

**THEORETICAL FOUNDATIONS OF SOCIAL NORMS:
The Co-Evolution of Social Norms, Intrinsic Motivation, Markets,
and the Institutions of Complex Societies**

Jean Ensminger
HSS 228-77
California Institute of Technology
Pasadena, CA 91125
jensming@hss.caltech.edu

Joseph Henrich
Departments of Economics and Psychology
University of British Columbia
Vancouver, B.C. VT6 1Z4
Canada
Joseph.henrich@gmail.com

Draft Date: April, 2008

Early Draft: Please do not quote, cite, or circulate

Chapter 2 (from):

Experimenting with Social Norms:

Fairness and Punishment in Cross-Cultural Perspective

Jean Ensminger and Joseph Henrich, Editors

Chapter 2

THEORETICAL FOUNDATIONS OF SOCIAL NORMS

The Co-evolution of Social Norms, Intrinsic Motivation, Markets, and the Institutions of Complex Societies

Jean Ensminger and Joseph Henrich

Though it seems surprising on its face, classical scholars long ago proposed a positive relationship between developed market economies and pro-social or fair-minded motivations.¹ One of the first and best known scholars to write in this vein was none other than Adam Smith (1997 [1759]), whose position is quite consistent with the findings of our cross-cultural sample (Chapter 4). Even before Smith, Montesquieu (1900:316-317 [1749]) was explicit on this subject:

Commerce has everywhere diffused a knowledge of the manners of all nations: these are compared one with another, and from this comparison arise the greatest advantages. Commerce is a cure for the destructive prejudices, for it is almost a general rule, that wherever we find agreeable manners, there commerce flourishes; and that wherever there is commerce, there we meet with agreeable manners.

Let us not be astonished, then if our manners are now less savage than formerly.

Commerce has everywhere diffused a knowledge of the manners of all nations: these are compared one with another, and from this comparison arise the greatest advantages....

The spirit of trade produces in the mind of a man a certain sense of exact justice, opposite, on the one hand, to robbery, and on the other to those moral virtues which forbid our always adhering rigidly to the rules of private interest, and suffer us to neglect this for the advantage of others.

Montesquieu anticipates the specifics of the trade-off between narrow economic self-interest and other regarding behavior that we are examining in the experimental games

¹ See Hirschman (1982) for an overview.

reported herein. Further, we examine these games in the societal evolutionary context he is describing. We are comparing societies with more and less commerce, and he anticipates the direction of our findings, with more fair-minded behavior in more market integrated societies.

While prescient, the early classical scholars did not have quantitative evidence, let alone controlled comparative experimental measures and game theory to back up their keen intuitions and observations. Further, predicting or observing a relationship is not the same thing as positing mechanisms of origin and maintenance. In this chapter we outline the broader theoretical significance of our empirical project for understanding the evolution of social norms and the institutions with which they pave the way to economic growth. We begin developing a theory of social norms by drawing on converging lines of work arising from both evolutionary and economic frameworks. In developing this foundation for social norms we consider evidence from sources as diverse as developmental psychology, neuroscience, and anthropological ethnography. Building on this, we discuss how pro-social or group-beneficial norms spread, and how this relates to the emergence of more centralized and formal institutions, which undergird the subsequent expansion of large-scale, complex societies.

The Evolution of Norms

As an empirical phenomenon, anthropologists, sociologists, and others from across the social sciences have long noted the existence of social norms, often evoking norms to explain behavioral similarities within groups, or to differentiate groups (Bendor and Swistak 2001, Bicchieri 2006). Despite a broad literature and the empirical importance of social norms, the concept has until relatively recently lacked sufficient micro-level theoretical foundations to be taken seriously by researchers in economic and evolutionary sciences. The concept of social norm that we analyze herein has the

following qualities—a norm represents a pattern of behavior with some degree of cultural consensus, and evidence for this is both in the statistical prevalence of the behavior and the expectation that sanctioning of a norm violation would be socially appropriate. As a result of these qualities, norms are self-enforcing and often highly decentralized in their origin and maintenance. That is, they are enacted by individuals in everyday life. We use the term institution somewhat differently, while accepting that social norms and institutions exist along the same continuum, and share many qualities. Following Greif (2005), we accept that institutions must also be self-enforcing in equilibrium. In the sense in which we use the term, institutions are both more centralized and more formalized. They need not have been the result of coordinated design, but they are more likely to be than are decentralized social norms. They are also more likely to be enforced by designated third party agents, but this does not necessarily mean that they are in any way less legitimate, nor that individual moral commitment to abide by them is any less than is the case for social norms. We return to institutions as a special case below.

Without at least plausible answers to key theoretical questions regarding how social norms (and institutions) emerge, why and how individuals' adopt norms that violate their self-interest, what "adoption" means, how individuals' decisions interact with group patterns, how norms spread across groups, and how and why norms have changed over human history, social norms cannot be readily incorporated into either economic or evolutionary frameworks. In recent decades, however, approaches arising from both evolutionary biology and economics are converging on a "ground-up" or "first principles" explanation of social norms. There are now plausible theoretical answers to the above questions, and in some cases there is an overabundance of plausible answers.

Learning Social Norms

The first step in approaching norms from an evolutionary perspective is to use the logic of natural selection, aided by formal evolutionary modeling, to consider what kinds of learning strategies or heuristics individuals, be they humans or guppies, would evolve to adapt to uncertain, novel and/or changing environments (Boyd and Richerson 1985, Aoki, Wakano and Feldman 2005, Henrich and Gil-White 2001). These environments include social interactions in which information about costs and benefits of alternative behavior is costly, inherently uncertain or incomplete, or impossible to acquire. In such environments these learning strategies, which include heuristics that integrate rules glossed as “copy the most successful” (prestige-biased transmission) and “copy the majority” (conformist transmission), can outcompete learning strategies that rely solely on direct evaluation of perceived costs and benefits (e.g., Boyd and Richerson 1988; Henrich and Boyd 1998).² This means that the direct evaluation of the costs and benefits of alternative actions is but one component in a suite of adaptive learning tools or heuristics that permit individuals to adapt to diverse informational environments. In this suite, cost-benefit evaluation operates best in stable, well-structured (clear choices), high information environments.

This line of evolutionary theorizing has led to many predictions about the kinds of learning rules or biases that individuals should use in calibrating their behavior to local environments and about resulting emergent population-level patterns. A substantial amount of evidence from both laboratory experiments and field observations documents these predictions in humans (Henrich and Gil-White 2001; Kameda and Nakanishi 2002; Coultas 2004; Kohler, Van Buskirk et al. 2004; McElreath, Lubell et al. 2005; Henrich

² This approach has long recognized and explored the influences of “representational content” (the content of what is socially learned) on transmission as well as the impact and importance of inferential processes in understanding imitative learning. See Boyd and Richerson (1985) for an early treatment.

and McElreath 2006; Henrich and Henrich 2007: Chapter 2; Mesoudi in press) as well as other species (Lachlan, Crooks et al. 1998; Day, MacDonald et al. 2001).

The second step in building an evolutionary theory of norms is to explore what happens when these evolved learning strategies are placed in formal game theoretical models that permit different kinds of social interactions with other individuals who are also trying to adaptively learn. The results of many such efforts indicate that stable equilibrium states at the population level often emerge from the individually adaptive learning strategies of interacting individuals. This often yields a reliable behavioral pattern for a given group. Often such analyses show that there are—in fact—many different stable equilibrium states for a given population. Which state emerges in a particular population is highly dependent on the initial conditions and details of particular shocks along the evolutionary path (*cf.* Greif 2005). This situation begins to look like social norms, at least in that these empirically-grounded learning heuristics give rise to stable statistical regularities and behavioral prescriptions that vary across social groups.

The situation gets even more interesting when one considers social interactions in which an individual has the possibility of exploiting others, such as in a public goods situation or other cooperative dilemma. One might expect such strategies to always dominate, but evolutionary learning models show that if individuals are using the above-mentioned adaptive learning heuristics (prestige-biased and conformist transmission), then pro-social or group-beneficial norms of cooperation, exchange, and respect for others' property can be maintained, even when the possibility of repeated interaction is low, or the future time horizons of individuals differ.³ In sustaining such group-beneficial equilibria, these approaches provide solutions to the well-known second and third order

³ It is an underappreciated theoretical finding that if populations consist of individuals that vary in the expected length of their interactions, the successful reciprocity strategies are much less cooperative than tit-for-tat and aren't "nice" (they don't cooperate initially). See Boyd (1992).

free-rider problems by (1) permitting, along with cooperation, the capacity to learn strategies that punish non-pro-social behavior (Boyd and Richerson 1992; Henrich and Boyd 2001), or (2) linking via reputation any norm-violating behavior to other types of social interaction, where there could be costs (Panchanathan and Boyd 2004; see below).

