A Review of *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective*, Jean Ensminger and Joseph Henrich (eds).

A puzzle of perennial interest across a variety of disciplines is the existence of the type of large-scale cooperation among strangers that is necessary for societies to outgrow the confines of kinship-based social groups. Why do individuals, as members of large, complex and often anonymous, societies routinely act against their own narrow self-interest to help their compatriots by, for example, donating to charity or paying their taxes? The puzzle deepens in the face of an ever-growing body of research from behavioral and experimental economists demonstrating that such prosocial behavior does not directly hinge upon the possibility of future interaction.

The central thesis of this book is that prosocial preferences and the (formal and informal) institutions permitting cooperation among large groups of strangers co-evolve. To shed light on this conjecture, the book both synthesizes (Part I) and reports separately (Part II) the combined studies of an interdisciplinary team of anthropologists and economists. The book represents the second volume produced by an on-going collaboration among a group of anthropologists and economists dating back almost two decades, the first volume being Henrich et al. (2004) and Henrich et al. (2001). While there is considerable overlap between the contributors to the first volume and the set of collaborators for this latest volume, there are also significant differences. The most apparent difference is the absence of several prominent economic theorists intimately associated with the previous volume: Ernst Fehr, Colin Camer, Samuel Bowles and Herbet Gintis. Even though several economists remain, the current team is consequently more heavily skewed toward anthropologists. The authors discuss the reasons for this general change in composition without naming names, chalking it up to differences in schedules and the inability of some researchers to meet deadlines and performance criteria, providing valuable insight into the pitfalls of working with a large team of academics across multiple cultures and continents.

As with the previous volume, this book draws heavily from previously-published articles resulting from the collaboration (Henrich et al., 2006; Henrich et al., 2010). One contribution of the book is the presentation of the constituent studies in detail, each in its own chapter, in Part II of the book. Most of these studies have not been previously published. The primary, and non-trivial, contribution of the book is the type of detailed and unified discussion of the totality of evidence for the co-evolution of prosocial preferences and institutions that was not possible in any one of the previously published articles alone.

For example, the authors spend an entire chapter providing a theoretical justification for their co-evolutionary hypothesis as well as outlining the extant experimental and empirical evidence, which is compelling. This same exercise, on the other hand, constituted only a few paragraphs in their *Science* paper (Heinrich et al., 2010). While too nuanced to do justice to here, essentially their argument has the following ingredients: i) adaptive learning strategies, like imitating successful others, convey an evolutionary advantage; ii) allowing for "adaptive learners," phenomena akin to social norms and institutions emerge as equilibria of formal evolutionary game theoretical models of societal interactions; iii) in particular, prosocial norms emerge and can be sustained even when individuals interact infrequently through direct and indirect punishment; iv) intergroup competition and comparison may select among the large set of possible social norms, favoring the social norms of successful groups; and finally, v) through adoption and repetition, norms may become internalized, so that their influence no longer relies on external punishment. It is in this last step that *preferences* may be thought of as being prosocial, exhibiting a taste for fairness directly.

This co-evolutionary argument implies a relationship between prosocial preferences at a societal level and the extent to which the institutions of society have developed to support the sort of trust- and fairness-based interactions among strangers that permit economic growth. In their *societal norms for complex market-integrated societies hypothesis*, the authors lay out the empirical heart of their endeavor, arguing that "norms that enhance fairness and trust among strangers are likely to be causally interconnected with … the expansion of both the breadth of and intensity of market exchanges … and the spread of universal religions with high moralizing gods" (pg. 134). The overarching goal of the book is to provide empirical evidence for this conjectured relationship.

Testing this conjecture requires measuring the intensity and breadth of market exchanges as well as the spread of moralizing gods and, of course, prosocial behavior across a wide-enough array of societies to yield identifying variation. This is where the study truly shines as a masterpiece of interdisciplinary collaboration and expertise. The combined knowledge of the entire team is on display as another entire chapter is devoted to convincing the reader, and one feels the team members themselves, of the reasonableness of the measures and methodology employed in the study, as well as in the choices of societies to study, all of which necessarily involved substantial compromise. The sampled populations and experimental games were chosen unusually deliberately and purposefully, building on insights gleaned from the collaboration's previous set of coordinated experiments, reported in Henrich *et al.* (2004). In the end, the proportion of calories typically purchased in the market was chosen as the best available measure of the intensity and breadth of market exchange, or "market integration," while the proportion of the society professing belief in Islam or Christianity was chosen as a measure of the spread of a universal religion with a moralizing god.

