Costly Punishment Across Human Societies

Joseph Henrich,[1]* Richard McElreath,[2] Abigail Barr,[3] Jean Ensminger,[4] Clark Barrett,[5] Alexander Bolyanatz,[6] Juan Camilo Cardenas,[7] Michael Gurven,[8] Edwins Gwako,[9] Natalie Henrich,[1] Carolyn Lesorogol,[10] Frank Marlowe,[11] David Tracer,[12] John Ziker[13]

Recent behavioral experiments aimed at understanding the evolutionary foundations of human cooperation have suggested that a willingness to engage in costly punishment, even in one-shot situations, may be part of human psychology and a key element in understanding our sociality. However, because most experiments have been confined to students in industrialized societies, generalizations of these insights to the species have necessarily been tentative. Here, experimental results from 15 diverse populations show that (i) all populations demonstrate some willingness to administer costly punishment as unequal behavior increases, (ii) the magnitude of this punishment varies substantially across populations, and (iii) costly punishment positively covaries with altruistic behavior across populations. These findings are consistent with models of the gene-culture coevolution of human altruism and further sharpen what any theory of human cooperation needs to explain.

For tens of thousands of years before formal contracts, courts, and constables, human societies maintained important forms of cooperation in domains such as hunting, warfare, trade, and food sharing. The scale of cooperation in both contemporary and past human societies remains a puzzle for the evolutionary and social sciences, because, first, neither kin selection nor reciprocity appears to readily explain altruism in very large groups of unrelated individuals and, second, canonical assumptions of self-regarding preferences in economics and related fields appear equally ill-fitted to the facts (*1*). Reputation can support altruism in large groups; however, some other mechanism is needed to explain why reciprocity should be linked to prosociality rather than selfish or neutral behavior (*2*). Recent theoretical work

suggests that substantial cooperation can evolve, even among non-kin, in situations devoid of reputation or repeat interaction if cooperators also engage in the costly punishment of non-cooperative norm violators (*3–10*). Consistent with these models, behavioral experiments have now confirmed the (i) existence of costly punishment, (ii) effectiveness of punishment in sustaining cooperation (*11*, *12*), and (iii) willingness by uninvolved third parties to punish in anonymous situa-

tions (*13*). Such experiments have even begun to probe the neural underpinnings of punishment (*14*, *15*).

These results are important, because the existence of costly punishment can explain important pieces of the puzzle of large-scale human cooperation. However, like previous experimental games used to study altruism, these experiments have been conducted almost exclusively among university students. We do not know whether such findings represent the peculiarities of students and/or people from industrialized societies or whether they are indeed capturing species characteristics. Our earlier research used experimental games in 15 diverse societies to measure other-regarding behavior (*1*, *16*). We found that canonical self-interest could not explain the results in any of the 15 societies studied. We also found much more variation in game behavior than previous studies with university students had found. Similarly, until costly punishment is studied in more societies and outside of university students, it is difficult to judge its importance for explaining human cooperation.

In addition to estimating how widespread it is, knowing whether costly punishment covaries with altruistic behavior is valuable. Models of the evolution of costly punishment suggest that societies in which costly punishment is common will exhibit stronger norms of fairness and prosociality, because the

[1]Department of Anthropology, Emory University, 1557 Dickey Drive, Atlanta, GA 30322, USA. [2]Department of Anthropology, Graduate Group in Ecology, Animal Behavior Graduate Group, Population Biology Graduate Group, University of California Davis, One Shields Avenue, Davis, CA 95616, USA. [3]GPRG, Department of Economics, University of Oxford, Manor Road, Oxford, OX1 3UQ, UK. [4]Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA. [5]Department of Anthropology, UCLA, 341 Haines Hall, Box 951553, Los Angeles, CA 90095, USA. [6]Department of Anthropology, College of DuPage, Glen Ellyn, IL 60137, USA. [7]Facultad de Economia, CEDE, Universidad de Los Andes, K1 No. 18A-70, Bogotá, Colombia. [8]Department of Anthropology, University of California, Santa Barbara, CA 93106, USA. [9]Department of Sociology, Guilford College, 5800 West Friendly Avenue, Greensboro, NC 27410, USA. [10]George Warren Brown School of Social Work, Washington University, 1 Brookings Drive, St. Louis, MO 63130, USA. [11]Department of Anthropology, Harvard University, 11 Divinity Avenue, Cambridge, MA 02138, USA. [12]Department of Anthropology, University of Colorado at Denver and Health Sciences Center, Post Office Box 173364, Campus Box 103, Denver, CO 80217, USA. [13]Department of Anthropology, Boise State University, 1910 University Drive, Boise, ID 83725, USA.

