately, this confuses two key aspects of biological explanation, namely, explanations about evolutionary history (or phylogeny) and explanations about evolutionary function. This is a distinction that was stressed nearly three decades ago by Tinbergen (1963) in a seminal paper. Building on an earlier analysis by Huxley (1942), Tinbergen (1963) pointed out that, in attempting to answer the question "Why is X the case?" when studying behaviour, we can produce four very different (and logically *independent*) kinds of answer: These would be answers to questions about motivation (or proximate causation), function, ontogeny, and evolutionary history. We conflate these at our peril.

The reason sociobiologists consider asociality the primitive condition against which to evaluate an observed social state is not that they presume our immediate ancestors to have been asocial, but that they want to know why a given behaviour persists in the population. What happened before is not necessarily relevant to why an organism is social now. (Indeed, Caporael et al. themselves point out that a character can evolve for one reason, and then be captured by another functional system and so turned to new purposes.) Whereas the question of origins has its own intrinsic interest (and may allow us to see why evolutionary history took one particular course rather than another), the question of evolutionary function (which is what lies at the heart of sociobiology; see Dunbar 1982) is concerned exclusively with the issue of why a given characteristic is maintained in the population in the here and now. This is especially important with respect to such a characteristic as social life, which incurs major costs. Group life, as all recent work on primates has emphasised (see, for example, Dunbar 1988; van Schaik 1983; Wrangham 1980), incurs serious costs in terms of increased social and ecological stresses. Without some positive advantage to grouping, groups will naturally disperse as individuals revert to the logically more primitive state of a solitary existence.

Hence, the correct question to ask, as Alexander (1979) and others have recognised, is why humans are *still* social. To argue, as Caporael et al. do, that it is a hangover from our evolutionary past is to beg questions about the forces that maintain genes in evolutionary equilibrium over time. More serious, perhaps, is the very pessimistic view of human nature they hold – a view, incidentally, that is both at odds with classical Darwinian theory and incompatible with the general thrust of contemporary sociobiology. A Darwinian perspective, above all, emphasizes the dynamic aspects of the evolutionary process, whereas sociobiology, with its emphasis on organisms' strategic behaviour, lays considerable stress on the flexibility of behaviour (especially in the more advanced species).

This source of confusion highlights the third misunderstanding, namely, the assumption that sociobiology necessarily implies genetic determinism. Caporael et al. speak, for example, of the "'innatists' who argue that the primary determinants of social behaviour are biological." Allow me, yet again, to disabuse those who persist in this fallacy, for it conflates yet another of Tinbergen's (1963) senses of "why," this time questions about ontogeny, with questions about function. That we can show that behaviour has an evolutionary function (by which I mean that it maximises genetic fitness) does not necessarily mean that it is genetically determined (see Dunbar 1982; 1988). Learning, as has so often been pointed out (see, for example, Plotkin 1988; Plotkin & Odling-Smee 1981; Pringle 1951), serves the same function as a mechanism of inheritance that genes serve. Indeed, this is the principal point of the last chapter of Dawkins's (1976) book, *The Selfish Gene*.

These criticisms do not, of course, invalidate the experimental results. Nor do they invalidate the inference that people do not necessarily behave in ways motivated by personal selfishness all the time. But, then, I would suggest that we might have anticipated that on sociobiological grounds anyway, the views of a minority of sociobiologists notwithstanding. I am not convinced, however, that these results, of themselves, would persuade a sceptic that humans are not innately selfish. After all, as Caporael et al. themselves note, punitive sanctions are imposed on cheats and hypocrites in most societies, and this surely implies that altruism is something we have to be taught. It is two purely methodological points, however, that oblige me to hesitate in going all the way with the authors' argument.

First, I wonder whether a "no loss" situation is really a strong enough test to uncover such innate selfishness as humans might possess? All sociobiological arguments hinge on the relative magnitude of the costs and benefits of a given course of action. Hence, it is important to distinguish two senses in which we commonly use the term "altruism": A weak sense in which the altruist incurs no significant cost by his action and a strong sense in which there is such a cost. Only the latter counts as altruism under sociobiology's rubric; even then, the costs have to be genetic, and it is difficult to see how donating five dollars is likely to incur a loss of genetic fitness for the donor. However, allowing for the moment that the donation of five dollars can be considered a *bona fide* cost, we cannot easily avoid the problem that the *net* cost to the donor in all these experiments is zero: Subjects merely pass on money that has just been given them. From a sociobiological point of view, we would expect individuals to behave more altruistically under these conditions than if they were being asked to donate five dollars of their *own* money, so that they risked being five dollars poorer by the end of the experiment. I would hazard the guess that the proportion of donors would be significantly lower in the second condition. Meanwhile, I suggest that we are obliged to conclude that these tests are not adequate to indicate an innate tendency to behave altruistically.

My second disquiet concerns the role that social learning plays in all human behaviour. As I understand them, Caporael et al. want to argue that people are essentially altruistic on the grounds that they do not behave selfishly in groups where discussion is permitted. Yet, the results clearly show that when decisions are made in the absence of any face-to-face knowledge about other group members (or, indeed, in the knowledge that the beneficiaries will be members of another group) people *are* more likely to behave selfishly. What is missing here is, I think, an appreciation of the extraordinary power of the processes of socialisation, especially when these are combined with the capacity for learning with which we are endowed.

I am equally unhappy about the implied change of motivation under these contexts: The results seem to suggest that subjects' self-confessed reasons for donating or withholding their money in the experiments in which discussion was permitted are that the group's decision obliged them to behave thus, whereas they openly admit to acting in self-interest when making their decision in isolation. Yet, as anthropologists have so often learned to their chagrin, people are notorious dissemblers when it comes to justifying their behaviour – and never more so than when that behaviour can be interpreted as being antisocial. It should surprise no one that those who are given a group sanction to behave antisocially should gratefully use that as an excuse to cover up their real motives. It seems to me that there are serious confounding variables hidden within the experimental design.

I suspect this means that we need to be very much more sensitive to the details of the social context than we have all (anthropologists, psychologists, and biologists alike) been inclined to be. This is surely one of the key lessons that sociobiology has to offer us, for sociobiological rules are always context-dependent. Very few can be expected to apply universally to all members of the same social group, never mind the same species. What to one individual may be a trivial cost, to another may be exorbitant. Their respective decisions on

whether or not to behave altruistically will vary accordingly. Hence, so far from the results of their experiments disproving Alexander's (1987) rule 5, as Caporael et al. suppose, I would suggest that they in fact reinforce rule 3 – "give if sanctions for the failure to give are likely." In a social context, sanctions perpetually hang over our heads; we are never entirely free of them. It is this complexity that makes human (and primate) sociobiology both so difficult and so challenging to study: We cannot afford to make simple-minded assumptions about how individuals will behave.

Biological underpinnings of social systems

Niles Eldredge

Department of Invertebrates, American Museum of Natural History, New York, NY 10024-5192

Sociobiology is a logical outgrowth of contemporary evolutionary theory – a form of "ultra-Darwinism" rooted in the notion that organisms are locked in a constant struggle to maximize their genetic representation in succeeding generations. Dawkins (1976), following this line of thought to its extreme, went so far as to imagine organisms as mere vehicles for genes – the real schemers engaged in the perpetual race to leave more copies of themselves than their rival alleles to the next generation. Biotic nature is portrayed in a curiously skewed manner under this world view: We are to believe that the structures we see – organisms, ecosystems, species, social systems – are mere epiphenomena of their underlying genetically encoded information. Moreover, ultra-Darwinism is shot through with teleology: The entire *purpose* of organisms (and in some formulations, of ecosystems, species, and social systems as well) is the preservation and further transmission of the underlying genetic information. Something is odd on the face of it when the instructions for assembly are deemed more significant than the assembled structure.

But such is the modern evolutionary perspective, which purports to deal with why biotic systems are the way they are from a historical, adaptive point of view. A functionalist perspective typically leads to rather different postulates. A "pure" ecologist (i.e., one not particularly concerned with the integration of ecology with evolutionary theory) is not constrained to view interactions among organisms as entirely competitive. There are wholly neutral, mutually advantageous, and parasitic possibilities as well – and this spectrum of interactive possibilities obtains whether the organisms are conspecifics (the issue here) or not.

Caporael et al. are suggesting that a purely ultra-Darwinian perspective on (specifically human) social organization manages to get the evolutionary story precisely backward. Their initial perspective is strongly functional, however. They observe that nonkin cooperation is a characteristic, functional aspect of many mammalian social systems (insects are something else again). They offer an experimental protocol and results that confirm that cross-genealogical cooperation is indeed part of the makeup of human social behavior. They then extend such functional considerations to an evolutionary perspective by suggesting that the "fitness" of individuals lacking such cooperative inclinations is likely to be correspondingly reduced. One cannot help feeling that theirs is likely to prove a more accurate depiction of human social structure than that typically found in standard sociobiological treatments, including those they cite.

More generally, recent analyses (see Eldredge, 1986, for references) have focused on the consequences for the establishment of biotic structures when organisms put their *economic* and *reproductive* attributes to use. In a nutshell, (sexual) reproduction leads to the formation and maintenance of local reproductive populations (demes), themselves parts of species. Species are parts of larger-scale genealogical systems: monophyletic taxa. On the other hand, economic activity (for organisms, matter-energy transfer for development, growth, and somatic maintenance – staying alive) leads to the assembly of local ecological populations, which are parts of local (and distinctly cross-genealogical) ecosystems; local ecosystems are parts of larger-scale, regional ecosystems. Organisms are inherently and simultaneously parts of two ontologically distinct, hierarchically structured systems – as a simple consequence of engaging in economic and reproductive activities. The point here is that social systems are manifestly hybrid; they result from peculiar and idiosyncratic (re-) integration of organismic economic and reproductive functions. Is the "nuclear family" economic or reproductive in nature? The answer can only be "both."

The version of evolutionary theory from which sociobiology springs severely elides the distinction between economic and reproductive organismic functions. Seeing all forms of behavior – even patently *economic* behavior – as just a manifestation of the competition to leave more of one form than another of genetic information, dampens the functional, economic side of any description of organisms or larger biotic systems. Seeing the organization of virtually any form of biotic system strictly from the perspective of the perpetuation and transmission of underlying genetic information, and understanding such processes to derive wholly from competition among organisms to maximize their genetic representation in succeeding generations (in part through the mechanism of "kin selection") begs the question of the significance, in a moment-by-moment functional sense, of economic activities. And it misses the main point raised by Caporael et al.: Relations among organisms, whether conspecific or not, or whether involving economic or reproductive activities, cannot be seen as wholly competitive lest essential ingredients of the fabric of living systems – especially social systems – be left out of our very description of nature.

Surely we require a sound, functional description of systems the origins of which we seek to explain in evolutionary terms. It is true, as so many evolutionists have been at pains to point out of late, that the causal pathways leading to the development of any given system may be quite tortuous; current functions may be a misleading guide to evolutionary history (cf. Gould & Vrba 1982). But Caporael et al. are reacting to the opposite, and arguably more pernicious, tendency: To frame descriptions of the structure of natural systems explicitly and exclusively in the light of a particular set of theoretical presuppositions. The resulting descriptions can only be as good as the theory on which they are based.

Cooperation is alive and well

C. Scott Findlay[a] and Charles J. Lumsden[b]

[a]*Department of Biology, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5 and* [b]*Department of Medicine, University of Toronto, Toronto, Ontario, Canada M5S 1A8*

Electronic mail: [a]*findlay@uohawa.bitnet* [b]*lumsden0@utormed.bitnet*

Research works best when theory and experiment communicate. The consequences of such a dialogue are apparent in the target article. Caporael and her colleagues have strong claims against the egoistic incentive (EI) perspective just because they have applied insights from game theory to their small-group designs. The sociobiology of EI and selfish genes will never be the same.

In concentrating on the putative distribution economy of early hominids, the authors' search for an explanation overlooks culture's presence per se. Surprisingly, culture learning is all that one needs to account for the evolution of non-EI altruistic behaviors that benefit one's group. When culture must be learned, *group selection* can be very effective in spreading non-EI behaviors through a population. Moreover, contrary to the authors' claims, the non-EI altruism so spread can be evolutionarily stable. This is so despite the traditional beliefs (for a

survey, see D. S. Wilson 1983) about how unimportant group selection is in populations where behavior is genetically determined. Biocultural populations, where learning and behavior depend on what others know and do, turn out to be much more susceptible to group selection effects.

As a result, we favor an evolutionary explanation structured along the following lines: Suppose that N social groups comprise the population, with p_ℓ, $\ell = 1,\ldots,N$ denoting the relative size of the ℓth group. Altruistic behavior is in part determined by a single diallelic locus. Take reproduction to be asexual, although the results for the diploid case are similar if there is no dominance. There are then two genotypes, A and a, with A denoting the altruistic genotype, and two phenotypes, B and b, with B denoting the altruistic phenotype, yielding four possible phenogenotypes: ab, aB, Ab, AB. We are interested in non-EI altruism, which benefits others at true fitness cost to the individual. If y_ℓ and z_ℓ denote the frequency of the altruistic gene and phenotype respectively in group ℓ, then

$$\bar{y} = \sum_l p_l y_l, \quad \bar{z} = \sum_l p_l z_l \tag{1}$$

give the corresponding mean frequencies in the population as a whole. Each generation, adults produce offspring who, after socialization and within-group selection, disperse according to the standard Wright island model, with a rate m at which each group exchanges members with the global population. Migration and reproduction take place continuously.

Let s be the fitness cost of altruism, measured in terms of decreased survival of the altruistic phenotype, and let f^j be the probability that an offspring with gene j develops the altruistic phenotype (i.e., adopts the non-EI altruistic behavior strategy from its congeners). Adoption probabilities will in general depend on both the offspring's genotype and the observed frequency of this altruistic behavior in the adult population, so that f^j has the form of a learning rule subject to gene-culture transmission (Lumsden & Wilson 1981): $f^j = f^j(z_\ell)$. If A denotes the altruistic genotype, we take this to mean that $f^A \geq f^a$ for all z_ℓ; that is, an A-genotype is more likely to adopt the altruistic phenotype than is the a-genotype. Furthermore, we assume that the greater the frequency of a trait in the parental generation, the more likely its adoption by the offspring – a form of cultural influence on behavior and development known as *trend watching* (Lumsden & Wilson 1981). A generic form of a trend-watching cultural transmission function is

$$f^j(z_\ell) = z_\ell[1 + \alpha^j(z_\ell - 1)], \tag{2}$$

where $-1 \leq \alpha^A \leq \alpha^a \leq 1$ represents each genotype's sensitivity to cultural transmission of the altruistic phenotype from the general social environment.

Group selection, as we will relate it to Caporael et al.'s findings here, takes the form of differential productivity: Whereas the non-EI altruists have a within-group fitness of $(1 - s)$:1 relative to selfish nonaltruists, the absolute survival rates of offspring in groups characterized by a high frequency of altruists among the parents exceed those of selfish groups. Let $\beta > 0$ be the regression of group fitness on the proportion of altruistic *phenotypes* in the parental generation. Then the rate-of-change in the relative size of group ℓ is just

$$\dot{p}_\ell = p_\ell\beta(z_\ell - \bar{z}), \tag{3}$$

where the dot denotes the time-rate of change dp_ℓ/dt. The corresponding rate of change for the frequency of the non-EI altruism gene in the population as a whole is then

$$\dot{\bar{y}} \approx V_{yz}[\beta - sd(2h - 1)] - sd[V_z + \bar{z}(\bar{z} - 1)][\bar{y}(1 - \bar{y}) - V_y]. \tag{4}$$

Here $h = \bar{y} + \bar{z} - 2\bar{y}\bar{z}$, $d = \alpha^A - \alpha^a \leq 0$, and V_y, V_z, and V_{yz} are the between-group phenotypic and genetic variances and covariance respectively, all of which are positive. In the case of purely genetic transmission, where the probabilities of adopting the behavior strategy are genetically determined and independent of the non-EI altruistic behavior frequencies z_ℓ in the group, the non-EI altruistic gene frequency in the population changes at the rate

$$\dot{\bar{y}} = D[V_y(s + \beta) - s\bar{y}(1 - \bar{y})]. \tag{5}$$

D is the difference in the penetrances of the two genes with respect to the altruistic phenotype and DV_y is the covariance of y and z under purely genetic transmission of the adoption probabilities.

To compare instantaneous rates of evolution under the two modes of transmission (gene-culture and purely genetic), we set the average difference in the adoption probabilities for gene-culture transmission across the population equal to the corresponding penetrance differential for pure genetic transmission. The between-group selection effects are then rendered identical for the two types of transmission (i.e. $DV_y = V_{yz}$), allowing comparisons to be drawn. The rate of evolution of the altruistic gene under this type of gene-culture transmission exceeds that under pure genetic transmission if

$$(2\bar{z} - 1)(1 - 2\bar{y})V_{yz} > 0. \tag{6}$$

Of particular importance to the data of Caporael et al. are the conditions under which a non-EI altruism gene will increase when introduced into a population initially monomorphic for the selfish-behavior gene (conditions of initial increase), and the conditions under which the same altruism gene can resist invasion by a selfish gene (conditions of evolutionary stability). The former corresponds to A increasing when rare ($\bar{y} \to 0$), the latter to A increasing when close to fixation ($\bar{y} \to 1$). For a small $\bar{y}$, we see from eq. (6) that less of a between-group (β) component of selection is required for initial increase under gene-culture transmission of the altruism gene if $\bar{z} > .5$, whereas for $\bar{y}$ near one, the corresponding criterion is $\bar{z} < .5$. When these inequalities hold, the range of conditions over which genetic evolution of the altruism gene can occur is expanded under gene-culture transmission relative to the purely genetic case, often dramatically so (Figure 1).

This kind of model for the evolution of non-EI altruism incorporates what we believe to be a critical element lacking in earlier theses about group selection's role, namely, the explicit characterization of how genetic and cultural information interact to mediate socialization. Such models suggest that, for populations with both genetic and cultural shaping of learning and behavior, group selection (and supraindividual selection in general) may be more important than previously realized. In particular, *genetic evolution of non-EI altrustic behavior can occur under conditions precluded if learning strategies were determined by genes alone, rather than by the interaction of genes and culture.*

Sociobiologists have for some time cautioned against taking selfish genes and inclusive fitness maximizing as the final answer, especially for human evolution, where the effects of co-evolving genes and culture are so poorly worked out (see, e.g., Lumsden & Wilson 1981, pp. 297ff. and the subsequent literature). Human sociobiology is best characterized as a field enlivened by competing answers to its "central problems," not monolithic consensus – popular summaries or news pieces to the contrary notwithstanding. Its continued vitality depends on mathematical theory working in partnership with such data as Caporael and her colleagues have provided here, data that cannot be ignored.

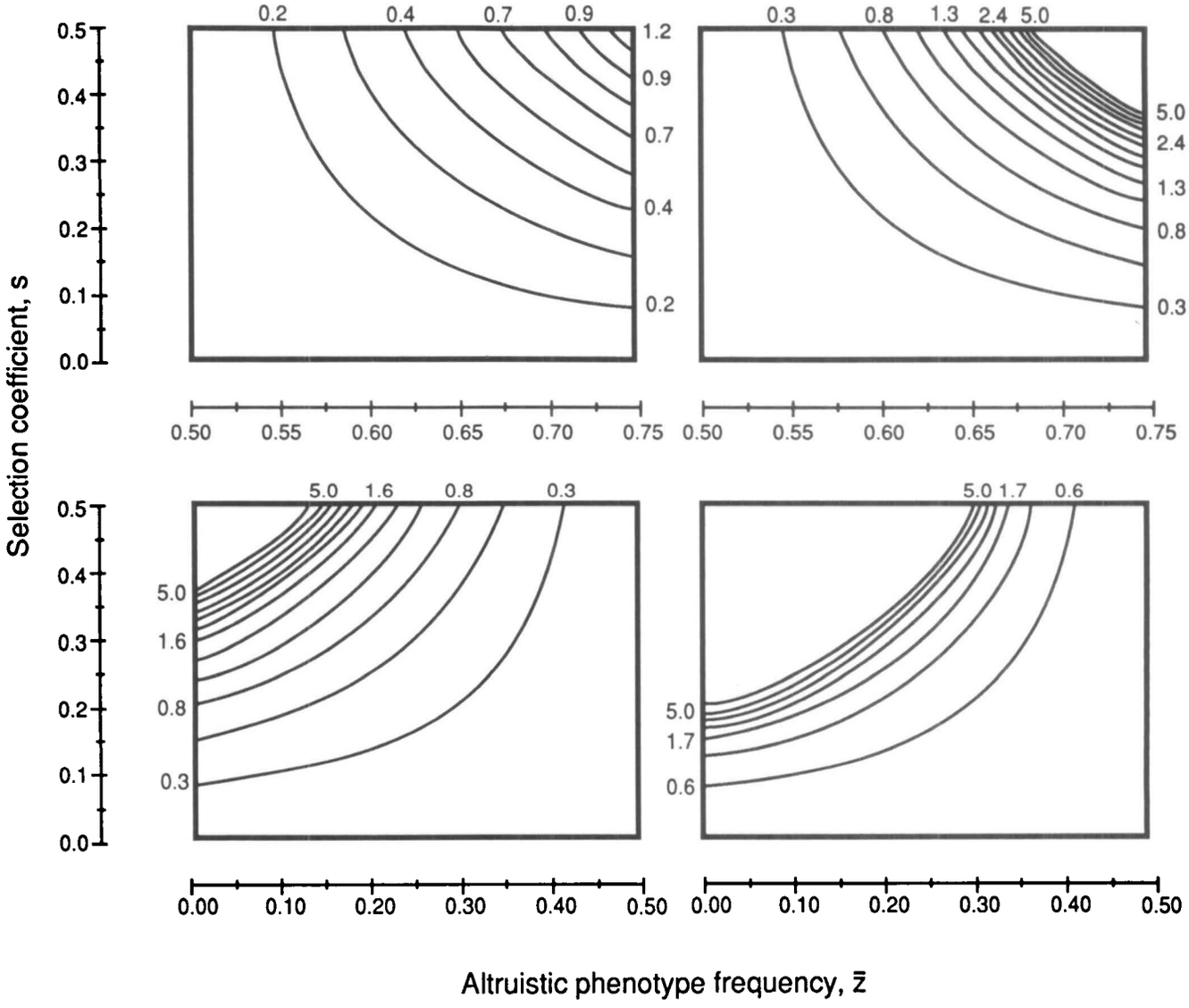

Figure 1. (Findlay and Lunsden). Enhanced efficacy of group selection under gene-culture (GC) versus pure genetic (G) transmission. Shown are contour plots of $\ln[\beta_{min}(G)/\beta_{min}(GC)]$, where $\beta_{min}(\cdot)$ is the minimum value of the between-group regression coefficient β required for the altruistic allele to increase in the total population, given a within-group selection coefficent s and altruistic phenotype frequency z. For all plots, $V_y = .005$, $V_z = .08$, $V_{yz} = .07$. For 1A and 1C, $d = -.40$, whereas for 1B and 1D, $d = -.80$, with A and B describing the situation when the altruistic allele is rare ($\bar{y} = .01$), C and D when $\bar{y}$ is close to fixation ($\bar{y} = .99$).

Ecological and social factors in hominid evolution

Robert Foley

Department of Biological Anthropology, University of Cambridge, Cambridge CB2 3 DZ, England
Electronic mail: *rafi10@phx.cam.ac.uk*

As the study of human evolution has become increasingly behavioural in orientation, a contrast has developed between social and ecological explanations for the patterns we observe in the archaeological and fossil record. The result has been that alternative models of social and ecological causality are seen as competing explanations for the same phenomena. Caporael et al. have opted for the primacy of social explanations, citing some of my work as examples of models that do not include social factors as primary causative factors in human evolution (Foley 1984; 1987). Particularly important here is my definition of the hominid "community" to exclude conspecific interactions (i.e., social relationships). That definition, however, refers specifically to the "ecological community," as used by ecologists studying interspecific relations. Such a community is not the same as a "social community" – i.e., repeated interactions between conspecifics – and therefore should not be taken to rule out the likelihood that many of the individuals of a species involved in an ecological community also operate within a social context, and therefore as part of a social community. As Caporael et al. point out, primates are strongly social; hence, social factors are bound to have been important during the course of human evolution. The critical question is not whether early hominids formed social groups, but what the specific character was of those groups (Foley & Lee 1989; Foley 1989).

The key question, then, concerns the nature of the interaction between social and ecological communities and how we move beyond social *versus* ecological explanations to discussions of how the one may be predicated upon the other.

This problem may be considered at two levels. First, the problem of communities: Early hominids existed in a matrix of other organisms (the ecological community). The extent to

which they were able to acquire resources in the context of competition from those interacting organisms was critical to individual survival and evolutionary success. This may seem to imply ecological determinancy and the primary importance of interspecific competition, which at one level is almost certainly true. Interspecific competition, however, is a more complex phenomenon. Although species may be described as competing with each other, in evolutionary terms this can be mediated through intraspecific mechanisms. For example, those individuals of Species A that are able to exist successfully in the competitive presence of Species B will do better (be selected for) relative to those individuals of Species A who suffer in the presence of Species B. If sociality reflects intraspecific relationships, it will also be dependent on the competitive context of the ecological community.

The second level arises from the foregoing conclusion: What is the role of sociality (which is essentially an expression of the nature of intraspecific competition) within this larger context? It may be argued that sociality will occur if it enhances an individual's reproductive success within its specific ecological context. This means that particular types and levels of sociality will reflect the extent to which the benefits of, for example, cooperation, exceed the costs. The high levels of sociality within the hominid lineage should therefore reflect the ecological conditions in which they evolved. Principal among these conditions are the use of high quality, patchily distributed resources (e.g., meat) and predatory risks in open environments (Foley 1987). In this sense, social explanations are not rivals of ecological explanations but depend on them, and Caporael et al.'s general hypothesis easily may be integrated in an ecological framework.

Many would argue (e.g., Humphrey 1976), however, that once sociality exists, it becomes its own driving force in evolution, and it will therefore evolve independent of ecological factors. This seems to be one of the main thrusts of Caporael et al.'s target article, and a view that is becoming increasingly popular in models of human evolution (e.g., Byrne & Whitten 1988). In particular, it has been argued that encephalization during the course of hominid evolution reflects increasing social complexity. The claim that social and other biological processes are independent is hard to maintain, however, when one is considering long-term evolutionary patterns. Although the enlargement of the human brain may well have been prompted by the need to cope with increasingly complex social problems and situations, this pressure would have existed for virtually all species that have evolved sociality. The brain has not evolved in other social lineages at the same rapid rate as in hominids, however, suggesting that social factors alone cannot account for the pattern we observe. [See also Glezer et al. "The "Initial Brain" Concept" *BBS* 11(1) 1988.]

The implication is that the advantages of sociality are held in check for most species, and that independent factors may inhibit the evolution of both large brains and complex sociality. Among primates and hominids, sociality seems to depend on a large brain, and a large brain is a major cost, in evolutionary terms. In particular, the maternal and growth costs may be such that, for most species, such an evolutionary development is strongly inhibited (Martin 1983). The conclusion to be drawn is that although the selective advantages of sociality, cooperation, and so forth, are high, so too are the costs, and it is only rarely that these constraints are escaped. Although the drive toward increased sociality is general, only among hominids did the ecological conditions exist that allowed this evolutionary drive to be realized. Defining those ecological conditions must remain a high priority in palaeoanthropological research. The value of such articles as Caporael et al.'s is that they show that this research must be carried out in conjunction with other approaches that emphasize the centrality of social factors in hominid evolution.

Honesty as an evolutionarily stable strategy

Robert H. Frank

Department of Economics, Cornell University, Ithaca, NY 14853

Caporael et al. have performed a valuable service with their meticulous survey of the evidence against the egoistic model. There is also considerable insights in their alternative account of human sociality. As they forthrightly acknowledge, however, a potential serious objection to their model is that automatic commitment to group goals is not an evolutionarily stable strategy (ESS). For example, a population of conformists could be invaded by individuals who obey group norms when others are looking, but behave opportunistically when they cannot be observed.

The authors respond that such a strategy may not be feasible because it places too great a burden on limited cognitive resources. And true enough, a person who lies and cheats frequently is likely to be caught sooner or later, no matter how clever or careful he may be. Yet we know that there are people who do break rules on selected occassions and seem to prosper in the process. Consider, for example, a person whose strategy is to follow all group norms except those that are manifestly impractical to enforce. Such a person would not tip at restaurants away from home, or make anonymous donations to charity, or vote; nor would he return a lost wallet found on a street corner. But where convenience dictated he would pour pesticide down his basement drain or toss litter on a deserted beach.

Such a person has little reason to fear direct group sanctions, even after a lifetime of such transgressions. And the strategy is simple enough to fall well within the bounds of human cognitive abilities. It follows, I believe, that the non-ESS critisism must be taken seriously. Without a satisfactory response to it, the Caporael et al. sociality model falls short as an account of nonegoistic behavior.

As I have argued in detail elsewhere, nonegoistic behavior becomes an ESS once we introduce a simple but important modification to Caporael et al.'s model (Frank 1988). That modification would be that the predisposition to behave nonegoistically should be observable, possibly at some cost. The following example captures the essence of the argument.

For simplicity, suppose the problem that tests people's egoism is a prisoner's dilemma like the one in Table 1. Suppose further that there are two kinds of people in the population, cooperators (Cs) and defectors (Ds), and that each type has a distinguishing feature (for instance, a sincere or sympathetic manner in the case of the Cs, and the lack thereof in the Ds). Suppose, finally, that this distinguishing feature can be observed only by those who incur a cost of 1 unit. People who incur this "cost of scrutiny" can tell Cs from Ds at a glance. To all others, the two types are indistinguishable.

