

SOCIAL EVOLUTION FORUM

The Evolution of Human Cooperation

Herbert Gintis

Santa Fe Institute and Central European University

The study of human cooperation today is the current state of a continuous line of intellectual inheritance from Adam Smith and David Hume, through Thomas Malthus, Charles Darwin, and Emile Durkheim, and more recently the biologists William Hamilton and Robert Trivers. The work of these thinkers represents fundamental contributions to the nature of altruism and fundamental prosociality in humans.

But Adam Smith led in another direction, through David Ricardo, Francis Edgeworth, and Léon Walras to contemporary neoclassical economics, that at least until recently recognizes only self-regarding behavior.

The twentieth century was an era in which economists and policy makers in the market economies paid heed only to the second Adam Smith, seeing social policy as the goal of improving social welfare by devising material incentives that induce agents who care only for their personal welfare to contribute to the public good. Ethics, in this paradigm, plays no role in motivating human behavior.

Contemporary research on human cooperation yields several insights. First, interdisciplinary research currently yields results that obeyed traditional disciplinary research goals. While the twentieth century was an era of increased disciplinary specialization, the twenty-first may well turn out to be an era of trans-disciplinary synthesis. Its motto might be: when different disciplines focus on the same object of knowledge, their models must be consistent where they overlap. Second, by combining economic theory (game theory in particular) with the experimental techniques of social psychologists, economists, and other behavioral scientists, we can empirically test sophisticated models of human behavior in novel ways. The data derived from this unification allows us to deduce explicit principles of human behavior that cannot be unambiguously derived using more traditional sources of empirical data.

The power of the experimental approach is obvious: it allows deliberate experimental variation of parameters thought to affect behavior

Corresponding author's e-mail: hgintis@comcast.net

Citation: Gintis, Herbert. 2012. The Evolution of Human Cooperation (*with comment*). *Cliodynamics* 3: 172–190.

while holding constant other parameters. Using such techniques, for example, experimental economists have been able to estimate the effects of prices and costs on altruistic behaviors, giving precise empirical content to a common intuition that the greater the cost of generosity to the giver and the less the benefit to the recipient, the less generous is the typical experimental subject (Andreoni and Miller 2002). The resulting ‘supply function of generosity’ and other estimates made possible by experiments, are important in underlining the point that other-regarding behaviors do not contradict the fundamental ideas of rationality. They also are valuable in providing interdisciplinary bridges allowing the analytical power of economic and biological models, where other-regarding behavior is big news, to be enriched by the empirical knowledge of the other social sciences, where it is not.

Biological approaches have been misled by the apparent explanatory power of two theories: inclusive fitness and reciprocal altruism (Hamilton 1964, Williams 1966, Trivers 1971). These theories convinced a generation of researchers that, except for sacrifice on behalf of kin, what appears to be altruism—personal sacrifice on behalf of others—is really just long-run material self-regard. Ironically, human biology has settled in the same place as economic theory, though from a quite different starting point, and using a quite contrasting logic.

The experimental evidence supporting the ubiquity of non-self-regarding motives, however, casts doubt on both the economist’s and the biologist’s model of the self-regarding human actor. Many of these experiments have in common a nexus of behaviors that we term strong reciprocity. Strong reciprocity is a predisposition to cooperate with others, and to punish those who violate the norms of cooperation, at personal cost, even when it is implausible to expect that these costs will be recovered at a later date.

Strong reciprocity contributes not only to the analytical modeling of human behavior, but also to the larger task of creating a cogent political philosophy for the twenty-first century. While the writings of the great political philosophers of the past were usually both penetrating and nuanced on the subject, they have come to be interpreted simply as having either assumed that human beings are essentially self-regarding (e.g., Hobbes and Locke) or, at least under the right social order, entirely altruistic (e.g., Rousseau, Karl Marx). In fact, people are often neither. Strong reciprocators are conditional cooperators, behaving altruistically as long as others are doing so as well, and altruistic punishers, applying sanctions to those who behave unfairly according to the prevalent norms of cooperation.

People cooperate not only for self-regarding reasons but also because they are genuinely concerned about the well-being of others, try to uphold social norms, and value behaving ethically for its own sake. People punish those who free-ride on the cooperative behavior of others for the same reasons. Contributing to the success of a joint project for the benefit of one’s group,

even at a personal cost, evokes feelings of satisfaction and pride. Failing to do so is often a source of shame or guilt.

We came to have these ‘moral sentiments’ because our ancestors lived in environments, both natural and socially constructed, in which groups of individuals who were predisposed to cooperate and uphold ethical norms tended to survive and expand relative to other groups, thereby allowing these prosocial motivations to proliferate. The first proposition concerns proximate motivations for prosocial behavior, the second addresses the distant evolutionary origins and ongoing perpetuation of these cooperative dispositions.

The Roots of Human Cooperation

Our Late Pleistocene ancestors inhabited the large-mammal-rich African savannah and other environments in which cooperation in acquiring and sharing food yielded substantial benefits at relatively low cost. The slow human life-history with prolonged periods of dependency of the young also made the cooperation of non-kin in child rearing and provisioning beneficial. As a result, members of groups that sustained cooperative strategies for provisioning, child-rearing, sanctioning non-cooperators, defending against hostile neighbors, and truthfully sharing information had significant advantages over members of non-cooperative groups.