*To whom correspondence should be addressed. E-mail: jhenric@emory.edu
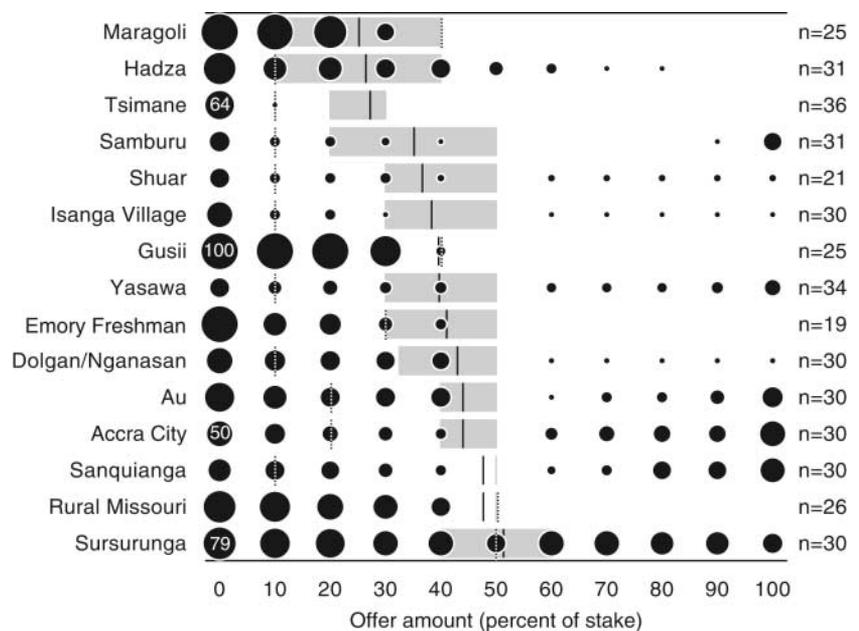
Fig. 1. UG results displayed as the distributions of rejections across possible offers in the UG, which overlay the mean offers and interquartiles. For each population labeled along the vertical axis, the areas of the black bubbles, reading horizontally, show the fraction of the sample of player 2s who were willing to reject that offer. For reference, inside some of the bubbles we noted the percentage illustrated by that bubble. The dashed vertical bars mark the IMO for each population. The solid vertical bars mark the mean offer for each population, with the gray shaded rectangle highlighting the interquartile of offers. Populations were ordered by their mean offers (from low to high). Counts on the right (*n*) refer to numbers of pairs of players.

existence of costly punishment is what allows such norms to remain stable against invading defectors. Thus, the status of costly punishment as a viable explanation of human prosociality depends on it being found outside of students as well as its association with cooperative norms and institutions across cultures.

In this article, we present a round of field experiments that address the nature of costly punishment. Field experiments are valuable tools because they allow for better comparability and control of causal factors, and these particular experiments were designed by economists to specifically test the predictions of a canonical model of narrow self-interest (16). We used two behavioral experiments, the ultimatum and third party punishment games, among 1762 adults sampled from 15 diverse populations from five continents, representing the breadth of human production systems. In addition to explicitly measuring costly punishment via a strategy method that provides more information about behavior than the method used in the previous study, this study represents greater methodological standardization than the previous round of experiments [Supporting Online Material (SOM) Text]. Although our findings revealed some consistent patterns of punishment in all populations, we also found substantial variation across populations in their willingness to punish, including several populations with a willingness to punish "excessive generosity" (a phenomenon not observed in typical student subject pools). By using a third experiment, the dictator game, we also show that punishment correlates positively with altruism across populations in a manner consistent with coevolutionary theories (4).