The first question that confronts a C in this environment is whether it makes sense to incur the cost of scrutiny. By incurring it, he removes all risk of interacting with a D, and is thus assured of a payoff of 3 units (the 4-unit payoff Cs receive when

Table 1 (Frank). *Payoffs in a prisoner's dilemma*

		Player X: Cooperate	Player X: Defect
Player Y	Cooperate	4 for each	0 for Y 6 for X
	Defect	6 for Y 0 for X	2 for each

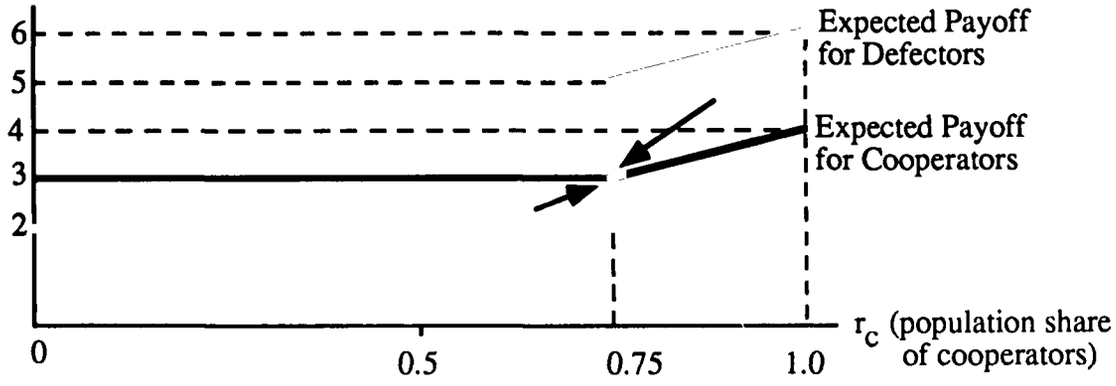

Figure 1 (Frank). Average payoffs for cooperators and defectors.

they interact, less the 1-unit cost of scrutiny). If he and other Cs do not incur the cost of scrutiny, he will interact at random with another member of the population, in which case his payoff will be either 4 or 0, depending on whether he happened to pair with a C or a D. If r_c denotes the share of Cs in the population, each C's expected payoff in the random-interaction case is given by

$$E_c = r_c 4 + (1 - r_c)0 = 4r_c.$$

In this example, it therefore pays for Cs to incur the cost of scrutiny whenever $r_c < 0.75$, the value for which $E_c = 3$. For example, in a population with $r_c = 0.5$, $E_c = 2 < 3$, which implies that Cs expect to do better by paying the cost of scrutiny. With $r_c = 0.9$, by contrast, $E_c = 3.6 > 3$, and so the Cs do better by simply taking their chances.[1] The expected payoffs for the two groups in this example are shown as functions of r_c in Figure 1.

Note in Figure 1 that the expected payoffs for Cs exceed those for Ds when $r_c < 0.75$, the region in which Cs incur the cost of scrutiny. Because of this difference in expected payoffs, the population share of Cs will tend to grow in this region. Conversely, the expected payoffs for Ds exceed those for Cs for values of r_c between 0.75 and 1.0, the region in which the two types interact at random. It follows that the share of Cs will tend to decline when $0.75 < r_c \leq 1.0$. The stable equilibrium in this example occurs when $r_c = 0.75$. At that value of r_c, both cooperation and defection are evolutionarily stable strategies.

This example is very much like the model suggested by Caporael et al. In both cases, people who behave honestly do so because they are predisposed to adopt group norms. The assumption I add is that people have the capacity to make accurate character judgments about at least some other people. In support of this assumption, there is evidence that the affective states that underlie a cooperative predisposition are accompanied by observable symptoms that are difficult to fake. The key to the cooperators' survival lies in these symptoms. Without them, cooperation is not an evolutionarily stable strategy. A blush may reveal a lie and cause great embarrassment at the moment, but in situations that require trust there can be great advantage in being known to be a blusher.

ACKNOWLEDGMENT

Research support was provided by NSF grant numbers SES-8707492 and SES-8605829.

NOTE

1. Whenever it pays Cs to incur the cost of scrutiny, it can never pay Ds to do so because Cs will be able to avoid them. Nor does it pay Ds to incur the cost of scrutiny when $r_c > 0.75$ in this example, because the Ds will always get a higher expected payoff in that region by simply interacting at random.

Selfish genes and ingroup altruism

Allan Gibbard

Department of Philosophy, University of Michigan, Ann Arbor, MI 48109-1003

Electronic mail: *allan_gibbard@ub.cc.umich.edu*

Caporael et al. urge evolutionists to consider "what psychological characteristics might have evolved as adaptations to living in small groups." In a beautiful set of experiments, they show that discussion can evoke willingness to contribute to one's group, even in the absence of any incentive we would normally think of as egoistic. All this I applaud.

These findings fit in well, I would insist, with talk of "selfish genes." Such talk is metaphorical: An organism, the claim is, will look as if its genes had designed it to replicate those very genes. Would metaphorically selfish genes make for literally selfish people? The paper contains a fine brief discussion of why they might not. Natural selection would produce not a perfect calculator assessing each action for its contribution to a person's inclusive fitness. Instead, it would shape psychic heuristics: Emotional and cognitive mechanisms that worked well in the conditions under which human beings evolved – where working well boils down to spreading one's genes in later generations. Now if defection was almost always fraught with grave reproductive risks and reliable mechanisms for identifying the exceptions would have been costly, selfish genes might favor simple heuristics that say to cooperate. Few of our ancestors, after all, left their groups one by one through a payout room. The authors' "sociality hypothesis" is that "the cognitive/affective mechanisms underlying certain human behaviors evolved in a social setting involving small face-to-face groups." This suggests the kinds of costs and opportunities selfish protohuman genes must have faced. The evolutionist's problem is to determine what kinds of psychic heuristics would pay off genetically in these circumstances.

Thus far I may be in full agreement with the authors (apart from the abstract with its quick disparagement of "selfish gene" theories). Caporael et al. avow, though, a "pessimism about the value of explaining contemporary human behavior in terms of fitness maximizing." If they mean explaining directly, in terms of current striving for reproductive success, I agree with this too. I would stress, though, the importance of careful thinking about fitness long ago. There is a place for game theory here, and the game theory will have the abstract structure of egoistic incentive theory. What recurrent kinds of situations among our ancestors might have borne most on the prospects of their genes' reproducing themselves? What would the "payoffs" to the genes have been from various combinations of possible strategies. What kinds of psychic mechanisms would these selection pressures shape?

Discussions in the target article leave puzzles that game theory might help unravel. "It was adaptive for ancestral humans to identify automatically with an ingroup and to accept its

goals as their own." It would be adaptive to do so in some cases and maladaptive in others, I would think. We should be on the lookout both for cues that prompt group feeling, and for cues that prompt alienation and self-protection. We should ask how heuristics tied to such cues would fare in terms of inclusive fitness. Moreover, even for group feeling, we need to fill in the evolutionary story. Group feeling was adaptive, we think, because it prompted a person to cooperate. Cooperation was adaptive because defection would be punished, perhaps by exclusion from the group. The story is incomplete, though, until we have explained tendencies to punish. To punish someone for not providing a group benefit may itself be costly. It can draw retaliation, and it can lose one opportunities for bilateral exchange. To enforce group solidarity is to provide a public good; why will anyone do so? My question is not what gain our ancestors saw in enforcing group norms. It is what payoffs their genes might have gotten, and what shape the punitive tendencies would have in order best to garner those payoffs. The game theory of small group enforcement may suggest something about our current psychic mechanisms: what kinds of cues they respond to and what kinds of actions they prompt.

I agree with the authors in rejecting egoistic incentive theory as the full story of our current proclivities. Let me say a word, though, about what egoistic incentive theory is. It is no single theory, but a loose collection of theories. The notion of a person's good is vague, and different specifications of what constitutes a person's good will give different content to the theory that each person pursues his own good. Pleasure and pain seem obvious cases of personal goods and ills, and money makes a good surrogate. Common sense is puzzled, though, when it turns to other things people want, such as fame, being valued, being worth valuing, or being father to a line of kings. Not everything a person will act to bring about counts as part of his good, by the commonsense standards that give the notion its meaning. Still, common sense is unsure where to draw the lines. Any particular egoistic incentive theory must specify what the goods are that each person is acting to obtain. What we learn from these experiments is that no such theory works if it keeps what will count as egoistic payoffs remotely within the vague confines of the ordinary notion.

Finally, a word on testability. Philosophers now tend to see red at intimations that respectably empirical theories must be directly testable. Scientists are in the business of testing, true enough, but a theory is never tested all by itself. The bearing of evidence on theory is more roundabout. A theory, after all, has no observational consequences taken alone. Deriving observational consequences requires auxiliary assumptions, tacit or explicit. The theory of the selfish gene is not directly testable, and neither is Caporael et al.'s sociality hypothesis. Neither is the worse in consequence; both are broadly empirical theories. How they can be made to respond to evidence must be a long story, but roughly, one assesses a theory both by its cogency and by its fecundity in contributing to fuller, more specific pictures that pass observational tests.

Folk psychology takes sociality seriously

Margaret Gilbert

Department of Philosophy, University of Connecticut, Storrs, CT 06268

Caporael et al. make several references to folk psychology, which they take to conflict with their position. I shall argue that, rather, they should see folk psychology as an ally.

I take "folk psychology" to be the account of psychological functioning implicit in everyday thought and talk. It includes both conceptual categories and forms of explanation. The human sciences would do well to characterize it accurately, for subjects will generally perceive themselves and others in folk-psychological terms, and these perceptions may be expected to help generate behavior (see Winch 1956; 1958; also Taylor 1971). If folk psychology envisages only selfish motives, this will provide a degree of support for the hypothesis that human behavior is always selfishly motivated. Those inclined to reject this hypothesis will need to explain why folk psychology diverges from the truth as they see it.

So, does folk psychology ally itself with those theorists who see human behavior as "basically selfish and individualistic?" There is more than one way to interpret this question. I shall consider two options and answer both in the negative.

Consider first the everyday meaning of the word "selfish." On this construal selfish behavior aims only to advance the subject's personal well-being. There may of course be some cynical folk who regard human behavior as basically selfish in this sense. But in everyday thought and talk people often give and accept explanations of behavior in terms of a person's generosity, concern for others, and so on, explanations that are understood to rule out the idea that the behavior in question was selfish. This confutes the idea that folk psychology assumes that people are basically selfish in the everyday sense under discussion.

Now consider a claim such as "she contacted him because she wanted to relieve his suffering." As it stands, this hardly imputes a selfish motive in the sense discussed above. (Compare, "She contacted him because she thought this would lead to her promotion.") One might, however, characterize it as an "individualistic" explanation, for it characterizes as crucial to the subject's motivation the fact that someone (the subject) *personally* wanted something. One might also characterize it as selfish in a highly technical sense: It is the subject's *own* personal wants, not another's, that are at issue. In the example, the subject *herself* wants such and such.

Does folk psychology assume that human behavior is basically selfish and individualistic in the sense that satisfaction of the subject's own personal wants (preferences, desires, goals) is the "organizing principle" of human behavior (sect. 1.5, para. 1)? One can see this as a question about what forms of reported "practical reasoning" are considered intelligible. So we should ask whether people judge that a premiss about what "I want" or "I prefer" is a necessary component in any course of reasoning that is sufficient rationally to motivate action. I shall now argue that this is not so. Folk psychology countenances a quite different, alternative structure of motivation. This allows that behavior need be neither "selfish" nor "individualistic" in the senses now at issue.

I have in mind certain forms of reasoning from premises involving the pronoun "we," including premises of the general form: "We seek such-and-such" or, "We value such-and-such." Concrete examples of such premises include, "We want to preserve the commons" and, "Our goal is cleaner air."

Here is a simple example of reasoning that would be judged self-sufficient and valid from a practical point of view:

Premise (1): We seek to preserve the commons.

Premise (2): The commons will be preserved if and only if I do not graze a cow on the commons.

Conclusion (3): I should not graze a cow on the commons. I do not claim that in all situations in which (1) is perceived to be true, (2) will be perceived to be true. I am arguing simply that (3) would normally be judged to follow from (1) and (2) as they stand. In particular, *there is no need to posit a suppressed premise* such as, "I seek to preserve the commons," or "I desire that we get what we seek" (see Gilbert 1989, p. 424 ff; see also Sellars 1963b and elsewhere, references in Gilbert 1989).

One might wonder whether premises of the "We seek . . ." form are decomposable, from a logical point of view, into a conjunctive premise involving a set of judgments about particular individuals such as "I seek, and each of these others seek." Though in certain cases "We seek . . ." could be interpreted as having this conjunctive logic, it can be argued that quite standardly it does not. Rather, people use "We seek . . ." of themselves and certain others when they understand that there is something like a tacit agreement among them jointly to pursue a

certain goal. (I argue for this at length in Gilbert 1989, especially pp. 167–203; see also Gilbert 1987.) If this is right, one can see how, from "We seek X," plus appropriate premises about the means to achieve X, a conclusion about what an individual should do can follow directly, without the interposition of any assumptions about what *that individual* wants or seeks. Indeed, no single *individual's* aims need be referred to.

Some further points, very briefly: It seems that part of the logic of "We seek such-and-such," is that it implies a commitment not to act for the sake of personal benefit if this will prejudice "our goal." In relation to a specified area of action, one has "given oneself over" to "our goal" (see Gilbert 1989, pp. 424–25).

A typical context in which "We" is used is a discussion: Participants are involved in a joint project (the discussion itself) and this may generate further joint projects. Given the logic of the corresponding "we" statements, it is clear that any joint project stands to generate a degree of self-abnegating and group-enhancing action.

The fact that folk psychology apparently allows for nonindividualistic motivation, and the specifics of its concepts, should not be overlooked in the course of the scientific debate in which the authors of "Selfishness examined" are engaged.

The case of the "redundant" donor: Neither egoistic nor altruistic

Gene M. Heyman

Department of Psychology, Harvard University, Cambridge, MA 02138

The assumption that individual behavior is governed by rational self-interest has maintained a powerful hold on social science, psychology, and biology. In economics, it is usually assumed that individuals choose in order to maximize utility; in psychology, recent accounts of learning and motivation have proceeded from the assumption that organisms maximize reinforcement rate (or some similar variable); and in biology, genes have been depicted as "selfish" and rational according to the criterion of promoting self-replication. However, egoistically rational motivational theories have not gone unquestioned. It has been argued that such theories are incomplete because they ignore fundamental psychological factors, such as emotions (e.g., Frank 1988) and cognitive biases (e.g., Kahneman et al. 1982). Researchers who have compared matching law and optimization predictions have often concluded that matching is the more general principle and that when optimization does occur, it can be shown to be a special case of matching (e.g., Herrnstein 1981).

In the target article, Caporael, Dawes, Orbell, and van de Kragt add their names to the list of critics who find that "economic man" provides too limited a model for human and subhuman motivation. They note that humans and other creatures often appear to behave altruistically rather than selfishly. Rational self-interest theories, which Caporeal et al. refer to as "egoistic," account for this by trying to show that selfless acts, such as heroism and team play, are misleading, and that upon analysis such behavior proves to be as egoistic as the more transparent maneuverings of the entrepreneur. Caporael et al. reject economic retranslations of altruism and offer instead a dualistic system. There are, they imply, two types of motives: selfish ones and altruistic ones. The problem, of course, is how to distinguish between them. Caporael et al.'s solution is implied by the dichotomy. They designed experiments that they claim eliminated – subtracted out – egoistic motives so that the resulting behavior was necessarily altruistic.

Caporael et al. ignore the possibility that human motivation is not exhaustively categorized by their two-part scheme: egoistic or altruistic. However, a large percentage of the subjects in the social-dilemma experiments described by Caporael et al. made decisions that appear to have been guided by a principle that was neither egoistic nor altruistic. These subjects, who comprised about one-half the sample, donated their initial dole, $5.00, even though they stated that the donation was unlikely to help anyone else and they knew that it would be subtracted from any profits they might gain. Caporael et al. label this sort of donation "redundant," but they do not give the phenomenon special emphasis. Redundant contributions, however, as will be argued below, are theoretically important and call into question the apparently straightforward dichotomy between egoistic and altruistic motivations. A summary of some aspects of the experimental procedures and of Caporael et al.'s approach will help show why this is so. Following this, it will be argued that redundant contributions were motivated by the desire for equitable social relations, and that this principle, equity, is neither simply altruistic nor egoistic.

In the standard procedure, each subject received a $10.00 bonus if a critical number of fellow participants (e.g., 5 of 9) donated $5.00. If less than the critical number donated, there was no bonus, and if more than the critical number donated, there was no additional benefit. Thus individual subjects who made what they believed to be a redundant donation absorbed a $5.00 loss and to the best of their knowledge helped no one else. Put somewhat differently, a redundant contribution was by definition not egoistic (the $5.00 loss), and the belief that the donation was redundant implies that the incentive could not have been altruistic.

Caporael et al. explain redundant cooperators in terms of a sense of allegiance among players. According to a process referred to as "in-group biasing" and "group identification," individuals react to themselves and others as exemplars of the group rather than as differentiated individuals. Caporael et al. write (and this is the whole of their account of redundant donors): "Social identity, by reducing the distinctions between one's own welfare and that of others, explains our subjects' willingness to contribute in the face of expectations of their contribution's being redundant." Caporael et al.'s explanation assumes that the motivation for redundant contributions was group welfare, but according to the pay-off structure, this could not have been the case – a redundant contribution, by definition, could not intentionally help anyone. One might argue that the subjects did not understand the social-dilemma contingencies; however, Caporael et al. claim that the experimental procedure ensured that subjects clearly grasped the consequences of cooperating and defecting. Thus, the question remains, why did so many subjects contribute if they did not expect the contribution to enhance the winnings of the other players? A plausible answer is provided by equity theory. Caporael et al.'s summary of egoistic theories suggests that they would label equity theory egoistic. However, below it is suggested that equity theory differs in significant ways from rational self-interest.

According to equity theory, social interactions are evaluated in terms of costs and benefits relative to the costs and benefits accruing to the other participants in the transaction (Brown 1986; Homans 1974). It is important to note that the scale is relative, not absolute. Thus, one prediction is that relative discrepancies rather than absolute levels of deprivation motivate action. Another, stronger, prediction is that under certain conditions (1) people can feel overcompensated and (2) they will make efforts to reduce personal undeserved advantage. Brown and Herrnstein (1975) cite an example. Adams and Jacobsen (1964) found that college students would work considerably harder at a proofreading task if they believed they were overpaid. The redundant donor in the Caporael et al. studies may be a second such example.

Some subjects may have felt that it was not fair to earn more ($15.00) for defecting (not donating) when others earned less ($10.00) for cooperating. Unlike the college students in the Adams and Jacobsen study, the social-dilemma subjects did not

have the opportunity to invest more work, so the only way to reduce the discrepancy in equity was to reduce their pay. Thus, donations occurred even when they believed they would be of no assistance to the group. This interpretation could be checked by comparing the proportion of equity-based verbal reports from those who defected and those who donated with the understanding that a donation was redundant.

The assumptions associated with equity theory and the approach advocated by Caporael et al. differ in interesting ways. Caporael et al. assume that social interactions are motivated by either strictly egoistic or strictly social incentives. For example, their basic methodological strategy is to remove egoistic incentives and test whether cooperation persists. If it does, then by the logic of their strict dichotomy, it must have been maintained by social incentives. In contrast, the desire for fairness or justice, as depicted by equity theory, is synthetic in that it entails both egoistic and social ends. Individuals tally their costs and benefits and are goaded into action if the ratio seems discrepant. However, whether a cost-benefit ratio is acceptable is determined by how others are doing. Social mediation has two consequences: First, if all parties in the social exchange come to a similar judgment, the disparity, except for its sign, is equally felt. Second, an action that reduces a disparity in equity for one party necessarily reduces disparities in equity for all parties. For example, by working harder or taking a self-imposed pay-cut, an overcompensated employee can reduce the resentment that fellow workers are likely to feel. Thus, in contrast to the dichotomy of social and egoistic ends assumed by Caporael et al., the desire for equity simultaneously serves both egoistic and group goals.

It should be pointed out, though, that equity may in some sense be a deficient or unstable equilibrium state. Note that the desire for equity does not necessarily lead to an optimal cost/benefit ratio, but simply the same ratio as others obtain. For example, if the participants in the social-dilemma experiments had not felt compelled to act fairly, they could have gone home at least $5.00 richer, and done so without cost to others (assuming that their estimates of redundancy were more often correct than not, which appears to be true). Thus, if equitable relations are generally not optimal, it is surprising that they are not upended by more nearly optimal social exchanges.

Are redundant contributions a unique consequence of noniterated social dilemmas or representative of a class of social motivations? The answer is not clear. On the one hand, the desire for equity is not the only incentive that promotes behaviors that enhance the welfare of more than one member of a social exchange. For example, good citizenship is rewarded and is in turn helpful for others. On the other hand, equity-based solutions are not necessarily optimal, and it is possible that equitable but nonoptimal social relations are unstable.

Selfishness reexamined: No man is an island

Alasdair I. Houston[a] and William D. Hamilton[b]

[a]King's College Research Centre, Cambridge University, Cambridge CB2 1ST, United Kingdom and [b]Department of Zoology, Oxford University, Oxford OX1 3PS, United Kingdom

Electronic mail: *houston@vax.oxford.ac.uk*

The target article combines experiments on human choice in a social setting with some general discussions on selfishness and cooperation in an evolutionary context. We find the experiments interesting and agree with the authors that they provide a challenge to sociobiological views of human behaviour. They do not seem, however, to imply such a radical difficulty for current sociobiological and evolutionary theory as Caporael et al. suppose. In the first place, the extension of theory that they themselves suggest is needed is not really outside the existing framework. This is the major point discussed below. In the second place, the experiments, although throwing light on a socially welcome aspect of human nature, may be no more puzzling, ultimately, than the maladaptive behaviour of a butterfly beating on a windowpane. In other words, psychologists' experiments, like other single-encounter social situations of civilised life, may be just as evolutionarily novel for us, and therefore just as confusing, as glass panes are to butterflies. This idea does not detract from the value of the experiments, which, if the idea is right, warn us that our behaviour is not stable at its present level of cooperativeness: Cheats, though uncommon, may be doing uncommonly well, and the proclivity to cooperate may be slowly declining (see also Hamilton 1971; 1975). Thus, the target article and its discussion may encourage us to consider remedial measures.

Turning now to Caporael et al.'s own interpretations of their findings, one of our major objections concerns the characterisation of egoistic incentive (EI) theories and an apparently alternative Darwinian framework. Selfish gene theory is of much broader scope than Caporael et al. suppose. It is in fact the basis of the current Darwinian approach. Although Darwin himself couldn't refer to genes, because they weren't known, the present theory has the same definite and individualistic spirit of interpretation that he used.

As follows from this, the authors' view of EI theory as part of sociobiology is open to objection. We clearly cannot speak for everyone who has used evolutionary considerations as a basis for speculation about human nature. To the extent that some people see selfish gene theory as a basis for only selfish behaviour, Caporeal et al. may be justified in the line that they take. We would argue, however, that their view of selfish gene theory and EI theories is misleading. They say: "According to EI theory, people will always choose the selfish strategy in social dilemmas" (sect. 1, para. 5). This is not true if kinship is involved or if there are repeated interactions. The authors go on to consider a variety of ways in which an individual's incentives may change in such a way that cooperation emerges. These ways are coercion, conscience, reciprocity, and inclusive fitness (IF) maximization. We feel that the last two categories do not really belong with the first two. The authors give the impression that IF theory attributes explicit incentives to individuals such that each rationally computes the IF associated with various options ("behavior . . . can be explained in terms of people's attempts to maximize their inclusive fitness" sect. 3, para. 12). Rationality is actually seen as a corollary of EI theory (sect. 1.4, para. 2). But IF says nothing about rationality, or the exact nature of the mechanistic and psychological processes involved.

Coercion and conscience may indeed change an individual's incentives so as to remove dilemmas, but reciprocity and IF theory do not rest on this basis. IF theory (Hamilton 1964) is a way of looking at the spread of genes, that is, a way of implementing the Darwinian approach that the authors themselves favour. Similar remarks can be made about reciprocity ("Darwin's emphasis on individual advantage has been formalized in terms of game theory. This establishes conditions under which cooperation based on reciprocity can evolve" [Axelrod & Hamilton 1981, p. 1,396].) The fact that such evolutionary conditions may underlie human behaviour does not necessarily remove a dilemma. We may in fact be conscious of a conflict between "our" interests and the maximization of IF.

Although we have argued that conscience differs from reciprocation plus IF as a potential explanation for cooperative behaviour, we are not barred from believing that conscience has been shaped by our evolutionary past so that certain sorts of behaviour are encouraged. (This form of evolutionary explanation may be an example of what Caporael et al. refer to in section 1.4, paragraph 3 as "positing intervening selective processes." The use of such a two-level approach is by no means unique to cultural evolutionists: In discussions of the evolution of behaviour it is common to distinguish the evolutionary advantage and the behavioural mechanisms – see Houston 1980; Houston & McNamara 1988, for further discussion and references.) In

their treatment of conscience (sect. 2.7), Caporael et al. ignore the possibility that conscience may be group-biased. People may be automatically applying rules for cooperating in a way appropriate to outdated circumstances whenever they find themselves in any face-to-face or generally "believed-in" group, treating this group as if it were not so ephemeral as the conditions of the experiment say – as if they couldn't believe that it could be so ephemeral because, in the Palaeolithic, for which their reactions are evolved, groups never were.

And in reality, when participants are allowed to discuss face to face they may also not want to be branded even possible or probable cheaters, in case the same people are met again in real life outside. If the experimenters want to get as close as possible to true rational motivations, maybe they should do everything by mail: Send out the dollar bills with a description of what the game is, how things will be arranged and grouped when and if the bills are sent back, and so on. The improved generosity in face-to-face discussion groups suggests to us that either the rationality of the subjects refuses to unlink the experiments from reputations valued in real life or else it reflects a quasi-innate reaction to the perceived formation of a group, a reaction that was adaptive long ago when groups were hardly ever ephemeral, and is now no longer adaptive. We suspect that the authors favour the second alternative, but they somehow seem to see it as contrary to sociobiology or selfish gene theory. This fundamental issue recurs in the context of revealed preference.

Caporael et al. says that the sociobiology of choice rests on the revealed preference approach. We do not agree with this claim. Revealed preference amounts to seeing a concept such as utility as inferred from choices; as long as choice is consistent, such a function can be constructed to describe choice. A sociobiological approach, whether it is called selfish gene theory, IF maximization, or a Darwinian framework, tries to find costs and benefits that would render the observed behaviour favoured by natural selection. Even if behaviour is consistent, no such explanation may be possible (see also Houston & Staddon 1981; Houston & McNamara 1988). In other words, behaviour can be consistent without making evolutionary sense. [We note in passing that a male elephant seal has been recorded as showing consistent and indeed seemingly thoughtful "psychopathic" behaviour towards members of another species (Best et al. 1981) and that dolphins sometimes extend altruistic behaviour to members of other species (Connor & Norris 1982).]

In the context of Caporael et al.'s emphasis on the importance of groups in human evolution, it is interesting that in his discussion of the prisoner's dilemma, Sen (1973) writes: "I would argue that the philosophy of revealed preference approach essentially underestimates the fact that man is a social animal, and his choices are not rigidly bound to his own preferences only" (pp. 252–53). We feel that Caporael et al. have underestimated the extent to which previous work has considered this idea. Wilson (1975) writes: "It is likely that the early hominids foraged in groups" (p. 567), and he goes on to mention that this may have given some protection from predators. Hamilton (1971; 1975) emphasises group living and attendant group selection in early man, pointing out that because group members would have been almost always related, group selection and IF theory offer equivalent interpretations.

The rationality of cooperation

Leonard D. Katz

Department of Philosophy, University of Connecticut, Storrs, CT 06269-2054
Electronic mail: *ldkatz@uconnvm.bitnet*

I shall suppose that Caporael et al. are right in their claims that people readily come to identify with a group, to respect what they perceive as group decisions or demands, and to contribute resources for group welfare, even at some cost to their individual advantage. But I shall challenge their statement, made most baldly in the Abstract, that their subjects' decisions to contribute resources are based "frequently on irrational grounds." Caporael et al.'s coupling of "individualistic, selfish" with "rational" might suggest that they accept as definitive of rationality the "economic man" view – a view they reject as descriptive of human behavior. But quite aside from any question of whether the unselfish pursuit of group welfare as an end is to count as irrational simply as such, in section 2.5 Caporael et al. raise a problem that seems to impugn the means-end rationality of many of their contributing subjects.

Our own decisions often count as part of our evidence about what others will decide when similarly situated. Accordingly, Caporael et al. suggest, their contributors' estimate of how many other group members will contribute (elicited *after* their decision has been made) has been raised owing to their own decision to contribute. But many should, in consequence, believe their contributions redundant for achieving the payoff for the group – in which case a present decision to contribute must, it seems, be irrational. Caporael et al. say that these subjects' behavior (based on a past decision) is "inconsistent" with their (postdecision) expectations. This, presumably, is the basis for the claim about "irrational grounds" in the Abstract.

But what should the poor subjects have done to be clear of this charge? Presumably, they should have considered beforehand now the fact of their decision would change their decision-relevant evidence. First, the need for doing this would not be salient to nonexpert subjects. Second, it would be no easy task. Subjects would have to keep track of their probability estimates and of their attitudes toward the different risks involved. Contributing will be a reasonable strategy for some of those who have group welfare as an end. And others may easily mistake their own case for one of these. Third, some sets of initial probability estimates and evidential weightings (for changing these estimates on the basis of new information about one's own decision) should result in oscillation between a tentative decision to contribute (which leads to greater confidence that one's contribution will be redundant, and hence to its withdrawal) and the resulting tentative decision not to contribute (which results in enough, but not too much, lowering of one's confidence in others' contributing, so that it once again seems rational to contribute, given the greater chance one will make the decisive difference . . .). It is at least not irrational for a person (especially given typical "real world" time, information, and processing constraints) to follow a general policy of sticking with an original decision in cases that "go on forever" (or seem to do so).

It is still more important that we be aware that socialized human subjects (whatever you tell them!) will have purposes and policies going beyond short-term individual and group profit. People don't like being taken for suckers, but they also like to pull their weight and do their share in even an incipient fair scheme of cooperation. (Dissonance reduction may be a further source of decision-dependent expectation bias.) They may even have autonomous desires to do their share *even when they believe their contribution to be redundant*, at least so long as they are not freed of their obligation to contribute by group decision or some other especially salient symmetry-breaking consideration. If nonmoral payoffs (whether for individuals or groups) are not our only ends, then even a contribution ill-designed to increase such nonmoral payoffs need not be an irrational choice of means.

Even absent any freestanding ultimate moral concern for justice as such, the best policies for imperfectly social animals such as ourselves will generally not be all that different from one based on an autonomous concern for fair play. They will give a high weight, especially when compared with trivial short-term individual or group advantage, to increasing the opportunities for future cooperation. Caporael et al.'s "stranger" subjects

were fellow students or fellow townsmen, and may have perceived each other as such. They certainly perceived each other as fellow human beings, and as such capable of communication and cooperation on an indefinitely extendible scale. The willingness to cooperate is a most important instrumental public good and one susceptible to erosion when expectations of cooperation are disappointed. Our ability to (reversibly) put aside our individual aims, regarding ourselves as constrained by social goals, decisions, or demands (either explicit or perceived as naturally arising from the circumstances of the case) seems an appropriate adaptation to our general social situation. The contributors presumably wanted to be perceived collectively (even when not individually) by each other and by themselves (even when not by others in the redundant contribution case) as people capable of sustained cooperation, in part because they also wanted to *be* people of this kind – in a world that is, after all, larger than the one-shot social dilemma of the experimental design. That design cannot prevent people from importing the policies they follow when they pay their taxes and make charitable contributions. Such policies can be approved by us as rational means given our mixture of group-biased (but indefinitely extendible) altruistic and selfish ends and the dependence of these ends on our own and others' dispositions to cooperate. But then these policies are not irrational.

