Science AAAS

www.sciencemag.org/cgi/content/full/312/5781/1767/DC1

Supporting Online Material for

Costly Punishment Across Human Societies

Joseph Henrich,* Richard McElreath, Abigail Barr, Jean Ensminger Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank Marlowe, David Tracer, John Ziker

*To whom correspondence should be addressed. E-mail: jhenric@emory.edu

Published 23 June 2006, *Science* **312**, 1767 (2006)
DOI: 10.1126/science.1127333

This PDF file includes:

Supplemental Materials for the

Costly Punishment across Human Societies

Methods

In this first section we detail our experimental procedures and protocols, and then describe the collection of some of our key economic variables, which are used in the regression analyses.

Experimental Procedures

Our standardized protocol and script tried to ensure uniformity across sites in a number of important dimensions. First, to encourage motivation and attention, we standardized the stake at one-day's wage in the local economy (except in the U.S., where $100 was used in urban Missouri to take account of the vastly higher cost of living in comparison with rural Missouri, where stakes were set at $50, consistent with minimum wage; $40 stakes were used with Emory students). Second, using the method of back translation, all of our game scripts were administered in the local language by fluent speakers. Third, our protocol design restricted those waiting to play from talking about the game and from interacting with players who have just played during a game session. Fourth, we individually instructed each participant using fixed (1) scripts (2) sets of examples, and (3) pre-play test questions. This guaranteed that all players faced the same presentation of the experiments and that they understood the game well enough to correctly answer two consecutive test scenarios.

Typically, the administration of the game went as follows: A randomly selected group of adults were invited, usually on the morning of the game or the night before, to the location of the experiment (often at a house or village school). Players were told nothing about the experiments before coming, except that (1) their participation was completely optional, (2) they would have an opportunity to obtain some money, and (3) the whole process would take several hours. Once all players had arrived, the game area was secured by the experimental team from the eyes and ears of non-players, a show-up fee was paid (20% of the stake/one-day's wage) and participants selected (randomly from a hat) to determine their order of play. The game script was then read to the whole group. The script included the following points (1) participation is purely optional and people should feel free to leave at any time, (2) people's decisions are entirely private, except to the lead experimenter who will not tell anyone (because most of our researchers were long-term field workers in these locales, players' trust of the experimenters was extremely high), (3) all games will be played only once, (4) players must not discuss the game (research assistants monitored the group for compliance) and (5) all the money is real and people will receive payments to take home at the end of the session. The description of the experimental situation and decision situation was followed by a fixed set of examples, which were illustrated to the group by manipulating bills or coins in the local currency.

After the instructions were read to the group, individual players were brought one-by-one into a separate area, where the game instructions were re-read and more examples were given. Again, examples were illustrated by manipulating cash on a masking tape layout (see image below). If the player confirmed that he or she understood the game and the experimenter agreed, they were given test questions that required them to state the amount of money that each player would receive under various hypothetical circumstances. Players had to correctly answer two consecutive test situations to pass, and be allowed to participate in the experiment (this actually requires four correct amounts to be stated for the DG and UG, and 6 correct amounts for the 3PPG). If a player could not do the required mathematics, they were permitted to manipulate the money according to the hypothetical examples, and then count the money in each pile to answer (thus, everyone had to have the ability to count to 10). After passing this test, players were told their role in the actual game (e.g., Player 2) and were asked to make the required decision(s). If a research assistant was present, he or she had to turn away and would not observe the actual decisions.

Figure 1. Third Party Punishment Game in the village of Teci, on Yasawa Island, Fiji (Photo by Robert Boyd).

As in most behavioural experiments, all participants knew everything about the experimental game, except who was matched with whom. Our script specified that players were matched with another person (or two people in the Third Party Punishment Game) from this village (or other relevant local grouping), but made clear that no one would know who was matched with whom. The script also made clear that the game would be played only once.

In our DG-UG protocol, players first played the Dictator Game through to completion and then immediately played the Ultimatum Game. Player 1s in the DG kept their role in the UG. The inert Player 2 in the DG, before finding out what they received in the DG, assumed the role of Player 2 in the UG. Players in the 3PPG were a fresh sample that had not participated in the prior two games—the 3PPG was usually done weeks later, and in the case of the Tsimane and Au, the 3PPG research was done in a different village from the DG and UG. Here, we largely avoid concerns of the effects of experience in game play by focusing on DG offers (the DG always came first), Player 2 in the UG (who was inert in the DG, and did not learn about how much money he got in the DG until after his UG decision), and Player 3 in the 3PPG, who were also usually first timers.

1762 individuals participated in the games we use in this paper (there were additional contextualized experiments that we do not present here). 962 individuals played only one of the three games. 652 played the DG and UG (in the same role--player 1 or player 2), but did not play the 3PPG. One (1) individual was player 2 in the DG but player 1 in the UG. 147 individuals played all three games. 91 individuals played all three games in the same role (player 1 or player 2). 17 (15) individuals played as player 2 (player 3) in the 3PPG, but were player 1 in the DG and UG. 9 (14) individuals played as player 1 (player 3) in the 3PPG, but were player 2 in the DG and UG. One (1) individual was player 1 in the DG, player 2 in the UG, and player 3 in the 3PPG.

All of the instructions, game scripts, data collection tools, and protocols are available at our project website: http://www.hss.caltech.edu/roots-of-sociality/phase-ii/docs. We encourage others to use these protocols and contribute to the database.