Experiments. In our first experiment, the ultimatum game (UG), two anonymous players are allotted a sum of real money (the stake) in a one-shot interaction (17). The first player (player 1) can offer a portion of this sum to a second player, player 2 (offers were restricted to 10% increments of the stake). Player 2, before hearing the actual amount offered by player 1, must decide whether to accept or reject each of the possible offers, and these decisions are binding. If player 2 specified acceptance of the actual offer amount, then he or she receives the amount of the offer and player 1 receives the rest. If player 2 specified a rejection of the amount actually offered, both players receive zero. If people are motivated purely by self-interest, player 2s will always accept any positive offer; knowing this, player 1 should offer the smallest nonzero amount. Because this is a one-shot anonymous interaction, player 2's willingness to reject provides one measure of costly punishment, termed second-party punishment.

In our second experiment, the third party punishment game (3PPG), two players are allotted a sum of real money (the stake), and a third player gets one-half of this amount (13). Player 1 must decide how much of the stake to give to player 2 (who makes no decisions). Then, before hearing the actual amount player 1 allocated to player 2, player 3 has to decide whether to pay 10% of the stake (20% of his or her allocation) to punish player 1, causing player 1 to suffer a deduction of 30% of the stake from the amount kept. Player 3's punishment strategy is elicited for all possible offers by player 1. For example, suppose the stake is $100: if player 1 gives $10 to player 2 (and keeps $90) and player 3 wants to punish this offer amount, then player 1 takes home $60; player 2, $10; and player 3, $40. If player 3 had instead decided not to punish offers of 10%, then the take-home amounts would have been $90, $10, and $50, respectively. In this anonymous

one-shot game, a purely self-interested player 3 would never pay to punish player 1. Knowing this, a self-interested player 1 should always offer zero to player 2. Thus, an individual's willingness to pay to punish provides a direct measure of the person's taste for a second type of costly punishment, third-party punishment.

To get behavioral measures of altruism, we also conducted dictator games (DG) in each population. The DG is the same as the UG except that player 2 cannot reject (18). Player 1 merely dictates the portions of the stake received by himself or herself and player 2. In this one-shot anonymous game, a purely self-interested individual would offer zero; thus, offers in the DG provide a measure of a kind of behavioral altruism that is not directly linked to kinship, reciprocity, reputation, or the immediate threat of punishment (19).

Here, we highlight several key aspects of our standardized procedures, protocols, and scripts (for further details, see SOM Text). First, to guarantee motivation and attention to the experiments, we standardized the stake of each game to 1 day's wage in the local economy. Players were also paid a show-up fee equal to 20% of a day's wage. Second, by using the method of back translation, all of our game scripts were administered in a local language by fluent speakers. Third, our protocol prevented those waiting to play from talking about the game and from interacting with experienced players during a game session. Fourth, individualized instruction using a fixed script, set of examples, and preplay test questions guaranteed that all players understood the game well enough to correctly answer at least two consecutive test questions (19).

We drew adults from a diverse set of populations scattered across the globe. Table 1

Table 1. Summary of populations studied. The column labeled "Economic base" classifies the production systems. Horticulturalists, for example, rely primarily on slash-and-burn agriculture, whereas pastoralists rely primarily on herding. "Residence" classifies societies according to the nature and frequency of their social groups' movements.