Selflessness examined: Is avoiding tar and feathers nonegoistic?

Douglas T. Kenrick

Department of Psychology, Arizona State University, Tempe, AZ 85287
Electronic mail: *atdtk@asuacad.bitnet*

Ecologist James Lovelock has noted that "new ideas follow a predictable progression. First people say, 'It's absurd!' Then they say, 'Maybe.' And finally they say, 'We knew it all along'" (quoted in Cowley 1988). Over the last decade and a half, sociobiology has progressed from the absurd through the maybe phases. Given the nomological network of evidence from cross-cultural, comparative, neurophysiological, and behavior genetic research, it is no longer defensible to hold the position that evolutionary analyses are irrelevant to human social behavior. However, the word "sociobiology" itself retains a poor marketing value among social psychologists. The mode du jour is to use the term "sociobiology" for "bad" evolutionary theory, and to distinguish it from informed evolutionary theory of the kind that any intelligent person should have known all along. But sociobiology is a very broad term, referring simply to the study of the biological basis of social behavior (Wilson 1975). The field of study is defined, plainly and simply, by the general assumption that evolutionary principles will apply not only to morphological structures, but also to behaviors. [See *BBS* multiple book review of Kitcher: "Vaulting Ambition" *BBS* 10(1) 1987.]

Pardon the language, but this sounds like sociobiology. Sociobiology implies no necessary commitment to one particular proximate fact or another. A "sociobiologist" (if anyone with a poor sense of public relations wanted to admit to being one) is certainly not committed to the view that every decision made by an individual will be "selfish." The notion of "inclusive fitness," which has generated ample supportive data (Trivers 1985), suggests one condition under which individual altruism might have been selected. Although Caporael et al. indicate some discomfort with this, evolutionary analyses imply "selfishness" not in a literal sense, but in an ultimate genetic sense. If particular classes of individual selfless acts would, on average, promote survival, reproductive success, or benefit for related individuals, they will be selected (for *genetically selfish* reasons). At this level, the theory is not a tautology. There have no doubt been animals (human and otherwise) whose genes inclined them to act in genetically unselfish ways. The likely selection pressures against those tendencies make it a good bet that few such inclinations will have survived. Nevertheless, it is perfectly consistent with sociobiological assumptions to consider conditions under which altruism could have been selected at the group level (Trivers 1985; Wilson 1975).

It seems to me that the evolutionary approach presented by Caporael et al. is a logical extension of the same reasoning that gave us the inclusive-fitness and reciprocity models of altruism. As the authors note, the "sanctions for uncovered hypocrisy in most societies are severe – occasionally extending to total ostracism from the group." Other evolution-based experiments suggest that humans are especially cognitively tuned to detect a cheater (Tooby & Cosmides 1989). Even in recent American history, the penalty for defecting from group rules was tarring and feathering. Thus, this evidence, and the evolutionary model offered by Caporael et al. is consistent with a view of selfishness at the level of the gene. As in the case of kin selection, we now have a nondilemma. The best heuristic for our ancestors was "be a group-player, or (on average) die."

Caporael et al. suggest that any tendency to welch on group commitments was selected against in our ancestors. Their reasoning that humans are selected for group-living is in line with arguments made by other evolution-minded psychologists (e.g., Hogan 1982). A reminder of our evolutionary history serves as a useful caution against the social psychological view of humans as selfish individualists, and the present effort is consistent with other work that uses an ultimate evolutionary model to explain seeming "irrationalities" in social psychological economics (Kenrick & Trost 1989). Although I therefore agree with their general framework, whatever the label, I think that selfishness may be found here at a level much more proximate than the gene.

Extrinsic and intrinsic egoistic incentives are not eliminated. Consider the dilemma faced by a potential Judas in one of the groups described by Caporael et al. The subject is given the chance to walk out with his 15 pieces of silver, and none of the other group members need ever know that he did his selfish deed. If the subject declines this "rational" opportunity, is it truly unselfish? I think not. There are several glaring sources of egoistic motivation here. To begin with, the potential Judas must face someone to collect his payoff. As a minimum, the subject can expect the experimenter to know of his treachery. Because people are notorious gossips, how is the subject to know that his selfishness is safely secret? And if he meets another group member later he may be forced to lie, with the possibility that his nonverbal responses will give him away. The authors themselves note the dangers of reprisals for breaking group rules. If the subject is found out, he may suffer consequences for which an extra five or ten dollars would be little solace.

And the egoistic incentives need not come from without. Research on children's socialization indicates that altruistic behavior goes through three stages: (1) for primary grade children, altruism is self-punishing; (2) for slightly older children, altruism is used instrumentally to elicit social approval; (3) for adults, the reward value of altruism is internalized, and adults will act prosocially even when no one is looking (Cialdini, 1981; Kenrick et al. 1979). So, money is not the only relevant egoistic incentive involved here. Subjects can feel good about themselves for "doing the right thing." This self-reward is no doubt maintained by intermittent extrinsic reinforcement, and modelling experiences. On the other side, the "unselfish" subject can avoid the guilt of going the Judas route.

Caporael et al. argue that "egoistic incentive notions are really untested metatheories" and they attempt to defend themselves against the sort of reasoning I used above by noting that "it is always possible to hypothesize arbitrary egoistic payoffs post hoc." However, there is another set of social psychological studies that bears on this issue, and that offers support for an egoistic viewpoint. Batson and his colleagues have found that

subjects who empathize with a victim will be more altruistic than nonempathic observers (Batson et al. 1983). In addition, empathic helping is not sensitive to certain variations in the cost of helping. At first glance, these findings argue against an egoistic incentive viewpoint. However, newer findings argue that empathic helping may be egoistic after all (Cialdini et al. 1987). Cialdini and his colleagues view altruism as a socialized self-reinforcer – people act altruistically to make themselves feel better, and to avoid the negative feelings associated with acting selfishly. These authors have replicated Batson's empathic helping studies and found evidence that egoism is lurking beneath the surface of this seemingly selfless behavior. Subjects in Cialdini's experiments indeed report unpleasant moods when they have been asked to empathize with a victim's plight, and they will help if that personally aversive state is not interrupted before the prosocial opportunity. However, if they are given a reward which dissipates that unpleasant mood, or if they are given a drug that they think will freeze their mood, subjects' helping drops to control levels. Thus, there are a priori reasons to argue that Caporael et al. have not eliminated the range of egoistic incentives that might have played a role in this research. Simply eliminating a few (even quite a few) incentives is not sufficient to seal the case for pure selflessness.

As a related point, the fact that subjects' self-reports do not support egoistic explanations is hardly strong evidence one way or the other. Jean Paul Sartre observed that sincerity is the mark of an actor who is taken in by his own act. Caporael and her colleagues note the research on nonconscious determinants of behavior which indicates that experimental subjects are not always able to describe the causes of their own behavior (Nisbett & Wilson 1977). And even if they know the causes of their behavior, subjects' explanations may be biased for self-presentational purposes (Tetlock 1981). Thus, a sincere expression of group interest is suspect for several reasons.

In sum, I believe that (a) Caporael et al. make an unjustified distinction between "sociobiology" and equivalent evolutionary approaches to social behavior, and that (b) they have not proven the case against selfish motivations, even at the proximate level. Despite these reservations, I believe they have presented a fascinating and thoughtful series of studies, and embedded them in a provocative framework. Their evolutionary model connects their work with other areas of social psychology, and with other developments in the life sciences. By moving beyond a narrow proximate analysis, they have made a worthy contribution to the emerging field of evolutionary social psychology.

Sociality versus self-interest in human evolution

Bruce M. Knauft

Department of Anthropology, Emory University, Atlanta, GA 30322

The "sociality hypothesis" forwarded by Caporael et al. provides an important potential alternative to egoistic incentive (EI) theories and deserves serious attention. The hypothesis is only nascently and unevenly developed, however, and rests in this target article on a quite narrow empirical and conceptual base.

Sociality and the study of human evolution. Though not noted by Caporael et al., the sociality hypothesis in the study of human evolution has explicit roots at least as far back as Kropotkin's *Mutual aid: A factor in evolution* (1902; see Gould 1988; cf. Vucinich 1988, Ch. 10). In anthropology, the central place of sociality in early human development has a legacy in the work of (among others) Levi-Strauss (1949) on primal structures of human kinship alliance; Leslie White (1949) on evolutionary features of symbolic communication; Geertz (1962) on culture and the evolution of mind; Hockett and Ascher (1964) on "the human revolution"; the late Glynn Isaac (1978; 1984) on hominid collective affiliation; and selected contemporary researchers on the evolutionary significance of intense sociality in simple human societies (Ingold 1987; Knauft 1988a; 1988b; 1988c; Testart 1985; cf. Lee & DeVore 1984).

The anthropological influence of group-collective perspectives on evolutionary theory has diminished, however, punctuated by William's (1966) trenchant critique of group selection and the practical problems of establishing early humans as an exception to his generalizations. In particular, the mental and communicative abilities of humans – the symbols of language that make human cultural transmission and sociality unique – don't fossilize directly and are difficult to assess operationally in terms of the early archeological and paleontological record. Hence, for example, Foley's (1987, pp. 4–5) recent suggestion that culture is not a useful concept in the study of human origins. This belies the fact that group-collective features quite dramatically distinguish the simplest human societies as ethnographically observed from nonhuman primates. These features, arguably quite crucial in the development and spread of *Homo sapiens*, if not *Homo* generally, include widespread symbolic communication in language, diffuse rule-of-thumb sharing of food and information (viz., beyond the parameters of inclusive fitness), and pronounced influence of culturally defined rules on human sexual behavior (Knauft 1988a, 1988b). Those nonintensive human foraging societies that have been ethnographically studied have placed extremely high emphasis on diffuse sociality and, with the exception of Australian aborigines, have actively inhibited the development of political dominance hierarchies among men (see Ingold 1987; Knauft 1987a; Leacock & Lee 1982; Lee & DeVore 1984; Woodburn 1982). The great potential significance of these features in early human development is regularly neglected by EI theorists, revealing an individualistic competitive bias in the study of human evolution (recent examples include Tooby & DeVore 1987 and Wrangham 1987). This bias deserves to be countered, not on the basis of a priori disagreement, but on the basis of alternative data analysis, hypothesis testing, and theory building. Caporael et al.'s article is an important step in this direction.

Empirical caveats and cross-cultural context. Significant weaknesses in the present study must be noted, however. The empirical basis for Caporael et al.'s assertions is narrow and the range of possible causal influences not adequately considered. The occurrence of in-group bias in Caporael et al.'s experimental groups under "discussion" conditions may be significantly influenced, for example, by (1) the ratio of students to townspeople (or men to women); (2) variations in subjects' perceived likelihood that they may encounter another group member at some future time; (3) differences in initial expectation of reciprocal altruism, e.g., students'/townspeople's history of reciprocity with their peers and others in the community; and (4) differences in subjects' degree of financial need. Such variables need to be empirically considered to show that variation among them cannot account for or significantly alter the experimental results.

More generally, it is virtually certain that differences in cultural and ethnic identity strongly effect the strength of in-group biasing: cross-cultural studies in social psychology show major differences in collectivism versus individualism as measured among subjects from different cultures (see Triandis 1988). At the same time, the trends generally established by these studies augur well for Caporael et al.'s basic argument: Strong individualism correlates robustly with a high Gross National Product and socioeconomic success (cf. ibid.). Considering these trends on a worldwide scale, Caporael et al.'s subject population is thus among those *less* likely to exhibit collectivist propensities on a priori grounds. The same may not be true intraculturally, however: College students, as a privileged nonwage-earning group, could perhaps be more prone to altruistic in-group biasing than, say, a group of inner-city drug addicts, a group of business entrepreneurs, or a group of people with

diverse ethnic affiliations. Presently, both Caporael et al. and the EI theorists assume uniform (if antithetical) patterns of individualism or collectivism – each projecting a single pattern onto human evolutionary history, typically based on direct experimental evidence from only a narrow segment of industrialized English-speaking populations. The cross-cultural data on variations in individualism/collectivism suggest that this is likely to be a great oversimplification.

This reflects an area in need of much future attention: the theoretical and empirical gap between Caporael et al.'s specific experimental study and their large-scale conclusions for evolutionary theory. The present study requires articulation with "middle-range" theories concerning stages of human evolutionary development, on the one hand, and paleoanthropological and ethnographic evidence, on the other. For example, Caporael et al. do correctly identify hominid foraging for dispersed and patchy resources as crucial in creating conditions that would have selected for diffuse human sociality (see Cashdan 1980; 1985; Kaplan & Hill 1985; Knauft 1987a; 1988b; Kurland & Beckerman 1985; Testart 1982; cf. Marshall 1979). In-group biasing as an evolutionary tendency among small groups of hominid kin, however, does not contravene EI theories. Indeed, concepts closely analogous to in-group biasing are at the heart of several current EI theories of ethnocentrism and warfare (e.g., Essock-Vitale & McGuire 1980; Reynolds et al. 1987; Shaw & Wong 1988). What is needed is to establish the theoretical and ethnographic likelihood that altruistic sociality among early humans extended outside the close kindred – that sociality is intrinsic to human interactions in general and not simply a latter-day extension of biogenetic in-group interest to ever-more-inclusive categories of people who assess themselves as having a common identity.

The suggestions of Boyd and Richerson (1985) are particularly important here: Imitation and social learning are plausible both theoretically and empirically as rule-of-thumb mechanisms explaining the early establishment of genuine altruism early in human evolution. Once begun, such patterns can become self-sustaining – group-level selection becomes possible – because cultural transmission can proceed faster than biogenetic deselection and can thus establish its own channel of phenotypic alteration and evolution. This articulates with strong empirical evidence of just such rule-of-thumb sociality among the simplest human societies known ethnographically, such as nonintensive foragers. These societies almost invariably stress both coresidential flexibility and cooperation among a large group of persons, unrelated or distantly related as well as closely related – e.g., through classificatory totemic, namesake, ritual friendship, fictive kin, and other affiliations (e.g., Balicki 1970; Guemple 1972; Lee 1979, Ch. 5; Meggitt 1962; Myers 1986; Turnbull 1965a; 1965b; 1984; Wiessner 1982; Woodburn 1984; cf. Gamble 1982; Yengoyan 1984). Such simple societies, it may be noted, are much less ethnocentrically subdivided and less warlike than "tribal" groups with relatively greater population densities, a more sedentary pattern and greater competitive leadership (e.g., Chagnon 1988; contrast Knauft 1985; 1987a).

In short, delineating concrete connections between general theory, specific social processes, and ethnographic evidence is central to establishing the applicability of Caporael et al.'s sociality hypothesis to various stages and contexts of human development. Conversely, neglect of ethnographic reality leads Caporael et al. to their most questionable statement (though not one crucial to their larger argument): that anthropomorphism of natural objects is a "default" mode of conceptualization, restored to when objective materialist explanations are weak or lacking. Levi-Strauss (1963) and many others have documented convincingly on the basis of ethnographic evidence that anthropomorphic projections do not inhibit (and cannot be easily explained by) materialist reasoning. Moreover, in line with Caporael et al.'s more general argument, Durkheim (1912) illustrated long ago on the basis of Australian ethnography that anthropomorphism itself stems from the strength of human sociality: Anthropomorphism "socializes" the natural world and harmonizes it with structures of ongoing societal functioning.

Conclusion. As an alternative to EI theories, Caporael et al.'s sociality hypothesis is bold and suggestive. Unsurprisingly, it will also need major empirical and conceptual development to make a significant contribution to the study of human evolution. Even without its present shortcomings, however, Caporael et al.'s announced disconfirmation of EI theory is unlikely to force EI adherents to cede much ground. That EI theories tend toward nonfalsifiability is strongly evident in the responses of its adherents to empirically based critiques of sociobiology (see Vining 1986; Knauft 1987a; 1987b; 1988c) and is hinted at by Caporael et al.: The teleological bent and unbridled empirical purview of EI theories make them particularly prone to recontextualizing data sets and time frames to avoid disconfirmation. That EI theories are largely self-sustaining makes innovative theoretical breakthroughs more likely to arise out of forced competition from rival theories than from within an EI perspective itself. Caporael et al.'s argument must thus be judged by the strengths of the alternative paradigm it proposes as well as by its ability to outmaneuver the inevitable explanatory defenses of EI theorists on their own terms.

Overall, Caporael et al.'s perspective points to the tip of an emerging body of alternative experimental and ethnographic theory that will increasingly both challenge EI theories and broaden our understanding of human evolution in years ahead. Though such defections from EI theory tend to be punished by EI adherents – auguring toward negative Tit-for-Tat – there may yet be hope: Boyd and Lorberbaum (1987) have shown that no pure strategy for the prisoner's dilemma game is evolutionarily stable; rare mutant strategies may have important impact as selective forces, and long-term cooperative advancement may in fact ultimately depend on them (cf. also Boyd 1988; Nowak & Sigmund 1989; Wilson & Sober 1989). The empirical lacunae and logical leaps of Caporael et al.'s present contribution should be confronted respectfully, keeping this positive metatheoretical possibility in mind.

Egoistic incentives in experimental games

Dennis Krebs

Department of Psychology, Simon Fraser University, Burnaby, B.C., Canada V5A-1S6

Electronic mail: *kathy-denton@cc.sfu.ca*

Caporael et al. set out to show that individuals may behave cooperatively even though it does not advance their self-interest, and therefore that egoistic incentive theories are invalid, or at least limited. Sociobiology is singled out for special criticism. The Caporael et al. target article is an ambitious attempt to integrate theory and research on experimental games, social cognition, group dynamics, and human evolution. It is a welcome antidote to research that purports to study altruism, but fails to attend to the motives underlying observed helping behavior (cf. Batson et al. 1986; Krebs & Miller 1985). Nonetheless, the target article falls short of its expressed goal, with good reason. Disproving the egoistic incentive assumption is equivalent to disproving the null hypothesis. You can't prove something doesn't exist; you can only show you haven't found it. Caporael et al. establish that people do not always maximize their monetary gain, but they fail to establish that the cooperative behavior of their subjects is not motivated by other selfish concerns.

In the studies the authors describe, individuals honor agreements to contribute money to enhance the monetary gain of their groups even though that entails the risk of losing five or six dollars. In the spirit of the revealed-preference conception the authors criticize, however, we must ask, "*Why* do individuals sacrifice the opportunity to make an extra five or six dollars; what

do they get in return?" My guess is that the main incentive driving the subjects' behavior is not monetary – five bucks is a relatively insignificant sum. Rather, the subjects are motivated to win, or more exactly to avoid losing, at the game. In conditions where there is no group discussion, their goal – the egoistic incentive – is to avoid being suckered, which they achieve by keeping their money. When discussion is allowed, group members' perspective changes from "me" to "we" – instead of me against the other members of the group, it becomes "us" against "them" (my group against the odds, the experimenter, another group). One reason designated contributors contribute is so they will win through their team (the "we" in "we're number one"). But there is another reason.

Subjects in discussion groups explained their cooperative behavior as a way to honor the group's decision. To understand what they have to gain from this, think of the alternative. How would you feel about yourself if you sat down with a group of cooperative people, made a democratic decision to maximize the group's net gain, then violated the agreement a few minutes later to maximize your own gains? Shades of Judas. Caporael et al. argue that subjects were not driven by the desire to maintain a clear conscience because they contributed only when their donations were directed toward the group with whom they reached a decision; however, that was the only condition in which they made a contract. Even though subjects were told they would never see the other group members again and no one would ever know how they behaved, *they* knew. The egoistic incentive driving their behavior would seem to be to uphold their conceptions of themselves as honorable people, and to maintain their self-esteem.

The authors argue that their results are inconsistent with sociobiological theory, but the recent work of Cosmides (1985) and Cosmides and Tooby (1987) show they are not. From an evolutionary perspective, the question becomes "who would propagate more genes, individuals who honor group contracts or individuals who break them?" Note that the question is voiced in terms of a general strategy; in Cosmides's words, a Darwinian algorithm; in the words of Caporael et al. a "rule of thumb." As acknowledged by the authors, the cognitive and affective dispositions that guide human behavior were selected eons ago in small groups of cooperating hominids. We can safely assume that none of our ancestors ever had the opportunity anonymously to violate a democratic decision about a division of resources reached in a group of strangers who would never encounter one another again. Indeed, the only setting in which a modern human would ever encounter this situation is a psychologist's lab. Thus, there is no reason to expect cognitive and affective mechanisms equipped to mediate the distinction necessary to maximize self-interest in such situations to have evolved. Cosmides and Tooby's work [see their accompanying commentary] "makes it clear that behavior in ancestral populations had only to be adaptive, on average, for the evolution of mediating mechanisms to occur" (Crawford 1989, p. 12).

Inclusive fitness is based on net gains. In the constricted social environments of our ancestors and in most contemporary social contexts, the small gains one might reap by cheating on group decisions are simply not worth the risk. I have advanced a similar argument to explain the evolution of altruism (Krebs 1987; Krebs & Miller 1985). In our ancestral environments, individuals guided by the general program or algorithm "help those who are familiar and similar to you" would have propagated more genes (by enhancing the fitness of relatives and by supporting systems of reciprocity – see Alexander 1987) than those who did not. Although this may not be as much the case in modern urban environments, these "anachronistic" dispositions may still mediate self-sacrificial behavior, which, because it is no longer supported by the original gains in fitness, may be more purely altruistic than it was in the environments in which it evolved.

Though poorly equipped to establish the negative case – the absence of egoistic incentives – the authors' experimental paradigm seems fertile for investigating the types of incentives that influence individualistic and cooperative behavior. If the general line of thought advanced in this commentary is valid, subjects should be willing to sacrifice money to win games and to uphold their conceptions of themselves. Our own research, however, and that of others, suggests that the processes involved in the maintenance of self-esteem and an adaptive conception of self are not necessarily rational (see Krebs et al. 1988 and Snyder & Higgins 1988). It would be interesting to compare the self-attributions and self-justifications invoked by subjects before they made selfish and cooperative decisions with those invoked after the decisions. We have observed significant differences between self and other attributions of moral behavior, captured in what we call the "self-righteous bias" – the tendency for people to assume that they are more likely than others to live up to their moral ideals. It would also be interesting to examine some of the sources of within-group variance in expectations, self-attributions, and behavior reported by Caporael et al., especially those stemming from moral development. Our research suggests that subjects at Kohlberg's Stage 2 would be most prone to justify their behavior in terms of "looking out for number one," subjects at Stage 3 in terms of the good of the group, and subjects at higher stages in terms of equity and justice (see Krebs et al., in press).

"Social man" versus "conscientious man"?

Vladimir A. Lefebvre

School of Social Sciences, University of California, Irvine, CA 92717

I would like to suggest an explanation for some of the experimental data presented in Caporael et al.'s target article. A formal approach to studying human moral feelings and their relation to decision making leads us to the assumption that man operates with evaluations "positive-negative" with the help of an *automated processor* (Lefebvre 1982; 1985). The work of this processor reveals itself most clearly when a person does not have any objective criteria on which to base a preference. In the framework of this approach, "conscience" is considered to be a special cognitive process that can be described in precise scientific terms. This process is not connected with any concrete moral prescriptions of the type "care about the well-being of another person." Caporael et al.'s critique of traditional intuitive approaches to consciences hence does not relate to our approach.

It is my impression that the experimental data presented by Caporael et al. support the hypothesis about the existence of a "processor of conscience." This processor can manifest itself both in experiments with bipolar choice and in experiments in which subjects used a continuous scale [0, 1] (where 0 means "negative" and 1 means "positive"). In the first case, the processor of conscience predetermines the probability of making a positive choice, and in the second case it predetermines the "degree of positivity" (Lefebvre 1987a; 1987b; Zajonc 1987).

A mathematical model of the processor of conscience has allowed us to explain the phenomenon of the golden section in binary choice experiments. In experiments in which subjects evaluated their acquaintances with the help of bipolar constructs (of the type "kind-stingy") they chose a positive pole with the frequency 0.62, which is close to the value $\frac{\sqrt{5}-1}{2} \approx 0.618$, known as the "golden section value" (Adams-Webber & Benjafield 1973; Benjafield & Adams-Webber 1976). A retrospective analysis demonstrated that the value 0.62 also appeared in other experiments with binary evaluations (Adams-Webber 1979; Benjafield 1984).

A mathematical model of the processor of conscience also

predicted the values of deviations from the golden section depending on the conditions of the experiment. These predictions coincided well with experimental data (Lefebvre et al. 1986). The model also explained why the value 2/3 appears when subjects repeatedly evaluate themselves with the help of one set of bipolar constructs (Benjafield & Green 1978; Lefebvre 1987a).

Let us look at some of the data of Caporael et al. in relation to the hypothesis of the processor of conscience. Consider the experiment with "super simple conditions." Each of the five subjects has two alternatives – to give or to keep – and they do not have any clear criteria to make them prefer one alternative over the other. (As a matter of fact, this is the essence of the dilemma.) In this situation, the subjects can completely *ignore the economical conditions of the game* and operate with the pure codes "positive-negative." The strategy "to give" is connected with activity, and the strategy "to keep" with passivity. In accordance with Osgood's (1965) ideas, we assume that "give" corresponds to a positive pole and "keep" to a negative pole. The model of the processor of conscience predicts for this case that the frequency of choosing "give" will be equal to 0.62. As is shown on Caporael et al.'s Table 3, the experimental value for "give" is 0.64. Of course, we cannot place too much importance on this coincidence because the number of subjects was quite small. (According to Table 3 there were five groups of five subjects each, that is, only 25 people.) However, we may expect that reproduction of this experiment with a larger number of subjects to produce a frequency of the strategy "to give" equal to 0.62.

Consider now how the subjects evaluated their own actions. This procedure consisted of subjects estimating (on a 100-point scale) the probability that their contribution would be (a) futile, (b) critical, (c) redundant. The data in Caporael et al.'s Table 2 demonstrate that neither experimental conditions nor the type of choice (contribution-noncontribution) influenced significantly the value of the estimated probability of "crucial." The estimated values of "futile" and "redundant," however, differ essentially. The authors noted that under conditions of "no-free-ride" and "money-back-guarantee" the ratio of contributors' estimations of being "redundant" to being "futile" is approximately 2:1.

Let us analyze this point in more detail. For the sake of convenience we construct Table 1 based on their Table 2. Consider a bipolar construct "futile-redundant." The data in Table 1 suggest that under conditions (1) and (3), a positive pole for contributors is "redundant," and for noncontributors it is "futile." Under condition (2), "redundant" is a positive pole for both contributors and noncontributors. Let us now find the ratios of "positive" shares to the sums of "positive" and "negative" for each line: They are close to 2/3, and their mean value is 0.68. This phenomenon can be explained by the hypothesis that when subjects indicated on a 100-point scale their likelihood of being "futile" or "redundant" they actually evaluated themselves (*their own actions*). And in this case, according to the mathematical model of the processor of conscience, such a self-evaluation must be equal to 2/3.

Table 1 (Lefebvre). *The data recalculated*

	Futile	Redundant	Ratio
(1) Standard dilemma			
contributors (N = 9)	.24(−)	.46(+)	46/70 = .66
noncontributors (N = 20)	.55(+)	.26(−)	55/81 = .68
(2) Money-back-guarantee			
contributors (N = 15)	.25(−)	.46(+)	46/71 = .65
noncontributors (N = 20)	.28(−)	.50(+)	50/78 = .64
(3) No-free-ride			
contributors (N = 27)	.20(−)	.57(+)	57/77 = .74
noncontributors (N = 8)	.51(+)	.24(−)	51/75 = .68
Mean value (taking into consideration the number of subjects in each line)			.68

All of the above considerations incline me to prefer the hypothesis of "conscientious man" to the hypothesis of "social man." However, these hypotheses are difficult to compare, because the first one is elaborated in great detail, whereas the second hypothesis is given only through the negation of others. We cannot exclude the possibility that the processor of conscience is, in fact, the very alternative to "economic man" and to some other approaches sought by Caporael et al.

Do we need two souls to explain cooperation?

Wim B. G. Liebrand

Department of Psychology, University of Groningen, 9718 JP Groningen, The Netherlands

Electronic mail: *lybra@hggrug5.bitnet*

As with all gregarious animals, "two souls," as Faust says, "dwell within his breast," the one of sociability and helpfulness, the other of jealousy and antagonism to his mates" (William James 1890).

The above quotation captures well the major thesis set forth by Caporael et al. The question remains, however, whether it presents a correct view of human nature. I doubt it. In my opinion, the parsimonious view provided by the assumption that human nature is fundamentally "selfish" has not yet been placed in serious jeopardy by the data or the arguments put forward by Caporael et al.

There are three reasons why I would question that one needs to posit a biological mechanism of sociality parallel to the selfishness mechanism, to account for cooperative choice behavior in human relationships.

The first is that many of the settings in which human decisions are made do not share the same interdependence characteristics as the commons dilemma, and in many, one does not have to posit sociality as a mechanism to account for cooperative behavior. The second reason is that many situations that are judged by an external observer to have the interdependence characteristics of a commons dilemma, are not viewed in this way by decision makers who assign different weightings to their own and others' outcomes. And third, I will argue that even when an individual is in a setting that objectively has the structure of a commons dilemma, and which the individual subjectively experiences as such, there exist strategies of self-interest or selfishness that may produce cooperative rather than competitive forms of choice behavior.

As regards the first reason cited above, Caporael et al.'s empirical evidence for their arguments is based on data coming from one particular kind of outcome interdependence relationship, namely, that observed in the n-person commons dilemma. There are obviously a far wider range of interdependence settings in which Faust's "soul" or "souls" of human decision making can be expressed. A useful way of illustrating this range is depicted in Figure 1.