One problem with these “pro-social solutions” is that the same mechanisms can stabilize any equally costly behavior, *independent* of its benefit to the group. Mechanisms that can stabilize behaviors in the group can hurt both the individuals themselves and the average payoffs of the group. This robust emergent aspect of the mathematical analyses is problematic if one is looking exclusively to solve the dilemma of large-scale cooperation or similar moral dilemmas. However, for building a theory of norms, this oddity is a feature, not a bug, because it better reflects the world we observe. Now these emergent phenomena are looking even more like the norms recorded by social scientists: we have behavioral regularities stabilized by either direct punishment or sanctioning (through reputational damage) that can be, but need not be, pro-social or group beneficial. Anthropologists and sociologists have recorded a wide range of regular stable behavioral patterns, though often idiosyncratic, that appear maladaptive (Edgerton 1992, Diamond 1997, 2005), but yet appear enforced by concerns about reputational damage or informal punishment. Conspicuous examples include Chinese foot binding (Mackie 1996), supercision of adolescent males, cranial deformation in infants, ritual cannibalism sustaining prion diseases, and female infibulation (Knight and Ensminger 1998). It is interesting that the same mechanisms that can theoretically account for the normative nature of the above practices are the mechanisms that can explain cooperation in large groups, including management of property rights (Ostrom 1990), and economic exchanges among strangers in the absence of formal third party enforcement (Ensminger 1992).

Internalization of Norms

There is at least one major empirically recognized aspect of norms missing from our “ground-up” development: the “emotional.” The desire to adhere to some norms, often termed “moral norms” by psychologists (Haidt 2001; Haidt 2007), appears to be internally motivated in some fashion. Once internalized, norms become ends, goals, or values in themselves. Economists would say these individuals have put the performance of the norm into their objective function (Greif 2005). By noting this, we mean to emphasize that internalizing a norm does not make one a mindless “norm-executer.” People have plenty of other internalized goals that compete constantly with internalized norms.

Why would natural selection build an organism to internalize social norms? To approach this issue, the evolutionary analysis focuses on the costs of information processing and making errors (violating norms), the temptation of not seizing immediate rewards, and the developmental circumstance of the adapting child. Natural selection could favor internalizing norms as ends in themselves if this saved on information processing costs and/or the associated errors. If norms are reliable and frequently sanctioned, an individual might be better off to “just do it” in most situations (motivated by having norm adherence as a goal in itself) rather than performing an analysis for each slightly different situation that would involve considering the probability of being judged in violation of a norm, and the resulting long-term reputation damage. Moreover, suppose that each time one runs a mental calculation like the one just described, actors occasionally make errors that lead to more sanctions. Internal motivation could help avoid unnecessary calculations, and avoid the errors of an inherently noisy environment. This, however, does not imply that other internal motivations (such as narrow economic self interest) don’t sometimes prevail (Ensminger and Knight 1997, Knight and Ensminger 1998, Gintis 2004, Gintis 2007).

Alternatively, internally motivated adherence to norms may be natural selection's way of psychologically overcoming the pull of an immediate reward versus long-term costs. If people overweight immediate rewards compared to future rewards or rewards amortized over years, as economists argue they do (Samuelson 1937, Laibson 1997, Frederick, Loewenstein, and O'Donoghue 2002), then internal motivation might provide that extra push to forgo the short-term gains in favor of long-term payoffs. Natural selection or adaptive learning could fix this problem by adjusting our temporal discounting, but since many non-human animals have the same problem, we assume that such mechanisms either provide other benefits that have allowed their preservation, or it was biologically too difficult to build a better estimator.

When organisms live in rapidly changing or variable environments, allowing proximate motivations to be internalized by learning can help the organism make adaptive decisions. To understand this, consider the acquisition of social norms from the perspective of children. Since norms vary among groups, but are generally locally stable, children can adapt to the local environment by first rapidly adopting and partially internalizing the local norms, thereby avoiding sanctions, and then begin evaluating the costs and benefits of norm violations once they have mastered the norm. The problem of errors in cost-benefit calculations is particularly acute for children since they lack the information possessed by adults to evaluate the consequences of violation.

Developmental work suggests that by adolescence, and often long before, children have solid knowledge and some internalization of many local norms. Children first acquire local norms in contextually specific circumstances and then rigidly apply them, while gradually expanding the norm's domain of applicability as observation and experience informs. Later, having mastered the knowledge of a norm, but perhaps not yet having *fully* internalized it, children start figuring out how and when they can violate it for their own benefit.

One further perspective on how norms are learned and internalized deserves attention. The creation of explicit rules, institutions, and laws, as we discuss below for more complex societies, helps to coordinate people's beliefs about what norm is applicable to a particular circumstance. Formalized rules also convey information about the incentives for adherence (sanctions); by influencing compliance, they may also influence the acquiring and internalization by learners trying to figure out what the local norm is (Cooter, Feldman, and Feldman 2007). Once internalized, guilt (in addition to fear of sanctioning) already provides a motivation to follow the norm, assuming that one knows what it is. The natural implication of this for our project is that the more highly developed these systems are, the better should be people's predictions about the behavior of others, thus reducing variance in compliance and creating a virtuous cycle (literally, in the case of pro-social norms) of norm internalization and compliance. In another paper, Bohnet and Cooter (2003) summarize this perspective as follows:

Our results bear on a disagreement among legal scholars about how law causes social change, such as the decrease in racial discrimination in the U.S. Some scholars argue that law has an "expressive function" that changes behavior, whereas other scholars deny that law has much influence on such phenomena as racial discrimination. Our research suggests that law changes society by changing beliefs more than preferences. According to this logic, laws imposing desegregation in the southern states may have changed behavior by changing beliefs about the willingness of others to integrate. Whereas beliefs changed relatively quickly, preferences probably changed relatively slowly.

McAdams (2000) has also stressed the unique impact of law as a focal point for belief about how others will behave. As he puts it (2000:1651), "law works by what it says in addition to what it does." The emphasis here is upon coordination via focal points. In this case, even sanction-less laws can serve this function by signaling information. The concept that social norms change behavior by changing expectations regarding the behavior of others is also supported by the literature on the power of

framing effects to change behavior in economic experiments (see Camerer 2003 and Ross and Ward 1996).

Experimental Support

This theoretical account of social norms permits us to unify an account for a diverse range of empirical findings. Here we will briefly summarize some particularly relevant experimental findings showing that (1) children can acquire individually costly pro-social behaviors via observational learning and automatically adopt sanctioning tendencies towards violators, (2) the pro-social behavior measured in economic experiments in the U.S. develops gradually over the first two or three decades of life, (3) behavior in experiments depends on population-specific contextual cues that aid individuals to map games to their local norms, and (4) subject's pro-social behavior (cooperation, fairness, and punishment) in experiments in the West activates the same reward circuits as does receiving cash payments. For a longer discussion of the evidence see Henrich and Henrich (2007).

A long line of research with children shows that context-specific, pro-social norms for altruistic behavior toward strangers can be acquired by observing others behaving altruistically in the same context. When exposed to either adult or peer models who donated more tokens to poor children, subjects between five and eleven years of age also donate more tokens themselves. These effects were (1) not influenced by whether the child was alone when making the donation decision, and (2) were not ephemeral, as they endured in re-tests months later in which no observations of models occurred. The altruistic effects extended to somewhat similar tasks, but did not generalize beyond this. When placed in the role of model, these child subjects continued to donate higher amounts and also spontaneously scolded younger children who did not donate sufficiently; this scolding behavior had not been modeled in their original exposure as

subjects. A comparison of ways to induce altruistic behavior towards strangers reveals that among children observation learning is the best method. Preaching and verbal instruction had little effect unless accompanied by costly donating actions by an adult or peer model. Incentive schemes can induce the pro-social behavior as long as the incentives persist, but they do not create internalized motivation (in fact, they seem to inhibit internalization).⁴

In experimental games similar to those used in this volume, we know from other research that children's offers begin to approach adult levels by around age 12, then there is a drop or plateau during adolescence, when they knowingly break the fairness norm. In the ultimatum game, in which the responder can punish a low offer, giving lower offers is rationally self-interested because adolescents' willingness to punish has not risen sufficiently high to discourage unfairness, as it will in adulthood (Ensminger and Cook, this volume; also see Henrich and Henrich 2007: Chapter 8). Experiments done over the developmental trajectory among American and European subjects show that the pro-social fairness norms associated with experimental play do not fully reach the adult plateau until the mid-twenties. The behavior of university students in experimental games continues to change from first to fourth year, and in some experiments such as the dictator game, the differences between students and fully socialized adults is dramatic (Ensminger and Cook, this volume, Henrich and Henrich 2007).