In the course of our subsequent history we created novel social and physical environments exhibiting similar, or even greater, benefits of cooperation, among them the division of labor coordinated by market exchange and respect of rights of property, systems of production characterized by increasing returns to scale (irrigated agriculture, modern industry, information systems with network externalities), and warfare. The impressive scope of these modern forms of cooperation was facilitated by the emergence in the last seven millennia of governments capable of enforcing property rights and providing incentives for the self-interested to contribute to common projects.

But prior to the emergence of governments and since, cooperation has been sustained also by motives that led some people to bear costs on behalf of others, contributing to common projects, punishing transgressors, and excluding outsiders. In the pages that follow we will advance three reasons why these altruistic social preferences supporting cooperation outcompeted unmitigated and amoral self-interest.

First, human groups have devised ways to protect their altruistic members from exploitation by the self-interested. Prominent among these is the public-spirited shunning, ostracism, and even execution of free-riders and others who violate cooperative norms. Other group activities protecting altruists from exploitation are leveling practices that limit hierarchy and inequality, including sharing food and information.

Second, humans adopted prolonged and elaborate systems of socialization that led individuals to internalize the norms that induce cooperation, so that contributing to common projects and punishing defectors became objectives in their own right rather than constraints on behavior. Together, the internalization of norms and the protection of the altruists from exploitation served to offset, at least partially, the competitive handicaps born by those who were motivated to bear personal costs to benefit others.

Third, between-group competition for resources and survival was and remains a decisive force in human evolutionary dynamics. Groups with many cooperative members tended to survive these challenges and to encroach upon the territory of the less cooperative groups, thereby both gaining reproductive advantages and proliferating cooperative behaviors through cultural transmission. The extraordinarily high stakes of intergroup competition and the contribution of altruistic cooperators to success in these contests meant that sacrifice on behalf of others, extending beyond the immediate family and even to virtual strangers, could proliferate. Modern-day nationalism is an example.

This is part of the reason why humans became extraordinarily group-minded, favoring cooperation with insiders and often expressing hostility toward outsiders. Boundary-maintenance supported within-group cooperation and exchange by limiting group size and within-group linguistic, normative and other forms of heterogeneity. Insider favoritism also sustained the between-group conflicts and differences in behavior that made group competition a powerful evolutionary force.

Why did humans, rather than chimps, lions, or meerkats, develop such exceptional forms of cooperation? The answer lies in the human cognitive, linguistic and what physical capacities that made us especially good at all of the above, and more. These capacities allow us to formulate general norms of social conduct, to erect social institutions regulating this conduct, to communicate these rules and they entail in particular situations, to alert others to their violation and to organize coalitions to punish the violators. No less important is the psychological capacity to internalize norms, to experience such social emotions as shame and moral outrage, and to base group membership on such non-kin characteristics as ethnicity and language, which in turn facilitates costly conflicts among groups. Equally essential was the developmental plasticity of humans and our long period of maturation, the latter initially a result of the particular feeding niche that early humans occupied. Also important is the unique human capacity to use projectile weapons, a consequence of which is to lower the cost of coordinated punishment of norm violators within a group, to reduce the costs of hunting large animals, with concomitant benefits accruing to groups with widely endorsed sharing norms, and to render intergroup conflicts more lethal. A result was to elevate group-level competition to a more powerful evolutionary

force. These exceptional aspects of human livelihoods and social interactions, we will show, have favored the evolution of an individual predisposition to cooperate with others and to punish those who exploit the cooperation of others. But more than individual-level motivation is involved. The regulation of social interactions by group-level institutions plays no less a role than altruistic individual motives in understanding how this cooperative species came to be. Institutions affect the rewards and penalties associated with particular behaviors, often favoring the adoption of cooperative actions over others, so that even the self-regarding are often induced to act in the interest of the group. Of course it will not do to posit these institutions a priori. Rather, the historical evidence indicates that they could have coevolved with other human traits in the relevant ancestral ecologies and social environments.

Cooperation and Competition

The tension between the relentless logic of self-interest and the ubiquity of collective action in real-world settings was eventually resolved by a series of experiments by psychologists and economists, most notably by Ernst Fehr and his colleagues (Fehr and Gächter 2000, Herrmann et al. 2008). The experiments confirmed that self-interest is indeed a powerful motive, but also that other motives are no less important. Even when substantial sums of money are at stake, many, perhaps most, experimental subjects were found to be fair-minded, generous toward those similarly inclined, and nasty toward those who violate these prosocial precepts. In light of these results, the evidence that the tragedy of the commons is sometimes averted and that collective action is a motor of human history is considerably less puzzling. The puzzle, instead, is how humans came to be like this.