In this figure, relationships of interdependence are depicted as falling on a continuum that at one end includes those that are purely competitive, and at the other those that are purely coordinative. In the former, the outcomes attained are negatively correlated, and the interactions tend to be highly competitive and conflictual. In the latter, outcomes are positively correlated, and the interaction between individuals tends to be

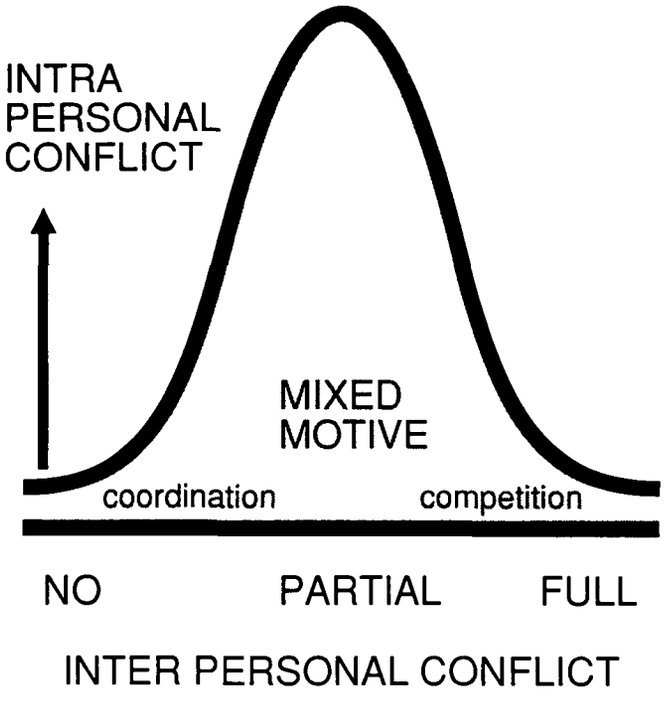

Figure 1 (Liebrand). Level of intrapersonal conflict for three degrees of interpersonal conflict.

trustful, cooperative, and coordinative. Finally, there are those relationships that do not fall strictly in either category.

The ordinate of Figure 1 depicts the level of *intrapersonal* conflict that is likely to be experienced by a given actor when making decisions in relations in which there are varying levels of *interpersonal* conflict. As can be readily seen, the highest level of intrapersonal conflict occurs in those settings where there is partial interpersonal conflict, that is, in mixed-motive relationships such as the commons dilemma. What characterizes such a relationship, of course, is that the interests of the outcome interdependent participants are partially correspondent and partially discordant.

So what implications does this have for the issue of whether evolution is driven by the motives of selfishness, of sociality, or both? As can be seen from Figure 1, the extremes of the continuum of interdependent relationships give rise to little or no conflict for the actor. *It is in their own selfish self-interest* to cooperate in coordinative settings and to compete in competitive ones. Caporael et al. do not pay much attention to these kinds of interdependence relationships. Yet they do exist and require no appeal to "sociality" as an explanation of the behavior of decision makers, whether cooperative or competitive.

The above, however, does not answer the question of why cooperative behavior, which Caporael et al. assume to be contrary to egocentric self-interest, is observed in the commons dilemma. I will argue that such cooperation can obtain for two reasons, neither of which requires that one posit an evolutionary mechanism of sociality. First, there exists a research tradition, issuing from the work on social values by Messick and McClintock (1968) and Kelley and Thibaut (1978), of explaining cooperative choice behavior in mixed-motive relationships without the need to posit an evolutionary mechanism of sociality.

Basically the social value approach assumes (and observes) that individuals transform any given objective outcome matrix, such as the one that obtains in a commons dilemma, into a subjective one. This transformation process occurs in a systematic manner consistent with the values the individual assigns to their own and others' outcomes. Thus, subjects making choices in experimental game settings can be playing one of a number of games. This obtains because different subjects apply different transformational rules to the various outcome contingencies that define the game. Thus, if I define my self-interest in terms of attaining coordinative outcomes, I will transform the outcomes differently than I would if I define my self-interest in terms of gaining a competitive advantage over others.

In either of the above cases, I am not likely to experience a high level of intrapersonal conflict because a coordinative transformation dictates cooperative choices, and a competitive transformation, competitive ones. In effect, these value transformations "unmix" the motivational structure of the game. The dilemma disappears with either transformation rule.

To date there are several studies supporting these predictable interpersonal differences in the way people approach, behave, and evaluate experimental games like the commons dilemma (McClintock & Liebrand 1988). To explain this divergence in cooperative and competitive behavior, does one posit that individuals with one value orientation were operating under the evolutionary principle of sociality, the others under the principle of selfishness?

Now it is also possible, if not likely, that following the subjective transformation of the given outcomes in a commons dilemma game, the dilemma will retain the properties of a mixed-motive game for some individuals. If this obtains, then the magnitude of the dilemma, as Caporael et al. rightly observe, is particularly strong if a decision to allocate valued resources between self and others can be made once and only once.

For one-trial prisoner's-dilemma or commons-dilemma games, game theorists have traditionally argued that the competitive choice is the more rational, self-interested one because it is dominant and minimizes the decision maker's maximum possible loss. For the subjects invited to the laboratory, life is not this easy, however. They are confronted with a conflict between two appealing options. Assume that they expect the other to be comparably sensible, as they ought to be according to game theorists. Such a perspective would make the mutually cooperative outcome cell resulting from a similar sensible choice by others more attractive than the deficient equilibrium cell achieved when they mutually decide to compete against one another.

Of course, one or more individuals may realize that in these circumstances a unilateral shift may be attractive if one assumes that others will view the cooperative choice as the mutually sensible one. Even those individuals who consider a unilateral shift under these circumstances must realize that others are likely to consider the same possibility. All in all, based upon the above reasoning for those individuals for whom a social dilemma is truly a dilemma, one would not expect the uniform selfish or defecting behavior that would be predicted from a game theoretical perspective. That is, given the specific structure of a commons-dilemma game, and some reasonable assumptions about the interpersonal expectations of decision makers in such forms of relations, one would expect to observe a certain level of cooperative choices even in the absence of "classic" selfish incentives or the application of transformation rules that convert the common's dilemma into a nondilemma.

In conclusion then, though I certainly do not have evidence to assert that sociality may not be an evolutionary principle comparable to egoistic self-interest, I do not find the arguments or the evidence presented for sociality as a criteria of evolutionary fitness compelling. For the time being, the single "soul" of rational self-interest or selfishness provides the most parsimonious explanation of cooperative and competitive behavior in the commons dilemma.

Love and duty: The new frontiers

Jane Mansbridge

Department of Political Science and Center for Urban Affairs and Policy Research, Northwestern University, Evanston, IL 60201

The Western social sciences began to evolve the fundamental assumption of egoistic incentives in the midseventeenth century, at a moment coterminous with the evolution of the sciences themselves. Economics and political science, the earliest modern social sciences, each adopted at their founding the assumption that human beings act only on the basis of self-interest. As the contemporary proverb had it, "Interest will not lie" (Gunn 1969; Hirschman 1977; Mansbridge 1990; Myers 1983).

Caporael et al.'s target article comes at a moment in intellectual history when in every discipline in the social sciences, and almost every subdiscipline, the tide is turning against self-interest. The tendency to postulate self-interest in American social science had hardened after World War II, when many practitioners aimed at duplicating the kind of logic found in Newtonian physics, or, failing that, in economics. But in 1977, Amartya Sen's now classic "Rational Fools" began to undermine the presumption of self-interest in economics, arguing that commitment to a principle, which involves counterpreferential choice, "drives a wedge between personal choice and personal welfare," although "much of traditional economic theory relies on the identity of the two." Sen's concept of a "meta-ranking" of preferences goes beyond "utility" to allow placing commitment above subjective preference.

In economics, Collard (1978), Schelling (1978), Hirschman (1985), Frank (1988), Kalt and Zupan (1984), Buchanan (1986), and others have followed Sen in suggesting that motives other than self-interest play an important role in human interaction. Elster (1990) has identified two distinct unselfish motives: "love" and "duty," variously called "sympathy" and "commitment" (Sen 1977), "empathy" and "morality" (Jencks 1979), and "affection" and "principle" (Hume 1741). In political science, Mansbridge (1980), Muir (1982), Maass (1983), and a spate of recent works (Derthick & Quirk 1985; Kelman 1987; Reich 1988; Vogler & Waldman 1985) have documented many important instances in which feelings of love or duty (solidarity with others or the commitment to a principle) have influenced the behavior even of elected politicians more than self-interest. In psychology, as cognitive psychologists have begun to explore bounded rationality, they have also joined economists in investigating which self-interested behaviors people think unfair, noting that consumers often try to punish firms they see as behaving unfairly (Kahneman et al. 1986). Political psychologists (e.g., Sears 1990) have shown that on many issues political principles have a greater effect than self-interest on political attitudes and behavior. Developmental psychologists (e.g., Hoffmann 1987) have uncovered what looks like primitive empathy in newborns, who cry when they hear other newborns crying but not when they hear other equally loud and aversive sounds, and more advanced empathy in one- and two-year-olds, who spontaneously offer help to others they see in distress (Radke-Yarrow et al. 1983). Even in biology, we now know that social contexts affect human and animal biological makeup. Menstruation in one woman causes menstruation in another (McClintock 1981; 1987). Eliminating a higher-ranking vervet monkey in a dominance hierarchy causes a rise in serotonin in the second-ranked one (McGuire et al. 1982). As evidence mounts that even the biological self is socially constituted, the very concept of self-interest becomes more complex. The key question concerns the contexts in which people are most likely to put the good of others ahead of their own.

Into this rapidly changing scene in the social sciences come Caporael et al. with a set of experiments cleverly designed to elaborate the contexts in which "sociality" has more effect on behavior than self-interest narrowly conceived. Their work is central to the development of this new frontier.

In interpreting their experiments, however, the authors neglect the sense of duty stressing only love. Their experiments actually demonstrate the positive effect on cooperation of *both* duty and love. In the experiment described in section 2.7, 20 to 59% of the participants gave to others against their own self-interest even when the recipients were members of another group (Table 5). when subjects contribute to a group other than their own, love (or sympathy, or empathy, or affection) for one's own group is unlikely to be the cause of the contribution. Caporael et al. indeed designed this experiment to distinguish the effects of conscience and sociality. They invite us to conclude that those who contributed to the other group acted out of conscience, or duty.

Even among the 32 to 79 percent who ended up contributing to their own group, conceptions of duty may well have played a role. In the 3- to 25-point increment we observe here over cooperation because of conscience, we might postulate the workings of "contingent conscience," in which members of a group act more morally toward members of their own group than toward others. Operationally, contingent conscience may not be distinguishable from pure group love, and may form one of the many motivations encompassed in "sociality" (footnote 2). Everyday experience with honor among thieves, colleagues, or members of any in-group certainly teaches us that conscience may *interact* with loyalty to an ingroup rather than being ruled out by it.

Is duty more self-interested than love? Caporael et al. seem to think so. Unlike Sen, who singles out duty instead of love as explicitly counterpreferential, the authors describe "conscience," or duty, as an internal "payoff," and therefore an "egoistic incentive." Sociality, on the other hand, is "totally unrelated to self-interest" and works "in the absence of egoistic incentives." The authors also interpret "sociality" in recurrently cognitive terms, concluding in footnote 2 that it is part of a "species-typical 'cognitive machinery'." They correctly point out the tautology of claiming that everything is motivated by self-interest. But the concept is clouded by their claim that whatever is responsible for the "own-group" increment in Table 5 is cognitive and not motivated by self-interest, whereas whatever is responsible for the "other-group" increment in the same table is an "egoistic incentive."

Innate selfishness, innate sociality

Susan Oyama

Department of Psychology, John Jay College, City University of New York, New York, NY 10019

There are several nice features of the explorations reported in the target article. Overapplied, "untested metatheories" certainly deserve scrutiny. Egoistic incentive (EI) is a kind of black hole explanation: It sucks everything into its maw, not only engulfing all objects on the landscape, but transforming the very landscape itself. EI becomes a universal explanation by redefining all alternatives in its own terms. (Schwartz, 1986, describes this powerful reductive process in economics, behavior theory, and evolutionary biology.) It's good to be shown how assiduously social-dilemma researchers have narrowed their subjects' range of possible responses, and Caporael et al. are ingenious at devising experimental situations that allow other choices, thus opening up an overrestricted theoretical framework.

In addition to performing these services, however, I suspect that the authors are up to something else. Some people, seeing the dominance of EI, have felt the need to combat it, not only because of the sorts of theoretical concerns voiced in the target article, but out of a certain dismay over the disappearance of everyday morality in a flurry of individualistic cost-benefit

analyses, as sweeping conclusions are drawn by evolutionary theorists about the nature, or rather, the radical absence, of genuine moral reasoning (Alexander 1987; Ghiselin 1974). In addition, rationality itself comes to be defined (even in the target article) by individual benefit. I have a hunch, developed by reading between the lines of this article, that an unspoken reason for these studies is the authors' discomfort with this explaining away of morality. I share that discomfort, but have some reservations about the tactic they have chosen to put things right.

Once biology is accepted as the arbiter of ultimate reality, those who object to EI's totalizing vision have a small, unattractive array of options. They can tell us to combat our inherent selfishness, to create virtue out of pure, ungrounded intention. Or they can contest the biology. Caporael et al. deny that they are simply countering innate selfishness with innate altruism (note 2). I see in their efforts, however, a tacit acceptance of the premise that biologically ratified descriptions of the moral life have more credibility than descriptions couched in other terms, as well as an attempt to contest the content of current biological characterizations. I do not object to their claim that human motivation involves more reasons than are dreamed of by EI theorists, but I am a bit concerned that their well-intentioned efforts may lend credence to an assumption that is better challenged than adopted.

Developments in evolutionary biology are certainly not the only reasons we have for doubting the prevalence of fellow-feeling in the world. Both current events and our available means of comprehending them often seem to point inexorably to EI-style thinking. Failure to "look out for Numero Uno" is even considered a bit psychologically unhealthy these days. For complex reasons, we are experiencing an extraordinary poverty of explanatory means. And, because our notions of intention are closely linked to our ways of explaining behavior, our store of possible reasons for acting is correspondingly impoverished.

Belief in sociobiological EI, then, springs from well-prepared ground. But so does the belief that biology defines the set of true motives. Caporael et al. dispute the first belief but embrace the second. Though they use the experimental apparatus of the social sciences rather than of biology, they attack the doctrine of innate selfishness by giving a different story about naturally selected characters.

Stories about natural selection require considerable circumspection, but they are not illegitimate in themselves. What is unfortunate is that they are often thought to reveal a profound underlying truth, one that allows us to glimpse an eternal genetic reality through the veneer of phenotypic appearance. Thus, EI theorists tell us what is *really* motivating behavior, no matter what a person's motives may *appear* to be. The danger of this kind of reasoning is that it compels us to use biology to reclaim the full complexity of our own experience. One gets the sense from the target article that it is not enough to demonstrate in-group bias; one must claim an evolved "innate tendency." This seems more than the usual psychological overgeneralization from limited samples; it seems to reflect a tacit belief that social motives must have natural selection on their side in order to go up against EI.

But is it true that reasons are somehow less than genuine if they do not appear to have a selective history? Whatever its other benefits, I would not want Caporael et al.'s work to reinforce the already widespread assumption that behavior about which an evolutionary story can be told is in some sense more morally significant, more fundamental, more real than behavior about which no such story is told. I would not want readers to conclude that a motive or act cannot be authentic unless it is seen to flow from selective advantage. Nor would I want them to believe that an evolutionary history must eventuate in "innateness," by most of its conventional definitions (that it be "fixed" or immutable in individuals, for instance, or "fixed" and universal or invariant in the species). A "bilevel approach," insofar as it incorporates these conventional notions of innateness, retains too much traditional conceptual baggage to be a satisfactory solution to the old innate-learned dispute. Better to eliminate the opposition of biology to culture and learning altogether. [See also Plotkin & Odling-Smee: "A Multilevel Model of Evolution" *BBS* 4(2) 1981.]

Having found all this between the lines of the target article, I am obviously risking being massively off base. If I am, I apologize, but it may be that some of these issues are worth thinking about even if I have misread Caporael et al. I also hope that my comments are not taken as a blanket criticism of their work; on the contrary, I find it interesting, whether or not I am right about their broader agenda.

A final note: In one of the cover letters from *BBS*, commentators are told that the target article will not be further revised, but that typographical errors will be corrected; in fact, we are asked to compile a list of such errors. Circling a typo, I thought about the extra effort it would take to give the article the close attention required for proofreading, and imagined, with scant pleasure, typing up a list. How many others would point out the same errors? Transfixed by my small dilemma, I wondered how many commentators, having had the same thought about multiple corrections, would be perfunctory in their checking, or forego it altogether. I was glad that no one was asking me to estimate the proportion who would be dutiful. It did occur to me, though, that whatever estimate I made, I would have imperfect faith in it, and would be inclined to risk being redundant, as well as inconsistent with expectations and therefore irrational, just to make sure that *someone* did it.

In-group bias is a kind of egoistic incentive

Howard Rachlin

Department of Psychology, State University of New York at Stony Brook and Russell Sage Foundation, Stony Brook, NY 11794

The target article presents an interesting and provocative series of experiments designed to demonstrate cooperative behavior without "egoistic incentives": (EIs) and even counter to them. Such nonegotistical behavior, Caporael et al. argue, is explicable in terms of an "in-group bias" (IB) which they say may have originated in the distant past of human biology as the product of egoistic incentives but which has long ago become a functionally autonomous motive of individual behavior.

There is nothing wrong in principle with the above position, but it depends critically on the success of the experiments in eliminating plausible egoistic incentives. In its normal functioning, in-group bias is conceived as a bias on top of egoistic incentives. Therefore, the burden of proof falls heavily on the authors. If their results could be explained plausibly in terms of egoistic incentives alone, the concept of in-group bias would be superfluous.

I will argue that egoistic incentives *can* plausibly explain the results of the experiments described in the target article and, moreover, that Caporael et al.'s concept of in-group bias (as an inherited cognitive mechanism separate from egoistic incentives) cannot plausibly explain the results.

These experiments pose two critical questions for an EI theory: (a) Why do any subjects cooperate in the standard procedure, given that the EIs "greed" and "fear" create a bias toward noncooperation? (b) Why does prior discussion among the members of a group increase the number of cooperators? Caporael et al. correctly state that "these laboratory games are not simulations of naturally occurring social dilemmas; they *are* social dilemmas" (sect. 2, para. 1). But this fact is lost in the analysis of the experiments. The authors claim to have eliminated "reputational concerns," defined in terms of EI, by keeping each player's choices a secret from the others. No player's choices, however, were secret from the experimenters.

Even if they were, it is far from clear that the subjects would believe they were. In any psychological experiment part of the subjects' "dilemma" is behaving correctly in the eyes of the experimenter. Thus "reputational concerns" were not eliminated as egoistic incentives in these experiments.

It might be argued that reputational concerns would not be sufficient to overcome the strong incentives (greed and fear) to defect. Table 2, however, provides some evidence that the incentives to defect (whatever they are called) were less strong for those who cooperated than for those who defected. There are large and apparently significant differences between the estimates of contributors (cooperators) versus noncontributors (defectors) in the standard dilemma condition. Contributors estimated the chance of their contribution being futile as .24 (versus .55 for noncontributors) whereas noncontributors estimated the chance of their contribution being redundant as .26 (versus .46 for contributors). These estimates can be used to calculate the expected value of contributing and noncontributing separately for the subjects who actually contributed and for those who did not contribute. The ratio of the expected values of contributing and noncontributing is (about) .75 for the contributors and .67 for the noncontributors. Regardless of whether these subjective estimates are causal factors or post hoc rationalizations, they indicate that the bias due to the (quite justifiable) fear of appearing selfish in the eyes of the experimenters was less difficult to overcome for some subjects than for others. Individual subjects' choices *and* estimates were consistent with each other (as "revealed-preference theory" would predict). Caporael et al. may claim that such consistency of choice and estimate is just as predictable from individual differences in in-group bias as from individual differences in reputational concern. As indicated above, however, to be successful, their explanation of their own experiments must not just be as good as an EI explanation, it must be better.

EI also has a simple explanation for why prior discussion increases the number of contributors in these experiments. The essential ingredient of discussion, regardless of how the contributors are chosen, is that some members of the group actually agree to contribute and explicitly or implicitly promise the others that they will contribute. It is true that if one considers egoistic incentives alone (ignoring reputational concerns), there is no reason for a person making such a promise to stick to it. This, however, disregards the effect of the promise on those who *hear* it. EI theory is an objective theory of behavior. There is nothing in EI theory that says the subjects must themselves subscribe to it regarding their estimations of other subjects' behavior. Each member who promises to contribute will lower the estimates of others that their own contributions will be futile. And, conversely, each member who does not promise to contribute will lower the others' estimates that their own contributions will be redundant. These are perfectly reasonable judgments based on past experience with other people's promises. Both raise the ratio of the expected value of contributing to that of not contributing, hence increasing the probability of contributing.

A test of the above speculation is whether discussion *without* promises made and understood has any effect on cooperation. According to Caporael et al.'s theory, it is discussion itself, not specific promises, that ought to create IB, and hence cooperation. In an experiment (Dawes et al. 1977) described at the end of the target article, it was found that "discussion of an irrelevant topic . . . [no promises] led subjects to contribute . . . at the same level as no discussion, and rates of contributing in these conditions were significantly lower than when subjects discussed their social dilemma [i.e., made and understood promises] (sect. 4, para. 4). Thus, in Caporael et al.'s experiments, it was not discussion alone but discussion relevant to the problem that increased cooperation. The authors do not speculate about what exactly makes a discussion relevant; it is not unreasonable to assume that the critical factor was the agreement on who would or would not contribute rather than the activation of a "limited-access program" or a bringing into consciousness of some "cognitively unconscious" mechanism.

This is not to say that in-group bias is unreal. These experiments do show that perceived group membership strongly biases individual decisions. But it is plausible that the bias group membership creates derives from individual egoistic incentives and is not a functionally autonomous motive to be set alongside them.

Egoistic incentive: A hypothesis or an ideological tenet?

Anatol Rapoport

University College, University of Toronto, Toronto, Ontario, Canada M5S 1A1

By marshalling experimental evidence, Caporael et al. attempt to refute the universal claim inherent in the egoistic incentives (EI) hypothesis: the assumption that apparently altruistic behaviour patterns can always be explained in terms of selfish motives. It seems, however, in the light of the authors' own observations, that this is a futile task. The proponents of EI are always ready to retreat to previously prepared positions by extending the concept of "selfishness" to include any demonstrable source of motivation not previously subsumed under the concept. It seems advisable, therefore, to concede the EI dogma once and for all, thus rendering the hypothesis unfalsifiable and, therefore, theoretically sterile. Then one can go about the business of examining the rich variety of sources of human motivation governing choices where the outcomes of those choices result in the distribution of costs and benefits to self and others.

From this point of view, the experimental findings reported in the target article are solid contributions to social psychology. The implied aim of refuting the universality of EI, however, is either a futile exercise (because of this unfalsifiability) or breaking through an open door, if EI is expressed, say, in the sociobiologists' explanation of moral systems: Give only to kin, or in expectation of reciprocity, or to avoid sanctions, or to project a profitable image.

There is another way of looking at the challenge to the EI hypothesis, however, namely, to focus on its ideological implications. Arguments about human nature are clearly rooted in ideological commitments. One important ideological rift is between those who derive satisfaction from winning in competition or combat and those who derive satisfaction from cooperative and nurturing relations; between those who strive to dominate others and those who tend to identify with others. Awareness of an incompatibility between commitment to power and commitment to love is reflected in myths, as in the tragedy of King Midas or in the fate of Alberich, who forswore love to acquire power. And it is reflected in the incompatability of power-oriented and integration-oriented global politics.

Throughout history, preoccupation with power or power struggles, being the preoccupation of the powerful, the vociferous, and the energetic, has dominated conceptions of human nature and of society. The Hobbesian war of everyone against everyone is the crassest example of this conception. Classical economics, appearing on the intellectual horizon at the dawn of the industrial revolution, chose a nonlethal version of the Hobbesian world as the framework in which processes that uniquely characterize the human species – production and exchange – take place. The same framework underlies the market model of national politics, where the epitome of democracy is depicted as the maintenance of a steady state resulting from competitive activities of interest groups. It underlies the realist conception of international relations, where peace is identified with the preservation of the status quo through balance of power.

All these models of society are expressed predominantly in the descriptive mode. In the context of social philosophy, however, the descriptive mode often merges with the normative. It is easy to pass from describing things as they are to concluding that this is the way they ought to be. Witness the ease with which Adam Smith passes from describing self-interest as the prime mover of economic activity to justifying it.

> It is not from the benevolence of the butcher, the brewer, and the baker that we expect our dinner, but from their regard to their own interest. We address ourselves not to their humanity but to their self-love and never talk to them of our own necessities but of their advantages. Nobody but a beggar chooses to depend chiefly on the benevolence of his fellow citizens (Smith 1776/1910, vol. 1, p. 13)

At times, this defense is implicit in resorting to a euphemism. Buchanan and Tullock (1962) replace "egoism" by "methodological individualism." Wicksteed (1933) replaces it by "nontuism."

The close relationship between cognitive and ethical evaluations in the social sphere points to the importance of coupling challenges to existing social, political, and economic systems with critiques of the images of society and of human nature that provide rationalizations of these systems. From this point of view, evidence of "sociality" as an important component of human motivation deserves serious attention, and the concept itself deserves further theoretical development in the light of rigorous experimental procedures.

Evidence of something that could be called "sociality" was obtained in single plays of the prisoner's dilemma with unknown coplayers by Terhune (1968), Pancer (1973), Hofstadter (1985), and Rapoport (1988). In some cases, cooperative choices were observed in up to 50% of the subjects. In these situations reciprocity, kinship, sanctions, and image projection can clearly be ruled out as contributing factors. The cooperative choice in single play prisoner's dilemma has at times been attributed to the salience of the cooperative outcome, which is not only Pareto-optimal but also symmetric. This factor was in effect excluded in Rapoport's (1988) experiments, where the prisoner's dilemma was presented in camouflaged form without displaying the payoff matrix. Moreover, a significant increase in cooperative choices was observed when the nature of the dilemma was explained in the instructions, that is, when the dominance of the defecting choice was made at least as salient as the Pareto-optimality of the cooperative choice.

Sociality can be introspectively recognized as the identification of self with others. In view of our biological heritage as mammals and social animals, it is not unreasonable to attribute this extension of the "self" to "empathy," the acute discomfort we often feel when witnessing the suffering of another. In view of the distinctive human "inheritance" mechanism, whereby transmission of culturally shaped modes of cognition and evaluation has overshadowed genetic transmission, it is not surprising that empathy is often readily extended to nonkin and even to nonhumans. On the same grounds, it is not unreasonable to assume that sociality is a component of prime importance in the system of human motivations.

Too important to be left to rational choice

Richard A. Shweder

Committee on Human Development, University of Chicago, Chicago, IL 60637

Electronic mail: *rshd@tank.uchicago.edu*

Although I am confident that Caporael et al. are on the side of the angels in their reluctance to glorify selfishness (see Etzioni 1988 for a recent treatment of the topic), I do worry a bit that a consideration of the useful distinction between *motivation* and (rational) *justification* may diminish the force of their critique of egoistic incentive (EI) theories.

This is what I have in mind. Larmore (1987, p. 136) discusses a doctrine known as "indirect consequentialism," which holds that "whereas actions must be *justified* consequentially [in terms of the goods or outcomes or ends they promote], they may have to be *motivated* by nonconsequentialist considerations" [e.g., doing what seems intrinsically right – fulfilling one's duty, helping a friend, keeping a promise, devotion to God].

In this commentary my interest is less in the distinction between consequentialist (teleological) versus nonconsequentialist (deontological) ethics, and more in the distinction between studying the motivations for actions versus studying the (rational) justifications for actions, and in the particular pattern of indirection described by Larmore (1987).

The lesson I take away from the doctrine of "indirect consequentialism" is that "setting A over B" may be "a better way of pursuing B than setting B first"; which is one way of saying that some things may be just too important to be left to rational choice. That lesson may apply to the prisoner's dilemma, which seems to be a paradigm case where rational self-interest is best served by *not* being motivated by it.

In the target article, Caporael et al. are concerned with the "cognitive machinery," the "mental structures and processes," the "underlying psychological factors" that motivate behavior. They argue that people behave cooperatively and promote group welfare even in the absence of the incentives that might stimulate and encourage egoistic motivations.

The main theme of the essay is in effect what the "Book of Job," a rather ancient document, is all about. The devil said to God, "Job loves you for your money. The only reason he is so virtuous is that you reward him for it, with wealth and prosperity." So God put Job to the test. He withdrew his favor and brought havoc on Job's life and family. And Job remained devout, for a very long time, even without egoistic incentives. In the end, God never gave Job a good consequentialist reason to motivate his righteousness, and Job never asked for one. I suppose we should leave it to the devil to suggest that virtue *shouldn't* pay.

I do not want to get into the details of Caporael et al.'s demonstration experiments, or quibble over the cogency of their argument that no conceivable egoistic motive (long-term or short-term) could explain their results. The genre of the target article – a narrow, contrived, artificially controlled experiment and a sweeping evolutionary speculation trying to be comfortable in each other's presence – may draw attention, but I can live with contrivance and speculation. I'll leave it to others to complain. Still, one might ask, in passing, was the experimental demonstration really necessary? After all, everyday life provides us with numerous examples of action motivated by a love of the virtues without regard for other, more "extrinsic" rewards; and, as we have known for a long time, many, perhaps most, people elect to cooperate in a single go at the prisoner's dilemma.

Of course, I know there will always be theoretical purists about, but they are incorrigible. And need we really be troubled if some incorrigible hedonist wants to argue that it is merely an egoistic or selfish motive to want to maximize feelings of decency, dignity, and personal sanctity, or to care about promoting the real interests of others, or to keep faith with the idea of community? So much the better for egoism, if it turns out that self-respect is in your self-interest (see Shweder & Much 1987 and Shweder et al. 1988 for a discussion of that tenet in Hindu ethics).

The real difficulty for Caporael et al.'s argument, as I see it, is that the denial of in-group identification or of social (or moral) motivations is not really entailed by any "sensible" EI theory. That is because "sensible" EI theories, *qua* EI theorists, are not concerned with what motivates individual action; they are concerned with the constraints (including the payoff matrix) that actions must satisfy, if those actions are to be justified as rational. And they assume that with competition and feedback there will be a premium on the evolution of some set of motives for action whose outcomes will satisfy those constraints.

That is why so many EI economists eschew any interest whatsoever in the motivations of individual behavior. For them a payoff matrix is a description of the constraints that actions must satisfy if they are to be justified as rational, and it is enough if people act "as if" they were rational, so that, in the aggregate or in the long-run, their real interests get served. Whether they act out of a sense of duty, or to maximize self-interest, or to feel righteous by asserting through action some fundamental value, or merely to please high status members of their ingroup is, from the justificatory perspective, beside the point.

So my basic worry is that to say that people make rational self-interested decisions is ambiguous between at least two senses: (1) (the position Caporael et al. properly reject) that all motivation is self-interested (and not other-regarding); and (2) (the position "sensible" EI theorists should endorse) that, whatever the motive, our decisions will, in effect, serve our real interests (even as they serve other interests as well).