This approach to norms gives us a means to anticipate and theorize about how different contextual cues in laboratory experiments, which do not directly impact the payoff structure of the game, can influence game play. Subjects arrive at experiments equipped with norms, which include contextually specific beliefs (including expectations of others' behavior) and internalized motivations, and then face a novel situation. They

⁴ We discuss the crowing out of intrinsic motivation with respect to our data in Chapter 4.

have to figure out how to behave, in part, by figuring out what—if any—of their norms apply in this situation. Since most experimental games involve both money and anonymity, players from some societies with norms that apply to such contexts will be influenced by both their norm-related beliefs (what they think others will do) and their internalized motivations (such as what's "fair").⁵ These norm-effects come through when experimentalists setting up the identical game (with the same payoff structure) vary the language or some other aspect of the process. Ross and Ward (1996), for example, used identical versions of a public goods game (a cooperative dilemma) and labeled one the "Wall Street" game and the other the "community" game. They found that university students in the Wall Street game contributed a bit less than those in the community game (see also Pillutla and Chen 1999, Hoffman, McCabe et al. 1994). In a different kind of variation, a variety of experiments show the positive effects on pro-social behavior of communicating before playing. Communicating, even if this is not explicitly about the game, can help players anticipate the norm-driven behavior of other players.

Hayashi et al. (1999) show that simple framing differences strongly affect rates of cooperation in an otherwise identical two-person prisoner's dilemma, *and* that these effects depend on whether one is from Japan or the U.S. This finding fits with observed differences between the U.S. and Japan in non-experimental contexts and shows that the same contextual manipulations yield different effects in different populations. This occurs in a manner that would seem linked to the differences between the larger societies.

⁵ See Ensminger (2004) for an example of a public goods experiment that did not use any specific framing, but which was immediately labeled the "*harambee*" game by local participants in reference to the local institution of public goods provision of which it reminded them. Behavior in the game tracked the locally accepted behavior pattern: the wealthy made higher contributions to the public good, as local social norms dictate for an exactly analogous institution. In contrast, in the dictator game, which triggered no particular local institutional reference, the wealthy did not contribute statistically more than others.

We emphasize here that we do not think that “norms” are the only thing influencing play in experimental games. Aspects of the games, such as the material costs and benefits, the possibility of cultivating a reputation, and the prospect of repeated interaction ought to influence game behavior in predictable ways, consistent with evolutionary considerations or rational choice, independent of the norms. Additionally, some seemingly contextual effects—that do not influence the actual payoff structure—likely affect game play by influencing players’ perception of possibilities for reputation formation or repetition (e.g., Haley and Fessler 2005) .

Finally, work in neuroscience and neuro-economics has recently contributed to this line of theorizing by showing that behaving in the manner demanded by local norms, by cooperating, contributing, or punishing in locally recommended or prescribed ways, activates the brain’s rewards or reward anticipation circuits in the same manner as when obtaining a direct cash payment (Fehr and Camerer 2007). Complying with local norms by cooperating, contributing, or punishing “feels good” to brains in the same way that personally getting money does. Cooperating and getting money (from the cooperation) feels better than just getting the same amount of money without this association (Rilling, Sanfey et al. 2004). Punishing by really hurting defectors (physically or monetarily) also activates these reward circuits (Sanfey, Rilling et al. 2003) more than punishing symbolically (de Quervain, Fischbacher et al. 2004). Giving money to charity activates the same reward circuits as receiving money. Activations of the brain’s reward circuitry in these experiments predict behavioral outcomes.

Convergence with Economics

Evolutionary and economic approaches have begun a convergence on both a unified conception of learning and a theoretical foundation for social norms. By considering both the impact of incomplete information and uncertainty on rational

decision-making, combined with the simplicity of ecologically rational heuristics for dealing with complex situations, bounded rationality is providing an understanding of human social behavior paralleling that derived via evolutionary theory (Young 1998; Gintis 2007). Economists have shown that copying successful people or copying the majority, are—under particular conditions—quite rational strategies, as well as fitness-maximizing (Ellison and Fudenberg 1993; Spencer and Huston 1993; Ellison and Fudenberg 1995; Schlag 1998; Schlag 1999). Such strategies are rational when information is costly to acquire or process, or when information about the costs and benefits of alternative behavior are noisy (error ridden)—circumstances common to many real life decisions. Economists have also led the way in exploring bounded learning strategies based on personal trial and error learning (Fudenberg and Levine 1998).

Theoretical models that place individuals deploying these learning strategies in social interaction yield stable behavior patterns that look like the emergence of norms. This is not surprising in some cases, since the underlying learning heuristics used by evolutionary theorists and economists are identical, but in other cases it re-affirms the phenomena that learning plus social interaction can yield a wide variety of fairly stable outcomes (Weibull 1995; Young 1998). This multiplicity of stable outcomes is even a feature of classic game theoretical models that assume perfect and free information and processing power—a finding enshrined in the folk theorem.⁶

Recently in economics, the emerging focus in game theory has led to some of the best experimental work on learning in social interactions. The experiments confirm

⁶ The social norms that arise from learning and social interaction are at least dynamically stable in the vicinity of the equilibrium behavior (when most people are sticking to the norm). Since dynamically stable equilibria are always also stable in classical game theoretic models in which actors are fully rational, selfish and omniscient norms can persist even when some members get better informed and begin to approach full rationality.

that, at least in laboratory experiments, learners do appear to be using learning heuristics both like “copy the successful” and “copy the majority,” as well as experienced-based learning rules. Many of these findings converge with findings from psychology, using quite different experimental tools, as well as those from field observations.

In returning to the inspirations of Adam Smith and others, economists are beginning to explore and theorize about the internalization of norms, or endogenous preference formation, and have specifically considered the effect of markets (Bowles 1998). Rather than waving off the question of where people’s preferences come from, an increasing number of economists are examining the possibility that preferences emerge in part from interaction with the local institutional environment (Greif 2005). People’s motivations or preferences partially calibrate to performing the local equilibrium (Alesina and Fuchs-Schuendein 2007).

Economists have also entered the debate about processes of internalization of norms; this is also of particular interest to them because it speaks to the origin of preferences and makes norms self-enforcing. As Greif (2005:37) puts it, “internalized norms are socially constructed behavioral standards that have been incorporated into one’s superego (conscience), thereby influencing behavior by becoming part of one’s preferences.” Even textbooks in microeconomics are now taking the need to consider internalized motivations seriously (Bowles 2004).

Scholars in the law and economics tradition have also taken an interest in the internalization of social norms. Cooter (1996a) has repeatedly emphasized the unique quality of social norms as distinct from formal laws in that they are socialized within individuals, and thus become self-enforcing. Cooter does not dispute the fact that we often have reputational reasons for abiding by social norms, and individual members of societies may also impose external sanctions on norm violators in the form of gossip and

even ostracism. But as Cooter notes, this is missing one of the unique features of social norms, which is that we internalize them and that includes a sense of guilt should we violate a norm that we believe to be morally just (1996b:1724, 1996b:152). Similarly, as we saw in our discussion of neuro-economics, people also get “mental rewards” from seeing a norm violator get punished. Thus social norms internally propel us (at least some of us) to engage in behavior that is not necessarily in our narrow economic self-interest. This also means that we may abide by norms even when no one is looking, as for instance, in an anonymous economic experiment. Indeed, Cooter operationalizes a definition of the degree of internalization of a norm as the price one is prepared to pay to conform to it (2000:1581), which dovetails nicely with economic experiments that are designed to measure exactly that dimension of behavior.

While economics does not generally concern itself with ultimate explanations of motivations or preferences, some efforts within economics have paralleled and even inspired some of the evolutionary explanations discussed above based on error management. If norm violations result in sanctions or miss-coordination and certain situations may tempt one to violate a norm, individuals should develop internalized motivations (preferences) that allow individuals to avoid norm violations that will cost them in the long run (*cf.* Frank 1988).

Multiple Mechanisms of Norm Stabilization

Recent models of the evolution of cooperative norms internalized by at least some fraction of the group challenge us also to account for the assumed linkage to costly punishment of norm violators. Earlier models, both those based on rational choice and evolutionary learning, solved the free-rider problem in larger-scale cooperative dilemmas by permitting the transmission of punishing strategies. The higher-order free-rider problem created by the cost of punishing was solved by a meta-punishing strategy

(punish all non-punishers) or because learners were using a combination of learning strategies under uncertainty including some conformist learning (Axelrod 1984; Boyd and Richerson 1992; Henrich and Boyd 2001). This is the kind of strategy that can use the diffusion of information about violations to create diffuse punishment (those who can most cheaply perform the function do so), or to allow those most able to do the punishing. These approaches will also work in the absence of reliable reputational information if direct observers of norm violation perform the sanctioning (Henrich and Henrich 2007).