Group	Continent	Nation, region	Environs	Economic base	Residence
Accra City	Africa	Ghana	Urban	Wage work	Sedentary
Gusii	Africa	Kenya, Nyamira	Fertile high plains	Mixed farming, wage work	Sedentary
Hadza	Africa	Tanzania	Savannah-woodlands	Foraging	Nomadic
Isanga village	Africa	Tanzania, Mbeya	Mountainous forest	Agriculture, wage work	Sedentary
Maragoli	Africa	Kenya	Fertile plains	Mixed farming, wage work	Sedentary
Samburu	Africa	Kenya	Semi-arid savanna	Pastoralism	Semi-nomadic
Emory freshman	N. America	U.S., southeast	Temperate forest, urban	Students	Temporary residence
Missouri	N. America	U.S., rural and urban midwest	Prairie	Wage work and farming	Sedentary
Sanquianga	S. America	Colombia, Pacific coast	Mangrove forest	Fisheries (fish, clams, shrimp)	Sedentary
Shuar	S. America	Ecuador, Amazonia	Tropical forest	Horticulture	Sedentary
Tsimane	S. America	Bolivia, Amazonia	Tropical forest	Horticulture-foraging	Semi-nomadic
Dolgan/Nganasan	Asia	Russian Federation, Siberia	Tundra-taiga	Hunting/fishing and wage work	Semi-sedentary
Au	Oceania	Papua New Guinea, West Sepik	Mountainous tropical forest	Horticulture-foraging	Sedentary
Sursurunga	Oceania	Papua New Guinea, New Ireland	Coastal tropical island	Horticulture	Sedentary
Yasawa	Oceania	Fiji, Yasawa Island	Coastal tropical pacific	Horticulture and marine foraging	Sedentary

provides the nation, region, environment, economic base, and predominant residence pattern for each population. As points of reference, we also ran these games with students at Emory University and nonstudent adults in both rural and urban Missouri. The Missouri samples provide the appropriate U.S. points of comparison with our diverse sample of societies, whereas the student sample links us to the subject populations used in most work. In considering the student data (vis-à-vis the nonstudent data), it is important to realize that behavior in these experiments continues to change through the university years and does not reach the adult plateau until the participants reach their mid-twenties (20–23). Thus, because we want to

explore the variation among adult populations and avoid confounding maturational effects, we used only the nonstudent samples in comparative analyses (24).

Punishment results. Our two measures of costly punishment revealed both a universal pattern, with an increasing proportion of individuals from every society choosing to punish as offers approach zero, and substantial differences across populations in their overall willingness to punish unequal offers. Figure 1 summarizes our UG data. For every population studied, the probability of rejection decreased as offers increase from 0% to 50%. At the lowest offer for which punishment is costly (10% offers), 56.5% of players rejected overall. However, the magnitude of

this effect varied substantially across groups. In five societies, the Tsimane, the Shuar, Isanga village, Yasawa, and the Samburu, less than 15% of the population were willing to reject 10% offers. In contrast, over 60% of the samples in four populations rejected such offers. Another, although indirect, measure of a population's willingness to punish is its income-maximizing offer (IMO), which is the offer that maximizes player 1's expected income, given the observed rejection probabilities in that society. Marked with a dashed line, the IMO varied from 10% (little punishment) in eight of the populations to 50% (strong punishment) in two.

To assess whether the observed variation in punishment between populations can be explained by demographic and economic differences among them, we conducted a set of three linear regression analyses using player 2s' minimum acceptable offer (MAO), the lowest offer between zero and 50% that a player would accept, as the dependent variable. For example, if an individual rejected an offer of zero but then accepted 10 through 50, the MAO is 10. First, regressing MAO on population dummy variables showed that 34.4% of the variation occurs between population means. Second, adding measures of players' sex, age, education, household size, income, and wealth increased the variance explained to 41.5%, implying that these capture about 7% of the variance within populations. Third, removing the dummies from the regression decreased the variance explained to 15.8%, indicating that a substantial portion of the between-population variance cannot be explained by these individual predictors (SOM Text).

Shown in Fig. 1 is that 6 of our 14 nonstudent populations also display a willingness to reject increasingly unequal UG offers as they rise from 50% to 100%, with upwards of half of the sample rejecting offers of 100% in two populations. Originally noted by Tracer in Papua New Guinea (25), this willingness to reject hyperfair offers (offers greater than 50%) now appears to be widespread, having also been documented in Russia (26) and China (27). Milder versions of this phenomenon have been detected among students in the U.S. and Europe by using more sensitive bargaining instruments (28, 29), and we argue that these hyperfair rejections are unlikely to result from players' confusion (SOM Text).

