www.sciencemag.org/cgi/content/full/327/5972/1480/DC1

Supporting Online Material for

Markets, Religion, Community Size, and the Evolution of Fairness and Punishment

Joseph Henrich,* Jean Ensminger, Richard McElreath, Abigail Barr, 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: henrich@psych.ubc.ca

Published 19 March 2010, *Science* **327**, 1480 (2010)
DOI: 10.1126/science.1182238

This PDF file includes:

SUPPORTING ONLINE MATERIALS FOR THE MARKETS, RELIGION, COMMUNITY SIZE AND THE EVOLUTION OF FAIRNESS AND PUNISHMENT

EXPERIMENT METHODS

In this first section we detail our experimental procedures and protocols. The procedures, protocols and scripts themselves, as well as data collection sheets and a data entry template (with error checking) can be found at http://www.hss.caltech.edu/~jensming/roots-of-sociality/ (scroll down to Phase II). We encourage others to use these protocols and contribute to the database.

BASIC GAME DESCRIPTIONS

In the Dictator Game (DG), two anonymous players are allotted a sum of real money (the *stake*) in a one-shot interaction (S*1*). The first player—Player 1—can offer a portion of this sum to a second player, Player 2; offers are restricted to 10% increments of the stake. In the DG, both players are told the total size of the stake given to the pair, that he/she is paired with an anonymous player, and that Player 1 has the job of deciding how the stake is divided between the two players. Player 2 is passive in this game and merely receives what is offered. In this one-shot anonymous game, a purely self-interested Player 1 would offer zero; thus, offers in the DG provide a measure of a kind of behavioral fairness that is not directly linked to kinship, reciprocity, reputation, or the immediate threat of punishment.

The Strategy Method Ultimatum Game (UG) also involves a stake that is allocated to the pair, with Player 1 given the first of two moves in deciding how the money will be allocated between the two anonymous players (S*2*). In this game the actual offer made can be rejected by Player 2, in which case neither player receives anything. The "strategy method" extension of this game refers to the fact that Player 2, before hearing the actual amount offered by Player 1, must decide whether to accept or reject each of the possible offers, and that these decisions are binding. We discovered during the pilot for this experiment in rural Missouri that there was a tendency for Player 2s to misunderstand the nature of the UG and engage in what appeared to be "demand" behavior. In other words, some players thought that they were actually influencing Player 1's behavior by their rejections. To drive home the fact that Player 1 had already made the offer, we explained that Player 1's offer was written on the turned over paper in front of the experimenter. So as we ran through the 11 possible offers that Player 1 might have made, and elicited Player 2's acceptance or rejection, they were reminded that the actual offer made was already marked on the paper in front of the experimenter and that they would be bound by the responses they were giving to each of the conceivable offers.

If Player 2 specified that he would accept the amount of the actual offer, then he received the amount of the offer, and Player 1 received the rest. If Player 2 specified that he would reject the amount actually offered, both players received 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 non-zero 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 third experiment, the Third Party Punishment Game (TPG), two players are allotted a sum of real money (the *stake*), and a third player gets one-half of this amount (S*3*). 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 20% of his allocation to punish Player 1, causing Player 1 to suffer a deduction of 30% of the stake from the amount he kept for himself. Like the strategy method in the UG, we elicited responses from Player 3 to the complete range of possible offers by Player 1. For example, suppose the stake is $100: if Player 1 gives $10 to Player 2 (and keeps $90 for himself), and Player 3 says he 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 be $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 his taste for a second type of costly punishment, *third party punishment*.

GAME PROCEDURES AND PROTOCOLS

GOALS AND SUMMARY

Our standardized protocols and scripts 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. This is much higher than that typically used in university labs. 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 had just played during a game session. Fourth, we individually instructed each participant using fixed scripts, sets of examples, and 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.

FIGURE S1. JOE HENRICH AND A FIJIAN ASSISTANT ADMINISTER THE THIRD PARTY PUNISHMENT GAME IN THE VILLAGE OF TECI, ON YASAWA ISLAND, FIJI

(PHOTO BY ROB BOYD).