Collection of Economic and Demographic Data

At each site we collected 25 different economic and demographic variables using standardized collection protocols and forms. Relevant to our analysis here we will briefly discuss our measures of income, wealth and household size. Income is an individual measure (unlike wealth and household size) and represents any flow of revenue available to the individual from legal, illegal, formal, and informal sources. Given the likely flux in seasonal income in many places, we measured this in an extensive interview for the previous year (see project website for interview protocol). Wealth is a measure of total productive assets owned by a household. These are revenue generating, or potentially revenue generating, assets, e.g.: farm acreage, livestock, farm equipment (plows, threshers), boats, commercial transport (trucks, ox and horse carts), firearms, etc. A household is defined as a group of people who share in the household estate—that is, a corporate body who may or may not live together (including absent school children, for example), but who share some household accounts and whose members are subject to some decision-making authority by the head/s of household. We used the standard ethnographic techniques of (1) cross-checking informant reports by asking multiple informants the same questions (e.g., independently asking fathers and sons at different times about wealth), and (2) checking free responses (which could be influenced by recall ability) by also listing possible sources of income or wealth from a master list gleaned from a combination of past interviews and observation.

Table S1 provides the summary statistics for the variables used in the regression analyses to follow.

Table S1. Summary Statistics for Demographic and Economic variables used below								
Population	Sex 1 = all female	Education (years)	Household Size	Income (USD)	Household Wealth (USD)	MAO UG	MXAO UG	MAO 3PPG
Accra	0.26 (0.44)	10.15 (3.35)	2.60 (2.09)	529.3 (543.7)	Index	13.00 (17.25)	87.93 (17.19)	26.15 (18.01)
Shuar	0.41 (0.50)	6.21 (3.72)	6.10 (2.23)	737.3 (955.8)	6086.67 (5873.63)	6.50 (13.87)	97.50 (11.18)	19.33 (22.19)
Sursurunga	0.50 (0.50)	6.63 (2.96)	5.53 (2.28)	276.5 (477.5)	5023.75 (5665.90)	24.35 (20.41)	83.68 (23.62)	10.31 (13.32)
Sanquianga	0.57 (0.50)	4.05 (3.14)	6.68 (2.93)	1894.7 (2321.8)	2234.98 (4383.50)	12.33 (18.13)	88.33 (16.21)	23.87 (21.55)
Rural Missouri	0.59 (0.50)	13.71 (2.13)	2.94 (1.22)	24085.4 (18792.7)	115,756 (180,875)	27.86 (19.50)	NA	NA
Urban Missouri	0.54 (0.51)	15.17 (1.95)	3.00 (0.00)	37083.3 (10417.0)	63,9357 (58,077)	NA	NA	NA
Tsimane	0.53 (0.50)	3.60 (3.58)	7.70 (4.00)	127.5 (207.3)	453.85 (290.77)	6.67 (5.40)	100.00 (0.00)	3.91 (7.83)
Maragoli	0.46 (0.50)	12.54 (1.20)	7.16 (1.75)	1192.5 (493.5)	1951.29 (373.39)	30.00 (7.64)	100.00 (0.00)	33.04 (16.63)
Yasawa	0.51 (0.50)	8.39 (2.27)	6.93 (3.24)	1158.7 (1111.7)	423.87 (510.01)	6.47 (13.46)	94.85 (13.26)	5.00 (9.23)
Samburu	0.56 (0.50)	1.38 (2.85)	8.73 (4.78)	359.1 (385.8)	2462.90 (3113.03)	6.13 (12.30)	97.10 (5.29)	18.93 (10.66)
Hadza	0.43 (0.50)	1.24 (2.01)	3.42 (2.01)	0.00 (0.00)	0.00 (0.00)	16.54 (17.42)	100.00 (0.00)	5.65 (13.76)
Isanga	0.53 (0.50)	7.57 (2.27)	5.86 (2.11)	203.9 (309.7)	152.71 (173.89)	7.33 (10.15)	98.33 (9.13)	31.00 (15.86)
Au	0.17 (0.38)	3.28 (3.21)	5.53 (2.07)	41.4 (142.6)	89.21 (52.61)	20.00 (21.01)	92.67 (13.63)	30.67 (19.99)
Emory	0.53 (0.50)	13.00 (0.00)		13859.2 (79199.2)	NA	20.53 (14.33)	100.00 (0.00)	16.00 (19.84)
Gusii	0.47 (0.50)	11.86 (2.55)	7.16 (1.75)	1520.1 (675.9)	6008.03 (1357.68)	38.00 (5.77)	100 (0)	41 (5.48)
Dolgan	0.63 (0.49)	10.10 (1.74)	4.66 (2.06)	1313.8 (1079.1)	Index	14.74 (20.38)	100.00 (0.00)	NA
Nganasan	0.43 (0.50)	9.04 (2.85)	4.71 (2.21)	1191.8 (1491.1)	Index	15.00 (19.58)	95.00 (15.81)	NA

Table S1. MAO-UG is the average minimum acceptable offer in the Ultimatum Game. MXAO-UG is the maximum acceptable offer. MAO-TPP is the minimum acceptable offer in the Third Party Punishment Game.

While we did obtain income and household size measures for all societies, we were able to generate only a 'wealth index' for the Dolgan/Nganasan in Siberia and wage workers of Accra, Ghana. This wealth index permits within-group analyses of wealth effects, but is not comparable with our wealth measurements from other groups, so these populations must be dropped from some of our analyses of wealth effects. Also, note that in the Third Party Punishment Game we did not obtain income or wealth data for the Tsimane'.

The student data from Emory is not used in any of our regression analyses below. Since it has been established that university students have not yet reached their adult-developmental plateau in these game measures (*S1-S3*), our adult (non-student) data from Missouri is the appropriate comparative dataset. Frey and Meier's analysis of a large natural experiment (*S5*) shows a similar age effect as the experimental approaches: controlling for other factors, older individuals are more pro-social than younger individuals. Introducing student data would potentially confound developmental variation with other sources of between-population variation, such as those arising from cultural or economic differences.