Social Preferences and Social Dilemmas

Social preferences are a concern for the well-being of others and a desire to uphold ethical norms. By contrast with self-regarding preferences, which are based on states concerning oneself alone, we stress other-regarding and ethical preferences, the former defined as valuations based at least in part on states that occur to others. Social preferences include not only generosity toward others and a preference for 'fair' outcomes, but also what Thomas Hobbes called the desire for 'eminence,' Thorstein Veblen's 'pecuniary emulation' exemplified by a desire to 'keep up with the Joneses' (Veblen 1899), Charles Horton Cooley's 'looking-glass self' according to which our self-esteem is dependent in part upon what others think of us, so we attempt to favorably impress others as a means of raising our subjective self-esteem (Cooley 1902, Brennan and Pettit 2004), and Aristotle's character virtues, such as honesty

and courage, which are personal values that promote prosocial behavior (Aristotle 2002[350BC]).

Social preferences assume special importance in interactions termed social dilemmas, that is, interactions in which the uncoordinated actions of individuals result in an outcome that is Pareto inefficient, meaning that there exists some other feasible outcome such that at least one member could be better off while no member would be worse off. Examples of social dilemmas modeled by game theorists are the prisoner's dilemma, the public goods game, sometimes termed an n-person prisoner's dilemma, the so-called war of attrition and other so-called arms race interactions, the tragedy of the commons and the common pool resource game in which contributing to the common project takes the form of forgoing the overexploitation of a jointly utilized resource such as a fishery, water supply, or forest. We say a person free rides if he benefits from the contributions of other group members while himself contributing less or nothing at all.

Another-regarding player cares about not only his own payoff, but that of his partner as well. Such a player might reason as follows. "I feel sufficiently positive toward a partner who cooperates that I would rather cooperate even if by doing so I forgo the larger payoff (\$15) I could have had by defecting. If my partner defects, I of course prefer to defect as well, both to increase my earnings, and to decrease the earnings of a person who has behaved uncharitably toward me." If Bob and Alice reason in this manner, and if each believes the other is sufficiently likely to cooperate, both will cooperate. Thus, both mutual cooperate and mutual defect are equilibria in this new game, transformed from the old by augmenting the material payoffs with the players' concerns about one another.

Genes, Culture, Groups, and Institutions

We define culture as the ensemble of preferences and beliefs that are acquired by means other than genetic transmission. Culture is an evolutionary force in its own right, not simply an effect of the interaction of genes and natural environments.

An alternative but we think incorrect approach holds that while preferences and beliefs that are transmitted culturally may constitute the proximate causes of behavior, they in turn are entirely explained by the interaction of our genetic makeup and the natural environment. It is of course true that natural environments and genes affect the evolution of culture. But it is also true that culture affects the relative fitness of genetically transmitted behavioral traits. C. J. Lumsden and Edward O. Wilson (1981), Luigi Luca Cavalli-Sforza and Marcus Feldman (1981), Robert Boyd and Peter Richerson (1985), William Durham (1991), Richerson and Boyd (2004) and others have provided compelling instances of these cultural effects on genetic evolution.

Recognizing the intimate interactions between genes and culture in humans, Edward Wilson, Charles Lumsden, Robert Boyd, Peter Richerson, Luigi Luca Cavalli-Sforza and Marcus Feldman began working in the 1970s on the parallels between genetic and cultural evolution and their interactions, their work initiating the modeling of gene-culture coevolution, the second concept underpinning a plausible explanation of the origins and nature of distinctive cooperation among humans. According to gene-culture coevolution, human preferences and beliefs are the product of a dynamic whereby genes affect cultural evolution and culture affects genetic evolution, the two being tightly intertwined in the evolution of our species.

In our gene-culture coevolution model of group-structured populations, the process of differential replication affects the frequency of both individual traits, generosity toward fellow group members, say, and group traits, a system of consensus decision making or property rights. Though inspired by biological approaches, especially those of Cavalli-Sforza and Feldman (1981), Boyd and Richerson (1985), and Durham (1991), like these authors, we do not privilege biological explanation. This approach may be summarized as follows.

First, while genetic transmission of information plays a central role in our account, the genetics of non-pathological social behavior is for the most part unknown. Knowledge of the genetic basis of the human cognitive and linguistic capacities that make cooperation on a human scale possible has expanded greatly in recent years, but virtually nothing is known about genes that may be expressed in cooperative behavior, should these exist. No ‘gene for cooperation’ has been discovered. Nor is it likely that one will ever be found, for the idea of a one-to-one mapping between genes and behavior is unlikely given what is now known about gene expression, and is implausible in light of the complexity and cultural variation of cooperative behaviors. Thus, when we introduce genetic transmission in our models, our reasoning operates at the phenotypic level.

Second, as is conventional in all models of selection, relative payoffs, whether in terms of fitness, material reward, social standing or some other metric, influence the evolution of the population shares of various behavioral types, higher payoff behaviors tending to increase their frequency in a population. The resulting so called payoff monotonic dynamic is often implemented using ‘as if’ optimization algorithms, though in doing this we do not attribute conscious optimization to individuals. Nor do we conclude that the resulting outcomes are in any sense optimal. In general they are not. The aggregation of individually optimal choices is universally suboptimal, except under highly unrealistic conditions.