SAMPLING

Some researchers already had village censuses for their sites. Others had to generate them in order to draw a random sample of adult participants for the experiments. In deference to community norms of fairness, some researchers limited the draw to one per family until all families were included, and then allowed random selection of second family members. Almost all invited individuals did participate unless they were away from the village. In societies where large numbers of residents were involved in inflexible work schedules that might preclude attendance, an effort was made to schedule games at times that would be convenient to more people. We did not have a village sampling strategy within sites; each researcher was left to their own discretion.

Overall, our samples are highly representative of the communities from which they are drawn, since it was rare that someone who was selected declined to participate if they were present. People were generally enthusiastic about participating.

COLLUSION AND CONTAGION

We knew from our collective experiences that collusion and contagion among closely-knit communities were potential risks in the kinds of experiments we were running. In experiments independent of this project that were run subsequent to Phase 1 of this project, several members of our team had experience with collusion among some players in large villages when they ran significant numbers of experiments over successive days in the same village. However, we never experienced collusion or contagion between villages. The good news is that once collusion occurred, behavior in the games changed so dramatically that it was immediately apparent to the experimenters that the games had been corrupted. Instead of appearing confused, unsure, tentative, and struggling to understand the game during the training, players announced immediately how they wished to play, and there was nearly complete compliance with the collusion, so the usual variation in offers vanished. The demeanor of the players changed in a manner that was instantly transparent to the experimenters. In these cases (not part of Phase 1 or Phase 2 of this project), some young village activists had organized the population and told everyone to play 50/50 in the Ultimatum Game. There was nothing subtle about it; the fact that virtually everyone complied with the collusion is interesting in itself. In these cases the games were terminated, the data were thrown out, and all experiments ceased in the affected villages. We suspect that not all small-scale societies are equally susceptible to this sort of community collusion, but we deemed it an extremely serious threat, and took the lessons of those experiences to heart when we set up the protocols for Phase 2 of the project.

FIGURE S2. DAVID TRACER ADMINISTERED THE THIRD PARTY PUNISHMENT GAME AMONG THE AU OF PAPUA NEW GUINEA

To avoid this problem, we took extraordinary care to ensure that the risk of collusion was minimized. The ideal solution is to play each game on only one day in each village, and to keep all participants isolated, or unable to communicate about the game for the duration of the actual play. But given the sample sizes we required, and the length of play, it was not possible to accomplish this in one day. We weighed the risks of contagion against the problems of using a second village for the second half of the sample. Sometimes village effects can be high, so we agreed to try to use one village for all of the DG/UG games. To minimize the likelihood of contagion, we did not inform players on the first day that there would be a second day of games, and we endeavored to provide as short notice as possible to those who were randomly selected to play the second day; these individuals were notified of their invitation either the evening before the game or the morning of the game. We detected no signs that contagion or pollution were problems. Furthermore, checks on order of play in the dataset revealed no signs of order or day effects.

INFORMED CONSENT

We knew in advance that many of the subjects would be illiterate and would not be able to read descriptions of the research and sign consent forms. So in place of this, at the start of each session, the participants were told that if at any point they became uncomfortable with any aspect of the games they were being asked to play they were free to leave at any time.

GATHER PARTICIPANTS IN ONE PLACE

All invited participants were told where and when to show up for the game. Where possible, we used community structures like schools; otherwise we used clusters of local homes. Numerous local research assistants were employed to control the logistical flow once the game began, to monitor the groups to prevent discussion of the game, and to conduct the requisite surveys. Those who had completed playing the game were allowed to depart if they were finished and had been paid, or were assigned to a separate waiting area so that they could not interact with those waiting to play. Participants were allowed to talk amongst themselves, but they were monitored constantly and not permitted to talk about the game.

BACK TRANSLATION OF SCRIPTS

Each researcher had to translate the game scripts into the appropriate local language. Researchers used the method of back-translation to obtain the best possible game translation. This involved having one bilingual assistant with no knowledge of the game translate the game instructions into the local language, and a second translate it back, thus identifying any problems in translation. Scripts are available on the aforementioned http://www.hss.caltech.edu/~jensming/roots-of-sociality/.