Additional methodological concerns

Readers will also have additional concerns about anonymity, how participants interpreted the experiments, and the specific contexts of each field site. The matter of anonymity and its interpretation in these experiments is already much discussed. Recently, several of us wrote about it at length in (*S6*), which details the previous round of experimental games. Many of those concerns stay with this study, so we refer the reader to that article and the citations within it.

For our previous round of experiments, we ultimately published an edited volume (*S7*) containing descriptions of each field site, methodological variations at each site, and individual researcher interpretations of differences among groups. We have already drafted chapters along this same plan for the new data and sites presented here, so we ask the reader curious about additional ethnographic details to wait for the book, as only a book will adequately address those details.

Supporting Analyses

In this section we present a series of supporting regression analyses that show (1) a substantial portion of variation among our population in their willingness to punish in both the Ultimatum Game and the Third Party Punishment cannot be explained by the main effects of measured economic and demographic differences and (2) "hyperfair rejections" (rejections of offers greater than 50%) in the UG likely do not result from confusion about the game. Along the way, we will highlight and discuss any economic or demographic variables that emerge and contribute to explaining the variation in punishment. In particular, we observe that population—not individual—differences in education (mean number of years of formal education) predict more willingness to punish. For rhetorical purposes, we will first focus on the UG data and then on the 3PPG data.

Ultimatum Game

For the Ultimatum Game (UG), our regression analyses examine the predictive capacity of six demographic and economic variables on individuals' willingness to reject both low and high offers. These analyses confirm that the observed variation between populations cannot be primarily explained by economic and demographic differences among our samples. We also assess the possibility that rejections of offers greater than 50% result from some form of confusion about the game by regressing the number of rejections each player made for offers above 50% on their education and the number of examples that were required for the individual to pass our test. This "confusion hypothesis" finds no support.

Minimum Acceptable Offers (MAO) in the UG

To explore the variation in people's willingness to punish low offers we used each Player 2's vector of accept/reject decisions to calculate their minimum acceptable offer (MAO). MAO is the lowest offer—between zero and 50%—that a person will accept. For example, if a player stated they would reject an offer of zero, but then accepted 10 through 50, their MAO is set at 10. If an individual accepted all offers up to and including 50%, their MAO was set at 0. If they rejected offers of 0% through 40% but accepted 50%, their MAO is 50. Under this restrictive scheme it is quite possible for people to produce sets of decisions that do not yield an MAO (e.g., reject 0, accept 10, reject 20…), 96% (434 out of 452) of Player 2s provided decision vectors that readily translated to MAOs—and the missing 18 are spread fairly evenly across our populations. Of these 18 deviant players, 11 were people who rejected everything between 0 and 50 (inclusive).

To study the effects of our economic and demographic measures on MAO, we followed a three step procedure. First, we regressed our MAO variable on the population dummy variables in order to establish the amount of variation among population means (the data from rural Missouri were used as the point of comparison for the other groups). This analysis shows that about 34.4% of variation in MAO arises from differences between population means. In step two we added our measures of sex, birth year (age), education, household size, income (U.S. dollars), and wealth (U.S. dollars) to see what fraction of the variation in MAO within populations can be captured by these variables. Table S2A shows that adding these variables explains an additional 7% of variation, bringing the total variance explained to 41.5%. Finally, we remove the population dummies in order to see how much of the total variation can be explained by our variables, both within and between populations. This explains about 15.8% of the variance (Table S2B), and indicates that a substantial portion of the between population variance is unaccounted for by our economic and demographic predictors.

Table S2A. MAO-UG Regression			$R^2 = 0.415$	$R^2(adj) = 0.380$
N = 302				
Effect	**Coefficient**	**Std Error**	**Std Coef**	**P(2 Tail)**
Constant	94.995	123.086	0.000	0.441
Shuar	-25.959	6.030	-0.354	0.000
Sursurunga	-8.182	6.055	-0.122	0.178
Sanquianga	-19.726	5.990	-0.331	0.001
Tsimane	-24.816	6.629	-0.397	0.000
Maragoli	-6.457	5.576	-0.100	0.248
Fiji	-27.497	5.342	-0.488	0.000
Samburu	-24.851	6.660	-0.417	0.000
Hadza	-13.963	6.627	-0.220	0.036
Isanga village	-26.299	5.600	-0.441	0.000
Au	6.935	8.100	0.054	0.393
Gussi	1.974	5.452	0.031	0.718
Birth Year	-0.033	0.063	-0.027	0.597
Sex (Female = 1)	0.756	1.764	0.021	0.668
Education (years)	0.476	0.341	0.130	0.163
Household Size	0.107	0.271	0.022	0.692
Income (USD)	-0.00032	0.00021	-0.124	0.125
Wealth (USD)	-0.000014	0.00015	-0.050	0.355

Table S2B. MAO-UG Regression			$R^2 = 0.158$	$R^2(adj) = 0.14$
N = 302				
Effect	**Coefficient**	**Std Error**	**Std Coef**	**P(2 Tail)**
Constant	250.295	136.588	0.000	0.068
Birth Year	-0.123	0.069	-0.098	0.078
Sex (Female = 1)	-1.879	1.936	-0.053	0.332
Education (years)	1.401	0.212	0.381	0.000
Household Size	-0.125	0.269	-0.026	0.642
Income (USD)	-0.000	0.000	-0.031	0.662
Wealth (USD)	-0.000	0.000	-0.033	0.598

Focusing first on Table S2B, we observe that the only potentially important predictors of MAO in the UG are education and birth year. An additional decade of formal schooling increases an individual's MAO by 14, while a decade in age increases MAO by 1.2. However, once the population dummy variables are introduced, neither birth year nor education emerges as a powerful predictor. Thus, these appear to be largely between-group effects. Perhaps mean education is correlated with some other variable not in the regression that varies between, but not within, populations.