To study the hyperfair rejections, we also calculated the maximum acceptable offer (MXAO), which is the highest offer above 50% that a player will accept. If a player accepted all offers above 50%, his MXAO was set at 100. First, regressing MXAO on our population dummies showed that 17% of the variation in MXAO occurs between populations. Then, adding age, sex, educa-
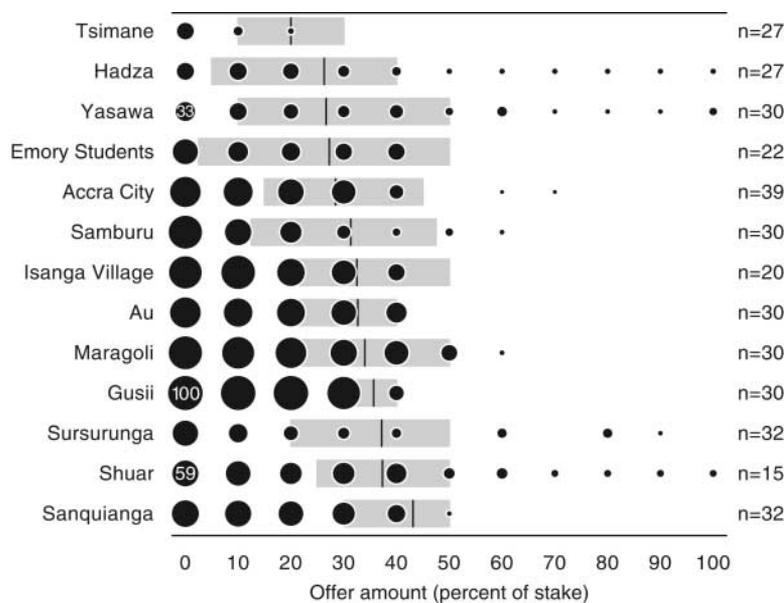
Fig. 2. 3PP results displayed as the distributions of decisions to punish across the possible offers in the 3PPG. For each population, the areas of the bubble display the fraction that was willing to punish at that offer amount. Counts on right (*n*) refer to numbers of triads of players. Inside some of the bubbles, we noted the precise percentage the bubble represents, for reference. For player 1 offers, the solid vertical bar marks the mean offer for each group, with the gray shaded rectangles highlighting the interquartiles. Populations ordered by mean offers (low to high). Emory students here are a general undergraduate sample, not the same freshman in Fig. 1.

Fig. 3. Relationship between 3PP and altruism shown as the relationship between mean MAOs in the 3PPG and mean offers in the DG. The different symbols indicate geographic proximity or continent. The size of each symbol is proportional to the number of DG pairs at each site. The dotted line gives the weighted regression line, with continental controls, of mean DG offers on mean MAO-3PP. The solid line gives the simple linear regression. Emory students were not included in the regression analysis, although we have plotted them for comparison.
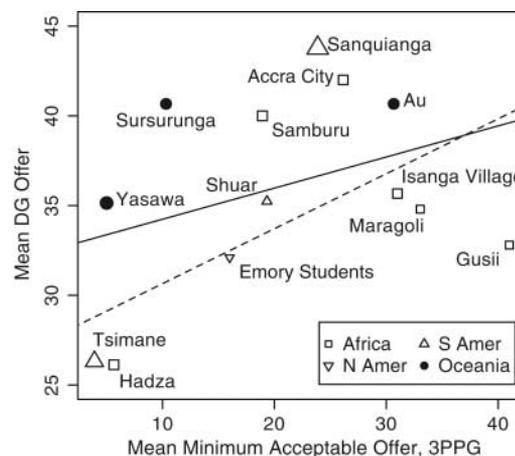
tion, household size, income, and wealth increased the variance explained to 24%. Third, removing the dummies dropped the variance explained to 5%, suggesting that very little of the variation among populations can be explained by measured differences in these economic and demographic variables.

Our findings for the Au, the Hadza, and the Tsimane largely replicated previous experimental work among these populations that used another version of the UG (*30–32*). Because these populations have provided some of the more unusual results, this robustness suggests that (i) at a population level, these findings are stable and (ii) basic patterns in the data are not substantially shifted by minor adjustments in the UG protocol.