GAME ADMINISTRATOR

All game instructions were read by native speakers unless the project researcher was fluent in the local language. If a local interpreter was used, he/she was asked to turn around when offers were made so that no local individuals had knowledge of the players' offers. This helped guarantee the anonymity that the games assured to participants.

SHOW-UP FEE

Before the game began, players in the Dictator, Ultimatum, and Punishment Games were all given a "show-up" fee paid in cash at a rate of approximately 20-25% of one day's wage in the local economy. In some sites, because of the vagaries of currency denominations, these percentages diverged a bit. It was made clear to the player that this money was strictly for their *participation* in the game, and it was not part of the game. Participants who failed to pass the required tests of game understanding were allowed to keep the show-up fee—which made it somewhat easier to reject them, if the need arose.

STAKES

The game stakes were set at roughly one day's minimum wage in the local community (i.e. the rate ordinarily paid for casual wage labor work if it were available). For rural Missouri this amounted to $50 games, while in many of the developing societies the stakes were in the range of $2. Because stakes had to be divisible into units of ten, some sites wound up with stakes that were marginally higher or lower than the daily wage rate. In many of our societies casual wage work is hard to come by, so if anything, our stakes were effectively more valuable than one day's casual wage.

TEACHING EXAMPLES

In both teaching and testing the participants, researchers used actual coins and paper currency to illustrate the game (see Figure S1). By presenting the arithmetic visually, people with limited or no arithmetic skill could still understand the game. If necessary, players could manipulate piles and count coins or bills during decision-making and testing. Specific teaching examples were scripted in the written protocols (see below). Analyses show that the number of examples a player required does not predict decisions (S*4*).

GAME LOGISTICS FOR THE CORE GAMES

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%-25% of the stake/one-day's wage), and participants randomly selected from a hat an ID number that determined their order of play. The use of the "hat" was meant to convey clearly that order and assignment to the role of Player 1, 2 or 3 were random.

The game script was then read to the whole group. The script included the following points (1) participation was purely optional and people should feel free to leave at any time, (2) people's decisions were entirely private, except to the lead experimenter who would not tell anyone (because most of our researchers were long-term field workers in these locales, players' trust of the experimenters was high), (3) all games would be played only once, (4) players must not discuss the game (research assistants monitored the group for compliance), and (5) all the money was real and people would receive payment 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 visually 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 table or cloth with a line indicating Player 1's versus Player 2's holdings. 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 TPG). If a player could not do the required arithmetic, 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 (that is, whether they were Player 1, Player 2, or Player 3) and were asked to make the required decision(s). If a research assistant was present, he or she had to turn away and did not observe the actual decisions.

As in most behavioral 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(s) from the same village, 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 passive 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 TPG were a fresh sample that had not participated in the prior two games—the TPG was usually done weeks later, and in the case of the Tsimane and the Au, the TPG research was done in a different village from the DG and UG.

If one is concerned about the effects of placing a DG before the UG, consider four things. First, this procedure was uniform across sites[1], and the findings within this project are comparable in this regard. Second, our UG findings in this project are comparable to our previous project despite the fact that our UG then was not preceded by a DG. Previously, the UG was the first and often only game played. Third, all our findings for offers hold up in both the DG and UG. Further, the market integration finding with the UG replicates a prior finding with the UG, without a preceding DG. Third, with regard to rejection in the UG, the second player in the UG was inert in the DG and had not yet been paid. Finally, in a control with university students in the U.S., the UG findings are quite similar to typical UG findings in the literature that were conducted without a preceding DG (S*5*).

GAME DAY 1

The goal for day 1 was two sessions with the *same* 20-30 players playing both the Dictator Game (DG) and the Strategy Method Ultimatum Game (UG). The morning session was the DG and the afternoon session was the UG. Each player assumed either the role of Player 1 or Player 2 in both games (Player 1 in the DG and Player 1 in the UG, or Player 2 in DG and Player 2 in UG). Individuals were not permitted to cross roles—e.g., Player 1 in DG and Player 2 in the UG. Two holding areas were utilized, one for players who had not played the current game, and one for those who had finished, but who were being held for the next game or waiting for payment.