In order to run the above regression analysis we had to drop the data from Accra (Ghana) and from the Dolgan/Nganasan (Siberia) because we only obtained a wealth index (suitable only for within population analysis), and not wealth measures equivalent to what we obtained elsewhere. To address this, we re-ran the above three step analysis dropping our wealth variable and including Accra and the Dolgan/Nganasan. The population dummies capture 29.2% of the variation—Missouri is again the point of reference. Adding the economic and demographic variables (not including wealth) increases the variation explained to about 35% (Table S3A). Dropping the population dummies shows that the economic and demographic variables alone explain 11% (Table S3B), and leave much of the variation between populations unexplained.

Table S3A. MAO-UG Regression (without wealth)			$R^2 = 0.346$	$R^2(adj) = 0.312$
N = 365				
Effect	**Coefficient**	**Std Error**	**Std Coef**	**P(2 Tail)**
Constant	41.778	121.298	0.000	0.731
Accra	-22.370	5.514	-0.344	0.000
Shuar	-27.376	6.227	-0.341	0.000
Sursurunga	-9.746	6.238	-0.133	0.119
Sanquianga	-20.661	6.158	-0.318	0.001
Tsimane	-25.943	6.657	-0.411	0.000
Maragoli	-6.332	5.830	-0.090	0.278
Fiji	-28.026	5.568	-0.456	0.000
Samburu	-26.236	6.770	-0.404	0.000
Hadza	-16.178	6.726	-0.233	0.017
Isanga village	-27.139	5.821	-0.418	0.000
Au	4.675	8.359	0.033	0.576
Gusii	2.133	5.703	0.030	0.709
Ngsn-Dolgan	-20.434	5.478	-0.305	0.000
Birth Year	-0.005	0.062	-0.004	0.939
Sex (Female = 1)	-0.378	1.655	-0.011	0.819
Education (years)	0.345	0.329	0.091	0.295
Household Size	0.003	0.273	0.001	0.992
Income (USD)	-0.0004	0.0002	-0.141	0.048

Table S3B. MAO-UG Regression (without wealth)			$R^2 = 0.11$	$R^2(adj) = 0.097$
N = 365				
Effect	**Coefficient**	**Std Error**	**Std Coef**	**P(2 Tail)**
Constant	197.821	131.671	0.000	0.134
Birth Year	-0.096	0.067	-0.074	0.151
Sex (Female = 1)	-2.807	1.790	-0.079	0.118
Education (years)	1.166	0.201	0.308	0.000
Household Size	0.047	0.253	0.009	0.854
Income (USD)	-0.000006	0.00015	-0.002	0.971

Again, focusing on Table S3B first, we observed that education remains an important predictor, while birth year, which was marginal at best above, has further weakened. A decade of education predicts an increase in an individuals MAO of 12. Once the population dummies have been entered in the regression (Table S3A), the effect of education again drops (indicating a between-population difference) while income moves into the marginal range, with an additional $10,000 of income creating a drop in MAO of 4. Additional Analysis suggests that this income effect is driven by the joint presence of Missouri and the Au. If the Au are dropped, the income coefficient essentially flips its sign. If Missouri is dropped, the effect vanishes.

Maximum Acceptable Offer in the UG

Using the same approach described above for the MAO, we 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 MXOA was set at 100. If he accepted 50, 60, 70, and 80, but rejected 90 and 100%, his MXAO was set at 80%. As explained above for MAO, it is quite possible for individuals to produce decision strategy vectors that do not fit the assumptions of our procedure. However, we were able to assign MXAOs to 96% of players. Unlike the assignment of MAO, only 2 of the 20 individuals who were unassignable to an MXAO rejected all offer amounts from 50 to 100% (inclusive). Ten of these unassignable players came from the Sursurunga of New Ireland, and five from the Hadza of Tanzania.

Following the three step analytical procedure above, we first regressed MXAO on our population dummies and found them to account for about 17% of the variation. Adding age, sex, education, household size, income, and wealth increases the variance explained to 24% (Table S4A). Removing the population dummies drops the variance explained to 5% (Table S4B), indicating that very little of the between group variance can be explained by differences in our economic and demographic variables. Note, here Missouri is not included because we did not initially consider the possibility that people would reject offers greater than 50%. Thus, the Shuar are used as a standard for comparison for the population dummies variables.[1]

Table S4A. MXAO-UG Regression			R^2 = 0.236	R^2(adj) = 0.188
N = 270				
Effect	**Coefficient**	**Std Error**	**Std Coef**	**P(2 Tail)**
Constant	-148.845	103.808	0.000	0.153
Sursurunga	-17.270	3.591	-0.374	0.000
Sanquianga	-9.255	3.401	-0.246	0.007
Tsimane	-2.479	3.683	-0.063	0.502
Maragoli	3.576	3.937	0.088	0.365
Fiji	-3.516	3.442	-0.098	0.308
Samburu	-3.274	3.520	-0.087	0.353
Hadza	-3.434	3.713	-0.086	0.356
Isanga village	-2.256	3.432	-0.060	0.511
Au	-12.089	5.242	-0.151	0.022
Gusii	7.807	3.869	0.192	0.045
Birth Year	0.129	0.053	0.146	0.016
Sex (Female = 1)	-2.267	1.450	-0.096	0.119
Education (years)	-0.288	0.268	-0.114	0.283
Household Size	-0.043	0.214	-0.013	0.839
Income (USD)	-0.002	0.001	-0.131	0.093
Wealth (USD)	-0.001	0.000	-0.182	0.008

[1] In Missouri we only asked players for their minimum acceptable offer and did not have them go through the entire response vector. Consequently, while it seems very likely that no Missourians would have rejected offer greater than 50% (given the data from Emory and elsewhere in the U.S.), we have omitted them from the following analysis. Assuming all Missourians have an MXAO of 100 only magnifies the above conclusions.

