Abstract

How does social order emerge among autonomous but interdependent agents? The expectation of future interaction may explain cooperation based on rational foresight, but the "shadow of the future" offers little leverage on the problem of social order in "everyday life" -- the habits of association that generate unthinking compliance with social norms. Everyday cooperation emerges not from the shadow of the future but from the lessons of the past. Rule-based evolutionary models are a promising way to formalize this process. These models may provide new insights into emergent social order -- not only prudent reciprocity, but also expressive and ritual self-sacrifice for the welfare of close cultural relatives.

Keywords:
cooperation, evolutionary models, artificial agents, altruism

The Enigma of Cooperation

1.1

Social order is easy to take for granted yet surprisingly difficult to explain. That human cultures have evolved with a predilection for group solidarity makes obvious sense. Evolution has left us physically incapable of self-reliant existence in a "state of nature." Hence we must depend on cooperation with one another for our very existence. Why, then, do scholars across the social sciences continue to expend such effort to explain a condition that everyone else takes for granted as unavoidable? A moment's reflection shows why. We need look no further than the proliferation of nuclear arms, the depletion of the ozone layer, or the greenhouse effect to see that the necessity for collective rationality is no guarantee that it will obtain.

1.2

That is perhaps the great paradox of the human condition. We cannot live without one another, but we may not be able to live with one another either. It seems Nature has played a cruel trick on our species -- we cannot survive alone, yet unlike social insects, we are not genetically hardwired for cooperative behavior. Sociobiologists suggest that humans may have a genetic predisposition for reciprocity and in-group altruism, as well as aggression and competition (Ruse and Wilson 1986;
Alexander 1987). This implies only a capacity for cooperation, not its necessity (Dawkins 1989). If nature has endowed us with highly supple instinctual programming at the genetic level, then sociality is not inherited but must be acquired (and re-acquired) through our capacity for behavioral learning (Cooley 1964). This plasticity makes the paradox of social order especially compelling. In the absence of innate cooperative propensities, can social order emerge entirely through our capacity for adaptation?

1.3

To find out, researchers have begun to experiment with evolutionary models of what Bainbridge et al. (1994) call "artificial social life." Computational models of emergent social order, based on an "evolutionary epistemology" (Schull 1996), provide a promising alternative to an older generation of models based on analytical game theory. First, most game-theoretic models of the evolution of cooperation are pointed in the wrong direction. The prospect of re-encounter may explain the forward-looking, calculated, and self-interested decision to cooperate for mutual gain, but this theory offers little leverage on the problem of social order in everyday life -- the habits of association that generate unthinking compliance with social norms. Everyday cooperation emerges not from the "shadow of the future" but from the lessons of the past.

1.4

Second, most game-theoretic models of self-interested cooperation cannot account for genuinely altruistic self-sacrifice. The problem here is the unit of adaptation. Until recently, the leading alternatives were the group or the individual. However, evolutionists in the life sciences have introduced a third approach, based on the gene, or more precisely, the genetically encoded rule, as the unit of adaptation. Gene-selectionist models explain the evolution of "kin altruism" in nature, based on the principle of inclusive fitness. Rule-based evolutionary models may also provide new insights into "in-group altruism," or self-sacrifice for the welfare of close cultural relatives.

The Janus-Face of Cooperation

2.1

The paradox of social order can be formalized as a "Prisoner's Dilemma," a game between two players, each with two choices, to "cooperate" or "defect." These two choices intersect at four possible outcomes, each with a designated payoff. R (for reward) and P (for punishment) are the payoffs for mutual cooperation and defection, respectively, while S (for sucker) and T (for temptation) are the payoffs for cooperation by one player and defection by the other. By definition, \( T > R > P > S \), a payoff inequality that creates tension between individual and collective interests. Consider what happens in a population of agents that always cooperate or always defect. The collectively optimal outcome (R) goes to two cooperators, and the collectively worst outcome (P) goes to two defectors. However, the best individual payoff (T) goes to a unilateral defector, while the victim receives the worst individual payoff (S). The dominant strategy in a single encounter is to defect, no matter what the other side chooses. The trap is that the optimal choice for each contestant can lead to what is often the worst possible collective outcome.