In the morning DG-session (which was shorter than the afternoon UG session), people were gathered together, paid the show-up fee, explained the game in a group, and brought individually into the gaming area. The random order of play was demonstrated overtly by picking names from a container. Players were told up front that they would be paid for this game after completing the next game.

After the DG session, the UG session was run as soon as possible. Researchers were encouraged to supply food and drink to keep players comfortable. In the UG session, players were told that this was an

[1] Only in Missouri, which was a pilot for this study, were the UG and the DG run on different samples.

entirely different game, and they would be playing with a different person. Again, researchers overtly established a new random order of entry into the gaming area. After playing, they sent all players to a holding area, and brought them in one-by-one for payment for both games after this second session was complete.

People were not informed that there would be more games on the following day. Either that evening or the following morning, notice was sent to a new group of 20-30 players inviting them to show up for a game.

DAY 2 AND DAY 3

Researchers repeated the same procedure as on Day 1. The minimum goal was 30 pairs total, so our estimates suggested that the games would run for 2 or 3 days, with a strong preference for only 2 days (15 pairs per day) to minimize contagion.

We believed that our post game questions had the potential to generate pollution, so we agreed to administer them only during the last (DG-UG) session, and only if we were not planning to run the TPG game in that village.

THIRD PARTY PUNISHMENT (BEGIN NEXT DAY OR WAIT 3 OR MORE WEEKS)

Our goal was 30 trios for the Third Party Punishment Game (TPG). Researchers used mostly fresh players who had not played any of the prior games. This required some researchers to use a new village. For those running short of fresh players, it was acceptable to substitute repeat players for Player 2 (who are passive in this game), but not for Player 1 or Player 3. This required a minimum of 60 fresh adults in one village, plus 30 repeat players. If this was not possible, we substituted another 30 repeat players in the role of Player 1, but only if they had played that role before. It was essential that only fresh players be used in the role of Player 3.

We used the same 2-session day to administer the games, with 8 trios per session, 2 sessions per day, and two days to completion.

ORDER OF SUPPLEMENTAL GAMES

Those running supplemental games finished all core games (DG, UG, TPG) before playing any of the supplemental games. If there was contagion, we wanted to be sure that the core games were as uncorrupted as possible. Our supplemental games included double-blind DGs (S*6*), contextualized games (“dressing” the game up in locally relevant contexts)(S*7*), and a simplified trust game combined with a social network analysis (S*8*).

TSIMANE CHECK

Among the Tsimane several methodological anomalies are potential causes for concern (S*9*). In one case, players received game instructions late in the day and had to be sent home, only returning the next day to actually play. In the second case, a game was administered on Sunday, directly after a church service. Since the Tsimane had particularly low offers and little punishment, this might be important. As a check, we have re-run all of our baseline regressions for UG, DG and TPG offers, as well as all offers together and both MAOs. These checks reveal no qualitative divergences from our baseline findings, with two exceptions. First, the *p*-value for the coefficient on MI in the baseline regression for UG offers increases to become only marginally significant at 0.13 when the Tsimane are dropped. Second, the *p*-value for the coefficient on WR in our baseline DG offer regression improved from 0.079 to 0.02.

While we felt it was important to perform these checks, we emphasize that independent of these analyses there are several reasons to believe that the Tsimane data were little impacted by these procedural anomalies. First, both DG and UG findings are directly in line with prior work using these experiments among the Tsimane, though with different protocols (S*9, S10, S11*). Second, prior work by a different researcher among a very similar Amazonian population yielded quite similar UG findings (S*12, S13*). Third, neither of these methodological anomalies is typically thought to drive offers down, which is what is necessary for them to be skewing our findings. Pre-play communication, which might have occurred when players went home for the night, typically drives offers up, not down, based both on much previous work with typical subjects (S*14*) and upon the experiences of two of our own researchers in work prior to this project. Similarly, given our findings and recent work in psychology (S*15*), attending a Catholic church service prior to playing ought to increase offers, if anything. In short, there is little reason to suspect these anomalies affected play, and if they did, there is reason to suspect those effects would work against our market integration hypotheses.