Table S4B. MXAO-UG Regression			$R^2 = 0.05$	$R^2(adj) = 0.032$
N = 270				
Effect	**Coefficient**	**Std Error**	**Std Coef**	**P(2 Tail)**
Constant	17.108	106.350	0.000	0.872
Birth Year	0.040	0.054	0.046	0.457
Sex (Female = 1)	-2.168	1.449	-0.092	0.136
Education (years)	0.278	0.166	0.110	0.095
Household Size	0.177	0.206	0.054	0.389
Income (USD)	-0.0013	0.0009	-0.096	0.156
Wealth (USD)	-0.0005	0.00018	-0.188	0.004

Focusing on Table S4B, we observe that both wealth and education contribute significantly to explaining the variance in MXAO. For every 1000 additional dollars of household wealth, individuals decrease their MXAO by 1 percent. For every decade of formal schooling, individuals increase their MXAO by 2.8 percent. When the population dummies are added to the regression, the wealth effect holds (keeping the same magnitude), birth year becomes significant at conventional levels, income rises to marginal significance, and any education effect again largely evaporate (Table S4A). The coefficient on income is twice the size of that of Wealth, so an additional $2000 in annual income decreases MXAO by 2 percent. An additional decade of life decreases MXAO by 1.3 percent.

For the same reasons described above, we ran the above analyses excluding the data from Accra and from the Dolgan/Nganasan. Here, for MXAO, we again drop our wealth measure and now include both of these populations in the analysis. The Shuar are used as the point of reference for the dummy variables. Regressing MXAO on the population dummies captures 16.6% of the variation. In Table S5A, adding our demographic and economic variables increases the variation explained to 18.5%. Then, in Table S5B, dropping the population dummies reduces the variance explained to 0.9%, demonstrating that little or none of the variation between populations is likely to be explained by our economic and demographic variables.

Table S5A. MXAO-UG Regression (without Wealth)			$R^2 = 0.185$	$R^2(adj) = 0.141$
N = 332				
Effect	**Coefficient**	**Std Error**	**Std Coef**	**P(2 Tail)**
Constant	-62.136	102.746	0.000	0.546
Accra	-9.098	3.578	-0.209	0.011
Sursurunga	-14.852	3.790	-0.280	0.000
Sanquianga	-8.399	3.532	-0.195	0.018
Tsimane	1.478	3.623	0.035	0.684
Maragoli	4.925	3.949	0.105	0.213
Fiji	-1.408	3.394	-0.034	0.679
Samburu	-0.801	3.674	-0.019	0.828
Hadza	0.878	3.793	0.019	0.817
Isanga village	1.007	3.459	0.023	0.771
Au	-7.553	5.480	-0.082	0.169
Gusii	5.691	3.974	0.122	0.153
Ngsan-Dolgan	2.528	3.608	0.057	0.484
Birth Year	0.083	0.052	0.087	0.116
Sex (Female = 1)	-1.155	1.350	-0.047	0.393
Education (years)	-0.195	0.261	-0.073	0.457
Household Size	-0.105	0.213	-0.031	0.623
Income (USD)	-0.001	0.001	-0.057	0.387

Table S5B. MXAO-UG Regression (Without Wealth)			$R^2 = 0.009$	$R^2(adj) = 0.000$
N = 332				
Effect	**Coefficient**	**Std Error**	**Std Coef**	**P(2 Tail)**
Constant	63.361	104.941	0.000	0.546
Birth Year	0.016	0.053	0.017	0.765
Sex (Female = 1)	-1.157	1.366	-0.047	0.397
Education (years)	0.087	0.160	0.033	0.588
Household Size	0.227	0.191	0.067	0.236
Income (USD)	-0.001	0.001	-0.062	0.308

Punishment of "hyperfair offers" does not result from confusion

Given the non-intuitive nature of hyper-fair rejections, we explored the possibility that despite our extensive instruction and rigorous one-on-one testing procedures, those who rejected high offers might have somehow misunderstood the game. For every Player 2 in the UG we summed up the number of rejections each individual made for offers greater than 50% and ran two regressions. First, we regressed this variable on education, suspecting that more educated people might better grasp some elements of the game missed in our testing procedure. The coefficient, standard error, and *p*-value for education were 0.02, 0.016, and 0.21, respectively. Adding population dummies (and using the Shuar as a reference point), to address any between-group variation yields a coefficient, standard error, and *p*-value of 0.04, 0.025 and 0.10 ($n = 383$), respectively. This increases the effect size of education in the direction opposite to that expected by the confusion hypothesis—here, more schooling, if it does anything, favors more hyperfair rejections. Second, we regressed our hyperfair rejections variable on the 'number of examples and test questions used', which records essentially how much effort was required in explaining the game, as it was conveyed through repeated examples and test questions. With dummies entered into the regression to control for any differences in the numbers of illustrative examples used by different researchers, the coefficient, standard error and *p*-value for this predictor are -0.022, 0.047, and 0.63, respectively. Again, the insignificant coefficient here is in the opposite direction to expect by a 'confusion explanation': people who required more examples and test questions had fewer rejections above 50%.