The most novel and compelling aspect of the book is how the study's authors collect measures of prosocial preferences, conducting coordinated but separate experiments in fifteen different, mostly small-scale, carefully chosen societies. Unlike a traditional laboratory implementation using a convenience sample of university students, the subject pool here constitutes a simple random, and hence representative, sample of each society. The experiments shared a common core of three games which have become a staple of behavioral and experimental economists in more traditional laboratory settings: the Dictator Game (DG), the Ultimatum Game (UG) and the third-party punishment game (TPG). In the DG, players are anonymously paired. Player 1 is given a fixed sum of money which she can share (or not) with Player 2. The UG and TPG add the option of punishing Player 1 for sharing too little, either by Player 2 himself (UG) or by a disinterested bystander (TPG). Since there are no material nor strategic incentives for sharing in the DG, it is commonly held to measure internalized fairness norms. Similarly, the latter two games measure different aspects of the willingness to punish unfair behavior. The authors are unusually careful in pointing out that all of these measures are probably only valid for the specific context of anonymous pecuniary interactions—suitable for their purposes but not necessarily generalizable to all types of fairness—turning context-dependence from a weakness into a strength of their study.

Overall, the results which are laudably carefully detailed in another entire chapter are largely consistent with the "societal norms for complex market-integrated societies" hypothesis. While this chapter draws heavily on the Henrich et al. (2006) and Henrich et al. (2010), the current chapter incorporates revised data to address a coding issue noted in a post-publication *erratum* to the 2010 *Science* paper. The most robust result, supported by a battery of regressions and robustness exercises, is a significantly positive correlation between market integration and the internalization of fairness norms as measured by sharing in the DG. A similar, though less robust, positive relationship is documented between DG sharing and the spread of a universal religion. Interestingly, while sharing is not affected by the specter of direct punishment (UG), it is significantly undermined by the presence of third-party punishment—evidence that fairness motives can be crowded out by pecuniary incentives or social observability. Finally, an intriguing positive correlation is documented between the size of a society and the willingness to punish unfair behavior at a personal cost. In larger societies, experimental participants had to share significantly more of their windfall in order to avoid being punished in both the UG and the TPG.

In total, the volume provides a detailed overview of a labor intensive but fruitful research collaboration which should set a new standard for interdisciplinary and cross-cultural studies investigating the evolution of prosocial preferences. The individual studies, each of which are reported as separate chapters in the second part of the book, as well as the previously-published articles (Henrich et al. 2010; Henrich et al. 2006), yield valuable insights on their own but when summarized and synthesized in the first part of the volume become exponentially more valuable. The body of results taken together provide compelling evidence in favor of the authors' co-evolutionary conjecture. Beyond the authors' central conjecture, however, the rich and varied data will also certainly prove useful for other inquiries into the nature of social norms and prosocial behavior. The relationship between punishment preferences and community size, as well as the evidence for "crowding out," warrant further investigation as they may have implications for the design of optimal institutions in disparate societies.

The book will be of considerable interest to any serious scholar of prosociality, theorist and empiricist alike. Researchers can ponder the stories the data may yield having confidence that they were collected and reported using the best available methods. Practitioners of development economics, behavioral and experimental economics, or even cultural anthropology, often mutually skeptical of each other's research, should also gain great insight into where common methodological ground can be reached and where hidden pitfalls lie in interdisciplinary research.

References:

Henrich, Joseph., Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis, eds., (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Cambridge: Oxford University Press.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath (2001). "In search of homo economicus: behavioral experiments in 15 small-scale societies." *American Economic Review*, vol 91, no. 2: pp. 73 - 78.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas et al. (2006). "Costly punishment across human societies." *Science* vol. 312, no. 5781: pp. 1767-1770.

Henrich, Joseph, Jean Ensminger, Richard McElreath, Abigail Barr, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas et al. (2010) "Markets, religion, community size, and the evolution of fairness and punishment." *Science*, vol. 327, no. 5972: pp. 1480-1484.

Jeffrey V. Butler
EIEF and UNLV