POPULATIONS

In this section we discuss the selection of our field sites and the immense diversity of life ways represented in our sample. Figure S3 presents the global distribution of our populations.

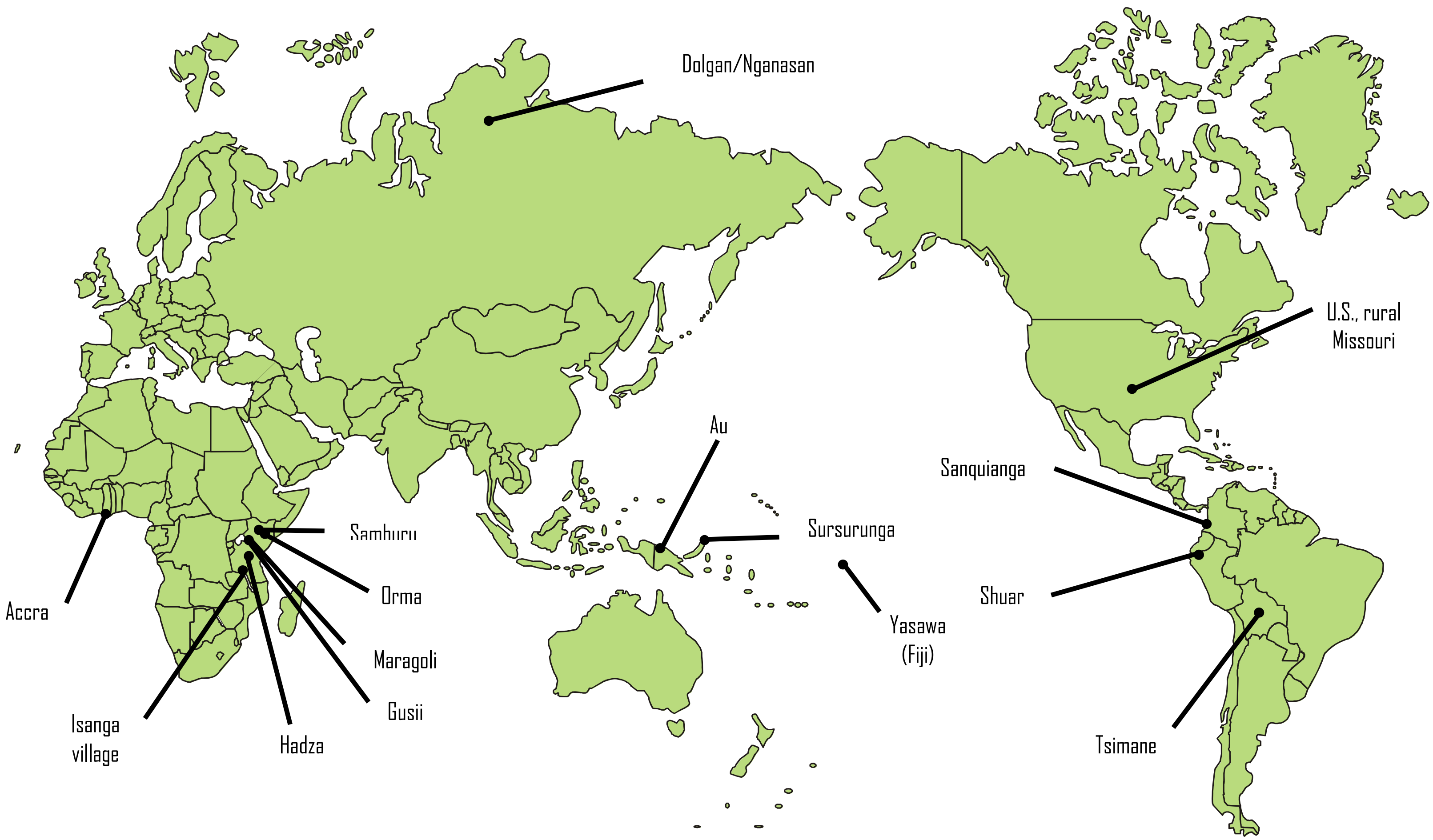


Figure S3. The global distribution of our populations.