More recent evolutionary models of cooperative norms illustrate two alternative routes to stabilize norms, which don't involve costly punishment. Both approaches link cooperating in larger-scale situations (doing something costly to contribute to a group benefit) to a two-person interaction in a different social context. Panchanathan and Boyd (2004) show how individual costly norms can be stabilized by attaching players' reputations in a dyadic helping game to their reputations in a larger-scale individually-costly interaction. If an individual fails to cooperate in the larger interaction he gets a "bad reputation" and other individuals can withdraw their help from him in the two-person game *without* getting a bad reputation themselves. Otherwise, individuals who refuse to help those with good reputations in the two-person helping game get a bad reputation. There is no free-rider problem here because individuals "sanction" by withdrawing help, and thereby not paying the cost of delivering help. Straightforward narrow self-interest drives sanctioning, so there is no costly punishment, though there remains a kind of sanctioning. Among pastoral nomadic Orma cattle herders, Ensminger has observed this phenomenon. Nomads who feel aggrieved by the behavior of their herding partners may secretly gather their donkeys in preparation for moving, an activity that may take several days. Once prepared, they decamp at dawn to the frustration of the deserted

family, now left to their own resources because with no donkeys collected they cannot join them.

Henrich (n.d.) has shown similar findings experimentally by attaching a larger-scale cooperative interaction to a two-person negative reciprocity game (a “no stealing game”). In the two-person game, individuals who profit by hurting (stealing from) those with good reputations, get a bad reputation. Profiting from someone with a bad reputation does not hurt the actor’s reputation. By attaching reputation in this two-person game to a larger-scale interaction, those not performing individually costly actions (defectors) get a bad reputation and people can steal from them with impunity (no reputational damage). Again, in the case of large-scale cooperative interactions, defectors get sanctioned, but there is no diffuse costly punishment. The sanction is motivated by narrow self-interest, and sanctioning is not costly. Greif (2005:75-77, 442) provides an illustration of a related mechanism from the economic history of Maghribi traders. Traders could withhold punishment of agents who cheated merchants who failed to punish an agent. For a similar formulation in rational choice theory, see Aoki (2001).

Both of these alternative routes can yield stable social norms. However, as we will argue in Chapter 4, unlike the models involving direct costly punishment, they require well-functioning, high-fidelity, reputational systems. Therefore, while they could potentially explain some norms in some groups, including pro-social norms, they cannot support cooperation and fair exchange with those outside the reputational system, for example, interaction in larger populations with strangers and infrequent exchange partners.

Cooter (1997) and Fehr and Gintis (2007) have argued that internalized norms make people more willing to directly punish norm violators at a cost because they

believe the normative behavior is the right thing to do. We will return to this debate in light of our findings and consider the circumstances in which the claim holds empirically.

The converging evolutionary and economic approaches described thus far suggest that social norms, whether they arise from rational decision-making mechanisms or some evolved adaptive learning processes, will cause different norms to emerge in different places and contexts, leading members of different groups to calibrate their beliefs and internal motivations differently. If our experiments are measuring norms for dealing with strangers in monetary exchanges or interaction, the above theorizing leads us to expect three features. First, we expect variation among populations both in our measures of fairness and punishment for interactions with anonymous others (*cf.* Henrich et al. 2004). We contrast these predictions with approaches that take university student behavior in experimental games to be a robust universal feature of our species and the product of cognitive adaptation based on repeated interactions (Hoffman, McCabe et al. 1998; Nowak, Page et al. 2000). Second, since norms based on costly punishment will cause both intrinsic motivation toward fairness and a willingness to punish unfairness to co-evolve, we expect a positive relationship across groups in our measures of fairness and punishment (see Henrich et al. 2006). Third, this variation among populations ought to be bounded by theoretically possible equilibria. It is not the case that “anything” goes cross-culturally, but only that multiple equilibria are possible. We don’t expect societies with normative expectations that favor, for example, giving all of one’s windfall to an anonymous other. Societies tolerate much maladaptive behavior, especially in non-competitive situations such as low population pressure, but societies are often subject to cultural group selection that favors those that have managed to develop pro-social norms to solve large-scale collective action problems.

The Spread of Pro-social Norms

And Institutions across Populations

Once a combination of expectations, motivations, and beliefs converges in a group to create a social norm, we have a somewhat sticky situation. Once different societies, or groups, converge on different social norms, due to the aforementioned path dependence and historical specificity of the process, is that the end? This problem is made even more poignant because, as we have argued above, models of norms all indicate that many different norms can be stable, and most of these are not pro-social or group-beneficial. So, what we have is a bunch of different groups, each with different norms, only a few of which involve any pro-sociality among strangers. Is there any way to select among these social norms? This is the classic problem of *equilibrium selection*, a serious problem that emerges in both dynamic evolutionary approaches and those rooted in narrow concepts of rationality (Henrich 2006).

Young (1998) suggests one approach based on the stochasticity inherent in any interaction. Different stable equilibria (norms) are more or less susceptible to the repeated shocks created by this stochasticity, meaning that in the long-run some equilibria will be substantially more common than others. Over the long run—and sometimes this is a really long run—some social norms will be more likely to collapse, and cause the group to evolve to a different norm. In his analysis, Young (1998:146-147) is focusing primarily upon small players, whose many thousands of decentralized interactions create the patterns we observe.

Up to this point, we have been largely concerned with these decentralized actions by small players, and the learning techniques they employ to bring innovation into their local social norms. But we also have many instances of important change in social norms and institutions that is driven by big players, as Young (1998) also readily

notes. Within the institutional literature there has been much discussion of whether such change is driven by those seeking distributional advantage (Knight 1992), or efficiency improvements (Williamson 1985, North and Thomas 1973). Representing an example of the latter, is Harsanyi and Selton (1988), who posit forward-looking individuals who recognize the long-term payoffs available at stable cooperative equilibria, assume others are similarly sensible, and choose the pro-social state.

It is here that we believe it is helpful to make a distinction between the end of the spectrum represented by social norms, and that represented by institutions. In societies lacking any form of centralized polity, including most hunter-gatherer societies and many horticultural societies, including those represented in our sample, there are few if any formal institutions beyond social norms, which are generally initiated by small players and most probably in an uncoordinated fashion. The techniques used in such local innovation may include the learning and mimicking strategies that we have discussed. They may also be motivated by cost/benefit calculations of distributional or efficiency advantage.⁷ Where they likely differ most from “institutional” change in larger-scale and more complex societies, is in the level of coordination and consensus achieved prior to experimentation.

Moving along the continuum of socio-political complexity from hunter-gatherer and horticultural societies, we find societies with councils of elders, formal chieftains,

⁷ Ensminger and Knight (1997) and Knight and Ensminger (1998) provide examples of how these processes might play out in a decentralized society. In these cases the relative bargaining power of the individuals leading the innovation in social norms can make a difference. In the case of a change in norm preference inspired by Islam and leading to a preference for dowry rather than bridewealth, the authors argue that it was the relative bargaining of those few individuals initiating the change that made the difference. The authors argue that people in the society were less likely to sanction norm violators of greater wealth and status, thus affording such people a greater ability to innovate social norms (either for their personal gain or to suit their values), without being punished. People often do not wish to forgo future interactions with such people, and will weigh such action differently than they would if the violators were less pivotal people in the social network. Similar efforts by elites led to changes in clan exogamy and failed attempts (due to poor coordination, not punishment) to eradicate female circumcision. We note that Mackie's (1998) example of Chinese pledge societies that were involved in foot binding eradication also involved elites as prime movers.

and the polities of archaic and modern states. At this end of the spectrum, coordinated and centralized innovation of institutions may occur regularly. It is here, we argue, that one is most likely to encounter coordinated manipulation both for strategic distributional ends and for the benefit of collective action in the interest of the common good.

However, in making this distinction between social norms and institutions, we do not wish to deny the fact that in many situations even within the most centralized societies, it is the strength of local norms—not formal institutions or laws—that govern behavior. In the developing world with which we are concerned, the sanctioning power attached to breaches of local social norms may be considerable, while the reach of state institutions may be limited and poorly internalized.

Economic history provides numerous examples of top down innovation in institutional structures that have led to considerable economic prosperity (see particularly North and Weingast 1989, and Greif 2005). Among ethnographic examples, Ensminger and Knight (1997) have examined the process by which sedentary elites in a livestock herding society managed to engineer a gradual change in the property rights institution from one of common grazing to more restrictive access that eased environmental degradation and favored sedentary elites (for additional examples of property rights change see also Ensminger 1997b and Ostrom 1990). This is an example of an institution driven by both distributional and efficiency motivations.