Individuals with higher payoffs may produce more copies of their behaviors in subsequent periods either through the contribution of their greater resources to differential reproductive success or because individuals disproportionately adopt the behaviors of the more successful members of

their group. The latter may occur voluntarily, as when youngsters copy stars, or coercively, as when dominant ethnic groups, classes, or nations impose their cultures on subjugated peoples. Of course, cultural transmission may also favor lower payoff behaviors (think of smoking or fast food).

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

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

Acknowledgment

These remarks draw upon Samuel Bowles and Herbert Gintis, *A Cooperative Press, Species: Human Reciprocity and its Evolution* (Princeton: Princeton University 2011).

References

- Alcock, John, *Animal Behavior: An Evolutionary Approach* (Sunderland, MA: Sinauer, 1993).
- Andreoni, James and John H. Miller, “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism,” *Econometrica* 70, 2 (2002): 737–753.
- Aristotle, *Nichomachean Ethics* (Newburyport, MA: Focus Publishing, 2002[350BC]).
- Axelrod, Robert and William D. Hamilton, “The Evolution of Cooperation,” *Science* 211 (1981): 1390–1396.

- Boehm, Christopher, "Conscience Origins, Sanctioning Selection, and the Evolution of Altruism in *Homo Sapiens*," 2007. Department of Anthropology, University of Southern California.
- Boesch, Christophe, "Cooperative Hunting in Wild Chimpanzees," *Animal Behavior* 48 (1984): 653–667.
- Boyd, Robert and Peter J. Richerson, *Culture and the Evolutionary Process* (Chicago: University of Chicago Press, 1985).
- Brennan, Geoffrey and Philip Pettit, *The Economy of Esteem: An Essay on Civil and Political Society* (New York: Oxford University Press, 2004).
- Cavalli-Sforza, Luigi Luca and Marcus W. Feldman, *Cultural Transmission and Evolution* (Princeton: Princeton University Press, 1981).
- Clutton-Brock, T.H., "Breeding Together: Kin Selection and Mutualism in Cooperative Vertebrates," *Science* 296 (2002): 69–72.
- Cooley, Charles Horton, *Human Nature and the Social Order* (New York: Charles Scribner's Sons, 1902).
- Dawkins, Richard, *The Selfish Gene* (Oxford: Oxford University Press, 1976).
- Durham, William H., *Coevolution: Genes, Culture, and Human Diversity* (Stanford: Stanford University Press, 1991).
- Eshel, Ilan and Marcus W. Feldman, "Initial Increase of New Mutants and Some Continuity Properties of ESS in Two Locus Systems," *American Naturalist* 124 (1984): 631–640.
- Eshel, Ilan, Marcus W. Feldman, and Aviv Bergman, "Long-Term Evolution, Short-Term Evolution, and Population Genetic Theory," *Journal of Theoretical Biology* 191 (1998): 391–396.
- Fehr, Ernst and Simon Gächter, "Cooperation and Punishment," *American Economic Review* 90, 4 (September 2000): 980–994.
- Frank, Steven A., *Foundations of Social Evolution* (Princeton: Princeton University Press, 1998).
- Gadagkar, Raghavendra, "On Testing the Role of Genetic Asymmetries Created by Haplodiploidy in the Evolution of Eusociality in the Hymenoptera," *Journal of Genetics* 70, 1 (April 1991): 1–31.
- Grafen, Alan, "Modeling in Behavioral Ecology," in John R. Krebs and Nicholas B. Davies (eds.) *Behavioral Ecology: An Evolutionary Approach* (Oxford: Blackwell Scientific Publications, 1991) pp.5–31.
- Hamilton, William D., "The Genetical Evolution of Social Behavior, I & II," *Journal of Theoretical Biology* 7 (1964): 1–16, 17–52.
- Hammerstein, Peter, "Darwinian Adaptation, Population Genetics and the Street car Theory of Evolution," *Journal of Mathematical Biology* 34 (1996): 511–532.
- Herrmann, Benedikt, Christian Thoni, and Simon Gächter, "Anti-Social Punishment across Societies," *Science* 319 (7 March 2008): 1362–1367.
- Hrdy, Sarah Blaffer, *Mothers and Others: The Evolutionary Origins of Mutual Understanding* (New York: Belknap, 2009).