Two additional facts support the claim that punishing hyperfair offers is not the result of confusion or misunderstanding. First, in the third party punishment game, which was generally more difficult to explain and took longer for players to apprehend, people did not punish hyperfair offers (main text, Figure 2). A look at the third party punishment game explains why: if Player 1 offered the full amount (100%) to Player 2, Player 3 cannot punish Player 1 because we did not allow negative payoffs (and he is not permitted to take money away from Player 2). Player 3 could pay 10%, but this would not take any money away from Player 1. If Player 1 had given 90% to Player 2, Player 3 could pay 10% to take 10% away from Player 1, but this is very costly punishment. It is not until Player 1 gives 70% to Player 2, when things aren't all that unequal, that Player 3 can administer the full brunt of his punishment to Player 1. Consequently, punishment was not expected for high offers in 3PPG, and very little was seen. However, if punishment of hyperfair offers in the UG was the result of confusion, we would expect similar confusion in the more difficult 3PPG. Second, post game interviews of players who punished high offers in the UG reveal both that people understood the game (answered factual questions about the game correctly) and made sensible responses as to why they rejected high offers, such as "it was too much, I cannot accept that much." Finally, our findings, that hyper fair rejections are not the product of confusion, are fully consistent with prior efforts to analyze similar findings from Tatarstan and Sakha-Yakutia, Russia (*S4*).

Third Party Punishment

Using the same technique described above for MAO in the UG, we calculated minimum acceptable offer (MAO-3PP) for each Player 3 in the Third Party Punishment Game. While it is quite possible for players to provide strategy vectors that defy assignment by our MAO process, we were able to assign 92% (317 of 346). Ten additional players could have been assigned if we would have permitted an MAO of 60 (94%). The remaining 6% who defy MAO assignment are scattered across all groups (with most groups having between zero and two), although seven such individuals can be found among the Maragoli.

Following the same procedure used above for the UG, we first regressed MAO-3PP on the population dummies and found that 38.2% of the variation occurs between groups (the Shuar were used as the point of reference). Adding our standard set of economic and demographic predictors increases the variation accounted for to 41% (Table S6A). Removing the population dummies drops the variance explained to 11% (Table S6B). Thus, as with second party punishment in the UG, third party punishment varies substantially among populations, and most of this variation cannot be accounted for by differences in our economic and demographic measures. Note, these regressions do not include the Tsimane or Accra, as we lack comparable wealth data from these. Also, recall that we do not have 3PPG data from the Dolgan, Nganasan, or Missouri.

Table S6A. MAO-3PPG Regression			R^2 = 0.41	R^2(adj) = 0.37
N = 241				
Effect	**Coefficient**	**Std Error**	**Std Coef**	**P(2 Tail)**
Constant	147.049	168.990	0.000	0.385
Sursurunga	-9.604	4.953	-0.170	0.054
Sanquianga	3.764	5.680	0.066	0.508
Maragoli	9.588	6.125	0.147	0.119
Fiji	-16.286	5.509	-0.264	0.003
Samburu	-0.450	5.973	-0.007	0.940
Hadza	-11.154	5.897	-0.171	0.060
Isanga	11.835	5.625	0.171	0.036
Au	14.924	5.973	0.194	0.013
Gusii	17.796	5.534	0.307	0.001
Birth Year	-0.067	0.086	-0.048	0.436
Sex (Female = 1)	0.557	2.260	0.015	0.805
Education (years)	0.367	0.436	0.089	0.401
Household Size	0.437	0.416	0.064	0.295
Income (USD)	0.001	0.001	0.037	0.579
Wealth (USD)	0.000	0.000	0.008	0.904

Table S6B. MAO-3PPG Regression			R^2 = 0.10	R^2(adj) = 0.08
N = 241				
Effect	**Coefficient**	**Std Error**	**Std Coef**	**P(2 Tail)**
Constant	397.9	181.3	0.000	0.029
Birth Year	-0.197	0.092	-0.141	0.034
Sex (Female = 1)	0.417	2.514	0.0109	0.868
Education (years)	1.01	0.269	0.245	0.00022
Household Size	0.479	0.437	0.070	0.274
Income (USD)	0.0016	0.00095	0.113	0.0876
Wealth (USD)	-0.00015	0.00028	-0.0359	0.583

Starting with Table S6B, we observed that three of our dependent variables show some predictive power. Paralleling the finding from the analysis of MAO-UG, an additional decade of formal schooling again predicts an increase in MAO-3PP of 10. An additional decade of life predicts an MAO-3PP increase of 2. Income is marginally significant, with an additional $1000 of income increasing MAO by 1.6. None of these predictions remain significant once the populations dummies are entered (Table S6A), suggesting that these variables are picking up between between-group differences.

To incorporate the Tsimane and Accra data, we dropped both the wealth and income variables. Regressing MAO-3PP first on the population dummy variables captures 37.9% of the variation. Adding the demographic variables increases the variance explained to 40.5% (Table S7A). Dropping the dummies reduces the variance explained to 9%. Thus, most of the variation between groups in MAO-3PP remains unexplained.