One can reconcile both the existence of maladaptive behavior and selection for more productive social norms and institutions with the introduction of cultural group selection acting upon social groups that have arrived at different stable equilibria. Where population pressure and inter-group competition is absent, anything goes and we see many examples of maladaptation. Where inter-group competition is strong, we may see favoritism given to the spread of individuals and/or practices from groups stabilized at equilibria that favor success in inter-group competition, which includes norms related

to larger-scale cooperation, in-group harmony and fairness, and out-group competitiveness. In humans, competition between groups can take the forms of warfare, demographic success, or more subtle forms in which individuals learn decisions and strategies by observing higher-payoff individuals—some of whom are from groups with higher payoff norms (Fehr and Fischbacher 2003; Henrich 2004). This between-group learning can lead to a differential flow of decisions, strategies, and even preferences from higher to lower payoff groups (Henrich and Boyd 2001; Boyd and Richerson 2002), or to differential migration (Boyd and Richerson 1990), favoring the spread of the high-payoff norms. “Migration” from groups with one norm to groups with another norm can occur without physical movement, such as when someone changes religions.⁸

There is both laboratory and field evidence supporting cultural group selection. In the laboratory, Guerek et al. (2006) permitted players to choose between one of two different “institutions.” In the first institution, players could contribute money to a group project. All contributions were increased and divided equally among all players, regardless of their contributions. Previous experiments (Fehr and Fischbacher 2003) have established that when this interaction is repeated, mean contributions to the public good drop to near zero (a “non-cooperative equilibrium”). The other “punishing institution” is very similar, except now, after players have contributed, they can pay to punish (reduce the payoff) or reward other players. When this interaction is played repeatedly (Fehr and Gächter 2000), a substantial fraction of players punish low

⁸ Some might question whether all of these examples of equilibrium selection process should be categorized under cultural group selection. There are two key requirements to assessing whether something is *cultural group selection*. The first is that the relevant behavior be influenced by learning—that’s the “cultural” part. The second and less well understood requirement arises from how evolutionary processes can be partitioned into their variance components such that contribution to change created with within-group processes are distinguished from between-group contributions (associated with group differences). Thus, when it’s the differences in payoffs based on variation between groups that influence the evolutionary processes, it is appropriate to call it group selection according to Price’s original formulation (Price’s 1970, 1972; McElreath and Boyd 2007). Since norms within groups are stable, the within-group component is zero on average, and there is no conflict between within and between group forces. The usual concerns regarding the evolution of altruism via group selection do not apply (Henrich and Henrich 2007).

contributors, causing mean contributions to rise and stabilize near full cooperation (a “cooperative equilibrium”) among samples of Western undergraduates (Herrmann, Thoni et al. 2007). Both laboratory institutions were run concurrently for 30 interactions and players could, initially and after each subsequent interaction (after seeing others’ payoffs), choose their institution for the next interaction.

Guerek et al.’s (2006) principle findings can be summarized simply: initially most players picked the institution *without* sanctioning possibilities. But in response to being exploited by free-riders, cooperators in the non-punishing institution began to reduce their contributions, which began to drive the total contributions toward zero. Meanwhile, punishers in the sanctioning institution started driving contributions up, despite the personal cost of punishing. After a few interactions, players from the non-sanctioning institution—presumably seeing the higher payoffs of those choosing the sanctioning institution—increasingly switched institutions. Notably, despite the incoming flow of migrants from the non-sanctioning institution, the mean contributions in the sanctioning institution consistently increased and/or held stable near full cooperation. In fact, most incoming migrants, consistent with local norms in their new setting, increased their contributions during their first interaction in the sanctioning institution, and a majority administered some punishment.

What does this tell us about equilibrium selection? First, the student subjects’ expectations of others’ behavior did not permit them to foresee the final outcome and select the higher payoff institution on the first interaction. This occurred despite the simplicity of these experiments compared to the real world. Most selected the lower payoff institution out of distaste for the possibility of being punished. Second, despite the stochasticity of human decisions, neither institution drifted into another equilibrium. There were only 30 interactions in this game, so one could hardly expect stochastic processes to begin selecting equilibria. But, that is the point: observational learning

across institutions occurs much faster than stochastic processes (at least in this one experiment). What did happen is that once players from the lower-payoff institution observed the higher payoffs of the other institution, they wanted to adopt the practices of that institution, or the decisions/strategies of those other players. In this experiment, players could only do that by "migrating" to the other institution. These migrants, however, do not appear to be merely uninformed payoff-maximizers who needed to adjust their beliefs about others. A majority of migrants into the sanctioning institution not only cooperated in their new institution, but also punished (cooperating but not punishing is the payoff-maximizing strategy).

The real world provides numerous case examples of this and other forms of cultural group selection (Richerson and Boyd 2000; Bowles 2004; Richerson and Boyd 2005; Henrich 2007). In particular, Ensminger (1997a) provides an example of this process of changing social norms in a discussion of conversion to Islam. Islam is known to have spread particularly fast along the long-distance trade routes of Africa and South Asia. Conversion was a necessary condition for access to trading networks. But while people may have originally been attracted to the economic benefits of conversion, there is no doubt that they fully internalized the norms and belief systems of the religion, which worked to everyone's advantage as they forged honest trading partnerships in which contracts were honored, thus vastly expanding the Islamic trade routes and simultaneously lowering transactions costs for all.

In the next section, we argue that a group's social norms influence how markets and other larger-scale institutions operate, especially in big populations. Social groups with norms that facilitate trust, fairness, cooperation, and direct costly punishment for norm violations in contexts involving strangers or low frequency interactions will succeed in establishing institutions that achieve higher payoffs or are more competitive in inter-group competition. These differences in group payoffs or competitiveness will favor the

spread of the associated norms via one of the above equilibrium selection mechanisms. Examples in the literature point to some combination of differential demographic expansion, conquest, assimilation, differential immigration into successful populations, or payoff-biased learning and adoption of norms from more successful groups.

Markets, Complex Societies,

And the Norms that make them Work

There is remarkably little consensus on two different, though inherently related, questions that researchers from diverse disciplines have confronted. The first is traditionally an anthropological question (Johnson and Earle 2000): until about 10,000 years ago our ancestors lived in relatively small, nomadic or semi-sedentary, populations dependent on hunting and gathering for subsistence. In the ensuing millennia, sedentary agriculture arose in several places, then larger towns emerged, then cities with monumental architecture, metal, militaries and writing, and the rest is literally history. The scale and intensity of human cooperation and exchange expanded dramatically during what, in human evolutionary terms, was a relatively short time. How can a species adapted to living in relatively small foraging groups, often dependent mostly on kin relationships, expand the sphere of cooperation and exchange to such an extent in such a short time, and why did this seem to occur at different rates on different continents (Diamond 1997; Richerson and Boyd 1998; 2000)? The second question, one of critical humanitarian importance, has long animated economic thinkers even before Adam Smith took up the challenge: Why are some societies rich and others poor?

Respected scholars from a variety of disciplines have considerably diverse perspectives on these most critical questions. Hypotheses concerning the role of

climate and geography have a long history in explanations of differential development, and have recently been further elaborated by Diamond (1997), Sachs (2001, 2003), Landes (1998), and Acemoglu, Johnson, and Robinson (2001). Such explanations place the emphasis more upon the fortuitousness of being in the right place at the right time, rather than upon motivations, beliefs, and expectations that create society-wide social norms, and that in turn generate institutions facilitating economic growth.

In contrast, scholars such as Plateau (2000), Sowell (1998), Putnam (1993) and North (2005), as well as much earlier modernization theorists of the fifties and sixties (for example, McClelland 1961), have stressed the role of cultural practices, beliefs, and values in the process of economic development and wealth generation. The theoretical framework for the evolution of social norms that we have laid out above is consistent with some of the arguments made in this literature, but we believe that better specification of the mechanisms and substantiation with experimental and ethnographic evidence alters and greatly clarifies this line of theorizing.

One way in which our data bear on the literature relating beliefs and values to economic growth has to do with religion. Religion plays a role in both our theory of the evolution of societal complexity and in our data. Religion in small-scale societies often lacks omniscient moralizing gods that actively reward and punish proper behavior. The idea of a lovely and desirable afterlife is relatively rare. Anthropologists have long noted a positive relationship between societal complexity and the presence of high moralizing gods. Societies with high moralizing gods have spread dramatically, as have their beliefs, in the last two thousand years. Henrich (2007) has argued that religion has often spread by one of the mechanisms discussed above, cultural group selection. Our analyses in this volume indicate, independent of a wide-variety of other factors like income, wealth, market integration and settlement size, a positive relationship between practicing a world religion (Catholicism, Islam, and Protestantism—Evangelical and non-

Evangelical) and pro-social behavior towards anonymous others. Notably, the relationship that we find does not support Weber (1958), who is the best known proponent of the relationship between religion and commercial development. Contra-Weber, it is not the Protestants who drive this relationship, but the Catholics and those who practice indigenous religions. Of course religions are major vehicles of socialization, so the processes at work may be similar to those described above for social norms, though psychologists have recently shown that unconsciously priming religious adults with “god” yields substantially higher dictator game offers (Sharif and Nor xxxx).