- Kappeler, P. K. and Carel P. van Schaik, *Cooperation in Primates and Humans* (Berlin: Springer, 2006). Kummer, Hans, In *Quest of the Sacred Baboon* (Princeton: Princeton University Press, 1995).
- Lumsden, C. J. and Edward O. Wilson, *Genes, Mind, and Culture: The Coevolutionary Process* (Cambridge, MA: Harvard University Press, 1981).
- Manson, Joseph H. and Richard W. Wrangham, "Intergroup Aggression in Chimpanzees," *Current Anthropology* 32, 4 (1991): 369–390.
- Milo, R., "Evidence for Hominid Predation at Klasies River Mouth, South Africa and Its Implications for the Behaviour of Modern Humans," *Journal of Archaeological Science* 25(1998): 99–133.
- Richerson, Peter J. and Robert Boyd, *Not by Genes Alone* (Chicago: University of Chicago Press, 2004).
- Sigg, Hans and Jost Falett, "Experiments on Respect of Possession and Property in Hamadryas Baboons (*Papiohamadryas*)," *Animal Behaviour* 33 (1985): 978–984.
- Singer, R. and J.J. Wymer, *The Middle Stone Age at Klasies River Mouth in South Africa* (Chicago: University of Chicago Press, 1982).
- Stephens, W., C.M. McLinn, and J.R. Stevens, "Discounting and Reciprocity in an Iterated Prisoner's Dilemma," *Science* 298 (13 December 2002): 2216–2218.
- Trivers, Robert L., "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology* 46 (1971): 35–57.
- Veblen, Thorstein, *The Theory of the Leisure Class* (New York: Macmillan, 1899).
- Williams, G. C., *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton: Princeton University Press, 1966). Wilson, Edward O., *Sociobiology: The New Synthesis* (Cambridge, MA: Harvard University Press, 1975).

Commentaries

Michael Doebeli: *Cultural Evolution as an Epidemiological Birth-Death Process*

University of British Columbia

e-mail: doebeli@zoology.ubc.ca

Gintis (2011) presents an interesting and broad overview of the problem of human cooperation, but some issues need clarification. The most crucial one is the general notion of ‘human evolution.’ The question is, what is evolving? Abstractly, evolution is a consequence of a birth-death process in a population of replicators with incomplete heredity. In classic, organismic evolution, genes are the replicators, and evolution is the temporal dynamics of gene frequencies. As a consequence, by default the term ‘human evolution’ refers to the dynamics of genetic change in a group of organisms that are called humans in their recent (geological) history.

But I don’t think this is what Gintis has in mind when using the term ‘human evolution.’ Instead, what is meant is the dynamics of cultural change. It is very likely that most cultural change is not driven by genetic change in humans (as Gintis points out, no genes for culture are known). Some have argued that cultural change may drive genetic change, as e.g. when lactose tolerance evolved, perhaps as a consequence of agricultural changes (Laland et al. 2010). But consider the explosion of cultural change that occurred in the last 400 years, i.e., since the dawn of modern science: it seems very unlikely that this cultural change was driven by genetic change, or that it generated marked genetic changes in humans in the short time span of c. 15 generations.

Culture can be viewed as an extremely plastic human phenotype, of which probably only basic core components are genetically hardwired (such as the property of having a large and complex enough brain to exhibit cultural complexity). So what is evolving, if not human genes? What are the replicators that undergo a birth-death process leading to cultural change? The answer is that the cultural content itself is undergoing a birth-death process. This idea goes of course at least as far back as Richard Dawkins’ notion of memes (Dawkins 1976), but this is a loaded term that seems to necessitate the existence of cultural ‘units’ of reproduction. Instead, the notion of cultural content is more flexible in general, and can for example easily accommodate replication of continuously varying cultural properties (Henrich et al 2008), as well as complex cultural traits, such as cooperation, or components of technologies and ideologies.

A conceptual problem in thinking about the evolution of cultural content is that such content is always tightly linked to the humans carrying it. As a consequence, cultural evolution is most often modeled in terms of 'fitness' of humans carrying different types of cultural content. But this not only leads to potential confusion of human genetic evolution with non-genetic cultural evolution, but it is also problematic from a theoretical point of view, as it is ultimately not the humans, but the cultural content, that is undergoing the birth-death process leading to cultural evolution: some cultural variants thrive, while others vanish.

A useful analogy may be to think of the flora of microbes colonizing humans. Every human carries in and on them more cells of microbes than actual human cells (essentially microbes are present on all interfaces of the human body with the external worlds, such as the gut, etc). These microbes undergo birth-death processes and evolve (according to classical organismic evolution), and when studying their evolution, one would naturally concentrate on the microbes as evolving replicators, rather than on their human hosts. Of course, the environment provided by the human hosts plays a central role for the evolution of these microbes. Similarly, humans provide the environment in which cultural content evolves, but evolution occurs at the level of the birth-death process of cultural content being transmitted among and between human hosts.

As Gintis explains, the current thinking about the evolution of cooperation in humans is dominated by the view that more cooperative and cohesive human groups outcompete smaller and less cooperative ones. Besides the fact that this line of thought equates human fitness with the fitness of the cultural replicators they carry, there are a number of problems with this group selection approach to cultural evolution. For example, it a priori assumes diversity among groups, and hence does not address the origin of cultural diversity within a given group of human hosts. Conceptually, the cultural group selection approach is equivalent to saying that a certain type of gut microbe evolved because its human hosts 'did better' than the hosts of another variant of gut microbe. This sounds reasonable, but it is not the default approach one would take when studying e.g. the evolution of virulence in gut microbes, where it would be misleading to consider survival of the host as the only determinant. Similarly, certain types of culture may evolve despite having a detrimental effect on their human hosts (Boyd and Richerson 1985, Yeaman et al. 2011). With regard to cooperation, it is possible that a culture of cooperation spreads for reasons other than group benefits for the human hosts. For example, cultural variants that are more conducive to cooperation, such as moral religions (Norenzayan and Shariff 2008), may have spread for other reasons, e.g. because they were less reliant on local traditions and hence more transmissible, thereby enabling larger and more cohesive societies. Thus,

size and cohesiveness of human groups may be a consequence, rather than the cause of the cultural evolution of cooperation.