Table S7A. MAO-3PPG Regression (without Wealth or Income)			R^2 = 0.405	R^2(adj) = 0.37
N = 305				
Effect	**Coefficient**	**Std Error**	**Std Coef**	**P(2 Tail)**
Constant	200.575	143.903	0.000	0.164
Sursurunga	-9.893	4.875	-0.159	0.043
Sanquianga	4.949	5.036	0.079	0.327
Maragoli	9.734	5.492	0.135	0.077
Fiji	-15.623	5.028	-0.233	0.002
Samburu	0.645	5.453	0.010	0.906
Hadza	-12.643	5.466	-0.176	0.021
Isanga	11.291	5.309	0.147	0.034
Au	13.942	5.628	0.164	0.014
Gusii	18.335	5.181	0.287	0.000
Tsimane	-15.842	5.158	-0.220	0.002
Accra	4.411	4.998	0.077	0.378
Birth Year	-0.093	0.073	-0.066	0.207
Sex (Female = 1)	-0.623	1.873	-0.016	0.740
Education (years)	0.365	0.320	0.091	0.254
Household Size	0.017	0.359	0.003	0.963

Table S7B. MAO-3PPG Regression			R^2 = 0.09	R^2(adj) = 0.078
N = 305				
Effect	**Coefficient**	**Std Error**	**Std Coef**	**P(2 Tail)**
Constant	532.877	158.649	0.000	0.001
Birth Year	-0.264	0.081	-0.187	0.001
Sex (Female = 1)	-0.331	2.164	-0.009	0.878
Education (years)	0.966	0.226	0.241	0.000
Household Size	-0.117	0.364	-0.018	0.747

Focusing on Table S7B, we observe that again education and birth year are significant predictors of MAO-3PP. An additional decade of formal schooling predicts an increase in MAO-3PP of about 10. An additional decade of life predicts an increase in MAO-3PP of 2.6 percent. Once the population controls are entered (Table S7A), none of our demographic predictors are significant, again suggesting that these variables are picking up between-group differences.

Dictator Game Results

Figure 2 shows the distributions of offers in the Dictator Game, our measure of altruism in one-shot anonymous interactions. The horizontal axis gives the possible offers as a percentage of the total stake, with the area of the circles at each offer amount displaying the proportion of the sample that made that offer. Overall, of our 428 DG offers (excluding Emory students), 5.4% (23) were zero, 30.4% (130) were 50/50 splits, 85.8% occurred between 10% and 50% (inclusive), and only 8.7% were greater then half the stake (38 offers, 21 of which were at 60%). Our populations differed in modes, means and standard deviations. Mean offers ranged from about 26%, among the Tsimane and Hadza, to the high 40's in Missouri and among the Sanquianga. Modal offers are zero among the Hadza, 10% for Tsimane, 30% for Gusii, and 50% for half of the societies studied (with some groups showing multiple modes). The standard deviation in offers varies across societies from 5.4 among the Gusii farmers in the highlands of Kenya to 24 among Emory freshmen and 25 among the Hadza—the mean standard deviation is 17.6.