In other words, to justify (or explain) conduct by saying that so-and-so does the rational self-interested thing in doing this or that is not the same thing as saying that this or that is done for self-interested motives. Being motivated to set A [others, community, virtue, God] over B [self] may be "a better way of pursuing B than setting B first."

It is noteworthy that Caporael et al. conclude by honoring Adam Smith's notion of sympathy or "fellow-feeling" (Smith described it as "changing places in fancy," 1969/1759, p. 202). Whether or not Adam Smith would qualify as an indirect consequentialist is perhaps a debatable issue, but in his "*Theory of moral sentiments*" (1969/1759, pp. 238–39), there is a discussion of "the governing principles of human nature" that is congenial to the doctrine. For Adam Smith it is the essence of the "moral faculty" (i.e., "conscience") to determine which of our preferences are preference-worthy, for, as Adam Smith knew very well, it is not rational to maximize your preferences unless they are "graceful and worthy" ("fit," "right," "proper to be done"; see Rescher 1989 for an important discussion of rationality that develops that point).

For Adam Smith, the *justification* for the moral faculty (what gets defined as fit, right, and proper) derives not from individual preferences but rather from a superior intelligence (a deity or sovereign who can "direct the conduct of his subjects" with an eye to the collective good). Yet Adam Smith recognized that the *motive* for conduct is far more immediate. For the "laws of the deity [are] promulgated by those vicegerents which he has set up within us" and they are "attended to with the sanction of reward and punishment," by which Adam Smith meant, quite explicitly, the reward of self-respect and mental contentment and the "torment of inward shame, and self-condemnation."

One might well ask: If you have Adam Smith, who needs Freud? Instead I ask: If you have Adam Smith, who needs to oppose moral *motivations* to EI *justifications?*

The fallacy of selfish selflessness

Edmund J. S. Sonuga-Barke

Department of Child and Adolescent Psychiatry, Institute of Psychiatry, University of London, London SE5, England

When social scientists phrase theory in terms borrowed from "ordinary" language they often create conceptual muddle. This is especially so when these terms are judgmental. An example of such a muddle is the idea that the rule of conscience, even when it promotes moral acts, is the most subtle form of human selfishness. EI theorists regard actions determined in this way as "selfish" when they can be traced back to some inconspicuous internal "psychological" reward. It is as though the tricks of those saints canonized for lifelong selflessness have at last been exposed.

Having accepted this argument, and ruled out the role of other social and material incentives for cooperation, Caporael et al. feel that the failure of EI theories of choice in social dilemmas is demonstrated by a bias toward "own-group" giving, which they argue cannot be traced back to the effects of the "psychological" rewards associated with a "clear conscience."

But the authors have mistaken a conceptual problem for an empirical one. The error that is actually made by EI theorists is more fundamental than the present authors realize. Caporael et al. charge EI theorists with making an incorrect statement about human nature, the falseness of which is demonstrated by the series of game-theoretic experiments described here, although their real error is to believe that the theory allows them to make any sensible and helpful statements at all about human nature.

This error is nicely demonstrated by the way conscience is conceptualized by some EI theorists. The term "conscience," used in this way, acts only as a figure of speech for those psychological mechanisms that organize the internal rewards that regulate interpersonal behavior. So the question for these EI theorists is not "Does conscience play a role in regulating social action (by definition it must)? but "How does it operate?" Given this, it is clear that EI theorists might reinterpret Caporael et al.'s final experiment as showing that the internal rewards associated with a "clear conscience" are organized on a more specific level (that of the "ingroup") than the present authors had at first imagined. Caporael et al. had argued that the actions of conscience should not disciminate between "own" and "other" group.

Indeed, implicit in the alternative sociality hypothesis presented by Caporael et al. is a recognition of the conceptual hole they have dug for themselves. Having argued that the "psychological payoffs" associated with acts of conscience could not explain "in-group" biases, they propose that "cognitive/affective" mechanisms which have evolved in "small face-to-face groups" form the basis for these biases. Clearly the term "cognitive/affective" mechanisms is just another figure of speech for what some EI theorists, and the present authors, had previously meant when they talked of conscience.

At this point, it will be helpful to compare the way the authors use the concept of EI in the theories they discussed with Rachlin's use of rationality in his Behavioral Economics [see also Rachlin et al.: "Maximization Theory in Behavioral Psychology" *BBS* 4(3) 1981.] At first glance, the two concepts appear to behave in a very similar way. For Rachlin (1980) "rationality" performs the function of an abstract working assumption rather than a hypothesis about the nature of human motivations. It provides the axiomatic basis from which falsifiable predictions are derived about the function maximized by particular behaviors. All behavior is by defiition "rational"; the aim of behavioral economics is to discover the maximand. This approach can provide an alternative way of looking at certain common psychopathological phenomena. For example, Sonuga-Barke (1989), has suggested that the core components of childhood hyperactivity (impulsiveness, inattention, and overactivity) can each be regarded as maximizing experienced levels of reward immediacy.

Similarly, the way EI theory uses the term "conscience" as a "catchall" to explain prosocial action highlights the unfalsifiable nature of EI theory. Like statements about rationality in Rachlin's theory, statements about human selfishness originating from EI theory are tautologous. The difference between the two approaches is that whereas Rachlin adopts the assumption of rationality to sidestep what many feel are uninteresting and perhaps even unanswerable questions about human nature (cf. Lea et al. 1987, Ch. 5), EI theorists make the mistake of thinking that they can answer these questions and hence make judgments about human nature's essentially selfish quality.

This clearly cannot be done. Given the unfalsifiable nature of these statements – thus, the logical impossibility of describing an unselfish action in terms of EI theory (its polar term) – we can see that the concept of selfishness is empty and redundant. In

fact, given all this, for EI theorists to say that all actions are selfish gives us no more information than to say all actions are actions and each has some cause. EI theory, as usually presented, cannot, by its very nature, make sensible judgments about the essential quality of human nature. The mistake of Caporael et al. is to imagine that it could.

So on what level do statements about "selfishness" make sense? Correct use of language in scientific discourse depends on a recognition of its meaning in ordinary language. For example, one meaningful use of the term rests on the idea that a selfish act is the product of calculations motivated by desire for personal gain. With this usage it is clearly possible to conceive of both selfish and unselfish acts. Of course, it is still possible that the apparently selfless saint is driven by such motivations, but at least it is not necessarily so.

Demonstrating unselfishness: They haven't done it yet

Stephen C. Stearns

Zoology Institute, University of Basel, CH-4051 Basel, Switzerland

The issue is important, the experiments are interesting, the results suggest optimism, but this research program is not yet strong enough to bear the weight of the claims the authors make. Caporael et al. have not shown that most people do not behave selfishly most of the time. They have shown that some people make small sacrifices to benefit group interest in a context that all involved clearly recognize as a "test situation." Whether that behavior will translate into larger sacrifices in larger groups is an unanswered question.

What is the maximum size of the group in which such effects occur, and how does the size of the sacrifice an individual is willing to make vary with the size of the group whose interests are at stake? Even if the effects are real and general across cultures for groups of 10 to 20 people, that does us little good if they dwindle or disappear in groups of hundreds of millions or billions of people.

War is, after all, possible only among groups that include some members who are willing to sacrifice themselves for the general welfare, but the same psychology that permits that sacrifice also identifies the opposing groups as the enemy. The tragedy of the commons might be avoided in a small group in which each member could be motivated on irrational grounds to undertake sacrifices for the good of the whole, but if the maximum size of the groups in which such sacrifices can be expected is much smaller than the population of the planet, then several large groups may simply organize themselves all the more efficiently in competing for the remaining resources. That is precisely the effect that one hopes to avoid.

Table 5 suggests that in-group, out-group effects are quite real. Does it do the cause of peace or the environment much good if an inevitable consequence of belonging to one group is being antagonistic to another such group? That just shifts the problem of selfishness to another scale or organization; it probably makes it worse.

Even if people can be convinced that they belong to a group defined as the population of the whole planet, whose interests they share and should promote, the question remains: How large an individual sacrifice will be made, on average, to promote those interests? Is the maximum likely sacrifice too small to be effective? Is peace or a healthy environment worth worldwide economic stagnation? Many would regard that as too high a price to pay for either. Even if a few large groups (nations) were willing to pay the price, history shows that a minority of defectors can destroy the whole system and engage the altruists in acts of self-protection whose byproducts are every bit as destructive as selfishness would have been in the first place. The lessons of history do not, in general, leave much room for optimism, and history itself can be regarded as a long series of psychological tests with much larger sample sizes and payoffs than those described here.

Defining the limits of selfishness and understanding just how much people are willing to do without is doubtless important. Much more important, however, is knowing how to educate people to understand what is in their long-term self-interest. Short-sighted altruism is almost certainly worse than long-sighted selfishness in terms of its practical impact on the environment and on the quality of life of the average citizen of the planet. The limits of our motivation are not so critical as the limits of our comprehension. Once the problems are understood, several different motivational systems could lead to solutions acceptable to all – even if one of the motivational systems is selfish and one of the groups that must accept the solution is altruistic.

Can we afford not to believe that man is selfish?

Nils Chr. Stenseth

Department of Biology, Division of Zoology, University of Oslo, Blindern N-0316, Oslo 3, Norway

Caporael et al. claim that man, *Homo sapiens*, is basically a cooperative creature – not a selfish creature as sociobiologists often suggest (e.g., Dawkins 1976). Caporael et al.'s claim may or may not be right.

I am by no means convinced that such a broad conclusion can be reached on the basis of the experiments they discuss. Neither am I convinced that we, on the basis of the material they present, need to ask for a "serious reexamination of the 'selfish human nature' assumptions" (sect. 4, para. 7) that they claim dominates the human sciences.

Space does not allow me to discuss fully the validity of the conclusions drawn by Caporael et al.

Whether or not humans are selfish, however, is an interesting problem, both to biologists and to human scientists, partly because the answers may help us understand the way we go about organizing society: Do we need egoistic incentives and constraints (enforced by, for instance, laws and traditions) to cooperate and not jeopardize mankind's future existence? Do we really need restrictions regulating individuals' behavior to make them cooperate and not destroy the future existence of our own species and not overexploit the common resources such as fish and wildlife, clean air, and clean water?

We need to ask what are the consequences of viewing *Homo sapiens* as a cooperative creature, should he really be egoistic. Unfortunately, these consequences may be rather disastrous. (Viewing humans as egoistic should they really prove to be cooperative by nature has no similar disastrous consequences.)

Currently, there is a tremendous overexploitation of our common resources (e.g., World Commission on Environment and Development 1987). Such natural resources as fisheries, much of the African savannas, tropical rain forests in South America, and so forth, as well as such common resources as clean air, clean water, and so forth, are currently being destroyed – they may not be available for future generations. It seems clear that these resources are being destroyed because of man's inability to cooperate to preserve/conserve his resources. Hardin's (1968) discussion of "The Tragedy of the Common" is a case in point.

With this background we may ask: Can we, under the current circumstances of resource destruction, take the risk of erroneously assuming man to be cooperative by nature and thereby take the risk of not imposing as many restrictions on individuals' behavior as we otherwise would do? In society at large and to policymakers in particular, taking such risks might seem rather rewarding on a short time scale (they might get voters'

support), but it could well turn out to be ruinous in the long run: It might bring us to the tragedy of the commons faster than ever by overexploiting fisheries, and so forth, and by polluting the air and the water. It is more important than ever not to make erroneous decisions on these issues. Whether it is natural for human beings or not, we must start cooperating more than we have until now. The view held by Caporael et al. makes us less willing (or likely) to provide the appropriate measures needed for cooperation to result if we are in fact selfish creatures.

The message conveyed by these authors could indeed be rather damaging if people take it seriously. Because their empirical basis for claiming that man is cooperative – not selfish – is rather meager, this is rather unfortunate: Indeed, it may prove disastrous.

As scientists we should, of course, have an open view about how the world functions. We should indeed be careful not to let our political views influence the way we think nature is structured. That is, we should not let our political attempts to protect the resources surrounding us influence our scientific conclusions about basic human nature. As citizens of this earth, however, we should also be aware that our theories are being used – consciously or not – as a basis for formulating policies.

Because we do not know the true nature of humans, we would minimize the likelihood of error by assuming that we exploit nature in an egoistic manner. I wish Caporael et al. had been clearer on this point.

Time is running short. We cannot afford to believe that man is basically unselfish if he really is selfish.

Counting contributions

Karl Halvor Teigen

Department of Cognitive Psychology, University of Bergen, Bergen, Norway

In studies of human behavior, the answers offered depend heavily on the questions raised. Which behaviors call for an explanation? According to Caporael et al., there is a bias in contemporary psychology toward regarding cooperative behavior as a deviant phenomenon to be explained, whereas selfishness seems to be accepted as the normal state of affairs, making the egoistic incentive (EI) theory an explanation by default. It is easy to agree with the authors that the primacy of the individual versus the group is a matter of metatheoretical assumptions, but one may well ask whether paradigmatic conflicts at this level can ever be settled by appealing to experimental results. The route taken by Caporael et al. is to demonstrate a class of phenomena that EI models are presumably unable to explain. It would perhaps be even more fascinating to start from the opposite assumption, namely, that people are basically social beings (not a very revolutionary assumption, after all), making cooperative behavior the norm rather than the exception. The burden of proof would then rest with the researcher who wanted to demonstrate instances of behavior that were purely individualistically motivated – which might well turn out to be an equally difficult task. Perhaps the most parsimonious model would be one that assumed human actions to be invariably determined by in-group considerations, but with wide margins for what could constitute an in-group, including "groups" with N = 1 as the limiting case. (Even then, the single "group member" whose welfare the actor is trying to maximize will not necessarily be the actor himself.) From this point of view, the important research questions will concern the possible determinants of in-group composition and size. It may not always be fruitful to draw a sharp line between individualistic and collective motives, not even for didactic purposes. Caporael et al.'s treatment of conscience is a case in point: There are obviously EI theories of conscience (e.g., the Freudian one). But the concept of conscience is not exclusively a matter of internal punishments and rewards. To do something according to one's conscience may well be to do it according to one's social nature, not out of selfish fears or feelings of guilt and remorse. Perhaps there is a need for a more inclusive analysis of what the concept of conscience could entail before it is dismissed as another subtle version of the EI doctrine.

So when does cooperative behavior fail to appear? My best guess, based on Caporael et al.'s data, is that this will occur when people have no identifiable target for their contribution, or when their contribution appears to be futile. The authors make a point of showing that a number of subjects (Ss) behave "inconsistently" with their beliefs, especially in being willing to contribute for the common good even if their contribution is not "critical." But it should be noted that Ss are never directly asked whether they believe their contributions are redundant or are serving a purpose. They are just asked how many of the other participants they think will make a contribution. This way of putting the question may be ill-suited to assess the importance that Ss attach to their own actions. Judgments of responsibility and causality are known to be influenced by which actor one is focussing on. Moreover, probability judgments, like those used in the present study, can be given several interpretations. Even if one thinks there is more than a 50% chance that a sufficient number of other participants will be paying, one's extra payment can be perceived as necessary for arriving at the "safe side." Actually, Ss may think that because anyone's contribution is equally important, it follows that each one's contribution really matters. As shown with verbal probability estimates, *equal* likelihoods tend to be interpreted as *great* likelihoods (even when the number of equiprobable alternatives ranges from 4 to 8), especially for the individual or the outcome that happens to be in focus (Teigen 1988). Without such collective illusions, very few people might have bothered to contribute their votes on election day. Although they have give their Ss a fair chance of showing their willingness to cooperate, Caporael et al. have given them less of an opportunity to express the importance they attach to it.

The selfishness-altruism debate: In defense of agnosticism

Philip E. Tetlock

Institute of Personality Assessment and Research, University of California, Berkeley, CA 94720

Caporael et al. launch a frontal assault on egoistic incentive theory – a loose intellectual alliance of reinforcement theorists, economists, sociobiologists, and political philosophers who share a belief in the basic selfishness of human nature. Caporael et al. claim that egoistic incentive theory rests on flawed evolutionary premises and is hard pressed to explain potentially disconfirmatory evidence without resorting to "just-so" stories. The assault is partially successful – a substantial achievement, given the well-entrenched historical status of the foe.

The proper theoretical response to the biological critique is to concede that it is indeed possible to weave a number of plausible evolutionary scenarios around our emergence as a species. Caporael et al. may well be right that the small face-to-face group (15 to 30 individuals) was more likely to have been the primary locus of natural selection pressures than the solitary individual. They may also be right that affective and cognitive mechanisms supportive of group living are genetically linked attributes that were highly conducive to the survival of early hominids wandering the savanna plains of Africa. Empathy, a predisposition to share, and punitiveness toward cheaters are likely natural consequences of this group life-style. But they are not the only possible consequences. It was probably no more "adaptive for ancestral humans to identify automatically with an in-group and to accept its goals as their own" (sect. 4, para. 6) than it is now for modern humans. Fanatics – those most eager

to leap to the defense of the group against predators or to lead the attack against rival groups – often have life expectancies too short to pass their DNA-encoded commitments onto the next generation. I suspect that there is an Aristotelian golden mean lurking here: Nature has probably smiled especially kindly on those who recognize that their long-term self-interest hinges on the viability of their group, but who also recognize that serious conflicts can arise between individual and collective interests and who appreciate the importance of knowing when to "defect" and how to do so without incurring the wrath of the collectivity. In short, human nature is complex and embodies many, often conflicting motives; evolutionary arguments are necessarily speculative; and agnosticism on the "selfishness" question probably remains the most prudent position.

Caporael et al. not only question the evolutionary hard core of the selfishness research program, they also raise an intriguing empirical challenge. In a fascinating series of public-goods experiments, they claim to have falsified the hypothesis that people will not contribute to group welfare in the absence of egoistic incentives to do so. Subjects in these experiments were strangers who made a single decision under conditions of anonymity and in the absence of interaction either before or after the experimental session. The basic design renders implausible many of the classic egoistic incentives (e.g., reciprocity, coercion) but not all of them. As the authors are well aware, advocates of egoistic incentive theory have numerous potential lines of theoretical defense. In various experiments, substantial percentages of subjects may have cooperated out of a desire to avoid guilt, to bolster their self-esteem, to help their team win, or just to put themselves in a good mood. Indeed, there is independent evidence that each of these motives promotes prosocial behavior (Krebs & Miller 1985). There are also potential methodological defenses. For example, subjects may not have believed the assurances of anonymity (why would the experimenter bother to conduct a study in which it is impossible to monitor what I've done?)

The key question is, of course, whether the results reported, taken as a whole, force egoistic incentive theory to invoke such an array of ad hoc hypotheses that, far from providing a parsimonious integrative framework, the theory begins to look like a rather tattered patchwork quilt. My view is that egoistic incentive theory is indeed worse for the wear but can still be rescued with a measure of dignity by forging new intellectual alliances – in particular, with social identity and role theorists who assume that people strive to create the most favorable identities for themselves within the constraints of social situations (cf., Hogan 1982; Schlenker 1982). *Homo economicus* must meet *Homo sociologicus* at least halfway. People are trying (not always successfully) to maximize socially defined or constructed utility functions. From this standpoint, independent variables can shape the willingness to contribute to public goods in two basic ways: (1) by affecting the identity implications of response options (people do not want to appear selfish or foolish but do want to appear loyal to the group, trustworthy, and generous); (2) by affecting the importance of different identity objectives or of different audiences (self vs. other, in-groups vs. out-group). Manipulations such as anonymity, no-free-ride, opportunity for group discussion, the minimum number of contributors, and group membership status of potential beneficiaries must influence willingness to contribute through one or the other of these mediational processes – a sweeping hypothesis that can only be tested by painstakingly measuring the self and social identity implications that participants and observers feel it is reasonable to draw from different response options in the original experimental situations (for examples, see the "interpersonal simulations" by Alexander & Knight 1971; Tetlock 1980).

In closing, it is reasonable to ask whether the proposed intellectual merger would negate most of what was provocative and interesting about the original versions of egoistic incentive theory. It is one thing to claim, *pace* Hobbes, that people are selfish to the core and will cooperate only under threat of centralized coercion; it is quite another to claim, *pace* Goffman, that people are so thoroughly socialized that they will readily make monetary sacrifices to avoid appearing to be the "wrong" type of person, either in their own eyes or the eyes of others. We can choose to call these two positions variants of egoistic incentive theory, but they are really very different. From a metatheoretical point of view, the key question concerns where we should draw boundaries around the powerful organizing metaphors – *homo economicus*, the actor, the scientist, the gene-maximizer – that still give intellectual impetus and direction to so much work in the behavioral and social sciences. As I have argued from the controversies over the merits of cognitive versus motivational explanations of social judgment (Tetlock & Levi 1982) and intrapsychic versus impression management explanations of interpersonal behavior (Tetlock & Manstead 1985), there is no nonarbitrary way to demarcate the explanatory range of research programs organized around competing metaphors. It is hard to tell the difference between a social-identity-theory variant of the egoistic incentive research program and the evolutionary-bases-of-group-altruism argument advanced by Caporael et al. In the end there will probably be no dramatic crucial experiment. The two positions may just gradually blur into each other. This theoretical convergence should, moreover, be taken as an encouraging sign that, notwithstanding the claims of the radical subjectivists, research programs anchored in starkly different first premises can be responsive to a common body of evidence.

Evolutionary psychologists need to distinguish between the evolutionary process, ancestral selection pressures, and psychological mechanisms

John Tooby and Leda Cosmides

Department of Psychology, Stanford University, Stanford, CA 94305
Electronic mail: *leda@psych.stanford.edu*

Although many of the experiments reviewed by Caporael et al. are interesting, and we are sympathetic to some of the intuitions that motivate the target article (e.g., that "selfishness" and "self-interest" as phrases can sometimes be misleading as characterizations of human psychological mechanisms), the authors' misconstrual of much of modern evolutionary theory makes their interpretation of their own work problematic. In particular, their categorization of modern evolutionary theory as an "egoistic incentive" (EI) theory misconstrues claims about the evolutionary process itself as claims about the psychological mechanisms that are the shaped product of the evolutionary process. Because Caporael et al. are not alone in this confusion, but rather hold misconceptions that persist both within and outside of the evolutionary community, it is worth dwelling on exactly where arguments about whether humans are "basically selfish" or (the authors' alternative) "basically social" go wrong.

In approaching a given species' behavior from an adaptationist perspective, evolutionary analysis requires three nested but distinct levels. These are:

(1) Models of the evolutionary process (involving definitions of fitness, selection, adaptation, genes, and the role of stochastic factors, and general models of such topics as kin-directed altruism, reciprocation, sexual recombination, and sexual selection);

(2) An analysis of how these principles were manifested as a species-specific array of selection pressures, refracted through the specific ecological, social, genetic, phylogenetic, and informational circumstances experienced along a given species' evolutionary lineage;

(3) a description of the species' innate adaptive specializations that evolved to solve the problems posed by the species-specific array of ancestral selection pressures.

The advances over the last three decades in the sophistication and power of evolutionary theory have justifiably generated an enormous amount of excitement, and although much still remains to be done, this first level is the best developed of the three (Dawkins 1976; 1982; Hamilton 1964; Maynard Smith 1982a; Williams 1966). The second level, the reconstruction of the array of selection pressures encountered over hominid evolution (in the human case), has made a limited amount of progress (see discussion in Tooby & DeVore 1986). The third level, the mapping of human psychological mechanisms as adaptations, is in its infancy and depends strongly on the other two for its success. The three levels fit together in a tightly structured way, and all are necessary to address any behavioral issue with confidence. The relative maturity of evolutionary theory compared to the other two levels has led the overeager to try to leap directly from evolutionary theory to psychology, in the belief that evolutionary theory by itself constituted a theory of psychology (Cosmides & Tooby 1987).

To understand exactly why one cannot leap from evolutionary theory to psychological theory, consider a thought experiment involving members of a hypothetical species of fig wasp whose life history leads them to enter, and spend all of their adult lives, encased in a single fig. Further assume that after recruiting a single mate during their juvenile phase, each pair colonizes a different fig. Adults would always interact only with their mates, and would never encounter another conspecific. The evolved mechanisms regulating altruism towards the mate would act as if they "valued" the welfare of the mate as highly as the individual they were in, because the death or disability of the mate would be reproductive death for the wasp (see discussion in Alexander 1987 on marriage). Moreover, such mechanisms would not evolve to dole out assistance differentiately depending on the recipient, but would be indiscriminately altruistic, because the only conspecific ever encountered by an adult was its mate, and so indiscriminate aid was never an error. Finally, imagine researchers who brought such wasps into the lab to investigate whether wasps were "selfish" inclusive fitness maximizers or whether the "selfish gene" theory was bankrupt. In raising such wasps with others, and running experiments on them in various social situations, they would discover that the wasps were, in contradiction to "theory," perfectly altruistic towards nonrelatives (thus ruling out inclusive fitness explanations of altruism) and toward "strangers" (ruling out reciprocation or mate cooperation as an explanation of altruism). Thus, a little hardheaded empiricism would dispel all these "selfish gene" theories. Similarly, Caporael et al. experiment on humans by putting them in situations where, for example, the "subjects were strangers," "their choices were anonymous," "they made a single decision," and "interaction among group members was prevented before and after each session" – i.e., in evolutionarily unprecedented situations, completely unlike anything humans would have regularly encountered during their Pleistocene evolution, and therefore situations to which their psychological mechanisms are not adapted.

Clearly, without a theory of the selection pressures operating in ancestral environments, evolutionary theory has little to say about psychology, and equally, psychology without this necessary intermediate cannot serve as a test for "selfish gene" approaches (as Caporael et al. appear to believe). This intervening level is necessary to make models of psychological mechanisms relate to evolutionary theory and vice versa. When these three levels are not kept clearly in mind, and models are not evaluated in terms of the appropriate level of analysis, confusion abounds. What Caporael et al. call "selfish gene" theories are the best existing characterizations of how the *process* of natural selection operates: to refute these theories, the authors would have to propose some new theory of the process of natural selection, or some overlooked element in present models of fitness, and this is exactly what the authors do not do. Instead, they misinterpret such theories as theories about motivation – something like: Under all circumstances, evolutionarily unprecedented or not, a human cognitively represents inclusive fitness, and "selfishly" has as a goal the desire to maximize it. In all fairness, one must admit that some evolutionary researchers treat evolutionary theory this way as well (i.e., they treat humans as fitness strivers), although more often this is simply a metaphorical conceptual shorthand, used to avoid cumbersome constructions (as Dawkins [1976; 1982] takes great pains to make clear). The requirement to find "individualistic payoff," that is, enhanced genetic propagation, as the designed consequence for any psychological adaptation is not a level 3 statement about psychological goals. Rather, it is the expression of the central requirement of the theory of natural selection (level 1): adaptations exist, and have the form they do, because they enhanced genetic propagation (fitness) in ancestral conditions. No matter what psychological mechanism is proposed, it is incumbent on the researcher to show how genes coding for it could have spread through the population. Discovering how they accomplished this is not a luxury (let alone a metatheoretical prejudice deriving from Western cultural beliefs), but is rather an inescapable logical necessity deriving from the structure of evolution by natural selection. Caporael et al.'s "sociality hypothesis" is a vaguely expressed theory of level 2 (species-specific selection pressures), introduced to account for the in-group bias phenomena they and others have encountered (a psychological phenomenon, level 3). They incorrectly see their level 2 hypothesis as exempting them somehow from showing how in-group bias (a level 3 phenomenon) leads to "individualistic payoffs," that is, enhanced fitness (the level 1 requirement) in ancestral environments (the level 2 context). The logically required attempt by others to find level 1 and level 2 explanations for group living (i.e., its fitness payoffs) is interpreted by the authors to mean that other researchers are somehow ignorant of the fact that humans evolved in social groups, and have extensive psychological adaptations to social life.

This confusion of levels does conceal several virtues in the target article: The authors are right that a direct mapping of evolutionary theory as a motivational theory is inadequate, and that not everyone recognizes this; their experiments add to our knowledge of the variables that our psychological mechanisms respond to in cooperation; in-group biasing is a phenomenon that must be addressed and explained (we favor a coalitional explanation in which small scale coalitional aggression was likely to have played a prominent role; Alexander 1971; Tooby & Cosmides 1988). Nonetheless, we find the claim that "it was adaptive for ancestral humans to identify automatically with an in-group and to accept its goals as their own" and that "human nature is basically social rather than selfish" as no more plausible a psychological characterization than the idea that humans cognitively represent fitness and selfishly pursue it as a goal. Human nature is not "basically social" or "basically selfish"; human nature is "basically" a collection of mechanisms designed to achieve genetic propagation in our environment of evolutionary adaptedness in ways that are sometimes characterizable as "selfish," sometimes as "altruistic."

Selfishness, sociobiology, and self-identities: Dilemmas and Confusions

Ian Vine

Interdisciplinary Human Studies, University of Bradford, Bradford, BD7 1DP, England
Electronic mail: *i.vine@cyber2.central.bradford.ac.uk*

The research programme of Caporael et al. certainly advances our understanding of immediate causes of prosocial choices, in intergroup as well as intragroup social dilemmas. But the authors interpret their data as refuting a loosely formulated egoistic incentive (EI) hypothesis and supporting a still more

ambiguous sociality (SO) hypothesis – each appearing to span evolutionary functional, as well as psychological, explanations of human nature. Their theoretical framework appears marred by confusing conceptual contrasts and elisions, by lapses in differentiating levels of empirical and even philosophical analysis, and by misrepresentations of others' claims. Because the main target for attack seems to be sociobiological theories founded on inclusive fitness concepts, I shall focus mostly on this part of their analysis.

The most obvious conceptual infelicities appear at the outset, with the category mistake in the SO hypothesis –that our nature "is basically social rather than selfish." Sociability or gregariousness should be contrasted with asociality or solitariness. Truly selfish acts or motives refer primarily to lack of appropriate concern for others' interests, and thus refer to *social* contexts. Competitive interactions are also social, and even individualism must normally be understood against the substantial background of social interdependencies found in all societies. Such continuous dimensions as selfish/altruistic, competitive/cooperative, and individualistic/collectivist are roughly orthogonal to asociality/sociality. A tendency to equate these dimensions, and to talk as if our nature must be *either* entirely selfish *or* fully altruistic, creates numerous ambiguities. Such dichotomous thinking may be useful in designing gaming experiments, but it does not clarify which questions the tests are resolving. If we take the SO hypothesis to be about *pro*sociality, then neither that hypothesis nor the EI alternative is an empirically plausible option (as conceded in section 4, para. 6) and in note 2) –even if their vagueness masks this fact.