2.2

Game-theoretic solutions to Prisoner's Dilemma require that the game be repeated. Defection is not the dominant strategy in the repeated game, which opens up the possibility for cooperation if both sides can engineer a tacit collusion. Repetition motivates cooperation by placing the game in the "shadow of the future" (Axelrod 1984: 12). In analytical game theory (as distinct from evolutionary game theory, considered below), the prospect of future encounters leads fully rational players to calculate the effect their choices may have on their partner's best response, assuming the partner is also fully rational and making a similar calculation of the optimal strategy in future play. A tacit collusion becomes possible if each side knows that the other will not tolerate defection (including
accidental defection), and each side knows that the partner knows this fact.

2.3

This game-theoretic analysis assumes strategic foresight based on complete information and a perfect grasp of the logical structure of a well-defined problem (Axelrod 1997: 47). Simply put, analytical game theorists explain self-interested cooperation between people like themselves (Binmore 1997: ix).

2.4

However, in everyday life, most games are played by lay contestants, not mathematicians. These games are often highly routinized, played according to internalized rules or instructions, with little conscious deliberation. The players rarely calculate the strategic consequences of alternative courses of action but simply "look ahead by holding a mirror to the past" (Macy 1993).

2.5

Like the tacit collusion engineered by skillful strategists, everyday cooperation also depends decisively on whether the game is repeated, but the reason has little to do with the strategic implications of the "shadow of the future." Unthinking cooperation depends not on the incentives created by the prospect of future interaction but on the habits of association distilled from prior exposure to a recurring problem. We learn to interact the same way we learn to speak -- we practice until we not only get it right, but do it without even thinking. Of course, we sometimes choose to cooperate deliberately, just as we sometimes "choose our words carefully." But mostly we do not think before we speak, we just talk. Similarly, we do not choose to cooperate, we simply know to take turns, stand in line, go to the polls, speak softly, show courtesy, defer to others, reply promptly, tell the truth, and so on. We mostly cooperate without realizing it, or intending it, or wondering if we might as easily "free ride" without being caught.

2.6

Although behavioral economists contend that rule-based behavior is mandated by cognitive limitations and the costs of information (Simon 1992), social psychologists point to a much broader and more basic requirement: the possibility of social life (Berger 1966; for a similar view from economics, see Hayek 1973). Among social species, we are unique in our plasticity. Unlike social insects, our genes leave us free to leave the swarm and chart our own individual course. Were we not creatures of habit, routine, and heuristic devices, effective coordination might be impossible -- a cacophony of inappropriate responses to unexpected reactions from others. Behavioral rules (including social norms and conventions) make social interaction predictable, so that interdependent individuals can influence one another in response to the influence they receive, thereby carving out locally stable patterns of interaction. Symbolic representations of meaning in communicative acts evolve through the repetitive use of signs, as interdependent actors try to coordinate effective interaction (Peirce 1955). Moral habits and social norms evolve in much the same way. "The majestic order of society," Heise contends, "emerges from repetitive application of evolved cultural resources to frame and solve recurrent problems" (1996:1). The solutions then become part of the cultural repertoire and are taken for granted by individuals. In short, heuristic devices are not simply analytic shortcuts that lower the cognitive costs of decision-making. They are the grammar that structures social life.

2.7

The rules that secure social order emerge not from the shadow of the future but from the lessons of the past. These lessons involve two types of experiential feedback. Reproduction alters the frequency distribution of strategies within a population of related individuals, while reinforcement, alters the probability distribution of rules within the repertoire of each individual. Through either mechanism, repetition, not calculation, brings the future to bear on the present, by recycling the lessons of the past.¹ For bees, these lessons are inscribed in their genes. For us, the lessons are
encoded in social norms, customs, conventions, routines, rituals, protocols, morals, habits, and heuristics. Their evolution is what we study when we build computational models of emergent cooperation.