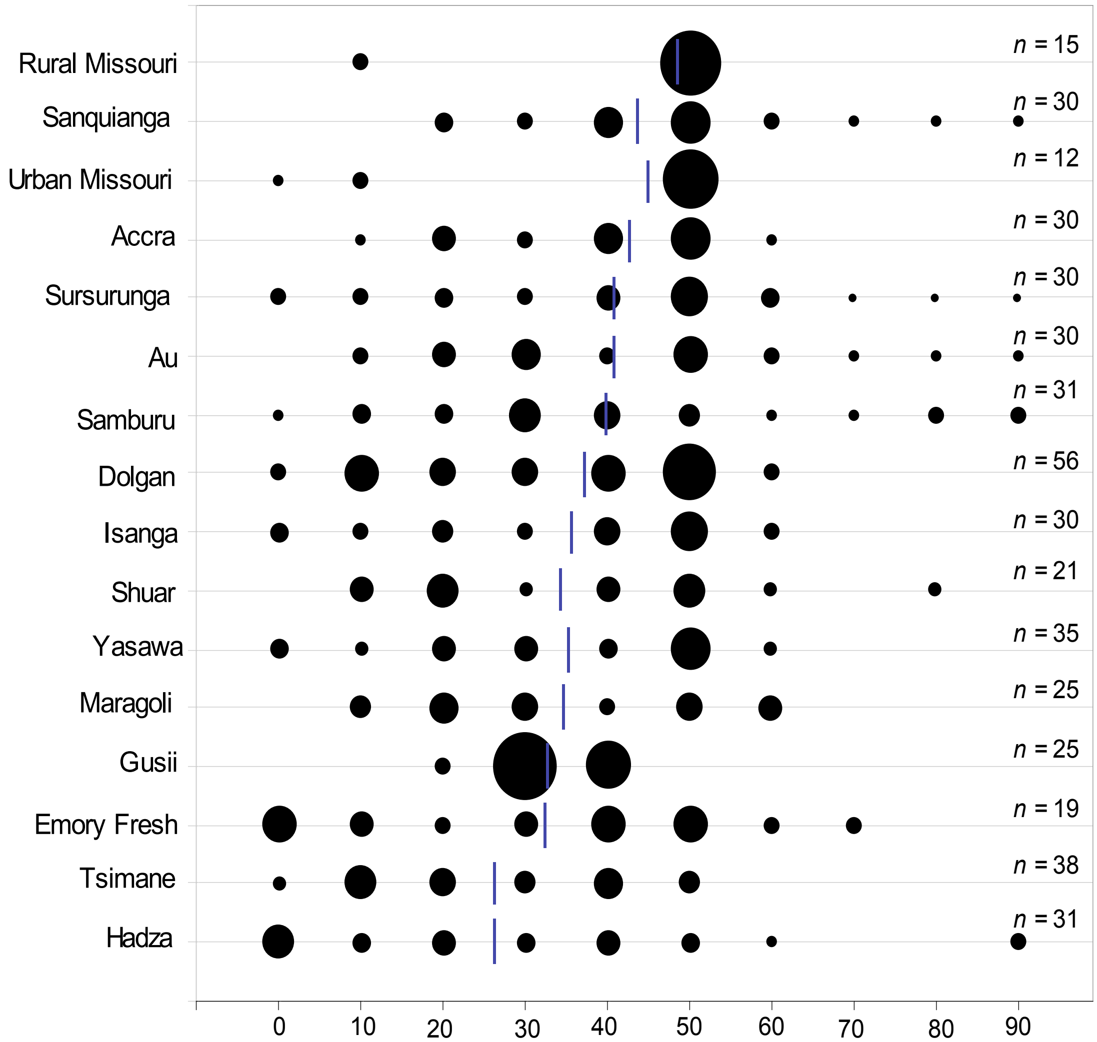

Figure 2. The distribution of offers in the Dictator Game.
Reading horizontally for each of the populations listed along the left vertical axis, the area of each bubble represents the fraction of our sample who made that offer, so each horizontal set of bubbles provides the complete distribution of offers for each population. The blue slash gives the mean offer for each population. The *n* values on the right side provide the number of pairs.

Relationship between punishment, fairness and altruism

To explore the possibility that a willingness to administer costly second and third party punishment may have culturally coevolved with notions of fairness and altruism, Table S8 presents the Pearson correlations between the mean offers for each population in each of our three experiments and three statistics measuring (in some fashion) each population's willingness to administer punishment, for both the UG and 3PPG. These three statistics are the mean MAO, the 80th percentile MAO, and the 90th percentile MAO. The last two statistics present the offer amounts at which a Player 1 could guarantee an eight or ninety (respectively) percent chance of not being punished. We have provided these statistics because it remains an unresolved theoretical issue as to what measure of a populations' willingness to punish should drive the emergence of fairness and altruism.

Table S8. Pearson correlations, sample sizes (in parentheses), and 95% confidence intervals (in brackets)

Population Statistics	Mean DG Offer	Mean UG Offer	Mean 3PPG Offer
Mean DG Offer	1.0	---	---
Mean UG Offer	0.81 (14) [0.46 - 0.88]	1.0	---
Mean 3PPG Offer	0.64 (12) [0.19 - 0.77]	0.57 (12) [0.18 - 0.75]	1.0
Mean MAOUG	0.14 (14) [-0.14 - 0.37]	0.14 (14) [-0.07 - 0.33]	0.33 (12) [0.0002 - 0.55]
80th Percentile MAOUG	0.39 (14) [-0.04 - 0.62]	0.42 (14) [0.07 - 0.63]	0.41 (12) [-0.07 - 0.66]
90th Percentile MAOUG	0.33 (14) [0.01 - 0.66]	0.41 (14) [0.18 - 0.65]	0.29 (12) [-0.04 - 0.71]
Mean MAO3PP	0.37 (12) [0.03 - 0.55]	0.17 (12) [-0.04 - 37]	0.52 (12) [0.17 - 0.70]
80th Percentile MAO3PP	0.57 (12) [0.14 - 0.73]	0.30 (12) [-0.007 - 0.47]	0.68 (12) [0.25 - 0.79]
90th Percentile MAO3PP	0.50 (12) [0.04 - 0.72]	0.25 (12) [-0.06 - 0.58]	0.63 (12) [0.23 - 0.79]

Before highlighting the relationship between punishment and fairness/altruism, we first note the general consistency of these findings. Results from both the UG and 3PPG show that the greater the punishment in a population, the higher the offers. In the 3PPG, the MAO3PP (punishment) statistics are all positively correlated with 3PPG offers, with values ranging from 0.52 to 0.68. Similarly in the UG, the MAOUG statistics are all positively correlated with mean UG offers, with values ranging from 0.14 to 0.42. Consistent with the idea that at least some players are seeking to avoid punishment, the 80th and 90th percentile statistics are substantially better predictors of offers than the mean MAO.

Using DG as a direct measure of altruism toward anonymous others, Table S8 shows that all six of our punishment statistics positively correlate with mean DG offers, with correlations ranging from 0.13 to 0.57. While all of the 3PPG measures show 95% confidence intervals that do not include zero, two of the MAO-UG statistics do slightly overlap zero. Since MAO in the UG (second party punishment) may combine both motivations for punishing norm violations (in this case equity norms) and motivations for revenge, due to a personal monetary loss, there is some reason to anticipate the stronger and tighter relationship between third party punishment (MAO in 3PPG) and altruism (DG offers). That is, because coevolutionary theories of cooperation are based on motivations for punishing norm/equity violation—and not for revenge for personal affronts—the additional potential for revenge motivations in the UG complicate the linkage between punishment and altruism. Our mean MAO-UG and MAO-3PP are correlated 0.55 (CI: 0.33-0.68). Thus, from the perspective of coevolutionary theory, our measure of third party punishment is the best measure to relate to DG-altruism.

References

S1. J. R. Carter, M. D. Irons, *Journal of Economic Perspectives* **5**, 171 (1991).

S2. W. T. Harbaugh, K. Krause, S. G. Liday. (2002).

S3. M. Sutter, M. Kocher. (Jena, 2003).

S4. D. L. Bahry, R. K. Wilson, *Journal of Economic Behavior & Organization* **60**, 37 (2006).

S5. B. S. Frey, S. Meier, *Economic Inquiry* **41**, 448 (2003).

S6. J. Henrich et al. *Behavioral and Brain Sciences* **28**, 795 (2005).

S7. J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, eds. Foundations of Human Sociality. Oxford (2004).