A readiness to take sociality itself as problematic presumably accounts for the loose collection of points in section 3. Eschewing the rigour of sociobiological models, which aim to consider the fitness of specific categories of social traits, the evolutionary account of hominid group-living sheds little light on which types and levels of social complexity and integration could benefit group members in which ways – and shape our own cognitive-affective mechanisms. Sociobiologists' views are misreported (e.g., Alexander 1979, p. 23–24) in the authors' determination to discredit the theory that features of sociality evolve to the degree that they serve individuals' fitness interests. Claiming that Hamilton (1964), Trivers (1971), and Alexander (1987) ignore our social primate origins, and "describe ancestors who begin in isolation and evolve to a social state in order to achieve the same egoistic goals" (sect. 3, para. 2) that they had before is a most peculiar misrepresentation. It suggests that Caporael et al. mistake for empirical claims the use of analytic fictions, simply designed to expose the formal structure of adaptive cost-benefit functions for varying modes of social life.

Uncharitable readings of what most sociobiologists insist upon are also required to claim that they "mistake evolutionary forces for the psychological mechanisms that result from them." In retrospect, the use of terms like "selfish" and "altruistic" in the original population-genetic models invited critics to make such errors, without due reference to the quite explicit technical meanings the terms had regarding resource transfers with fitness implications. Dawkins (1976) is somewhat atypical in emphasizing "competitive" selection and differing "interests" at the level of genes themselves, because he holds that a "gene's eye" perspective is less vulnerable to error than thinking of individuals or groups in working through the issues about fitness relevant to evolutionary change. Yet, he asserts clearly that human behavour is in no direct sense specified as motivationally selfish by the genes that help construct us. It is not reformulations of Darwin's insights about selection at the genetic level that "have provided the ultimate justification" for "economic man" and EI notions. Some crude early applications of fitness models to humans in genetic determinist ways, by Wilson (1975), Barash (1979), and others, are a rather thin excuse for caricaturing the models themselves. [See *BBS* multiple book review of Lumsden & Wilson's *Genes, Mind, and Culture BBS* 5(1) 1982.] And here it is Caporael et al. who refer to concepts like "self-interest" without properly elucidating their quite different senses in varying analytic contexts. They are nominally alert to the relative independence of distal and proximal explanations of behaviour, but they do not reliably differentiate between variants of the EI or SO hypotheses at different levels.

In terms of conscious goals it is clear that prosocial acts can be truly unselfish if motivated by sympathy or reasoned moral concern. They can be functionally self-sacrificial, in an immediate material sense, even if recompense is anticipated or subjective satisfactions are thereby sought. Merely experiencing unexpected "positive psychological payoff" after acting cannot suffice to show an "egoistic incentive" (as stated in sect. 1.3). Tricky cases arise where functional benefits to oneself have contributed to acquisition of a trait, but may be actively repressed from consciousness while one is exhibiting it. Where self-sacrificial traits reflect promptings that can be traced back to naturally selected genes, or to cultural forces that also tend to serve our "ultimate" and normally unconscious "genetic self-interest" (in the extended sense of inclusive-fitness theory), the sense in which we act "selfishly" in EI terms is extremely attenuated. A mother risking her life to protect the child she loves is surely altruistic in moral terms, quite irrespective of having furthered her unconscious fitness interests (Vine 1983). Although sociobiology does *not* presuppose direct "genetic determinism" for higher animals (e.g., Dunbar 1987), theorists like Alexander (1987) hypothesize that our phenotypic traits are quite tightly constrained to restrict prosocial acts to what does serve inclusive fitness. But where our motives are those of a subjective self-system, its plasticity and susceptibility to social demands and manipulations (Crook 1980; Marková 1987) suggest that the voluntary overriding of fitness constraints *is* humanly possible. I have proposed a model that accepts the social-identity theory of Tajfel and Turner (cf., pp. 18–19), so that when in-group identity encompasses others, we can assimilate their interests to our own interests. Although some genetic biasing in favour of our somatic self is likely to have evolved, we can learn to *transcend* fitness constraints (Vine 1987). This model fits the evidence that our nature is both self-interested *and* prosocially oriented.

The remaining variant of the EI hypothesis emerges in the "revealed-preference" assumption – which no competent sociobiologist should take to be absolute fact or evolutionary necessity. Caporael et al. sometimes appear worried by the philosophical egoism that they cannot in principle refute empirically. The arguments against it are conceptual (e.g., Midgley 1978). It is easily confused with the "unconscious self-interest" theory of sociobiologists like Alexander – and Caporael et al. sometimes appear to do this. That view supposes the extreme potency of natural selection to be empirically true. If so, then what we find subjectively satisfying will indeed simply reflect our own genetic interests – and no results from social-dilemma games will refute that EI hypothesis. But the philosophical variant essentially holds that every rational choice implicitly anticipates satisfactions that by definition are the agent's own. That makes EI inevitable – but trivial. It should not fool us into supposing that sacrifices that *result* in our feeling good are coextensive with those with "feeling good" as their selfish *goal*.

How best to critique egoism?

Lise Wallach and Michael A. Wallach
Department of Psychology, Duke University, Durham, NC 27706

The target article by Caporael et al. is a welcome sign that questioning is now occurring regarding the selfishness or egoism assumption – an assumption pervasive and dominant enough that its criticism has been a long time coming. It is especially heartening to see that three of the authors raising

questions have until only recently been EI theorists themselves. And we particularly welcome the elaboration of evolutionary mechanisms for the development of nonegoistic behaviors, because evolution is so often regarded as implying that such behavior does not exist.

It is also important to emphasize the target article's point that egoism cannot be tested by seeing whether egoistic incentives "work." Such has been a longstanding practice in psychology. For example, studies in the social-learning theory tradition demonstrating that children's donation behavior can be increased by having them observe models whose altruistic behavior is rewarded are repeatedly cited as evidence that expected personal returns underlie children's altruism (Wallach & Wallach 1983, pp. 183–85). As the target article rightly insists, however, if one is to test the assumption of egoism, one must determine, not whether prosocial behavior occurs in the presence of egoistic incentives, but whether it can occur in their absence.

This is what the authors have attempted to do. Unfortunately, however, in our opinion the experiments they report are not successful in that aim. These experiments provide good evidence that human beings do not act only on incentives "that would be just as highly valued by an isolated nonsocial animal as by a social one." (sect. 1.5, para. 2). It would be very difficult to understand why people who cared nothing about others or about social norms, who had no conscience, and whose self-esteem had no connection with whether they met or failed to meet their commitments, would cooperate in the social dilemmas described.

The experiments, however, do not seem to provide evidence against the sufficiency of egoistic incentives as these incentives are commonly understood, as they are implied to operate by most psychological theories, or as the authors themselves treat them throughout most of their article. According to Caporael et al. – and they would find wide agreement here – a payoff is no less egoistic for being internal rather than external, and to act in such a way as to have a clear conscience or avoid guilt is to act on an egoistic incentive. Guilt avoidance, however, appears a very likely explanation of the cooperation in their experiments.

Caporael et al.'s account seems to neglect the direct effects of what goes on when discussion is allowed. It would be reasonable for subjects, even from the most narrowly egoistic point of view, to offer to, and elicit from, one another promises to cooperate. What this would represent would be a kind of reciprocity of promises. Each subject, after all, stands to gain if they all (or the required ones) cooperate. This is indeed how people behave in discussions in social-dilemma games, as has been shown in an earlier article by Orbell, van de Kragt, and Dawes (1988) themselves.

But once a joint commitment to cooperate has been made during discussion, the meaning of actually deciding to contribute is very different from what it is without discussion. It seems clear that after promising to contribute, failing to do so – even if no one would know and there were never to be any further interactions – would be a violation of trust. Except for psychopaths, a subject who agreed to cooperate and then defected could hardly avoid feeling some guilt. The egoistic incentive of guilt avoidance is thus sufficient to account for such a subject's actually making the contribution.

Caporael et al. attempt to reject the possibility that conscience accounts for the cooperation obtained with discussion by arguing that if satisfying one's conscience is the incentive for contribution, then individuals should be as likely to contribute to strangers as to members of their own group. This, they show, is not the case. Subjects did not contribute to other groups as to their own, which the authors regard as supportive of in-group biasing. Caporael et al., however, seem to be assuming that the only possible bearing that discussion could have on the workings of conscience would be to enhance the salience of its demands. They are ignoring that discussion is also capable of *creating* demands of conscience – which is what seems to happen when promises are made. That subjects did not contribute to other groups as they did to their own may well have had nothing to do with in-group biasing but may simply be a reflection of the fact that it is only among the members of the group in which discussion occurred that promises could have been made. There would be no violation of trust in not contributing to groups other than one's own – one has made no commitments to these groups.

Conscience, then, cannot be ruled out as an incentive for the cooperation obtained under discussion conditions in any of the experiments reported here. Rather, it seems extremely likely that subjects who committed themselves to contribute during discussion were later motivated to carry out their promise to avoid feeling guilty. Cooperation in the absence of guilt avoidance has not been shown in this research, and thus it does not seem to provide evidence against the senses of egoism the authors mean to eliminate. We agree with Caporael et al. that EI theories give only a limited picture of our natures. It may, however, be very difficult, if not impossible, to exclude experimentally the operation of all internal as well as external egoistic incentives. Perhaps social-psychological experiments, or at least social-dilemma games, are not a fertile ground on which to critique egoism. Improved understanding of evolution may be more helpful.

Authors' Response

Thinking in sociality

Linnda R. Caporael,[a] Robyn M. Dawes,[b] John M. Orbell,[c] and Alphons J. C. van de Kragt[c]

[a]*Department of Science and Technology Studies, Rensselaer Polytechnic Institute, Troy, NY 12181,* [b]*Department of Social and Decision Sciences, Carnegie-Mellon University, Pittsburgh, PA 15213, and* [c]*Department of Political Science, University of Oregon, Eugene, OR 97403*

Electronic mail: [a]*caporael@mts.rpi.edu,* [b]*dawes@andrew.cmu.edu, and* [c]*jorbell@oregon.uoregon.edu*

We begin with a discussion of certain misconceptions about the experiments.

The Experiments. Although **Rapoport** has some kind words to say about the experiments, we (unsurprisingly) disagree with his interpretation that they constitute walking through an open door. Some people cooperate – (to their own disadvantage – some of the time, as in contributing to public broadcasting. We all knew that. The purpose of the experiments, however, was not to demonstrate that fact, but to determine how cooperation rates vary systematically as a function of social setting with payoffs constant, anonymity of choice, and no iteration. (While Rapoport is right that other single play experiments have been reported in the literature, these studies have not systematically varied social setting in the way ours did; in fact, many were designed simply to show that people do cooperate in such anonymous single play situations, or evaluate the propensity to cooperate as a function of payoffs or personality characteristics of the players – e.g. "internationalism.")

Our procedure was to assign subjects to conditions as randomly as possible. ("Randomization" is only an ideal in actual experiments; with the actual complexities of recruitment and scheduling, we can ensure only that we

have not violated random assignment in some specifiable way.) Thus, personality and (distant) kinship (which would have to be assumed on the basis of phenotypic similarity) were not relevant to the results. Nor were such constant factors as the fact (pointed out by **Dunbar**) that subjects made choices with money "given to them" rather than "their own" (a distinction not recognized by traditional economic theory, but of crucial importance to human subjects committees). Nor is it important that in all conditions the experimenters knew what choice the subjects made (as again pointed out by **Dunbar, Houston & Hamilton,** and **Knauft**). It is the *difference* between conditions that is important.

The more plausible EI alternative explanations presented by the commentators involve (1) keeping promises because the experimenters knew whether they had been kept (**Houston & Hamilton, Katz, Wallach & Wallach**), (2) "implicit" concern with reputation (**Dunbar, Houston & Hamilton, Krebs, Rachlin, Wallach & Wallach**) and (3) conscience, perhaps interacting with group identity (**Houston & Hamilton, Mansbridge, Sonuga-Barke**). None of these alternative explanations is consistent with data.

Promises. As we pointed out in the target article, manipulations other than discussion produce cooperation since that cannot involve promising. Of particular importance is the work of Brewer and Kramer (1986) and Kramer and Brewer (1986), in which the level of payment based on the same (collective identity) or different (individual identity) lottery for converting points resulting from the choices to money resulted in greater restraint or less restraint, respectively, in the use of a collective resource (points that could be exchanged for money).

In addition, promising cannot interact with the knowledge of the experimenters in the experiments discussed in the target article because the experimenters did not know who promised to cooperate and who did not. More important, however, are the experimental results of Orbell et al. (1988) mentioned by **Wallach & Wallach.** (We present these findings here, in part, because Wallach & Wallach's summary is misleading.) These experiments involved a trinary choice. Subjects first met in discussion groups of 14 and then were divided into two subgroups of 7 – at which point (with or without further discussion) subjects anonymously decided to keep \$5 for themselves, to give \$12 to other members of their subgroups (\$2 to each), or to give \$21 to the members of the other subgroup (\$3 to each). These alternatives were known prior to discussion in the 14-person groups. In contrast to the earlier experiments, the experimenters monitored the discussions, and hence knew who promised and who did not.

In 13 of the 24 such groups all subjects promised to give away the \$21. In these universal promising groups, 84% of the subjects gave away the money, while 58% did so in the remaining 11 groups. That appears to support the importance of promising, perhaps interacting with knowing that the experimenter knew about it. A crucially important finding, however, was that there was *no* relationship between promising and behavior in the groups in which not everyone promised. First, the sample correlation across groups between the number of people who promised to give away the \$21 and the number who later did so was only .09. Second, the correlation (phi value) at the individual level between promising to give the money away and actually doing so was .03; finally, the correlation between each individual's choice whether or not to give away the \$21 and the number of others in the group who promised to give it away was, again, .03. If either promising were binding (**Krebs**'s self-esteem through promise keeping, **Lefebvre**'s processor of conscience), or were binding because the experimenters knew which choice was made (**Houston & Hamilton**), or were kept because we have "sanctions perpetually hanging over our heads" for keeping promises (**Dunbar**), promises should be every bit as predictive of choice when not everyone else promised as when everyone else did, but they were not. In fact, they were totally nonpredictive.

We have two interpretations: Either promises merely reflect group identity (see target article) and failure of unanimity in promising indicates a failure to establish such identity, or promises interact with unanimity. Atiyah (1981) has argued that when two people are considering mutual promises, these are binding only in the presence of reciprocity, and if our second interpretation is correct, subjects in this experimental situation may be treating the entire remainder of the group as a single entity – one that fails to reciprocate in the absence of unanimity. Either view is consistent with our interpretation of our previous experiments.

Many commentators raise the possibility of conscience interacting with group identity: "Group based conscience" (**Houston & Hamilton, Mansbridge, Sonuga-Barke**). A "conscience" that does not transcend group boundaries is suspect, however. Rudolf Hoess, for example, the commandant at Auschwitz who systematically murdered 2,900,000 people, claimed in his autobiography (1959) that his actions were based on the Platonic subjugation of his "softer emotions" to his rational belief in behaving justly in the interests of his group – much as a pilot who bombs civilian territory. People who wish to term this type of rationale (rationalization, retrospective conscious and unconscious distortion – probably all three) to be "acting out of conscience" are free to do so, as is Humpty-Dumpty. Most observers, however, would not, using ordinary language, refer to conscience in analyzing such statements as: "Wouldn't it be great if we all kept our money and they all gave theirs to us." As **Stearns** points out, the group identity "solution" to social dilemmas easily creates social dilemmas between groups. We agree; we even agree with Olson's (1982) analysis that it is precisely "self-sacrificial loyalty" to subgroups that can create dilemmas between them, leading to consequences far worse than any that would occur through individual selfish behavior. If so, we see little distinction between group based conscience and simple group identity. Group based equity (**Heyman**) has the same definitional problem.

Finally, we wish to argue that the importance of the experiments lies in their empirical findings. The general field has suffered from a failure to blend mathematical models with empirical results (**Findlay & Lumsden**) – due to the peculiar paucity of the latter. **Frank**'s commentary is an example of theorizing without findings, from our perspective a bad example, unfortunately, since we admire his other work. Frank asks us to suppose that people can correctly differentiate cooperators from defectors (at least at an above chance level). If they could, all sorts of

nice results would indeed follow. But where is the evidence they can? In his book Frank (1988) cited Experiment 1 in Dawes et al. (1977); in that experiment, subjects were able to differentiate cooperators from defectors at a level very slightly above chance (resulting from the fact that some conditions elicited more cooperation than others); the residual percentage was .03. Dawes et al. unfortunately provided a "significance level" along with this result, although they cautioned that it was post hoc and small and should not be taken seriously in the absence of replication. (Currently, Dawes advocates not stating a significance level at all in such situations.) Experiment 2, which is ignored in Frank's book, did provide an opportunity to replicate the finding. In this experiment, the residual correlation was .02. In Frank's own experiments, reported in his book, he claims to find an ability to differentiate after a half-hour's interaction. It turns out, however, that people can do this only on the basis of predicting that the other person will do what they do, that is, cooperate after a pleasant interaction or defect after an unpleasant one (Frank, personal communication). There is no evidence that people have any ability to predict cooperation or defection when the prediction is a nonconsensus one – a prediction that is absolutely crucial to the theory of identifiability. Instead, Frank's hypothesis is based on an analogy with the croaking of frogs (1988). In contrast, **Heyman**'s hypothesis that subjects are concerned about (within-group) equity does have empirical support, although we cannot agree that it explains redundant contributions, because in contexts where subjects have greater certainty that their contributions would be redundant, they do not contribute, for example, when they are not part of a "designated set" of contributors (van de Kragt et al. 1983).

This importance of empirical findings is again emphasized in the speculations about perfectly successful defectors, who only fake devotion to group goals. The question in considering constraints imposed by evolutionary theory is not whether such people *could* be fitter than those who automatically identify with group goals, but whether such people exist. No commentator has presented evidence that they do. As Simon (1983) points out, evolution is a "hill climbing" activity, and the fact that various hills exist (or rather could exist) does not imply that they have been climbed. Believing that all conceivable hills *must* have been climbed is the basic fallacy in arguing directly from evolutionary theory to an egoistic incentive (EI) theory of human motivation.

Two major points. 1. *The target article does not claim that humans are "basically" prosocial or cooperative or selfless.* According to **Stenseth,** however, we claim that *homo sapiens* "is basically a cooperative creature" and according to **Mansbridge** we say that sociality is totally unrelated to self-interest. (What we *did* say in the beginning of our target article was that our "experimental" manipulations were unrelated to self-interest, as will be elaborated in the next section of this Response.) We were not trying to "seal the case for selflessness" (**Kenrick**). We did not suggest a dichotomy of egoistic versus altruistic motivations (**Heyman**). We agree with **Tooby & Cosmides** and with **Tetlock** that – speaking loosely – people have a "mixed bag" of motives (if we may be permitted to paraphrase their arguments with a colloquialism). Yet even such "basically" sympathetic commentators as **Vine,** and as **Tooby & Cosmides,** seem to have made what **Brewer** terms "the error of unidimensional thinking": Applying the law of the excluded middle not to negation but to opposites, such as selfish versus prosocial. Hence we can hardly be faulted for failing to show that most people do not behave selfishly most of the time (**Stearns**).

2. *The target article did not dismiss Darwinian evolution or propose an alternative to inclusive fitness as an explanatory variable for evolutionary history.* What we attempted to make clear in our article was that we both accept standard evolutionary theory and argue that it does *not* automatically provide an explanation for observed behavior. Here, we go somewhat beyond the doubt of **Houston & Hamilton** (that "behavior. . . can be explained in terms of people's attempts to maximize their inclusive fitness"), the pessimism of **Gibbard** ("about the value of explaining contemporary human behavior in terms of fitness maximization"), and the judgment of **Tooby & Cosmides** that such an explanation ("a direct mapping of evolutionary theory as a motivational theory") is "inadequate." Instead, we propose that evolutionary theory does not provide such explanations but rather *constrains* them by casting doubt on any that would apparently lead to a decrease in inclusive fitness. It was precisely our acceptance of Darwinian evolution combined with this conclusion, in fact, that led to our attempts to integrate our research findings with an evolutionary hypothesis of "sociality" (not "*pro*sociality"), which constrains unbridled self-interest.

What we maintain is that egoistic incentive (EI) theory is an "overeager" attempt "to leap directly from evolutionary theory to psychology." (**Tooby & Cosmides**), and the attempt is there. Despite the suggestion by several commentators that really sophisticated sociobiologists wouldn't make such a leap, we agree with **Brewer** and **Heyman** about the "pervasiveness of individual self-interest assumptions in the behavioral sciences." In traditional economic theory, for example, hedonistic rationality is so strong an assumption that it is generally referred to as "rationality" per se. In social psychology, it is found in "social exchange" theories, and in personality psychology – as **Oyama** points out – as a criterion of "mental health."

Motives. A large number of authors (**Gilbert, Krebs, Mansbridge, Oyama, Stearns, Shweder, Tooby & Cosmides, Vine,** and perhaps **Kenrick** and **Knauft**) interpreted us as proposing an evolutionary theory of motives. (In contrast, **Brewer, Gibbard,** and **Eldredge** explicitly read us correctly.) A careful reading of the target article will show why we were so often interpreted as presenting a motivational theory. We challenged EI theory on its own grounds. To do this, we had to use its motivational language for the critique in section 1 and the experiments in section 2. From section 3 to the end – where we write as sociality theorists – there is only one use of the term "motive," and that is to mention their erroneous attribution to inanimate objects.

Our interest is with the mechanisms with which humans select information from the environment, categorize it, evaluate it, and act on it. Contra **Sonuga-Barke,** not all ordinary language has useful referents in scientific

discourse. The casual use of *motive* is unashamedly teleological. In addition, the word connotes egoism independently of any observations about how organisms *actually* behave. The best example of this circularity is found in **Liebrand**'s analysis: A partitioning (conscious or unconscious) of "social motive" into concern for own (*ego*) payoff and concern for other's (*alter*) payoff is termed "egotistic" because, after all, such partitioning occurs within the actor. (Where else?) We agree with **Oyama** that when words are used in this way "egoistic incentive" theory becomes a "black hole." (We are less sympathetic to **Krebs**'s analogy with the "null hypothesis," since so many people so often base a proclamation of positive findings on a claim of refuting it!) The social sciences, however, have a dreary history of using words in a tautological fashion – e.g. "reinforcement," which was expanded to include "self-reinforcement" (Skinner 1957) when no reinforcement could be found elsewhere. Whereas **Rapoport** suggests that the extension of the concept of "selfishness" to "include any demonstrable source of motivation not previously subsumed under the concept" should lead us to "cede" its validity, we argue that granting such an extension should lead us to abandon it.

A world view. Our remaining response to the commentaries will be based on the sociality hypothesis. This begins with the understanding that the human mind/brain evolved for being social (and for learning what that means in our cultures), and *not* for doing science, philosophy, or other sorts of critical reasoning and discourse (Caporael 1987). We expect and find cognitive limitations especially under conditions of uncertainty (Dawes 1988). These limitations contribute to and interact with various sociocultural constructions, including folk psychological notions of "human nature." Without the approach **Rapoport** recommends (i.e., the analysis of social images and ideology accompanied by rigorous experimentation), human evolutionary psychology is bound to be a potpourri of myths confused with facts. Cognitive limitations and the ruses of culture may be overcome to some extent by education, environmental feedback, or "collective rationality" (Campbell 1986a), the process of target article, commentary, and response being an example of the last.

This view of human reasoning is quite different from the traditional "objectivist view" described by Lakoff (1987), in which thought is considered to be abstract, disembodied, logical, and ahistorical. Taking an analogy from **Shweder,** we believe that "setting over" the objectivist view by recognizing human limitations at the outset may be a better way of pursuing objectivity. We hope to show by implication how our view *functions* differently from the traditional view in critical discourse.

Levels of explanation. We were criticized by **Dunbar, Krebs,** and **Tooby & Cosmides** for confusing proximate and ultimate levels of explanation; **Kenrick, Eldredge,** and **Gibbard** praised us for making the distinction. **Caraco** appears to do both; **Vine** allows that we were "nominally alert" to it. We are not the first authors in the literature, including the pages of *BBS*, to be accused of this deficiency. Walter (1989) complains that if Kitcher (1987) only understood the distinction, Kitcher's objections to human sociobiology would vanish. The complaint is so commonly the first one directed at critics that it verges on cant. What is going on here? How is it that competent specialists, evaluating the identical text (in the case of our target article) come to such wildly opposing interpretations?

We believe a number of factors account for this impasse. Before describing them, let us make clear that our target article attempted to maintain the distinction between proximate and ultimate levels. In footnote 2, we made clear that our interest was in species-typical "cognitive machinery," and we hypothesized that it evolved because it enhanced fitness in the environments in which humans evolved. **Caraco** and **Vine** explicitly drew attention to this footnote. In section 1.4, we developed our conceptual objections to conflating genetic selfishness with individual selfishness. As **Dunbar** points out, human sociobiologists *unlike those who work on mammals,* perpetuate this "depressingly common" conflation. Not only was our discussion a criticism of these human sociobiologists, but we used the same research (by Cheney et al. 1986) cited by Dunbar to make the identical point: The simplistic view equating genetic and individual levels of explanation is in error. Nevertheless, as we noted earlier, we took the claims of sociobiologists as serious hypotheses. The experiments show that even if our conceptual objections to equating proximate and ultimate levels carried no weight, equating the two is not empirically tenable. In section 3, we claimed that mechanisms for sociality evolved because they enhanced fitness. We considered a description of cognitive evolution (Rozin 1976) that further removes the equation between genetic and individual mechanisms, and we also explained why we believe that fitness explains the origins of psychological mechanisms, but not necessarily their functions in novel domains of action. Is it impossible to adopt an evolutionary approach and simultaneously criticize sociobiology? If so, sociobiology is vacuous.

We believe that several factors contribute to such communication at cross-purposes between human sociobiologists (hereafter just sociobiologists) and their critics. Among these are (1) folk psychology, (2) distortions evoked by the "selfish gene" metaphor, and (3) the use of the proximate/ultimate distinction, specifically in human sociobiology.

Folk psychology. We sometimes envy our colleagues in the nonbehavioral sciences. For example, we may meet a cell biologist on a plane and at best listen with attention to stories about assaying methods about which we know nothing. But how many people confess that they know nothing about psychology, or that they are bad judges of character, or that they find humans to be an enigmatic species? Human sociobiologists often operate on their intuitions of psychology, not with the empirical literature on "their" species, humans.[1] Almost universal are descriptions of animal research linked by mere proximity on the printed page to casual observations of human behavior. Even where the animal literature is not invoked, the human literature may still be overlooked (Caporael 1989a).

The dearth of psychological citations in the human evolutionary literature raises a question. If authors are not giving evolutionary explanations for scientific findings, what are they explaining? The answer, we believe, is their folk psychological constructions of casually observed behavior.

Gilbert has given a lucid account of what folk psychology is. It includes the conceptual categories and forms of explanation implicit in everyday thought and talk. A close reading of her commentary also reveals the preferred form of explanations in everyday talk – motives. We have already explained why we reject everyday talk about motives – they are circular and inherently egoistic. In addition, although Gilbert does not mention it, when everyday talk of motives is generalized from a situation to a person, then from a person to a group, the result is a blurring of discourse from motives to everyday talk about traits. (Consider: "He was looking for a fight," and "He's an aggressive person," "Men fight," and "Humans are an aggressive species.") Traits are the conceptual categories of folk psychology and motives are the forms of explanation. Traits and motives are also the language of sociobiology.

The problem is not that the human sciences need to characterize folk psychology, as **Gilbert** recommends, but that they need to keep the distinction between folk psychology and scientific psychology in mind. Even now, much of psychology, particularly social psychology, blurs the distinction. Although the folk may have the luxury of choosing their beliefs, prejudices, and intuitions over research results, when scientists speak of their findings they are already committed, at least in principle, to the overthrow of naivete. Accumulating research data *does* explain why folk psychology diverges from the "scientific truth."

There is little reason to believe that folk psychology will bear much resemblance to scientific psychology. Folk psychology is derived partly from individual histories of social relationships and partly from *culturally shared* conceptual categories and forms of explanation. Nothing in it suggests a valid account of underlying mechanisms. To paraphrase **Gilbert,** however, sociality does take folk psychology seriously. We view folk psychology as a means of negotiating the social-symbolic environment (and if worst comes to worst, sometimes the nonsocial environment; see below). Of course, folk psychology should not be ignored entirely in scientific discourse but be treated as a source of hypotheses, rather than as evidence for their validity.

Anthropomorphizing the gene. The use of terms like "altruistic" or "selfish" in some of the early population genetics models, writes **Vine,** invited critics of sociobiology to "mistake evolutionary forces for the psychological mechanisms that result from them" (quoting the target article). **Tooby & Cosmides** assert that when some theorists appear to treat evolutionary theory as motivational theory, they are simply using a metaphorical shorthand. We have already discussed the difficulty of using folk psychological terms to describe human psychology. The problems are magnified when we use them to describe genes.

We recognize that serious difficulties are bound to arise when we humans have to describe processes that are beyond our sensory ken – geologic time, DNA interactions, cognitive processes for which we have no introspective access. There are also problems in dealing with supervenient properties (e.g. fitness or probability; Sober 1984) that are not themselves physical. Human minds need some way to bring such ideas down to human scale. More problems are found to arise when the desire to understand is greater than the understanding. Circumstances exist that invite anthropomorphism as a default metaphor. Note that evolutionary processes are not brought into the realm of human ken by comparing them to a computer or a molecular process. Like the Australian aborigines described by **Knauft,** "selfish gene" theorists "socialize" evolutionary processes.

Tetlock is right that the demarcation of the explanatory range of research programs organized around competing metaphors is arbitrary. Some metaphors are better than others for structuring our thinking, however. Take the metaphor "argument is war" (Lakoff & Johnson 1980). We win and lose arguments by strategic attacks on weak positions that were thought to be invincible by their defenders. The concept, the language, and the activity of argument is structured by the metaphor of war. But do evolutionists really want the concept, language, and processes of natural selection structured by the metaphor of persons? Even so widely used and innocent a construction as "natural selection favors" draws attention away from natural selection as a process that probabilistically *eliminates* variants rather than picking them out for special attention. If we focused on the culling aspects, theorists would not be so eager to assume that every characteristic is an adaptation (cf. Tooby & DeVore 1986, who make this argument).

Personification of genes is no more neutral conceptually than the "argument is war" image. Anthropomorphic metaphors provide a very specific way of thinking about nonhuman entities, in "terms that we understand on the basis of *our own motivations, goals, actions, and characteristics*" (Lakoff & Johnson 1980, p. 34, emphasis added). Such metaphors direct how we think and behave toward a particular nonhuman entity by picking out the human attribute in it that we want to highlight. The problem with the selfish gene metaphor is that it structures the way we think about mind and behavior in restrictive ways. Anthropomorphizing the gene invites the very form-function error **Brewer** warns against in her commentary.