The 3PPG revealed patterns similar to those seen in the UG, with all societies showing a decreasing frequency of punishment as offers increase to 50% (*19*), as well as substantial differences between populations (Fig. 2). Overall, two-thirds of player 3s were willing to pay 20% of their endowment (half of 1 day's wage) to punish player 1 for allocating zero to player 2. However, this fraction varied from around 28% among the Tsimane and the Hadza to over 90% among the Gusii and the Maragoli.

By using the same technique described above for MAO, we calculated the MAO-3PP for each player 3. Regressing MAO-3PP first on the group dummies showed that 38.2% of the variation occurs between groups. Adding our standard set of predictors increased the variation explained to 41%. Then, removing the dummies dropped the variance explained to 11%, indicating that a substantial portion of the between-population variation cannot be captured by our economic and demographic measures.

If costly punishment culturally coevolves with an intrinsic motivation for certain forms of altruism, societies with high degrees of punishment will also exhibit more altruistic behavior. Figure 3 plots the relationship between punishment, based on the mean MAO-3PP, and altruism, based on mean offers in the DG. To examine this relationship, we first regressed the mean DG offer from each population on the respective mean MAO-3PP. This yielded a coefficient of 0.17, with a 95% confidence interval (CI) from 0.031 to 0.51 (*33*). Second, because the population means are derived from somewhat different sample sizes, we re-ran the same regression weighted by our sample sizes. The coefficient increased to 0.23 (CI from 0.075 to 0.38). Lastly, because our samples may be correlated due to shared history (shared cultural phylogenies), we added continental dummy variables to our weighted linear regression. The coefficient for MAO-3PP then increased to 0.31 (CI from 0.13 to 0.51). In addition,

measures of punishment other than the group means also correlate with mean DG offers (SOM Text).

Conclusions. We have shown three things about costly punishment as measured in one-shot anonymous experiments. First, costly punishment is present across a highly diverse range of human populations and emerges in a patterned fashion in each population. In every population, less-equal offers were punished more frequently. Second, we also find substantial variation among populations, with some societies showing very little overall willingness to punish, others demonstrating substantial willingness to punish, and still others revealing a willingness to punish offers that are either too generous or too stingy. Given the critical importance of costly punishment in maintaining cooperation in experimental studies (*12, 34*), the observed variation here suggests that the same institutional forms may function quite differently in different populations (*33*). Third, at the population level, this willingness to punish covaries with a behavioral measure of altruism.

These three results are consistent with recent evolutionary models of altruistic punishment (*3, 4, 9*). In particular, culture-gene coevolutionary models that combine strategies of cooperation and punishment predict that local learning dynamics generate between-group variation as different groups arrive at different "cultural" equilibria (*36, 37*). These local learning dynamics create social environments that favor the genetic evolution of psychologies that predispose people to administer, anticipate, and avoid punishment (by learning local norms). Alternative explanations of the costly punishment and altruistic behavior observed in our experiments have not yet been formulated in a manner that can account for stable between-group variation or the positive covariation between altruism and punishment (*38, 39*). Whether the coevolution of cultural norms and genes or some other framework is ultimately correct, these results more sharply delineate the species-level patterns of social behavior that a successful theory of human cooperation must address.