2.8
"Emergent cooperation" captures the idea that very simple local interaction patterns can generate highly complex, and often surprising, global solutions. The process can be modeled using evolutionary game theory, a method that assumes a population of rule-based and myopic adaptive agents. Applied to multi-agent dynamical systems, the evolutionary approach can identify basins of attraction on an ecological landscape (Epstein and Axtell 1996). However, the nonlinear and stochastic properties of many dynamical systems often preclude analytical modeling and provides a promising application for recent advances in the simulation of artificial life (Simon 1992:45; Axelrod 1997:48).

2.9
In sum, computational models of emergent cooperation represent an important advance in the effort to explain the creation and recreation of social order in everyday life. Because repetition, not foresight, links game payoffs back to the choices that produce them, these models need not assume that the payoffs are the intended consequences of action. Thus, the models can be applied to expressive and righteous behavior that lacks a deliberate or instrumental motive. For example, Frank's (1988) evolutionary model of trust and commitment formalizes the evolution of emotions like vengeance and sympathy. An angry or frightened actor may not be capable of deliberate and sober optimization, yet the response to the stimulus has consequences for the individual, and these in turn can modify the probability that the behavior will be repeated, through reinforcement, imitation, or some combination of learning and reproduction.

2.10
In a rule-based model, not only may actors be unaware that their actions are self-serving, it is not even necessary to assume any benefit whatsoever to the individual. The evolutionary approach is thus applicable not only to habitual and expressive cooperation but also to behavior that is genuinely altruistic. Altruism entails more than a prudent detour in the pursuit of self-interest (e.g., a well-publicized charitable donation). By definition, altruistic behavior requires sacrifice -- not just an investment that pays back with a compensating benefit, but a net loss -- an outcome that is inferior to an available alternative. How can this evolve in a payoff-driven process of adaptation?

2.11
Part III addresses this question by comparing three specifications of the unit of adaptation: the group, the individual, and the underlying rule or strategy. I conclude that rule-based computational models are uniquely capable of explaining self-sacrificial in-group altruism.

Altruism from the Very Bottom Up

3.1
Analytical game theory cannot account for genuinely altruistic behavior without the troublesome epicycle that altruists are simply egoists with other-regarding preferences. In this case, all behavior is instrumentally rational by definition and analytical interest shifts to the origins of preferences, a problem outside the theoretical domain of game theory.

3.2
While genuine altruism cannot be explained using the self-interested individual as the unit of analysis, it is readily amenable to functionalist explanations that import the sociobiological principle of group-selection, in which the *deme*, not the individual, is the starting point (Wynne-Edwards 1962). In group selection, demes that lack altruistic mores may be more prone to extinction than are
competing groups whose members are more willing to sacrifice for the greater good (Boorman and Levitt 1980:5). This makes altruism trivially easy to explain. Nevertheless, the group-selectionist account has fallen out of favor among sociobiologists for reasons that may apply to social and cultural applications as well. The problem is the need for "nearly complete isolation as between demes. Otherwise, demes rapidly lose their diversity as genetic units, and group selection through differential extinction fails to find continuing genetic variance on which to act" (Boorman and Levitt 1980:7). Moreover, differential extinction searches much less efficiently than reproduction. Extinction "is inherently a negative force, acting only to eliminate demes having a relatively less adaptive genetic composition. In particular, it cannot in the first instance, create populations having a highly adaptive genetic makeup, but (if opposed by individual selection) must rely on relatively inefficient random effects such as drift to create such favored demes" (Boorman and Levitt 1980:8). If individual selection favors egoism and is more efficient than group selection, then altruism can be expected to lose the evolutionary footrace.2

3.3

If altruism cannot be explained using either the individual or group as the unit of analysis, what is left? The evolutionary game theorist Maynard-Smith (1979) has proposed a provocative answer: the underlying strategy or rule (whether a genetic or heuristic instruction). By "rule," I mean the smallest possible unit of effective instruction. Rules need not be genetically hardwired. Culturally evolved rules are patterned behavioral responses to stimuli, carved out through repetition, stored in neural pathways, and continually tested in local interaction. Whether genetic or cultural, rules can be formalized as input-output functions, where the input is a set of conditions of varying complexity and the output is an action. The consequences of the action, which may or may not have been consciously anticipated, then modify (1) the probability that the action will be repeated the next time the input conditions are met, and (2) the probability that the associated rule will be replicated and diffused.