To disentangle these effects, it seems promising to develop a theory of cultural epidemiology, in which the cultural content itself undergoes a birth-death process based on colonizing human hosts, and in which incomplete transmission of cultural variants, as well as de novo variants occurring in human hosts, generate cultural evolution. This approach has already been proposed in the foundational work of Cavalli-Sforza and Feldman (1981), but deserves renewed attention, not only for understanding the cultural evolution of cooperation, but also for understanding the origin of cultural diversity (Hochberg 2004, Doebeli and Ispolatov 2010) and the evolution of ‘bad’ culture (Yeaman et al. 2011). Indeed, such an epidemiological approach might be useful for controlling cultural epidemics, such as fast food and terrorism (Lafferty et al. 2008).

References

- Boyd, R., Richerson, P.J., 1985. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.
- Cavalli-Sforza, L., Feldman, M., 1981. *Cultural Evolution*. Princeton University Press, Princeton.
- Dawkins, R., 1976. *The Selfish Gene*. Oxford University Press, Oxford, UK.
- Doebeli, M. and Ispolatov, I. 2010. A model for the evolutionary diversification of religions. *J. Theor. Biol.* 267: 676–684
- Gintis, H. 2011. *The Evolution of Human Cooperation*.
- Henrich, J., Boyd, R., Richerson, P.J., 2008. Five misunderstandings about cultural evolution. *Human Nat.* 19, 119–137.
- Hochberg, M.E., 2004. A theory of modern cultural shifts and meltdowns. *Proc. R. Soc. B* 271, 313–316.
- Lafferty, K.D., Smith, K.F., Madin, E.M.P., 2008. Natural security: a Darwinian approach to a dangerous world. In: Sagarin, T.T.R.D. (Ed.), *University of California Press, Ltd., London, England*, pp. 186–206.
- Laland, K., Odling-Smee, J. and Myles, S. 2010. How culture shaped the human genome: Bringing genetics and the human sciences together. *Nature Review Genetics* 11, 137–148.
- Norenzayan, A. and Shariff, A. F. 2008. The Origin and Evolution of Religious Prosociality. *Science* 322, 58–62.
- Yeaman, S., Schick, A. and Lehmann, L. 2011. Social network architecture and the maintenance of deleterious cultural traits. *J. R. Soc. Interface* doi:10.1098/rsif.2011.0555

Jessica C. Flack: *A Future for Social Evolution*

Center for Complexity and Collective Computation at the Wisconsin Institute for Discovery and Santa Fe Institute

e-mail: jflack@santafe.edu

A powerful framework in the study of social evolution has been game theory. This involves identifying a set of strategies, positing payoffs, assuming a game structure (e.g. tit for tat with repeated interactions in an n person setting), and calculating the evolutionary stable strategy, or distribution of strategies that should evolve in a population given certain stability assumptions. Empirical research in social evolution has been closely connected to this theoretical framework. One of two approaches is typically adopted. The experimenter asks of observational data whether there is evidence in nature for tit-for-tat or some other interaction rule. Studies of reciprocity in animal societies that dominated the 1980s and 1990s provide examples (reviewed in Schino and Aureli, 2009). A second, now more common, approach is to recapitulate the game in an experimental setting by asking a set of subjects to play, for example, the prisoner's dilemma or ultimatum game in a controlled context (e.g. Henrich et al. 2004). A goal is to determine whether the predicted distribution of strategies is recovered in the experiment. If it is not, then the goal shifts to studying how the basic assumptions of the game might be realistically modified in models to arrive at the observed distribution of strategies.

Increasingly this behavioral-mechanistic approach has been coupled to a physiological-mechanistic approach in which the neurophysiological states of subjects are studied using various imaging techniques and physiological measures, while subjects play a game. Experimenters measure the biological response to cheating, conflict and cooperation, the capacity for empathy, numerosity, and so forth. The goals are to determine whether the subjects are self-regarding or other regarding, what the biological basis for these dispositions might be, and whether the subjects have the computational capacity to use the cooperation mechanisms posited in game theoretic models.

This research program has been a powerful source of insight as is reviewed in Gintis's essay. However, there are some intriguing components and ideas missing from this approach. Whereas we now have compelling theories for the evolution of cooperation, we know little about the evolution and development of the social organization in which the cooperation takes place. The importance of social structure has been recognized to some extent. This recognition is reflected in the increasing prevalence of models that include elements of spatial structure and by the concerted development of modeling approaches, like evolutionary set theory (e.g. Tarnita, Antal, Ohtsuki, and Nowak, 2009),

that allow for the incorporation of social network structure into cooperation games in interesting, principled ways.