Another tradition close to what we propose here has focused upon the consequences of institutional change as the driver of economic growth and development. In particular, Locke (1980), Smith (1778), and Hayek (1960) all emphasized the importance of property rights for stimulating economic development. More recently, Rosenberg and Birdzell (1986), de Soto (2003), North (1981 and 1990), Greif (2005), and Acemoglu et al. (2002), have considered the evolution of efficient institutional forms, including the development of property rights, as crucial for understanding differential global development. Acemoglu et al. (2002) have argued that the reversal of the economic fortunes of many regions after 1500 resulted from an interaction of European institutional forms with the spread of the industrial revolution out of Britain.

These diverse perspectives on the differences in the fortunes of nations are not necessarily mutually exclusive, and can be linked by examining the first question, the evolution of societal complexity since the beginning of the Holocene. Consider the origins of agriculture. Certain geographical regions had better combinations of natural endowments in the form of more easily-domesticated grains and animals and climates better suited to cultivation, population expansion, and long-term settlement. But as empirical data suggest, sustaining relatively stable, harmonious, large, sedentary settlements built around fixed resources (land, water, and pasture) requires at least

some agreement on rudimentary property rights, access agreements, storage systems, and defensive collaborations, before agriculture can take-off. It seems likely that at least in some regions the technical know-how of domestication may have been poised for a takeoff long before the incentives existed for individuals to go to the bother of planting and nurturing a crop through harvest (North 1981). Who would take on such a task absent some confidence in their ability to access the fruits of their labor? Working out cooperative property rights, or defense capabilities, may have taken far longer than evolving the technical routines and know-how of stable productive agricultural packages. Merely learning to live together long enough in large settlements with significant numbers of non-kin to sow and reap a harvest may have required the development of norms of fairness that facilitated submission to an authority. Initially, such authority systems may have been nothing more than diffuse consensus among elders with the legitimacy to settle the disputes that arose from co-habitation and the additional strains of dividing surplus. The development of more elaborate religious beliefs with professional specialists may have facilitated such submission to authority, as well as making the potential consequences of punishment more biting. Ethnographically, many groups, including some of the societies we have studied, lack such religious professions, and don't linger in large settlements, as disagreements and dispute cause dissolution; they don't submit to authorities above the head of household (Henrich 1997; Henrich 2000; Johnson and Earle 2000; Johnson 2003, Murphy and Murphy 1986).

Consistent with this, the archaeological record indicates that the knowledge of domestication existed long before its deployment on a broad scale. Implementation of more intensive farming practices may have had to wait for the social and institutional systems to catch up (Richerson, Boyd et al. 2001). The problem is not dissimilar from what we observe in the modern world today. Clearly there are many societies aware of and in possession of the technological capacities of higher performing economic

systems that they wish to emulate, but for which they have not yet mastered the institutional capacity to do so.

Expanding the division of labor and trade, sustaining effective political and judicial decision-making, controlling corruption, maintaining public safety and effective policing require the evolution of an immense variety of institutions for dealing pro-socially to solve the inherent collective action problems that arise in interacting with strangers or low-frequency partners. Solutions to these problems require institutional foundations. Groups, or their norms, tend to spread because they allow larger populations to achieve higher payoffs in the competitive ecologies of other groups. Such norms can, and did, spread, probably by some version of the equilibrium selection mechanisms discussed above. But these processes are often slow, stochastic, reversible, and often lead to the eventual collapse of specific societies (Diamond 2005).

Part of this process involves the emergence of the norms and institutions that create the environment to permit market exchanges among strangers or low-frequency partners. Certain forms of market exchange and their associated norms for dealing with strangers co-evolve. Groups with norms that allow profitable exchange among strangers will achieve, on-average, greater success than those lacking such norms. Such market norms can spread as individuals seek to join the group, impressed by its success, and adopt the norms (just as they did in Guerek et al.'s experiment), or as members of other groups adopt the norms of the successful group and shift the balance of their own group toward new norms. Successful groups may proliferate demographically, as did early Christians and Muslims, or the successful group may apply its excess wealth to military or commercial expansion (as occurred in Europe during the industrial revolution). At the same time, markets and related institutions developed elsewhere can spread into groups already possessing sufficiently appropriate or applicable norms (even if non-optimal), though markets won't spread readily into groups lacking the appropriate norms without

the operation of one of the aforementioned equilibrium selection mechanisms. One implication of this reasoning is that more market integrated societies ought to, *ceteris paribus*, have more pro-social norms for dealing with strangers, anonymous others, or low frequency exchange partners.

Independent of markets, larger populations suffer from the daily need to interact with, cooperate with, not steal from, and exchange with other individuals, well beyond one's circle of friends and family. Sustaining larger populations will require similar pro-social norms, though if the population is not too big, reputation can sustain the norms in the absence of direct costly punishment (Panchanathan and Boyd 2004). Thus, we predict a relationship between settlement size and punishment behavior in our experiments. Large populations require costly punishment to sustain pro-social norms, while small populations can use either reputation or costly punishment. Dixit (2004:76) has formalized this proposition and argues that it is systems of intermediate size that are at the greatest disadvantage, operating as it were between the advantages of face to face direct reputation based systems of punishment, and those at the other end of the institutional spectrum with strong mechanisms of third party enforcement. Indeed, there is some indication that our results are consistent with this proposition, though our sample size is too small to actually test it.

Conclusions

We believe that the data from this project shed some light on the mechanisms by which social norms and institutions evolved through human history. Our experiments were designed to test for the existence and strength of social norms of fairness and punishment that sustain exchange relationships, maintain harmony, and facilitate collective action and exchange among unrelated individuals not engaged in a long-term relationship. Our findings from the first phase of this project (Henrich et al. 2004), now

replicated and extended in this second phase with new experiments and a new sample, indicate a positive relationship between pro-social behavior and market integration, as well as a positive relationship between settlement size and the punishment of norm violations.

The sample of societies from which we draw the data for this project is highly unusual in that it runs the gamut from almost pure hunter-gatherer (absent most traces of modern development and material possessions), through numerous horticultural societies (almost equally remote from modern markets), to cash cropping farmers, urban African workers, and small town residents in rural American. As such, it offers a rare opportunity to address some core questions about the relationship through time between social norms governing fair-minded behavior, institutional complexity, and level of market exchange.

In this chapter we have laid out a theoretical framework, together with experimental and empirical evidence to support it, in which we attempt to explain the mechanisms by which pro-social behavior can arise in the simplest of human societies. From there we describe the means by which we believe pro-social behavior can proliferate in a virtuous cycle of reinforcement as institutional structures become more complex. We recognize that multiple equilibria persist, and many, if not most, are maladaptive. But as societies come in competition with one another, those where individuals and groups have drifted or designed their way into equilibria that stabilize higher levels of pro-social behavior have selective advantages over their neighbors. Quite commonly, aspects of successful systems are recognized and copied by neighbors, but never with exactly the same result, given path dependence. The level of economic exchange that is supported by complex societies today has its foundation in the earliest human societies, where some pro-social behavior is also maintained. But as institutions become more complex, they have the capacity to specify and enforce higher

levels of pro-social behavior that can allow societies to realize the economic benefits of overcoming huge collective action problems.