Our rejection of anthropomorphized gene theories ("selfish gene" theories) is most emphatically not a rejection of natural selection – as **Caraco** interpreted it. His objection is that we do not make a distinction between "ultimate-level 'selfishness'" and the mechanisms that result from selective processes. We do, but we view selfishness as a characteristic of humans, not of "ultimate levels," genes, or natural selection. His misinterpretation, we suggest, may be related to anthropomorphizing the gene.

Ultimate/proximate revisited. For many human sociobiologists the "ultimate cause" of behavior is reproductive success: The "proximate causes" are everything else – from habitat conditions through ontogeny, physiological systems, conscious subjective states, and unconscious cognitive states (e.g., Symons 1979). Understanding the ultimate cause, inclusive fitness maximization, is the key to unlocking the secrets of human nature (Alexander 1979). The operation of ultimate cause is described by natural selection and such subtheories as kin selection and reciprocity. Proximate causes are merely "tools or servants" in the service of natural selection

(Barash 1977). These sociobiologists claim that their interests are in the ultimate cause of behavior, not the proximate causes. We have two objections to the sociobiological blurring of ultimate and proximate causes. First, echoing Kitcher (1985, pp. 281–82): If natural selection does not shed light on the proximate mechanisms of human psychology, then what does it illuminate? Second, lacking a bridge between models and proximate mechanisms, to apply population-genetic models to human behavior is to use them as *de facto* psychological theories. But they aren't. We view population-genetic models as models of the alterations of gene frequencies in a population given such parameters as selection coefficients and allelic frequency. These models do not describe, nor are they reducible to, psychological mechanisms. The inappropriate transition between gene frequency alterations and mechanisms is most apparent in **Dunbar**'s commentary when he asserts that the correct question to ask concerns "why humans are *still* social" (his emphasis). As opposed to what? The "logically more primitive state of solitary existence" Dunbar describes? Even if Alexander's (1979) "extrinsic causative factors" for group living were removed, humans would still group; they (we) are unable to develop and function in isolation. Obligate interdependency is an outcome of our species history, not cost-benefit calculations in the present on the advantages for being social or solitary.

It is one thing for evolutionists to take cooperation or selfishness as a supervenient property of organisms; it is quite another to assume that a motive or cognitive mechanism exists that resembles the property. We take **Findlay & Lumsden**'s model to be a model in the former sense. They assume that learning can be an effective means of spreading non-EI behavior throughout a population, and they track the resulting changes in gene frequency. In principle, the model describes any species in which learning from conspecifics is important.

The model **Findlay & Lumsden** propose may justify our position to some sociobiological critics, but it does not make predictions about the mechanisms resulting in prosocial behavior, nor does it suggest interesting directions for empirical study. Especially in the human case, the relationship between the model and what it actually represents is unclear. For example, is "cultural learning" to be equated with social learning? If so, then the "altruistic" property in the model cannot be general across organisms, because most animal social learning is limited to direct observation whereas humans can also learn symbolically.

More important, the group selection model **Findlay & Lumsden** propose is very different from the scenario we proposed. In particular, we hypothesize that phenotypically selfish individuals have lower fitness in the group relative to nonselfish individuals (because of constraints on group size), and that within-group selection, rather than between-group selection, is the primary locus of selection (because interdependent individuals exploit the habitat). The difference between the two approaches can be appreciated by taking Findlay & Lumsden's model literally as an explanation for heroic (or fanatic *qua* **Tetlock**) self-sacrificial altruism. Under the model, such behavior might be explained as an outcome of an "altruistic impulse." We view it as a relatively rare "error" that occurs when distinctions between one's own and others' outcomes are completely lost in certain stressful conditions. We venture to suggest that fewer men or women have sacrificed their lives for their country than for their comrades. While we appreciate the praise of Findlay & Lumsden, we maintain that the utility, in more than a very general sense, of population genetics models for research on human evolution has yet to be established. We would like to see more explicitly social models that incorporate social structure and ecological parameters. Foley and Lee (1989) have made a step in this direction.

Human evolutionary research. It is argued by **Tooby & Cosmides** that "an adaptationist perspective" requires three levels of analysis. In general, we agree with their description – it was, after all, the strategy we used. We argued that given the conditions of hominid evolution, we would expect adaptations for sociality (i.e. mechanisms to support the development and maintenance of group membership) to enhance fitness in ancestral populations. In the context of the target article, ingroup bias is clearly proposed as one such adaptation. Nevertheless, Tooby & Cosmides erroneously assert that we see the sociality hypothesis as somehow exempting us from showing how ingroup bias leads to enhanced fitness in ancestral environments. Their misperception may be related to differences in approaching the past. Tooby & Cosmides accurately portray their perspective as adaptationist. They rely on a backward projection from modern metaphors and behavior to infer the characteristics of ancestral environments (cf. Tooby & DeVore 1986). From this line of reasoning, what people do now must have had adaptive advantages in the past – a "hindsight analytic framework" (Foley 1984a). In contrast, we believe that evolutionary psychology requires greater attention to paleoanthropological research than is paid by sociobiological research.

We do argue that precisely because the ingroup bias enhanced fitness in *ancestral* environments we need not claim that it enhances fitness under current environments. Hence, we may sound like sociobiologists (**Kenrick**), but to the extent that they posit a direct link between inclusive fitness maximization and contemporary human behavior (Alexander 1979; 1987; Barash 1977; Smith 1987), we do not qualify (as is evident from some of the commentary). As we claimed in the target article, mechanisms adapted for one function can be extended or captured (**Dunbar**) by another functional system. Not only can this functional system be psychological; it can also be cultural: hence, we speak of mechanisms being engaged in cultural contexts. Many mechanisms have multiple functional contexts, but only some contexts suggest the nature of the adaptive fit. Identifying evolved mechanisms involves such indirect approaches as the location of "mismatches" or "anachronisms" (**Krebs**) in modern life. Demonstrating the disjunction between adaptation and cultural engagement was a singular virtue in Cosmides's (1985) own research, in which she showed that an arguably adaptive mechanism responsive to rule-breaking in ancestral social domains functioned poorly in abstract domains where subjects would have to follow the rule of *modus tollens*.

An "evolutionary prediction" that corresponds with predictions based on casual observation, folk psychology,

revealed preference analysis, economic factors, and structural conditions is the weakest possible explanation for behavior. These are all factors that have to be treated as potential threats to validity in evolutionary psychology (Caporael 1989b). Both scenarios and experiments are subject to such confounding factors.

Scenarios. Evolutionary reconstructions are a necessary evil for evolutionary psychology – at the present time, at least. A scenario provides the raw material for formulating hypotheses about psychological mechanisms, but it is subject to a number of distorting influences. First, it is very difficult to interpret fossil remains and artifacts. Before inferences about behavior can even be drawn from the existing data, it has to be established that a particular assembly was not created by natural processes (Binford 1983). Other problems arise from using animal models for hominid behavior.

Second, the ideological factors **Rapoport** describes influence the construction of evolutionary scenarios. Some researchers and theorists are increasingly aware of the mythic properties in their scenarios (Isaac 1983; Landau 1984), but if anything, scenarios have become *more* embroiled in ideological positions rather than less. Even a cursory look at models for the evolution of hominid social behavior suggests that scenarios have been "captured" by current ideological conflicts (cf. Foley & Lee 1989, Table 1, or summaries in Richards 1987; Tooby & DeVore 1986). Hypotheses about the social structure characteristic of early hominids range from monogamous pair-bonds to polygynous harems to female-offspring groups to male kin territorial groups. Most of these scenarios deal with relationships of dependency and autonomy between males and females – the same issues that mark many current social debates, from abortion to equal pay. There is a need to develop reasonable constraints on theorizing (Foley & Lee 1989).

Third, scenarios are especially subject to cognitive limitations that are collectively known, appropriately, as the "scenario effect" (Dawes 1988) or the simulation heuristic (Kahneman & Tversky 1982). The greater the ease in constructing a story and the more detail that can be added to it, the more likely that people will believe the story to be true. In contrast, rational analysis quickly reveals that greater detail makes scenarios *less* likely to be true; the probability of combination of components cannot be greater than the probability of the least likely component, and equal only if the probability of all other components is 1.00. **Tooby & Cosmides** identify the scenario in the sociality hypothesis as "vague"; it was intentionally constructed to be conservative.

We disagree, however, with **Tetlock**'s implication that all evolutionary scenarios are equal. Some scenarios are worse than others. The best that evolutionary researchers are likely to be able to do is to set constraints on the constituents of a possible scenario. One set of constraints requires the analysis of social images, ideologies (**Rapoport**), and metaphors (**Tetlock**) that potentially distort the scenario. We envision this analysis to be no different from the process of attempting to control (or explain away) confounding variables in an experiment. Another requires sensitivity to research on cognitive limitations, another confounding variable. A third, of course, is that we consider the species' history, morphology, and ecology. We expect that incorporating the results of psychological research will also help constrain scenarios (Caporael 1989b).

The target article was written prior to the appearance of the articles of Foley (1989) and Foley and Lee (1989), otherwise we would not have called attention to Foley's (1984; 1987) definition of hominid community. Our reason for believing that the ecological definition of community should include conspecifics is that, in some cases (particularly humans and their immediate ancestors), conspecific relationships are as great a source of selective pressure as interspecific relationships (or even greater). In arguing that sociality becomes its own driving force, however, we are not claiming that encephalization occurs only to solve social problems but that solving the problems presented in the ecological community has become a social problem for humans, and perhaps for a few other species. That is, other individuals mediate selection pressures in the habitat and at the same time become an independent source of selection pressures on individuals. We do not believe that preexisting sociality will give rise to the "intense sociality" characteristic of humans in the absence of appropriate ecological conditions. Thus, not only do we agree with **Foley**'s comment that our approach can be integrated within an ecological framework, we believe such an integration is essential for evolutionary psychology.

The relevance of the experimental approach. We are asked by **Teigen** whether paradigmatic conflicts at the metatheoretical level can be settled by appeal to experimental results. Probably not. Our purpose was to point to an alternative paradigm; researchers do not have to accept EI as the only explanatory framework. Still, **Tooby & Cosmides**'s imaginative fig wasp story notwithstanding, we think there is a great deal to be said for "a little hardheaded empiricism."

Krebs, Houston & Hamilton, and **Tooby & Cosmides** object to experiments in "evolutionarily unprecedented environments" such as ours, where subjects are strangers making single anonymous choices. It is quite true that the focus on "stranger psychology" is a deficiency in psychological experiments (Caporael 1987a), but this is hardly an indictment of the experimental approach for evolutionary studies. A variety of factors can account for observed behavior, and these need to be eliminated before we can resort to an evolutionary account. It is precisely because there is no reason to assume that cognitive/affective mechanisms evolved to function in situations like a psychology laboratory that the laboratory is a suitable situation for studying evolved mechanisms. Such factors as those noted by **Knauft** as empirical caveats can be randomized (as they were in these studies) across conditions. Intervening variables can be controlled, and alternative explanations can be eliminated. The laboratory is one means of creating a "mismatch" between a proposed adaptation and a context that will illuminate the nature of the mechanism. If the behavior occurs even in the wrong context, or is distorted in some other fashion in the context, the claim for an *evolved* mechanism is strengthened. As is the case with scenarios, experiments are a means of developing constraints on theorizing.

We applaud **Knauft**'s call for a broadly collaborative effort and better articulation with theories of human

evolutionary development, ethnographic evidence, and paleoanthropological data. As he points out, the sociality hypothesis does have various incarnations in the anthropological literature. These certainly merit a comprehensive review and critique, much along the lines of Kuper's (1988) trenchant review of the invention of primitive society. Where the sociality hypothesis differs from the ethnographic and paleoanthropological work is in its emphasis on human psychology as an adaptation to group living, as we suggested in the target article. Psychology experiments on purported outcomes of human evolution can be devised to test some aspects of theories in these fields.

In general, the constraints we urge on both scenarios and experiments are to limit the ultra-Darwinism described by **Eldredge.** As is clear from Eldredge's commentary, however, such constraints also allow evolutionists to broaden their scope in interpretating organisms' activities. In the case of humans (and perhaps some other species), we would add information-gathering activities. We are uncertain how such activities would fit into the hierarchically structured systems described by Eldredge, but we believe they are important to the functionalist perspective.

Mechanisms. A number of commentators either elaborated on the ingroup bias mechanism we proposed or offered alternative explanations, many of which were grounded in self-interest as a mechanism. The commentaries by **Brewer, Mansbridge,** and **Teigen** pointed to self-conception in scientific psychology. Scientific psychology increasingly recognizes that the locus of self can shift with group identification (cf. Turner 1987; Vine 1987). Brewer identifies the shifts as a cognitive extension of self-interest, which suggests a notion of "self" that departs from the traditional individualistic conception. Teigen makes the same departure, but from another direction: He recommends assuming that human actions are invariably determined by ingroup considerations with a "group" of one person being the limiting case. We agree with Brewer that clique-selfishness may be the most intractable form of selfishness. It may not be useful to make sharp divisions between individualistic and collective behavior (Teigen): The determination of "selfish" becomes relative in cases of clique selfishness/altruism (e.g. the fanatic/hero who sacrifices himself for the group).

Tetlock proposes that EI theory can be rescued by forging new intellectual alliances with social identity and role theorists. The egoistic incentive here at issue is positive social identity. We believe two notions are embedded in Tetlock's description of social identity theory. One is related to people's goals (a marriage between incentives and the individual subjective self-system referred to by **Vine**). The other is a notion of social identity that has more to do with identity as simply an outcome of fairly automatic cognitive processes involved in classifying and evaluating the "stimulus stream" to which humans are exposed. The former can be married to EI theory (if it will take it) in the justificational sense **Shweder** discusses. The latter suggests not only goals, but the epistemological framework we have adopted for this Response. This is the major difference between the social identity theoretic variant of the EI research program and the sociality argument we advanced.

It is not clear to us exactly how far **Brewer**'s characterization will take us vis-à-vis other mechanisms. For example, empathy was suggested by **Kenrick, Krebs,** and **Rapoport** as a mechanism mediating helping behavior. On one hand, both ingroup bias and empathy may be considered as cognitive extensions of the self; on the other, ingroup bias may be a cognitive extension of self-*interest,* and empathy may be an identification with others' suffering in instances where the suffering can be personalized. For example, a familiar story is told about two enemy soldiers suddenly surprising each other. Under ordinary conditions, one should kill the other; yet he does not because the soldier on the opposing side takes some quick action – perhaps flashing a picture of his wife and children – that makes their common predicament salient.

All these commentaries suggest that research on the evolution and description of self would be well worthwhile.

Conclusion. In reading between the lines, **Oyama** presupposes that we are seeking a motivational theory and grounding it in biology. Our interests are in the cognitive mechanisms from which social motives are constructed through development, experience, and learning in particular cultural milieus. This is the "bi-level" approach we advocate. Although we tried (unsuccessfully) to stay away from terms such as "innate tendency," we do not want to continue an opposition between biology and culture. The issues Oyama raises are indeed worth thinking about. Our "between the line" agenda, such as it is, is a struggle against the extraordinary poverty of explanatory means for discourse on moral life. Eliminating the inherently egoistic incentive language may force us to find new means of authenticating moral life. We hope that the kind of wit Oyama displays in her final note will have a place in a richer discourse on morality. We are grateful to the commentators who, like Oyama, risked being redundant to save us from ourselves. (If it is any comfort, we rarely found the errors pointed out to us redundant.)

NOTE

1. Some sociobiologists do cite anthropological literature (e.g., Alexander 1979; 1987; Chagnon & Irons 1979), but that work is not integrated with the psychological literature (Jahoda 1982).

References

Letters "a" and "r" appearing before authors' initials refer to target article and response respectively.

Abelson, R. P. (1976) Social psychology's rational man. In: *Rationality and the social sciences,* ed. S. I. Benn & G. W. Mortimore. Routledge and Kegan Paul. [aLRC]

Adams, J. S. & Jacobsen, P. R. (1967) The effect of wage inequities on work quality. *Journal of Abnormal and Social Psychology* 69:19–25. [GMH]

Adams-Webber, J. (1979) *Personal construct theory: Concepts and applications.* Wiley. [VAL]

(1987) Comments on Lefebvre's model from the perspective of personal construct theory. *Journal of Social and Biological Structures* 10:177–89 [VAL]

Adams-Webber, J. & Benjafield, J. (1973) The relation between lexical marking and rating extremity in international judgment. *Canadian Journal of Behavior Science* 5:234–41. [VAL]

Alexander, C. H. & Knight, G. W. (1971) Situated identities and social psychological experimentation. *Sociometry* 34:65–82 [PET]
Alexander, R. D. (1971) The search for an evolutionary philosophy of man *Proceedings of the Royal Society (Victoria)* 84:99–120. [aLRC, JT]
(1974) The evolution of social behavior. *Annual Review of Ecology and Systematics* 5.325–83. [aLRC]
(1979) *Darwinism and human affairs*. University of Washington Press. [arLRC, RIMD, IV]
(1987) *The biology of moral systems* Aldine de Gruyter. [arLRC, RIMD, DK, SO, JT, IV]
Atiyah, P. (1981) *Promises, morals and law*. Oxford University Press [rLRC]
Axelrod, R. (1984) *The evolution of cooperation*. Basic Books. [aLRC]
Axelrod, R. & Hamilton, W. D. (1981) The evolution of cooperation. *Science* 211:1390–96. [TC, AIH]
Balikci, A. (1970) *The Netsilik Eskimo*. Natural History Press [BMK]
Barash, D. P. (1977) *Sociobiology and behavior*. Elsevier. [arLRC]
(1979) *Sociobiology: The whisperings within*. Harper & Row. [IV]
Batson, C. D., Bolen, M. H., Cross, J A & Neuringer-Benefiel, H. E. (1986) Where is the altruism in the altruistic personality? *Journal of Personality and Social Psychology* 30:212–20. [DK]
Batson, C. D., O'Quin, K., Fultz, J., Vanderplas, M. & Isen, A. (1983) Self-reported distress and empathy and egoistic versus altruistic motivation for helping. *Journal of Personality & Social Psychology* 45:706–18. [DTK]
Beecher, C. (1847/1972) On the peculiar responsibilities of American women. In: *Root of bitterness*, ed. N. F. Cott. E. P. Dutton. (Original work published in 1847.) [aLRC]
Benjafield, J. (1984) On the relation between the Pollyanna and golden section hypotheses. *British Journal of Social Psychology* 23:83–84. [VAL]
Benjafield, J. & Adams-Webber, J. (1976) The golden section hypothesis. *British Journal of Psychology* 67:11–15. [VAL]
Benjafield, J. & Green, T. R. G. (1978) Golden section relation in interpersonal judgment. *British Journal of Psychology* 69:25–35. [VAL]
Best, P. B., Meijer, M. A. & Weeks, R. W. (1981) Interactions between a male elephant seal *Mirounga leonina* and Cape fur seals *Aretociphalus pusillus*. *South African Journal of Zoology* 16:59–66. [AIH]
Binford, L. (1983) *In pursuit of the past*. Thames and Hudson. [rLRC]
Boyd, R. (1988) The evolution of reciprocity in sizeable groups Journal of Theoretical Biology 132.337–56. [BMK]
Boyd, R. & Lorberbaum, J. P. (1987) No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. *Nature* 327:58–59. [BMK]
Boyd, R. & Richerson, P. J. (1985) *Culture and the evolutionary process* University of Chicago Press. [aLRC, BMK]
Brace, C. L. (1962) Comment on "Food transport and the origin of hominid bipedalism. *American Anthropologist* 64:606–7. [aLRC]
Brewer, M. B. (1976) Comment on Campbell's "On the conflicts between biological and social evolution and between psychology and moral tradition." *American Psychologist* 31:372. [MBB]
(1979) Ingroup bias in the minimal intergroup situation. A cognitive-motivational analysis. *Psychological Bulletin* 86:307–24. [aLRC]
(1981) Ethnocentrism and its role in interpersonal trust. In: *Scientific inquiry and the social sciences*, ed. M. B. Brewer & B. E. Collins. Jossey-Bass. [aLRC]
(1985) Experimental research and social policy. Must it be rigor versus relevance? *Journal of Social Issues* 41(4):159–76. [aLRC]
Brewer, M. B. & Campbell, D. T. (1976) *Ethnocentrism and intergroup attitudes: East African evidence*. Halsted. [aLRC]
Brewer, M. B. & Kramer, R. M. (1986) Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of Personality and Social Psychology* 50.543–49. [arLRC, MBB]
Brewer, M. B. & Schneider, S. (1987) *Effects of group composition on contributions to a public good*. Unpublished manuscript. [aLRC]
Brown, R. (1986) *Social psychology, the second edition*. The Free Press. [GMH]
Brown, R. & Herrnstein, R. J. (1975) *Psychology*. Little, Brown. [GMH]
Buchanan, J. M. (1986) Then and now, 1961–1986: From delusion to dystopia. Presented at the Institute for Humane Studies, November. [JM]
Buchanan, J. & Tullock, G. (1962) *The calculus of consent*. University of Michigan Press. [AR]
Byrne, R. & Whitten, A. (1988) *Machiavellian intelligence*. Oxford University Press. [RF]
Campbell, D. T. (1975) On the conflicts between biological and social evolution and between psychology and moral tradition. *American Psychologist* 30:1103–26. [aLRC]
(1982) Legal and primary-group social controls. *Journal of Social and Biological Structures* 5:431–38. [aLRC, MBB]
(1983) The two distinct routes beyond kin selection to ultrasociality. Implications for the humanities and social sciences. In: *The nature of prosocial development. Theories and strategies*, ed. D. Bridgeman. Academic Press. [aLRC]
(1986a) Rationality and utility from the standpoint of evolutionary biology. *Journal of Business* 59:S355–64. [aLRC]
(1986b) Science's social system of validity – enhancing collective belief change and the problems of the social sciences. In: *Metatheory in the social sciences. Pluralisms and subjectivities*, ed. D. W. Fiske & R. A. Shweder. University of Chicago Press. [rLRC]
Caplan, A. L., ed. (1978) *The sociobiology debate*. Harper & Row. [aLRC]
Caporael, L. R. (1983) An evolutionary context for human behavior. Paper presented at the Second International Conference on Event Perception, Nashville, TN. [aLRC]
(1987) Anthropomorphism and mechanomorphism: Two faces of the human machine. *Computers in Human Behavior* 2:215–34. [aLRC]
(1987a) *Homo sapiens, Homo faber, Homo socians:* Technology and the social animal. In: *Evolutionary epistemology. A multiparadigm program*, ed. W. Callebaut & R. Pinxten. Reidel. [rLRC]
(1989a) Man: Darwin's – or Richard's – best subject? [Review of *Darwin and the emergence of evolutionary theories of mind and behavior.*] *Science, Technology, and Human Values* 14:215–19. [rLRC]
(1989b) Mechanisms matter The difference between sociobiology and evolutionary psychology. *Behavioral and Brain Sciences* 12:17–18. [rLRC]
Cashdan, E. A. (1980) Egalitarianism among hunters and gathers. *American Anthropologist* 82:116–20. [BMK]
(1985) Coping with risk: Reciprocity among the Basarwa of Northern Botswana. *Man* 20:454–74. [BMK]
Chagnon, N. A. (1988) Life histories, blood revenge, and warfare in a tribal population. *Science* 239.985–92. [BMK]
Chagnon, N. A. & Irons, W , eds. (1979) *Evolutionary biology and human social behavior*. Duxbury. [arLRC]
Cheney, D., Seyfarth, R. & Smuts, B. (1986) Social relationships and social cognition in nonhuman primates. *Science* 234·1361–66. [arLRC, RIMD]
Cialdini, R. B., Baumann, D. J. & Kenrick, D. T. (1981) Insights from sadness: A three-step model of the development of altruism as hedonism. *Developmental Review* 1.207–23 [DTK]
Cialdini, R. B., Schaller, M., Houlihan, D., Arps, K., Fultz, J. & Beaman, A. L. (1987) Empathy-based helping· Is it selflessly or selfishly motivated? *Journal of Personality & Social Psychology* 52:749–58. [DTK]
Collard, D. A. (1978) *Altruism and economy· A study in non-selfish economics*. Martin Robinson [JM]
Connor, R C. & Norris, K. S. (1982) Are dolphins and whales reciprocal altruists? *American Naturalist* 119:358–74 [AIH]
Coombs, C. (1973) A reparameterization of the Prisoner's Dilemma game. *Behavioral Science* 18:424–28. [aLRC]
Cosmides, L. (1985) *Deduction or Darwinian algorithm? An explanation of the "elusive" content effect on the Wason selection task*. Unpublished doctoral dissertation, Harvard University. [arLRC, DK]
Cosmides, L. & Tooby, J. (1987) From evolution to behavior: Evolutionary psychology as the missing link. In: *The latest on the best: Essays on evolution and optimality*, ed. J. Dupre. MIT Press. [DK, JT]
Cowley, G. (1988) The earth is one big system. *Newsweek* Nov. 7:98–99. [DTK]
Crawford, C. (1989) The theory of evolution: Of what value to psychology? *Journal of Comparative Psychology* 131.4–22. [DK]
Crook, J. H. (1980) *The evolution of human consciousness*. Clarendon Press. [IV]
Dawes, R. M. (1980) Social dilemmas. *Annual Review of Psychology* 31:169–93. [aLRC]
(1988) *Rational choice in an uncertain world*. Harcourt Brace Jovanovich. [rLRC]
Dawes, R. M., McTavish, J. & Shaklee, H. (1977) Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation *Journal of Personality and Social Psychology* 35.1–11. [arLRC, HR]
Dawes, R. M., Orbell, J. M., Simmons, R T. & van de Kragt, A. J. C. (1986) Organizing groups for collective action. *American Political Science Review* 80.1171–85. [aLRC]
Dawkins, R (1976) *The selfish gene*. Oxford University Press. [aLRC, RIMD, NE, NCS, JT, IV]
(1982) *The extended phenotype: The gene as the unit of selection*. W. H. Freeman. [JT]
Derthick, M. & Quirk, P. J. (1985) *The politics of deregulation*. Brookings. [JM]
Deutsch, M. (1961a) The effect of motivational orientation upon trust and suspicion. *Human Relations* 13:123–39. [aLRC]