References and Notes
1. J. Henrich *et al.*, *Behav. Brain Sci.* **28**, 795 (2005).
2. K. Panchanathan, R. Boyd, *Nature* **432**, 499 (2004).
3. R. Boyd, H. Gintis, S. Bowles, P. J. Richerson, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3531 (2003).
4. E. Fehr, U. Fischbacher, *Nature* **425**, 785 (2003).
5. A. Gardner, S. A. West, *Am. Nat.* **164**, 753 (2004).
6. H. Gintis, *J. Theor. Biol.* **206**, 169 (2000).
7. H. Gintis, *J. Theor. Biol.* **220**, 407 (2003).
8. K. Sigmund, C. Hauert, M. A. Nowak, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10757 (2001).
9. J. Henrich, R. Boyd, *J. Theor. Biol.* **208**, 79 (2001).
10. R. Boyd, P. Richerson, *Ethol. Sociobiol.* **13**, 171 (1992).
11. E. Fehr, S. Gachter, *Am. Econ. Rev.* **90**, 980 (2000).
12. E. Fehr, S. Gächter, *Nature* **415**, 137 (2002).
13. E. Fehr, U. Fischbacher, *Evol. Hum. Behav.* **25**, 63 (2003).
14. A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, J. D. Cohen, *Science* **300**, 1755 (2003).
15. D. J. de Quervain *et al.*, *Science* **305**, 1254 (2004).
16. C. F. Camerer, E. Fehr, in *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, J. Henrich *et al.*, Eds. (Oxford Univ. Press, Oxford, 2004), pp. 55–95.
17. W. Guth, R. Schmittberger, B. Schwarze, *J. Econ. Behav. Organ.* **3**, 367 (1982).
18. R. Forsythe, J. L. Horowitz, N. E. Savin, M. Sefton, *Games Econ. Behav.* **6**, 347 (1994).
19. See methods discussion in the supplemental online materials.
20. M. Sutter, M. Kocher, "Trust and reciprocity across different age groups," 2005 (www.mpiew-jena.mpg.de/esi/alumni/sutter/pdfs/children_trust_36.pdf).
21. J. Carter, M. Irons, *J. Econ. Perspect.* **5**, 171 (1991).
22. W. T. Harbaugh, K. Krause, *Econ. Inq.* **38**, 95 (2000).
23. W. T. Harbaugh, K. Krause, S. G. Liday, working paper, 2002 (http://econpapers.repec.org/paper/oreuoecwp/2002-04.htm).
24. We only have DG data for urban Missouri and no 3PPG data for rural Missouri. Because the rural and urban Missouri samples cannot be distinguished, we have combined them in all analyses. Similarly, we have treated the Dolgan and Nganasan as one sample. The Dolgan/Nganasan did not play the 3PPG.
25. D. Tracer, in *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, J. Henrich *et al.*, Eds. (Oxford Univ. Press, Oxford, 2004), pp. 232–259.
26. D. L. Bahry, R. K. Wilson, *J. Econ. Behav. Organ.* **60**, 37 (2006).
27. H. Hennig-Schmidt, Z.-Y. Li, C. Yang, working paper, 2004 (http://ideas.repec.org/p/bon/bonedp/bgse22_2004.html).
28. S. Huck, *J. Econ. Psychol.* **20**, 183 (1999).
29. J. Andreoni, M. Castillo, R. Petrie, *Am. Econ. Rev.* **93**, 672 (2003).
30. F. W. Marlowe, in *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, J. Henrich *et al.*, Eds. (Oxford Univ. Press, Oxford, 2004), pp. 168–193.
31. M. Gurven, in *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, J. Henrich *et al.*, Eds. (Oxford Univ. Press, Oxford, 2004), pp. 194–231.
32. D. Tracer, *Curr. Anthropol.* **44**, 432 (2003).
33. All confidence intervals here were calculated by resampling (10,000 times) the means from each population and then recalculating the regression coefficient.
34. O. Gurerk, B. Irlenbusch, B. Rockenbach, *Science* **312**, 108 (2006).
35. S. Gächter, B. Herrmann, C. Thoni, *Behav. Brain Sci.* **28**, 822 (2005).
36. R. Boyd, P. Richerson, *J. Theor. Biol.* **215**, 287 (2002).
37. J. Henrich, *J. Econ. Behav. Organ.* **53**, 3 (2004).
38. M. Price, J. Tooby, L. Cosmides, *Evol. Hum. Behav.* **23**, 203 (2002).
39. K. Haley, D. M. T. Fessler, *Evol. Hum. Behav.* **26**, 245 (2005).
40. We first thank our study participants, who welcomed us into their homes, lives, and communities. This research was funded primarily by the NSF in the United States (grant BCS-0136761), with additional support from the MacArthur Foundation's Norms and Preferences Network. Thanks also to our many research assistants.