References Cited

- Acemoglu, D., S. Johnson, et al. (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation." The American Economic Review **91**(5): 1369-1401.
- Acemoglu, D., S. Johnson, et al. (2002). "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." The Quarterly Journal of Economics **117**(4): 1231-1294.
- Ainslie, G. (2005). "Precis of Breakdown of Will." Behavioral and Brain Sciences. **28**(5):635-.
- Alesina, and N. Fuchs-Schuendein (2007) "Good Bye Lenin (or not?) – The Effect of Communism on People's Preferences." American Economic Review. **97**:1507-1528.
- Aoki, K., J. Y. Wakano, and M. W. Feldman (2005) "The Emergence of Social Learning in a Temporally Changing Environment: A Theoretical Model." Current Anthropology **46**(2):334-340.
- Aoki, M. (2001). Modeling Aggregate Behavior and Fluctuations in Economics: Stochastic Views of Interacting Agents. New York: Cambridge University Press.
- Axelrod, R. (1984). The Evolution of Cooperation. New York, Basic Books, Inc.
- Bendor, J. and P. Swistak (2001). "The Evolution of Norms." American Journal of Sociology **107**(2): 1493-1545.
- Bicchieri, C. (2006). The Grammar of Society: The Nature and Dynamics of Social Norms. New York: Cambridge University Press.
- Bohnet, I. and R. Cooter (2003). "Expressive Law: Framing or Equilibrium Selection?" John F. Kennedy School of Government, Harvard University Faculty Research Working Papers Series RWP03-046.
- Bowles, S. (1998). "Endogenous Preferences: The Cultural Consequences of Markets and other Economic Institutions." Journal Economic Literature **36**: 75-111.
- Bowles, S. (2004). Microeconomics: Behavior, Institutions, and Evolution. Princeton, Princeton University Press.
- Boyd, R. (1992). The Evolution of Reciprocity when Conditions Vary. Coalitions in Humans and Other Animals. F. Dewaal and A. H. Harcourt. New York, Oxford University Press: 473-492.
- Boyd, R. and P. J. Richerson (1985). Culture and the Evolutionary Process. Chicago, IL, University of Chicago Press.
- Boyd, R. and P. J. Richerson (1988). An evolutionary model of social learning: the effects of spatial and temporal variation. Social Learning: Psychological and Biological Perspectives. T. R. Zentall and B. G. Galef. Hillsdale, NJ, Lawrence Erlbaum Associates: 29-48.
- Boyd, R. and P. J. Richerson (1990). "Group Selection Among Alternative Evolutionarily Stable Strategies." Journal of Theoretical Biology **145**: 331-342.
- Boyd, R. and P. J. Richerson (1992). "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups." Ethology & Sociobiology **13**(3): 171-195.
- Boyd, R. and P. J. Richerson (2002). "Group Beneficial Norms Can Spread Rapidly in a Structured Population." Journal of Theoretical Biology **215**: 287–296.
- Camerer, C. F. (2003) Behavioral Game Theory: Experiments in Strategic Interaction. New York: Russell Sage Foundation.
- Cooter, R. (1996a). "Symposium: Decentralized Law for a Complex Economy: The Structural Approach to Adjudicating the New Law Merchant." University of Pennsylvania Law Review **144**: 1643-1696.

- Cooter, R. (1996b). "The Theory of Market Modernization of Law." International Review of Law and Economics **16**: 141-172.
- Cooter, R. (1997). "Symposium: Normative Failure Theory of Law." Cornell Law Review **82**(947): 955-957.
- Cooter, R. (2000). "Do Good Laws Make Good Citizens? An Economic Analysis of Internalized Norms." Virginia Law Review **86**(8): 1577-1601.
- Cooter, R. and M. A. Eisenberg (2001). "Symposium: Norms and Corporate Law: Fairness, Character, and Efficiency in Firms." University of Pennsylvania Law Review **149**(6): 1717-1733.
- Cooter, R., M. Feldman, et al. (2007). "The Misperception of Norms: The Psychology of Bias and the Economics of Equilibrium." Berkeley Program in Law and Economics, Working Paper Series 206.
- Coultas, J. (2004). "When in Rome... An Evolutionary Perspective on Conformity." Group Processes & Intergroup Relations **7**(4): 317-331.
- Day, R. L., T. MacDonald, et al. (2001). "Interactions between shoal size and conformity in guppy social foraging." Animal Behaviour **62**: 917-925.
- de Quervain, D. J., U. Fischbacher, et al. (2004). "The Neural Basis of Altruistic Punishment." Science **305**: 1254-1258.
- de Soto, H. (2003). The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else. New York: Basic Books.
- Diamond, J. M. (1997). Guns, Germs, and Steel: The Fates of Human Societies. New York, W.W. Norton & Co.
- Diamond, J. M. (2005) Collapse: How Societies Choose to Fail or Succeed. New York, Viking.
- Dixit, A.K. (2004). Lawlessness and Economics: Alternative Modes of Governance. Princeton, Princeton University Press.
- Edgerton, R. B. (1992). Sick Societies: Challenging the Myth of Primitive Harmony. New York, The Free Press.
- Ellison, G. and D. Fudenberg (1993). "Rules of Thumb for Social Learning." Journal of Political Economy **101**(4): 612-643.
- Ellison, G. and D. Fudenberg (1995). "Word-of-Mouth Communication and Social Learning." The Quarterly Journal of Economics **110**(1): 93-125.
- Ensminger, J. (1992). Making a Market: The Institutional Transformation of an African Society. New York: Cambridge University Press.
- Ensminger, J. (1997a). "Transaction Costs and Islam: Explaining Conversion in Africa." Journal of Institutional and Theoretical Economics **153**(1): 3-29.
- Ensminger, J. (1997b). "Changing Property Rights: Reconciling Formal and Informal Rights to Land in Africa." In The Frontiers fo the New Institutional Economics, J. N. Drobak and J. V. C. Nye, eds. New York, Academic Press.
- Ensminger, J. (2004) "Market Integration and Fairness: Evidence from Ultimatum, Dictator, and Public Goods Experiments in East Africa" In: Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies. J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis, eds. London: Oxford University press.
- Ensminger, J. and J. Knight (1997). "Changing Social Norms." Current Anthropology **38**(1): 1-24.
- Fehr, E. and C. F. Camerer (2007). "Social neuroeconomics: the neural circuitry of social preferences." Trends in Cognitive Sciences **11**(10): 419-427.
- Fehr, E. and H. Gintis (2007). "Human motivation and social cooperation: Experimental and analytical foundations." Annual Review of Sociology **33**: 43-64.

- Fehr, E. and S. Gächter (2000). "Cooperation and punishment in public goods experiments." *American Economic Review* **90**(4): 980-995.
- Fehr, E. and U. Fischbacher (2003). "The nature of human altruism." *Nature* **425**: 785-791.
- Frank, R. (1988). *Passions within Reason: The strategic role of the emotions*. New York, W. W. Norton & Company.
- Frederick, S., G. Loewenstein, et al. (2002). "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature* **40**(2): 351-401.
- Fudenberg, D. and D. Levine (1998). *The Theory of Learning in Games*. Cambridge, MA, MIT Press.
- Gintis, H. (2004). "The genetic side of gene-culture coevolution: internalization of norms and pro-social emotions." *Journal of Economic Behavior & Organization* **53**(1): 57-67.
- Gintis, H. (2007). "A framework for the unification of the behavioral sciences." *Behavioral and Brain Sciences* **30**(1): 1-+.
- Greif, A. (2006). *Institutions and the Path to the Modern Economy: Lessons from the Medieval Trade*. New York, Cambridge University Press.
- Guerek et al. (2006)
- Haidt, J. (2001). "The emotional dog and its rational tail: A social intuitionist approach to moral judgment." *Psychological Review* **108**(4): 814-834.
- Haidt, J. (2007). "The new synthesis in moral psychology." *Science* **316**(5827): 998-1002.
- Haley, K. J. and D. M. T. Fessler (2005). "Nobody's watching? Subtle cues affect generosity in an anonymous economic game." *Evolution and Human Behavior* **26**(3): 245-256.
- Harsanyi, J. C. and R. Selton (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge, MA, MIT Press.
- Hayek, F. A. (1960). *The Constitution of Liberty*. Chicago: University of Chicago Press.
- Hayashi, N., E. Ostrom, et al. (1999). "Reciprocity, Trust, and the Sense of Control." *Rationality and Society* **11**(1): 27-46.
- Henrich, J. (1997). "Market Incorporation, Agricultural Change and Sustainability Among the Machiguenga Indians of the Peruvian Amazon." *Human Ecology* **25**(2): 319-51.
- Henrich, J. (2000). "Does Culture Matter in Economic Behavior: Ultimatum Game Bargaining Among the Machiguenga." *American Economic Review* **90**(4): 973-980.
- Henrich, J. (2004). "Cultural group selection, coevolutionary processes and large-scale cooperation." *Journal of Economic Behavior & Organization* **53**: 3-35.
- Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J. Camilo Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker (2006) "Costly Punishment Across Human Societies." *Science*. June 312(23): 1767-1770.
- Henrich, J. (2006). "Cooperation, Punishment, and the Evolution of Human Institutions." *Science* **312**(5770): 60-61.
- Henrich, J. (2007). *The Evolutionary Foundations and Cultural Evolution of Inferentially Potent Displays*, University of British Columbia.
- Henrich, J. (n.d.) Negative Reciprocity stabilizes costly behavior, including larger-scale cooperation. In preparation.
- Henrich, J. and F. Gil-White (2001). "The Evolution of Prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission." *Evolution and Human Behavior* **22**(3): 165-196.