These mathematical advances are clearly steps in the right direction as they give us insight into how social structure influences the evolution of cooperation and other behaviors. Yet, we know little empirically about the diversity of network structures constituting social systems, or how emergent, functionally significant, aggregate social properties are encoded in these networks. We also know little empirically or formally about the timescales on which these social structures and their associated statistical properties change—hence the extent to which social structure can influence behavior through feedback. These are questions about the construction or development of social systems and, more generally, pattern formation and collective behavior (see Flack and Krakauer, 2011).

Research into the evolution of institutions might in principle seem relevant to these developmental questions. In practice this body of literature rarely addresses issues of construction of complex aggregate social traits. Instead ‘institution’ more often than not is a code word for counts or ratios of strategies in a given equilibrium distribution. Although simplifying the problem of institutions in this way makes models tractable and may be justifiable in some cases, it is not fully satisfactory. Many of the institutions observed in human and other social systems have a more complex character and this needs explaining.

When the models and statistics used to operationalize an institution are not just counts over strategies but require a more elaborate computation, and when the inputs are not simply individual traits (cooperate, defect, etc.) but network data, then we need to consider explicitly the mapping between behavioral strategies at the individual level and social organization (Flack and Krakauer 2011). How do these strategies get collectively combined by multiple individuals to produce aggregate social properties? Answering these questions requires study of the mesoscopic scale—the causal networks that specify how different combinations of strategies produce different institutions. Once we can describe how an aggregate social property is produced, we can study how the social process producing it might have evolved. The parameters in our game theoretic models will also become more empirically grounded.

The argument put forward in this essay should sound familiar to readers who know the history of the debate in evolutionary theory surrounding the genotype-phenotype map (see Laubichler and Maienschein, 2009). Two long-standing assumptions in population genetics are that the g-p map, as it is called, is simple and that the timescale on which the environment changes is slow enough compared to evolutionary (or behavioral) change that it can be treated as static (the adiabatic assumption).

We now know that the first assumption is fundamentally wrong for most organisms – the gene activation patterns underlying phenotypic traits are

modulated by complex regulatory machinery that itself evolves—the work of Eric Davidson and colleagues on echinoderm development stands as an excellent example (e.g. Davidson, 2010). And, the second assumption, which if correct would justify studying development and evolution independently, is problematic in any system in which organisms can modify environmental variables and by modifying them change the selection pressures to which they are subject, as in ecological (e.g. Odling-Smee, Laland, and Feldman, 2003) and social niche construction (e.g. Flack, Girvan, de Waal and Krakauer, 2006). The consequences of softening these assumptions are now being explored by researchers who study the evolution of development. With these advances we are seeing the beginnings of an evolutionary theory that can account for the origins and diversity of complex forms, as well as for causes of gene change.

The role of developmental dynamics has long been debated in the larger evolutionary theory, and so research programs emphasizing developmental mechanisms have been pursued in parallel to population genetics. Hence the current merger of development and evolution was in a way poised to happen as the data to give momentum to the merger have been (partly) collected. In social evolution, on the other hand, there has only been the game theoretic-population genetics trajectory with no sizable quantitative research program on the developmental dynamics of social organization running in parallel (one exception is the work on social insect societies).

To catch up we need to collect behavioral time series and social network data from model social systems. Once we have these data, we can empirically derive the natural scales of the system by extracting the strategies and decision-making rules that individuals use during social interactions, and using these to build the causal networks that specify how these rules when collectively implemented produce aggregate social properties (see DeDeo, Krakauer, and Flack, 2010; Flack and Krakauer, 2011). With quantitative, mechanistic descriptions of the microscopic, mesoscopic and macroscopic scales and their feedbacks we will be in a position to theorize about the evolution of the development of social systems and answer the question of why different types of societies arise.

References

- E H Davidson. Emerging properties of animal gene regulatory networks. *Nature*, 468(7326):911-920 (2010).
- S DeDeo, D C Krakauer and J C Flack. Inductive game theory and the dynamics of animal conflict. *PLoS computational biology*, 6(5):e1000782 (2010).
- J C Flack, M Girvan, F B M de Waal and D C Krakauer. Policing stabilizes construction of social niches in primates. *Nature*, 439(7075):426-429 (2006).

- J C Flack and D C Krakauer, Challenges for complexity measures: A perspective from social dynamics and collective social computation. *Chaos*, 21: 037108 (2011).
- J Henrich, R Boyd, S Bowles, C Camerer, E Fehr and H Gintis. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small- - Scale Societies*. Oxford University Press (2004).
- M D Laubichler and J Maienschein, editors. *From Embryology to Evo-Devo: A History of Developmental Evolution (Dibner Institute Studies in the History of Science and Technology)*. The MIT Press (2009).
- J Odling - Smee, K N Laland and M W Feldman. *Niche Construction: The Neglected Process in Evolution (MPB-37) (Monographs in Population Biology, 37.)*. Princeton University Press (2003).
- G Schino and F Aureli. The relative roles of kinship and reciprocity in explaining primate altruism. *Ecology Letters*, 12: 1-6. (2009).