(1961b) Trust, trustworthiness, and the F scale. *Journal of Abnormal and Social Psychology* 61:138–40. [aLRC]
Downs, A. (1957) *An economic theory of democracy.* Harper & Row. [aLRC]
Dunbar, R. I. M. (1982) Adaptation, fitness and the evolutionary tautology. In. *Current problems in sociobiology*, ed. King's College Sociobiology Group. Cambridge University Press. [RIMD]
(1984) *Reproductive decisions: An economic analysis of gelada baboon social strategies.* Princeton University Press. [RIMD]
(1987) Sociobiological explanations and the evolution of ethnocentrism. In. *The sociobiology of ethnocentrism*, ed. V. Reynolds, V. S. E. Falger & I. Vine. Croom Helm/University of Georgia Press. [IV]
(1988) *Primate social systems.* Croom Helm & Cornell University Press. [RIMD]
(1989) Social systems as optimal strategy sets The costs and benefits of sociality. In: *Comparative socioecology*, ed. V. Standen & R. Foley. Blackwell Scientific. [RIMD]
Durkheim, E. (1912) *The elementary forms of the religious life.* (Reprinted 1964, Free Press) [BMK]
Eldredge, N. (1985) *Time frames· The rethinking of Darwinian evolution and the theory of punctuated equilibria.* Simon & Schuster. [aLRC]
(1986) *Information, economics and evolution. Annual Reviews of Ecology and Systematics* 17:351–69. [NE]
Elster, J. (1990) Selfishness and altruism. In: *Beyond self-interest*, ed. J. J. Mansbridge. University of Chicago Press [JM]
Essock-Vitale, S. M. & McGuire, M. T. (1980) Predictions derived from the theories of kin selection and reciprocation assessed by anthropological data. *Ethnology and Sociobiology* 1:233–43 [BMK]
Etzioni, A. (1988) *The moral dimension: Toward a new economics.* Free Press. [RAS]
Fleagle, J. G. & Kay, R. F. (1985) The paleobiology of catarrhines. In. *Ancestors: The hard evidence*, ed. E. Delson. Alan R. Liss. [aLRC]
Foley, R., ed. (1984) *Hominid evolution and community ecology.* Academic Press. [aLRC, RF]
(1984a) Putting people into perspective. An introduction to community evolution and ecology. In: *Hominid evolution and community ecology*, ed. R. Foley. Academic Press. [rLRC]
(1987) *Another unique species. Patterns in human evolutionary ecology.* Longman/John Wiley. [arLRC, RF, BMK]
(1989) The evolution of hominid social behaviour. In: *Comparative socioecology*, ed. V. Standen & R. Foley. Blackwell Scientific. [RF]
Foley, R. & Lee, P. C. (1989) Finite social space, evolutionary pathways and reconstructing hominid behavior. *Science* 243.901–6. [rLRC, RF]
Frank, R. H. (1988) *Passions within reason: The strategic role of the emotions.* W. W. Norton. [arLRC, RHF, GMH, JM]
Freud, S. (1965) *The interpretation of dreams* (J. Strachey, trans.). Avon Books. (Original work published in 1900.) [aLRC]
Friedlander, J., Cook, B. W., Kessler-Harris, A. & Smith-Rosenberg, C. (1986) *Women in culture and politics.* Indiana University Press. [aLRC]
Gamble, C. (1982) Interaction and alliance in Palaeolithic society. *Man* 17:92–107. [BMK]
Geertz, C. (1962) The growth of culture and the evolution of mind. In. *Theories of the mind*, ed. J. Scher. Free Press [BMK]
Ghiselin, M. T. (1974) *The economy of nature and the evolution of sex.* University of California Press [SO]
Gilbert, M. (1987) Modelling collective belief. *Synthese* 73.185–204. [MG]
(1989) *On social facts.* Routledge. [MG]
Gould, S. J. (1988) Kropotkin was no crackpot. *Natural History* 7:12–21. [BMK]
Gould, S. J. & Vrba, E. S. (1982) Exaptation – a missing term in the science of form. *Paleobiology* 8:4–15. [NE]
Gowlett, J. A. J. (1984) Mental abilities of early man: A look at some hard evidence. In: *Hominid evolution and community ecology*, ed. R. Foley. Academic Press. [aLRC]
Grether, D. M. & Plott, C R. (1979) Economic theory of choice and the preference reversal phenomenon. *American Economic Review* 69:623–38. [aLRC]
Guemple, L., ed (1972) *Alliance in Eskimo society.* University of Washington Press (for AES). [BMK]
Gunn, J. A. W. (1969) *Politics and the public interest in the seventeenth century.* Routledge and Kegan Paul [JM]
Hamilton, W. D. (1964) The genetical evolution of social behavior, I & II. *Journal of Theoretical Biology* 7:1–52. [aLRC, AIH, JT]
(1971) Selection of selfish and altruistic behavior in some extreme models. In: *Man and beast: Comparative social behaviour*, ed. J. F. Eisenberg & W. S. Dillon. Smithsonian Press. [AIH]
(1975) Innate social aptitudes of man: An approach from evolutionary genetics. In: *Biosocial anthropology*, ed. R. Fox. Malaby Press. [AIH]
Hardin, G. R. (1968) The tragedy of the commons. *Science* 162:1243–48. [aLRC, NCS]
Harris, M. (1979) *Cultural materialism: The struggle for a science of culture.* Vintage. [aLRC]
Herrnstein, R. J. (1981) A first law for behavioral analysis. *Behavioral and Brain Sciences* 4:392–95. [GMH]
Hirschman, A. O. (1977) *The passions and the interests· Political arguments for capitalism before its triumph.* Princeton University Press. [JM]
(1985) Against parsimony. *Economics and Philosophy* 1:7–21. [JM]
Hobbes, T. (1958) *Leviathan.* Bobbs-Merrill. (Original work published in 1651.) [aLRC]
Hockett, C. F. & Ascher, R. (1964) The human revolution. *Current Anthropology* 5·135–68. [BMK]
Hoess, R. (1959) *Commandant at Auschwitz: Autobiography.* Weidenfeld and Nicholson. [rLRC]
Hoffman, M. L. (1987) The contribution of empathy to justice and moral judgment. In: *Empathy and its development*, ed. N. Eisenberg & J. Strayer. Cambridge University Press. [JM]
Hofstadter, D. R (1985) *Metamagical themas* Basic Books. [AR]
Hogan, R. T. (1982) A socioanalytic theory of personality. In: *Nebraska Symposium on Motivation*, ed. M. Page & R. Dienstbier. University of Nebraska Press. [DTK, PET]
Homans, G. S (1974) *Social behavior: Its elementary forms.* Harcourt Brace Jovanovich. [GMH]
Houston, A. I. (1980) Godzilla v. the creature from the black lagoon. In: *Analysis of motivational processes*, ed. F. M. Toates & T. R. Halliday. Academic Press. [AIH]
Houston, A. I. & McNamara, J. M. (1988) A framework for the functional analysis of behaviour. *Behavioral and Brain Sciences* 11.117–54. [AIH]
Houston, A. I & Staddon, J. E. R. (1981) Optimality principles and behavior: It's all for the best. *Behavioral and Brain Sciences* 4:395–96. [AIH]
Hume, D. (1741) *Essays moral, political, and literary*, vol. 1, ed. T. H. Green & T. H. Grose. Longmans, 1898. [JM]
Humphrey, N. K. (1976) The social function of intellect. In: *Growing points in ethology*, ed. P. P. G. Bateson & R. A. Hinde. Cambridge University Press. [aLRC, RF]
Huxley, J. (1942) *Evolution: The synthetic theory.* Oxford University Press. [RIMD]
Ingold, T. (1987) *The appropriation of nature: Essays on human ecology and social relations.* University of Iowa Press. [BMK]
Isaac, G. (1978) The food-sharing behavior of protohuman hominids. *Scientific American* 238:90–108 [BMK]
(1983) Aspects of human evolution. In: *Evolution from molecules to men*, ed. D. S. Bendall. Cambridge University Press. [rLRC]
(1984) The archaeology of human origins: Studies of the lower Pleistocene in East Africa 1971–1981. *Advances in world archaeology* 3:1–89. [BMK]
Jahoda, G. (1982) *Psychology and anthropology. A psychological perspective.* Academic Press. [arLRC]
James, W. (1890) *The principles of psychology.* Henry Holt. [WBGL]
Jarvenpa, R. & Brumbach, H. (1988) Socio-spatial organization and decision-making processes: Observations from the Chipewyan. *American Anthropologist* 90:598–618. [aLRC]
Jencks, C. (1979) The social basis of unselfishness. In: *On the making of Americans. Essays in honor of David Riesman*, ed. H. J. Gans, N. Glazer, J. Gusfield & C. Jencks. University of Pennsylvania Press. [JM]
Jerdee, T. & Rosen, B. (1974) Effects of opportunity to communicate and visibility of individual decisions on behavior in the common interest. *Journal of Applied Psychology* 5:712–16. [aLRC]
Johanson, D. C. & Edey, M. A. (1981) *Lucy: The beginnings of humankind.* Warner. [aLRC]
Kahneman, D., Knetch, J. L. & Thaler, R. (1986) Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review* 76:728–41. [JM]
Kahneman, D., Slovic, P. & Tversky, A. (1982) *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press. [aLRC, GMH]
Kahneman, D. & Tversky, A. (1982) The simulation heuristic. In: *Judgment under uncertainty: Heuristics and biases*, ed. D. Kahneman, P Slovic & A. Tversky. Cambridge University Press. [rLRC]
Kalt, J. & Zupan, M. A. (1984) Capture and ideology in the economic theory of politics. *American Economic Review* 74.279–300. [JM]
Kaplan, H. & Hill, K. (1985) Food sharing among Ache foragers: Tests of explanatory hypotheses. *Current Anthropology* 26:223–45. [BMK]
Kelley, H. H. & Thibaut, J. W. (1978) *Interpersonal relations.* Wiley. [WBGL]
Kenrick, D. T., Baumann, D. J. & Cialdini, R. B. (1979) A step in the socialization of altruism as hedonism: Effects of negative mood on

children's generosity under public and private conditions. *Journal of Personality & Social Psychology* 37:747–55. [DTK]
Kenrick, D. T. & Trost, M. R. (1989) A reproductive exchange model: Putting proximate economics in ultimate perspective. In: *Review of Personality & Social Psychology* (10:92–118), ed. C. Hendrick. Sage Press. [DTK]
Kitcher, P. (1985) *Vaulting ambition: Sociobiology and the quest for human nature.* MIT Press. [arLRC]
(1987) Précis of *Vaulting ambition: Sociobiology and the quest for human nature. Behavioral and Brain Sciences* 10:61–99 [rLRC]
Knauft, B. M. (1985) *Good company and violence: Sorcery and social action in a Lowland New Guinea society.* University of California Press. [BMK]
(1987a) Reconsidering violence in simple human societies: Homicide among the Gebusi of New Guinea. *Current Anthropology* 28.457–500. [BMK]
(1987b) Divergence between cultural success and reproductive fitness in preindustrial cities. *Cultural Anthropology* 2:94–114. [BMK]
(1988a) Culture and cooperative affiliation: The evolutionary emergence of simple human societies. Paper presented at the Annual American Anthropological Association Meetings, Phoenix. [BMK]
(1988b) Violence in the simplest human societies: Comparison among great ape, simple, and not-so-simple pre-state human societies. In: *Proceedings of the 1987 Annual Meeting of the American Political Science Association.* University Microfilms. [BMK]
(1988c) On reconsidering violence in human evolution. *Current Anthropology* 29:629–33. [BMK]
Kramer, R. M. & Brewer, M. B. (1984) Effects of group identity on resource use in a simulated commons dilemma. *Journal of Personality and Social Psychology* 46:1044–57. [MBB]
(1986) Social group identity and the emergence of cooperation in resource conservation dilemmas. In: *Psychology of decisions and conflict.* Vol. 3: *Experimental social dilemmas*, ed. H. Wilke, D Messick & C. Rutte. Verlag Peter Lang. [arLRC]
Krebs, D. (1987) The challenge of altruism in biology and psychology. In: *Sociobiology and psychology: Ideas, issues, and applications*, ed. C. Crawford, M. Smith & D. Krebs. Lawrence Erlbaum. [DK]
Krebs, D., Denton, K., Carpendale, J , Vermeulen, S., Bartek, S. & Bush, A. (in press) The many faces of moral judgment. In: *Psychological development. Perspectives across the life span*, ed. M. A Luszcz & T. Nettelbeck. Elsevier. [DK]
Krebs, D. L., Denton, K. & Higgins, N. C. (1988) The evolution of self-deception and self-knowledge. In. *Sociobiological perspectives on human development*, ed. K. MacDonald. Springer-Verlag. [DK]
Krebs, D. L. & Miller, D. (1985) Altruism and aggression. In: *Handbook of social psychology*, vol. 2, ed. E. Aronson & G. Lindzey. Lawrence Erlbaum. [DK, PET]
Kropotkin, P. A. (1902) *Mutual aid: A factor of evolution.* (Reprinted 1972, New York University Press.) [BMK]
Kuper, A. (1988) *The invention of primitive society* Routledge. [rLRC]
Kurland, J. A. & Beckerman, S. J. (1985) Optimal foraging and hominid evolution: Labor and reciprocity. *American Anthropologist* 87:73–93. [aLRC, BMK]
Lakoff, G. (1987) *Women, fire, and dangerous things: What categories reveal about the mind.* University of Chicago Press. [rLRC]
Lakoff, G. & Johnson, M. (1980) *Metaphors we live by.* University of Chicago Press. [rLRC]
Landau, M. (1984) Human evolution as narrative. *American Scientist* 72 262–68. [arLRC]
Larmore, C. E. (1987) *Patterns of moral complexity* Cambridge University Press. [RAS]
Lea, S. E. G., Tarpy, R M. & Webley, P. (1987) *The individual in the economy.* Cambridge University Press. [EJSS-B]
Leacock, E. & Lee, R. B. (1982) Introduction. In: *Politics and history in band societies*, ed. E. Leacock & R. B. Lee. Cambridge University Press. [BMK]
Lee, R. B. (1979) *The Dobe !Kung.* Holt, Rinehart, and Winston [BMK]
Lee, R. B. & DeVore, I., eds. (1984) *Man the hunter*, 2nd ed. Aldine. [BMK]
Lefebvre, V. A. (1982) *Algebra of conscience.* Reidel. [VAL]
(1985) The golden section and algebraic model of ethical cognition. *Journal of Mathematical Psychology* 29:289–310. [VAL]
(1987a) The fundamental structures of human reflexion. *Journal of Social and Biological Structures* 10:129–75. [VAL]
(1987b) Responses to Zajonc. *Journal of Social and Biological Structures* 10:237–38. [VAL]
Lefebvre, V. A., Lefebvre, V. D. & Adams-Webber, J. (1986) Modeling an experiment on construing self and others. *Journal of Mathematical Psychology* 30:317–30. [VAL]
Levi-Strauss, C. (1949) *The elementary structure of kinship.* (Reprinted 1969, Beacon Press.) [BMK]
(1963) *Totemism.* Beacon Press. [BMK]
Lewin, R (1987) The origins of the human mind. *Science* 238:667–71 [aLRC]
Lichtenstein, S. & Slovic, P. (1971) Reversals of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology* 89.46–55. [aLRC]
Liebrand, W. B. G. (1984) The effect of social motives, communication, and group size on behavior in an N-person multi-stage mixed-motive game. *European Journal of Social Psychology* 14:239–64 [aLRC]
Logue, A. W. (1984) Is it possible to be optimal? *Behavioral and Brain Sciences* 7:111. [aLRC]
Lovejoy, C. O. (1981) The origins of man. *Science* 211.341–50. [aLRC]
Lumsden, C. J. & Wilson, E. O. (1981) *Genes, mind, and culture: The coevolutionary process.* Harvard University Press. [CSF]
Maass, A. (1983) *Congress and the common good* Basic Books. [JM]
Mansbridge, J. J. (1980) *Beyond adversary democracy.* University of Chicago Press. (Reprinted in 1983.) [JM]
ed (1990) *Beyond self-interest.* University of Chicago Press. [JM]
Marková, I. (1987) *Human awareness. Its social development.* Century Hutchinson. [IV]
Marshall, L (1979) Sharing, talking, and giving: Relief of social tensions among !Kung Bushmen. In: *Kalahari Hunter-Gatherers*, ed. R. B. Lee & I. DeVore. Harvard University Press. [BMK]
Martin, R. D. (1983) *Human brain evolution in ecological context* 52nd James Arthur lecture on the Evolution of the Brain. American Museum of Natural History. [RF]
Maynard Smith, J. (1982a) Introduction. In: *Current problems in sociobiology*, ed. King's College Sociobiology Group. Cambridge University Press. [RIMD]
(1982b) *Evolution and the theory of games.* Cambridge University Press. [JT]
(1984) Game theory and the evolution of behaviour. *Behavioral and Brain Sciences* 7:95–125. [aLRC]
McClintock, C. G. & Liebrand, W. B. G. (1988) The role of interdependence structure, individual value orientation and other's strategy in social decision making. A transformational analysis. *Journal of Personality and Social Psychology* 55:396–409. [WBGL]
McClintock, M. (1981) Social control of the ovarian cycle. *American Zoologist* 21:243–56. [JM]
(1987) A functional approach to the behavioral endocrinology of rodents In. *Psychology of reproductive behavior: An evolutionary perspective*, ed. D Crews. Prentice-Hall. [JM]
McGuire, M., Raleigh, M. & Brammer, G. (1982) Sociopharmacology. *Annual Review of Pharmacological Toxicology* 22:643–61. [JM]
Meggitt, M J. (1962) *Desert people A study of the Walbiri Aborigines of Central Australia.* Angus and Robertson. [BMK]
Messé, L. A. & Sivacek, J. M. (1979) Predictions of others' responses in a mixed-motive game. Self-justification or false consensus? *Journal of Personality and Social Psychology* 37.602–7. [aLRC]
Messick, D. M. & McClintock, C. G. (1968) Motivational basis of choice in experimental games. *Journal of Experimental Social Psychology* 4:1–25. [WBGL]
Midgley, M. (1978) *Beast and man: The roots of human nature.* Harvester Press/Cornell University Press. [IV]
Muir, W. K., Jr. (1982) *Legislature: California's school for politics.* University of Chicago Press. [JM]
Myers, F. R. (1986) *Pintupi country, Pintupi self. Sentiment, place, and politics among Western Desert Aborigines.* Smithsonian Institution Press. [BMK]
Myers, M L (1983) *The soul of modern economic man Ideas of self-interest, Thomas Hobbes to Adam Smith.* University of Chicago Press. [aLRC, JM]
Nisbett, R. E. & Wilson, T. D. (1977) Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84:231–59. [aLRC, DTK]
Nowak, M. & Sigmund, K. (1989) Oscillations in the evolution of reciprocity. *Journal of Theoretical Biology* 137:21–26. [BMK]
Olson, M. (1982) *The rise and decline of nations: Economic growth, stagnation, and social rigidities.* Yale University Press. [rLRC]
Orbell, J. M., van de Kragt, A J C & Dawes, R. M. (1988) Explaining discussion-induced cooperation. *Journal of Personality and Social Psychology* 54:811–19. [arLRC, LW]
Osgood, C. E. (1965) *Psycholinguistics.* Indiana University Press. [VAL]
Ostrom, T. M. (1984) The sovereignty of social cognition. In: *Handbook of social cognition*, vol. 1, ed. R. S. Wyer, Jr. & T. K. Srull. Erlbaum. [aLRC]

Pancer, M. (1973) Approval motivation in the prisoner's dilemma game. Unpublished bachelor's thesis, University of Toronto. [AR]
Parker, S. T & Gibson, K. R (1979) A developmental model for the evolution of language and intelligence in early hominids. *Behavioral and Brain Sciences* 2.367–408. [aLRC]
Plotkin, H. C. (1988) Learning and evolution. In: *The role of behaviour in evolution*, ed. H. C. Plotkin. MIT Press. [RIMD]
Plotkin, H. C. & Odling-Smee, F. J. (1981) A multiple-level model of evolution and its implications for sociobiology. *Behavioral and Brain Sciences* 4:225–68. [RIMD]
Potts, R. (1984) Hominid hunters? Problems of identifying the earliest hunter/gathers. In: *Hominid evolution and community ecology*, ed. R. Foley. Academic Press. [aLRC]
Pringle, J. W. S. (1951) On the parallel between learning and evolution. *Behaviour* 3.174–215. [RIMD]
Rachlin, H. (1980) Economics and behavioral psychology. In: *Limits to action*, ed. J. E. R. Staddon. Academic Press. [EJSS-B]
Radke-Yarrow, M., Zahn-Waxler, C. & Chapman, M. (1983) Children's prosocial dispositions and behavior. In. *Handbook of child psychology*, vol. 4, 4th edition, ed. P. H. Mussened. John Wiley. [JM]
Rapoport, Amnon (1967) Optimal policies for the prisoner's dilemma. *Psychological Review* 74:136–48. [aLRC]
Rapoport, Anatol (1974) Prisoner's dilemma: Recollections and observations In *Game theory as a theory of conflict resolution*, ed. Anatol Rapoport D. Reidel. [aLRC]
(1983) *Mathematical models in the social and behavioral sciences* Wiley. [aLRC]
(1988) Experiments with n-person social traps I. Prisoner's dilemma, weak prisoner's dilemma, volunteer's dilemma, and largest number. *Journal of Conflict Resolution* 32:457–72. [AR]
Rapoport, Anatol & Chammah, A. M. (1965) *Prisoner's dilemma*. University of Michigan Press. [aLRC]
Rapoport, Anatol, Guyer, M. J. & Gordon, D. G. (1976) *The 2 × 2 game* University of Michigan Press. [aLRC]
Reich, R. B., ed. (1988) *The power of public ideas*. Ballinger. [JM]
Rescher, N. (1989) *Rationality*. Clarendon Press [RAS]
Reynolds, V., Falger, V. & Vine, I., eds. (1987) *The sociobiology of ethnocentrism: Evolutionary dimensions of xenophobia, discrimination, racism and nationalism*. Croom Helm. [BMK]
Richards, G. (1987) *Human evolution*. Routledge & Kegan Paul. [rLRC]
Richards, R. J. (1987) *Darwin and the emergence of evolutionary theories of mind* University of Chicago. [aLRC]
Rozin, P. (1976) The evolution of intelligence and access to the cognitive unconscious. In: *Progress in psychobiology and physiological psychology*, vol. 6, ed. J. M. Sprague & A. N. Epstein. Academic Press. [arLRC]
Rushton, P. J. (1980) *Altruism, socialization, and society*. Prentice-Hall. [aLRC]
Sahlins, M. (1976) *Culture and practical reason*. University of Chicago Press. [aLRC]
Schelling, T. C. (1978) Altruism, meanness, and other potentially strategic behaviors. *American Economic Review* 68.229–30 [JM]
Schlenker, B. R. (1982) Translating actions into attitudes: An identity-analytic approach to the explanation of social conduct. In: *Advances in experimental social psychology*, vol. 15, ed. L. Berkowitz Academic Press. [PET]
Schwartz, B. (1986) *The battle for human nature: Science, morality and modern life*. Norton. [aLRC, SO]
Sears, D. O. (1990) Self-interest in the mass public's political preferences. In: *Beyond self-interest*, ed. J. J. Mansbridge. University of Chicago Press. [JM]
Sellars, W. (1963a) *Science, perception, and reality*. Routledge and Kegan Paul. [aLRC]
(1963b) Imperatives, intentions, and the logic of "ought." In: *Morality and the language of conduct*, ed. G. Nakhnikian & H.-N. Castaneda. Wayne State University Press. [MG]
Sen, A. (1973) Behaviour and the concept of preference. *Economica* 40:241–59. [AIH]
(1977) Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy and public affairs* 6:317–44. [JM]
Shaw, R. P. & Wong, Y. (1988) *Genetic seeds of warfare: Evolution, nationalism, and patriotism*. Unwin Hyman. [BMK]
Shubik, M. (1964) *Game theory and related approaches to social behavior*. Wiley. [aLRC]
Shweder, R. A., Mahapatra, M. M. & Miller, J. G. (1988) Culture and moral development. In: *The emergence of morality in young children*, ed. J. Kagan & S. Lamb. University of Chicago Press. [RAS]
Shweder, R. A. & Much, N. C. (1987) Determinations of meaning. Discourse and moral socialization. In: *Moral development through social interaction*, ed. W. M. Kurtines & J. L. Gewirtz. Wiley. [RAS]
Simon, H. (1980) The behavioral and social sciences. *Science* 209:72–78. [aLRC]
(1983) *Reason in human affairs*. Stanford University Press. [rLRC]
Skinner, B. F. (1957) *Verbal behavior*. Appleton-Century-Crofts. [rLRC]
Smith, A. (1776/1910) *The wealth of nations*. E. P. Dutton. [AR]
(1969/1759) The theory of moral sentiments In. *British moralists 1659–1880*, vol. 2, ed. D. D Raphael Clarendon Press. [RAS]
Smith, P. K. (1987) Is human sociobiology a progressive or a degenerating research programme? *Behavioral and Brain Sciences* 10:86–87. [rLRC]
Smuts, B. B. (1985) *Sex and friendship in baboons*. Aldine. [RIMD]
Snyder, C. R. & Higgins, R. L. (1988) Excuses: Their effective role in the negotiation of reality. *Psychological Bulletin* 104:23–35 [DK]
Sober, E. (1984) *The nature of selection*. MIT Press. [rLRC]
Sonuga-Barke, E J. S. (1989) Viewing hyperactivity as delay aversion. Paper presented at Institute of Psychiatry, London, June. [EJSS-B]
Symonds, D (1979) *The evolution of human sexuality*. Oxford University Press. [rLRC]
Tajfel, H. (1970) Experiments in intergroup discrimination. *Scientific American* 223·96–102. [aLRC]
(1981) *Human groups and social categories. Studies in social psychology*. Cambridge University Press. [aLRC]
Tajfel, H. & Billig, M. (1974) Familiarity and categorization in intergroup behavior. *Journal of Experimental Social Psychology* 10:159–70. [aLRC]
Tajfel, H. & Turner, J. C (1979) An integrative theory of intergroup conflict. In: *The social psychology of intergroup relations*, ed. W. Austin & S. Worchel. Brooks/Cole. [aLRC]
(1985) The social identity theory of intergroup conflict. In: *Psychology of intergroup relations*, ed. S. Worchel & W. Austin. Nelson-Hall. [MBB]
Taylor, C (1971) Interpretation and the sciences of man. *The review of metaphysics* 35.3–51. [MG]
Teigen, K. H. (1988) When are low-probability events judged to be "probable"? Effects of outcome-set characteristics on verbal probability estimates. *Acta Psychologica* 68:157–74. [KHT]
Terhune, K. W. (1968) Matrices, situations and interpersonal conflict within the prisoner's dilemma. *Journal of Personality and Social Psychology*. Monograph 8, no. 3, part 2;1–24. [AR]
Testart, A. (1982) The significance of food storage among hunter-gathers: Residence patterns, population densities, and social inequalities. *Current Anthropology* 23:523–37. [BMK]
(1985) *Le communisme primitif· Economie et ideologie*. Maison des Sciences de l'Homme. [BMK]
Tetlock, P E. (1980) Explaining teacher explanations for pupil performance: A test of the self-presentation position. *Social Psychology Quarterly* 43:283–90. [PET]
(1981) The influence of self-presentation goals on attributional reports. *Social Psychology Quarterly* 44:300–311. [DTK]
Tetlock, P. E & Levi, A. (1982) Attribution bias: On the inconclusiveness of the cognition-motivation debate. *Journal of Experimental Social Psychology* 18:68–88. [PET]
Tetlock, P. E. & Manstead, A. S. R. (1985) Impression management versus intrapsychic explanations in social psychology: A useful dichotomy? *Psychological Review* 92:59–77. [PET]
Tinbergen, N. (1963) On the aims and methods of ethology. *Zeitschrift für Tierpsychologie* 20:410–33. [RIMD]
Tooby, J. & Cosmides, L. (1988) The evolution of war and its cognitive foundations. *Institute for Evolutionary Studies Technical Report* 88-1. [JT]
(1989) Evolutionary psychology and the generation of culture, part I. Theoretical considerations. *Ethology and sociobiology* 10:29–49. [DTK]
Tooby, J. & DeVore, I. (1987) The reconstruction of hominid behavioral evolution through strategic modeling. In: *The evolution of human behavior: Primate models*, ed. W. G. Kinzey. State University of New York Press. [rLRC, BMK, JT]
Triandis, H. (1988) Collectivism v. individualism: A reconceptualization of a basic concept in cross-cultural social psychology. In: *Cross-Cultural Studies of Personality, Attitudes, and Cognition*, ed. G. K. Verma & C. Bagley. Macmillan. [BMK]
Trivers, R. L. (1971) the evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35–57. [aLRC]
(1985) *Social evolution* Benjamin/Cummings. [aLRC, DTK]
Turnbull, C. M. (1965a) The Mbuti Pygmies: An ethnographic survey. *Anthropological Papers of the American Museum of Natural History* 50(3):139–282. [BMK]
(1965b) *Wayward servants: The two worlds of the African Pygmies*. Natural History Press. [BMK]

(1984) The importance of flux in two hunting societies. In: *Man the hunter*, 2nd ed., ed. R. B. Lee & I. DeVore. Aldine. [BMK]
Turner, J. C. (1987) *Rediscovering the social group: A self-categorization theory*. Basil Blackwell. [arLRC, MBB]
Turner, J. C. & Giles, H. (1981) *Intergroup behavior*. Basil Blackwell. [aLRC]
Tversky, A. & Kahneman, D. (1981) The framing of decisions and the psychology of choice. *Science* 211:453–58. [aLRC]
van de Kragt, A. J. C., Orbell, J. & Dawes, R. M. (1983) The minimal contributing set as a solution to public goods problems. *American Political Science Review* 77:112–22. [arLRC]
van de Kragt, A. J. C., Orbell, J. & Dawes, R. M. with Braver, S. R. & Wilson, L. A., II (1986) Doing well and doing good as ways of resolving social dilemmas. In: *Experimental social dilemmas*, ed. H. Wilke, D. Messick & C. Rutte. Lang GmbH. [arLRC]
van Schaik, C. P. (1983) Why are diurnal primates living in groups? *Behaviour* 87:120–44. [RIMD]
Vine, I. (1983) Sociobiology and social psychology: Rivalry or symbiosis? The explanation of altruism. *British Journal of Social Psychology* 22:1–11. [IV]
(1987) Inclusive fitness and the self-system: The roles of human nature and sociocultural processes in intergroup discrimination. In. *The sociobiology of ethnocentrism*, ed. V. Reynolds, V. S. F. Falger & I. Vine. Croom Helm/University of Georgia Press. [rLRC, IV]
Vining, D. R. (1986) Social versus reproductive success. The central theoretical problem of human sociobiology. *Behavioral and Brain Sciences* 9:167–216. [BMK]
Vogler, D. J. & Waldman, S. R. (1985) *Congress and democracy*. Congressional Quarterly Press. [JM]
von Neumann, J. & Morgenstern, O. (1947) *The theory of games and economic behavior*, 2nd ed. Princeton University Press. [aLRC]
Vucinich, A. (1988) *Darwin in Russian thought*. University of California Press. [BMK]
Wallach, M. A. & Wallach, L. (1983) *Psychology's sanction for selfishness* Freeman. [aLRC, LW]
Walter, A. (1989) Pop goes the weasel. *Behavioral and Brain Sciences* 12.185–86. [rLRC]
Washburn, S. L. & Lancaster, C. S. (1968) The evolution of hunting. In: *Man the hunter*, ed. R. B Lee & I DeVore. Aldine. [aLRC]
Washburn, S. L. & McCown, E. R. (1978) Human evolution and social science. In: *Human evolution: Biosocial perspectives*, ed. S. L. Washburn & E. R. McCown. Benjamin/Cummings. [aLRC]
White, L. A. (1949) The symbol: The origin and basis of human behavior. In: *The science of culture*. Farrar, Straus & Co. [BMK]
Wicksteed, P. H. (1933) *The common sense of political economy*. George Routledge. [AR]
Wiessner, P. (1982) Risk, reciprocity, and social influences on !Kung San economics. In: *Politics and history in Band societies*, ed. E. Leacock & R. B Lee. Cambridge University Press. [BMK]
Williams, G. C. (1966) *Adaptation and natural selection. A critique of some current evolutionary thought*. Princeton University Press. [aLRC, BMK, JT]
Wilson, D. S. (1983) Individual and group selection: An historical and conceptual review. *Annual Review of Ecology and Systematics* 14:159–87. [CSF]
Wilson, D. S. & Sober, E. (1989) Reviving the superorganism. *Journal of Theoretical Biology* 136:337–56. [BMK]
Wilson, E. O. (1975) *Sociobiology: The new synthesis*. Harvard University Press. [aLRC, AIH, DTK, IV]
Winch, P. (1956) Social science. *British journal of sociology* 7:18–33. [MG]
(1958) *The idea of a social science*. Routledge. [MG]
Wolpoff, M. H. (1987) Competitive exclusion among Lower Pleistocene hominids: The single species hypothesis. In: *Primate evolution and human origins*, ed. R. L. Ciochon & J. G. Fleagle. Aldine de Gruyter. (Reprinted from *Man* 6:601–14, 1971.) [aLRC]
Woodburn, J. C. (1982) Egalitarian societies. *Man* 17:431–51. [BMK]
(1984) Stability and flexibility in Hadza residential groupings. In: *Man the hunter*, 2nd ed., ed. R. B. Lee & I DeVore Aldine. [BMK]
World Commission on Environment and Development (1987) *Our common future*. Oxford University Press. [NCS]
Wrangham, R. W. (1980) An ecological model of female-bonded primate groups. *Behaviour* 75.262–300. [RIMD]
(1982) Mutualism, kinship and social evolution. In: *Current problems in sociobiology*, ed. King's College Sociobiology Group. Cambridge University Press. [RIMD]
(1987) African apes: The significance of African apes for reconstructing human social evolution. In. *The evolution of human behavior. Primate models*, ed. W. G. Kinzey. State University of New York Press. [BMK]
Wrightsman, L. S. (1966) Personality and attitudinal correlates of trusting and trustworthiness in a two-person game. *Journal of Personality and Social Psychology* 4:328–32. [aLRC]
Yengoyan, A. A. (1984) Demographic and ecological influences on Aboriginal Australian marriage sections. In: *Man the hunter*, 2nd ed., ed. R. B. Lee & I. DeVore. Aldine. [BMK]
Zajonc, R. B. (1980a) Cognition and social cognition: A historical perspective. In: *Retrospections on social psychology*, ed. L. Festinger. Oxford University Press. [aLRC]
(1980b) Feeling and thinking: Preferences need no inferences. *American Psychologist* 35:151–75. [aLRC]
(1987) Comments on interpersonal affiliation and the golden section. *Journal of Social and Biological Structures* 10:212–14. [VAL]