3.4

The theory that rules, not groups or individuals, are the units of adaptation, not only provides a more rigorous explanation of altruism but also yields empirically plausible predictions as to the identity of the beneficiaries. The theory can be applied to an explanation of "kin altruism" in nature, the strategy to sacrifice for the benefit of close genetic relatives (Ruse and Wilson 1986; Alexander 1987; Hamilton 1964). A gene for kin altruism can improve its viability by directing a transfer of vital resources from its agent to another organism that carries the same gene. From a "rule's-eye" view of the problem, the altruist serves the interests -- not of the beneficiary -- but of the "selfish gene" that controls its behavior (Dawkins 1989). The "self-interest" of a gene refers to its evolutionary stake in the outcomes of the phenotypes that it instructs, insofar as these outcomes influence the odds that the gene will flourish in the face of competitive selection pressures.

3.5

Allison (1992) has extended the kin altruism model to benevolence based on cultural relatedness, such as geographical proximity or a shared cultural marker (such as dress or language). The logic of kin altruism may explain the importance of group boundaries around tightly knit "in-groups" and the need for cultural markers that signify membership. As in the genetic specification, a "self-interested rule" need not imply a program that instructs egoistic behavior by its agent. Suppose a rule could propagate faster by ordering its carrier to export life-chances to those who follow the same rule. Then the rule could be said to have a "self-interest" in the altruistic behavior of its carrier.3 Solidary expressions of identity can thus be reduced to rational self-interest, but it is the evolutionary interest of the rule, not the interest of its agent, that is the effective cause. Hence, cultural evolution favors a tendency to associate with those who are similar, to differentiate from "outsiders," and to defend the in-group against social trespass with the emotional ferocity of parents defending their offspring.
In-group altruism may in turn come to be imitated through the logic of sexual selection. "Sexual selection" refers to the evolutionary advantage of traits that increase the probability of mating. For example, sexual selection explains the generous and colorful plumage of the male peacock. The cumbersome tail increases mortality but provides a net gain in fertility by making the males more sexually attractive, thus increasing their overall reproductive chances (Dawkins 1989). Much the same thing may happen in the selection of role models or mentors. Conspicuous prosocial behavior may increase vulnerability to predators, but also make it easier to attract a protégé -- someone who can then be influenced to disseminate the emergent norm.

3.7
Cultural markers may also explain cooperation in one-shot Prisoner's Dilemma. Following Frank (1988), John Skvoretz and I (forthcoming) model the evolution of trust among strangers who must learn to read telltale signs of character. In this simulation experiment, cooperation in a one-shot game required the coordination of shared meanings of symbolic cues. Conventions were created by an adaptive process tied to the outcomes of compliance, while the consequences of compliance depended on the emergence of a local consensus about the meaning of the telltale signs. Under certain conditions, we found that cooperation between strangers could evolve, even though this is a dominated strategy in the one-shot game.

3.8
This points to a decisive difference between analytical and evolutionary game theory. Analytical game theory applies to the individual as the unit of analysis, and therefore requires the axiom of individual self-interest. Evolutionary game theory applies to rules (or strategies) as the units of analysis, as well as to the individuals who act on these rules. This means that the behavior the rules instruct need not necessarily advance the interests of the individual agent. Individually self-interested behavior can then be modeled as empirically variable, not axiomatic.

3.9
It must be stressed, however, that the self-interested individual is clearly not precluded in evolutionary models. Consider, for example, the evolution of "reciprocal altruism," such as the trading of favors. Reciprocal altruism is fully compatible with individual self-interest and not a form of genuine altruism. Rather, it is "a straightforward illustration of prudent behavior," Frank notes, "enlightened prudence, to be sure, but self-interested behavior all the same" (1988:34-5). An enlightened egoist may transfer resources to another, expecting that this behavior will trigger sufficient compensation.