- Henrich, J. and R. Boyd (1998). "The Evolution of Conformist Transmission and the Emergence of Between-Group Differences." Evolution and Human Behavior **19**(4): 215-242.
- Henrich, J. and R. Boyd (2001). "Why People Punish Defectors: Weak Conformist Transmission can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas." Journal of Theoretical Biology **208**(1): 79-89.
- Henrich, J. and R. McElreath (2006). Dual Inheritance Theory: The Evolution of Human Cultural Capacities and Cultural Evolution. Oxford Handbook of Evolutionary Psychology. R. Dunbar and L. Barrett. Oxford, Oxford University Press.
- Henrich, N. S. and J. Henrich (2007). Why Humans Cooperate: A Cultural and Evolutionary Explanation. Oxford, Oxford University Press.
- Herrmann, B., C. Thoni, et al. (2008). "Anti-Social Punishment Across Societies." Science **319**(5868): 1362-1367.
- Hirschman, A. O. (1982). "Rival Interpretations of Market Society: Civilizing, Destructive, or Feeble?" Journal of Economic Literature **XX**:1463-1484.
- Hoffman, E., K. McCabe, et al. (1994). "Preferences, Property Rights, and Anonymity in Bargaining Games." Games and Economic Behavior **7**(3): 346-380.
- Hoffman, E., K. McCabe, et al. (1998). "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology." Economic Inquiry **36**(3): 335-352.
- Johnson, A. (2003). Families of the Forest: Matsigenka Indians of the Peruvian Amazon. Berkeley, University of California.
- Johnson, A. and T. Earle (2000). The Evolution of Human Societies. Stanford, Stanford University Press.
- Kameda, T. and D. Nakanishi (2002). "Cost-benefit analysis of social/cultural learning in a non-stationary uncertain environment: An evolutionary simulation and an experiment with human subjects." Evolution and Human Behavior **23**(5): 373-393.
- Knight, J. (1992). Institutions and Social Conflict. New York, Cambridge University Press.
- Knight, J. and J. Ensminger (1998). Conflict over Changing Social Norms: Bargaining, Ideology, and Enforcement. The New Institutionalism in Sociology. M.C. Brinton and V. Nee. New York, Russell Sage Foundation: 105-126.
- Kohler, T. A., S. VanBuskirk, et al. (2004). "Vessels and villages: evidence for conformist transmission in early village aggregations on the Pajarito Plateau, New Mexico." Journal of Anthropological Archaeology **23**(1): 100-118.
- Lachlan, R. F., L. Crooks, et al. (1998). "Who follows whom? Shoaling preferences and social learning of foraging information in guppies." Animal Behaviour **56**(1): 181-190.
- Laibson, D. (1997). "Golden Eggs and Hyperbolic Discounting." The Quarterly Journal of Economics. **112**(2):443-477.
- Landes, D. S. (1998). The Wealth and Poverty of Nations: Why Some Are So Rich and Some Are So Poor. New York, W.W. Norton & Company, Inc.
- Locke, J. (1980) [1690]. Second Treatise of Government. Indianapolis: Bobbs-Merrill.
- Mackie, G. (1996). "Ending Footbinding and Infibulation: A Convention Account." American Sociological Review **61**(6): 999-1017.
- McAdams, R.H. (2000). "A Focal Point Theory of Expressive Law." Virginia Law Review. **86**(8): 1649-1729.
- McClelland, D. C. (1961) The Achieving Society. Princeton: Van Nostrand.
- McElreath, R., M. Lubell, et al. (2005). "Applying evolutionary models to the laboratory study of social learning." Evolution and Human Behavior **26**(6): 483-508.

- McElreath, R. and R. Boyd (2007). Modeling the Evolution of Social Behavior. Princeton University Press.
- Mesoudi, A. (in press). "The cultural transmission of Great Basin projectile technology: An experimental simulation." American Antiquity.
- Montesquieu, B. (1900). The Spirit of Laws. New York, P. F. Collier & Son.
- Murphy, R. and Y. Murphy (1986). "Northern Shoshone and Bannock." In Handbook of North American Indians: 11. Great Basin, edited by Warren L. d'Azevedo, pp. 284-307. Washington, D.C.: Smithsonian Institution.
- North, D. C. (1981). Structure and Change in Economic History. New York, W. W. Norton & Company, Inc.
- North, D. C. (1990). Institutions, Institutional Change and Economic Performance. Cambridge, UK, Cambridge University Press.
- North, D. C. (2005). Understanding the Process of Economic Change. Princeton, Princeton University Press.
- North, D. C. and R. P. Thomas (1973). The Rise of the Western World: A New Economic History. Cambridge, Cambridge University Press.
- North, D. C. and B. R. Weingast (1989). "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." The Journal of Economic History **XLIX**(4): 803-832.
- Nowak, M. A., K. M. Page, et al. (2000). "Fairness versus reason in the Ultimatum Game." Science **289**(5485): 1773-1775.
- Ostrom, E. (1990). Governing the Commons: The Evolution of Institutions for Collective Action. New York:Cambridge University Press.
- Panchanathan, K. and R. Boyd (2004). "Indirect reciprocity can stabilize cooperation without the second-order free rider problem." Nature **432**: 499-502.
- Pillutla, M. M. and X. P. Chen (1999). "Social Norms and Cooperation in Social Dilemmas: The Effects of Context and Feedback." Organizational Behavior and Human Decision Processes **78**(2): 81-103.
- Platteau, J. P. (2000). Institutions, Social Norms, and Economic Development. Amsterdam, Harwood Academic Publishers.
- Price, G. (1970). "Selection and Covariance." Nature **227**: 520-521.
- Price, G. (1972). "Extensions of Covariance Selection Mathematics." Annals of Human Genetics **35**: 485-490.
- Putnam, R. D. (1993). Making Democracy Work. Princeton, Princeton University Press.
- Richerson, P. J. and R. Boyd (1998). The Evolution of Ultrasociality. Indoctrinability, Ideology and Warfare. I. Eibl-Eibesfeldt and F. K. Salter. New York, Berghahn Books: 71-96.
- Richerson, P. J. and R. Boyd (2000). Climate, Culture and the Evolution of Cognition. The Evolution of Cognition. C. M. Heyes. Cambridge, MIT Press: 329-345.
- Richerson, P. J. and R. Boyd (2000). "Complex Societies: the evolutionary dynamics of a crude superorganism." Human Nature **10**: 253-289.
- Richerson, P. J. and R. Boyd (2005). Not by Genes Alone: How Culture Transformed Human Evolution. Chicago, University of Chicago Press.
- Richerson, P. J., R. Boyd, et al. (2001). "Was Agriculture Impossible During the Pleistocene But Mandatory During the Holocene? A Climate Change Hypothesis." American Antiquity **66**(3): 387-411.
- Rilling, J. K., A. G. Sanfey, et al. (2004). "Imaging the social brain with fMRI and interactive games." International Journal of Neuropsychopharmacology **7**: S477-S478.
- Rosenberg, N. and Birdzell, Jr., L. E. (1986). How the West Grew Rich: The Economic Transformation of the Industrial World. New York, Basic Books, Inc.

- Ross, L. and A. Ward (1996). 'Naïve realism: implication for social-conflict and misunderstanding.' In Values and Knowledge. E. S. Reed, E. Turiel and T. Brown, eds. New Jersey: Lawrence Erlbaum Associates.
- Sachs, J. (2001). "Tropical Underdevelopment." NBER Working Paper No. 8119.
- Sachs, J. (2003). "Institutions Don't Rule: Direct Effects of Geogrphahy on Per Capita Income." NBER Working Paper No. 9490.
- Sanfey, A. G., J. K. Rilling, et al. (2003). "The Neural Basis of Economic Decision-Making in the Ultimatum Game." Science **300**(5626): 1755-1758.
- Schlag, K. H. (1998). "Why Imitate, and If So, How? A Boundedly Rational Approach to Multi-Armed Bandits." Journal of Economic Theory **78**(1): 130-156.
- Schlag, K. H. (1999). "Which one should I imitate?" Journal of Mathematical Economics **31**(4): 493-527.
- Sharif and Nor (xxxx).
- Smith, A. (1976). An Inquiry into the Nature and Causes of the Wealth of Nations. Chicago, University of Chicago Press.
- Smith, A. (1997). The Theory of Moral Sentiments. Washington, DC, Regnery Publishing, Inc.
- Sowell (1998). Conquests and Cultures: An International History. New York: Basic Books.
- Spencer, R. W. and J. H. Huston (1993). "Rational forecasts: On confirming ambiguity as the mother of conformity." Journal of Economic Psychology **14**(4): 697-709.
- Weber, M. (1958). The Protestant Ethic and The Spirit of Capitalism. New York, Charles Scribner's Sons.
- Weibull, J. W. (1995). Evolutionary Game Theory. Cambridge, Mass, MIT Press.
- Williamson, O. (1985). The Economic Institutions of Capitalism. New York, Free Press.
- Young, H. P. (1998). Individual Strategy and Social Structure: An evolutionary Theory of Institutions. Princeton, N.J, Princeton University Press.