Reply

Herbert Gintis: *Gene-Culture Coevolution and the Roots of Human Cooperation*

Santa Fe Institute and Central European University

e-mail: hgintis@comcast.net

I welcome the insightful remarks of Michael Doebeli and Jessica Flack. I agree wholeheartedly with Professor Flack's stress on the complexity of human evolution, and especially the need for more work on spatial structure and, even more important, social network structure. I welcome her work on inductive game theory and complexity measures as contributions in these areas. I want to stress that my coworkers and I have worked on the premise that a 'behavioral-mechanistic' analysis of human behavior gives strong insights not otherwise available, but is surely not sufficient to capture the panoply of regularities in human social behavior. In particular, we also need ethnographic and historical studies, as well as plausible models of the evolution of the behaviors we describe in laboratory and field (Bowles and Gintis, 2011, Greif, 2006, Aoki, 2010, Bowles, 2004).

Professor Doebeli's remarks concerning cultural evolution are correct and useful, but he does not fully represent my argument. "The term 'human evolution'," he asserts, "refers to the dynamics of genetic change in a group of

organisms that are called humans. Instead [what Gintis means] is the dynamics of cultural change.” In fact, I argued in favor of a model of human evolution in which genes and culture are *causally interrelated*, genetic evolution in humans being as much a product of cultural evolution as vice-versa. Of course, I am not thinking of cultural evolution over short periods of time, such as years or centuries, but rather over the long period of human evolution as hunter-gatherers in the Pleistocene.

Consider, for instance, the evolution of the physiology of speech and facial communication in humans. The increased social importance of communication in human society rewarded genetic changes that facilitate speech. Regions in the motor cortex expanded in early humans to facilitate speech production. Concurrently, nerves and muscles to the mouth, larynx and tongue became more numerous to handle the complexities of speech.

In short, humans have evolved a highly specialized and very costly complex of genetically rooted physiological characteristics that both presuppose and facilitate sophisticated aural and visual communication. This example is quite a dramatic and concrete illustration of the intimate interaction of genes and culture in the evolution of our species (Gintis, 2011)

Similarly, hunter-gatherer groups developed *social norms* to govern their social interactions. The viability of social norms, however, depends on the group punishing norm-violators, and we know that this form of punishment is often severe (Boehm, 2000, Wiessner, 2005). This punishment thus rendered more fit individuals predisposed genetically to conform to social norms. This gave rise to the psychological propensity of individuals to *internalize norms*, so that individuals do not conform out of fear of punishment, but because they recognize the moral basis of the norms and they are predisposed to follow them voluntarily, simply because it is the right thing to do (Simon, 1990, Gintis, 2003, Bowles and Gintis, 2011).

Another example is the predisposition to cooperate in social dilemmas, such as the of public goods game described by Ernst Fehr and his colleagues (Fehr and Gächter, 2000, Herrmann et al., 2008). In these and other experiments (described and analyzed in Gintis, Bowles, Boyd and Fehr, 2005 and Gintis, 2009), when subjects are allowed to punish other subjects, many choose to punish, at a cost to themselves, free-riders who have contributed little or nothing to the collective effort, even under conditions where there is no possibility of the punishers being monetarily compensated for their actions. We term this behavior *strong reciprocity*, and we argue that it too is the product of gene-culture coevolution (Gintis, 2000).

References

Aoki, Masahiko, *Corporations in Evolving Diversity: Cognition, Governance, and Institutions* (Oxford: Oxford University Press, 2010).

- Boehm, Christopher, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Cambridge, MA: Harvard University Press, 2000).
- Bowles, Samuel, *Microeconomics: Behavior, Institutions, and Evolution* (Princeton: Princeton University Press, 2004).
- _____ and Herbert Gintis, *A Cooperative Species: Human Reciprocity and its Evolution* (Princeton: Princeton University Press, 2011).
- Fehr, Ernst and Simon Gächter, "Cooperation and Punishment," *American Economic Review* 90,4 (September 2000):980-994.
- Gintis, Herbert, "Strong Reciprocity and Human Sociality," *Journal of Theoretical Biology* 206 (2000):169-179.
- _____, "The Hitchhiker's Guide to Altruism: Genes, Culture, and the Internalization of Norms," *Journal of Theoretical Biology* 220,4 (2003):407-418.
- _____, *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences* (Princeton: Princeton University Press, 2009).
- _____, "Gene-culture Coevolution and the Nature of Human Sociality," *Proceedings of the Royal Society B* 366 (2011):878-888.
- _____, Samuel Bowles, Robert Boyd, and Ernst Fehr, *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life* (Cambridge, MA: MIT Press, 2005).
- Greif, Avner, *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade (Political Economy of Institutions and Decisions* (Cambridge University Press, 2006).
- Herrmann, Benedikt, Christian Thoni, and Simon Gächter, "Anti-Social Punishment across Societies," *Science* 319 (7 March 2008):1362-1367.
- Simon, Herbert, "A Mechanism for Social Selection and Successful Altruism," *Science* 250 (1990):1665-1668.
- Wiessner, Polly, "Norm Enforcement among the Ju/'hoansi Bushmen: A Case of Strong Reciprocity?," *Human Nature* 16,2 (June 2005):115-145.