3.10
Although reciprocal altruism can be modeled using the forward-looking individual as the unit of analysis, it may nevertheless be instructive to consider a rule-selectionist formulation. Again taking a "rule's-eye" view of the problem, the manifestations of self-interested rules can be altruistic, so long as the outcomes ultimately promote the propagation of the rule. When these benefits to the rule obtain immediately from the behavior, we classify the behavior as self-interested. When the programmatic benefits are separated from the behavior in time and space, we classify the behavior as altruistic. With reciprocal altruism, the benefit is temporally removed in that the rule has learned how to export life-chances as a way to trigger a later return on the investment. With kin altruism, the benefit is spatially removed, in that the rule has learned how to export life-chances from one of its agents to another, such that the "inclusive fitness" (Hamilton 1964) of all its carriers is greater than before. Given that space and time are alternative measures of distance, it seems reasonable to classify both reciprocity and kin-benevolence as altruistic behaviors (or phenotypes) that provide indirect benefits to the self-interested rules (or genotypes) in which these behaviors are encoded. The only difference is whether the benefits must leap over time or space to make their way back to the rules that generated them. Indeed, the two dimensions can be combined, as in "generalized reciprocity" (Ekeh 1974), the rule to help one's in-group in the expectation that others in the group
Summary and Conclusion

4.1 Recent contributions using "bottom up" evolutionary models have relaxed the assumption of calculated self-interest in models of emergent social order. Rule-governed behavior implies that individual self-interest is empirically variable, depending on the rule that governs the host's behavior. Altruistic rules are those that have learned how to replicate by transferring reproductive chances from the present to the future and from one host to another. Reciprocal altruism typifies cooperation in an ongoing social exchange, while kin altruism may account for the importance of cultural markers and collective identity in successful social movement mobilizations.

4.2 Social movement theorists debate the relative explanatory power of the consequences and meaning of participation in the mobilization of a group. Consequences include the costs of contribution, value of the public goods, and access to selective incentives. The meaning of participation refers to the symbolic affirmation of shared social classifications and normative protocols that regulate interaction and structure social life. Yet expressive, symbolic behavior entails consequences, even if these are unintended by the actors. Computational models of emergent cooperation can restore the explanatory power of the unintended consequences of symbolic collective action, in which egoism is a variable instead of a constant. A dynamical theory of microsocial interaction, formalized using artificial agents, appears to be a viable and promising new direction for theoretical research on prudent cooperation, as well as solidarity mobilized around emergent norms and identities.

Notes

1 Of course, it makes an enormous difference whether the lessons are distilled from a series of one-shot games with "strangers" or from a series of repeated games with "neighbors." Hence, one might learn the lesson to "cooperate with reliable neighbors but never with strangers." This is not equivalent, however, to the "shadow of the future." The latter requires foresight and an indefinite endpoint (Axelrod 1984, p. 10), while the benefit of reciprocity with neighbors requires neither (1984, p. 22).

2 Boyd and Richerson (1990) have discovered a process of quasi-Lamarckian cultural evolution, characterized by individual learning and conformist cultural transmission, that appears to preserve group boundaries sufficiently to allow altruistic selection pressures to operate at the level of the group.

3 Care must be taken not to read purpose or intent into this characterization. By definition, rules instruct actions by individuals (and organizations) that have adopted the rule, and these actions have consequences that may alter the viability of the rule. Hence, like organizations, rules have interests, even though they are not purposive and do not pursue interests in the way that a purposive individual might. A rule-centered approach is thus incompatible with assumptions of purposive action and is strictly limited to applications where iteration, not intention, is the link between outcome and action.
A fox is more likely to catch a male peacock with a large and colorful tail, but a fancy tail also makes it easier for the male peacock to catch a "fox." Some readers may be bothered by the claim that rules can learn how to spread more effectively, as if they were intelligent viral entities that infect and manipulate humans for their own nefarious purposes. That might make interesting science fiction, but the point here is only to introduce a "rule's-eye" view of social life, so that our attention is directed to the evolutionary consequences of behavior, not for the agent, but for the rule that instructs the agent. Evolutionary game theorists often write about the fortunes of competing "strategies" using similar animating language.

References


HEISE, D. 1996. "Social order through macroactions: an interactionist approach." Panel on micro-