SITE SAMPLE SELECTION

Spurred by our preliminary finding in Phase I of this project, we again hypothesized, as we did in the proposal for Phase I, that market integration would be a crucial variable in explaining fair mindedness and punishment behavior in our study. We were cognizant of this as we attempted to build a sample that captured the full range of market integration; however, there was a considerable degree of opportunism in our selection of researchers. We needed seasoned researchers with particular scientific skills and preferably with long-term field sites with which they were familiar and would therefore be capable of executing a complex protocol. We also needed researchers eager and available to undertake the research within a relatively tight time frame. These are severely limiting constraints and narrowed the candidates to a small number of anthropologists and an even smaller selection of economists. This resulted in more of an opportunistic, rather than a systematically stratified, sampling of societies. That said, we got lucky, and did end up with a nicely diverse sampling of small-scale human societies.

We were in no way attempting to make our sample of sites representative of the world today, but we did wish to create a distribution that spanned the range of variation in institutions such as the market and supporting institutions such as the rule of law, as well as capturing different types of economic systems (foragers, herders, subsistence farmers, cash crop farmers, etc.). Given that our new formal measure of market integration would not exist until after we had sent the researchers to the field, we had to hope that we managed to guess correctly and create a reasonable distribution. As it turned out, our sites nicely spanned the range of market integration (more on this variable below). The mean level of market integration for our Phase 2 sample sites, as measured by the percentage of calories purchased in the market is shown in Table S1. Given the constraints enumerated above, this probably represents as good a distribution across the spectrum as we could have hoped to achieve in the absence of such measurements *ex ante*.

In our site selection we also gave consideration to geographic and ethno-linguistic distribution, lest we confound our sample with over-representation of a unique geographical or cultural idiosyncrasy. As was the case in Phase 1 of the project, we achieved considerable geographic and ethno-linguistic dispersion in this phase (see Figure S3). We have seven African societies, two Papua New Guinean, one Oceanic, one Siberian, three Latin American, and one rural U.S. Africa is dominant in this phase of the project, just as Latin America was more prominently represented in Phase 1 of the project. However, our African sample also includes great diversity: one fully urban population (Accra), one peri-urban farming society (Isanga), two rural cash crop farming societies (Maragoli and Gusii), two pastoral societies (Samburu and Orma), and one hunter-gatherer society (Hadza).

In one significant respect, we went out of our way to alter the sample selection from the criteria specified above. In Phase 1 of the project, Tracer (S*16, S17*) produced an unusual finding among the Au and the Gnau of Papua New Guinea. In the Ultimatum Game people made some offers of greater than fifty percent of the pot to their partners, and these offers were refused. Tracer (S*17*) interpreted these findings as culturally consistent with the norms in a gifting society where people attempt to incur the indebtedness of others by giving out unsolicited gifts. By refusing such high offers, the Au and the Gnau are resisting the indebtedness that would typically be associated with the receipt of such gifts. We found this result intriguing. It ran counter to our overall market integration finding because the Au and the Gnau are not highly market integrated, yet they made quite high offers in the UG. The Au were retained in the sample during Phase 2 of the project, and we also considered the case interesting enough to warrant the inclusion of a second society from New Guinea (from the Melanesian island of New Ireland) to see if we would find the same behavior there. Bolyanatz (S*7*) was recruited and ran experiments with the Sursurunga, who are located quite geographically far from the Au, but share this cultural trait. As reported herein, not only did the Au behave in the same unusual manner, replicating Tracer's earlier experiments (S*18*), but the Sursurunga behaved in the same unusual manner (S*7*). These two sites represent two out of our sample of fifteen. Consequently, we significantly over-represented these societies in our world sample, and because they run counter to our overall finding of the relationship between market integration and fairness, they make it more difficult for us to achieve statistical significance in this result, though once again we have done so.

Table S1. Sites and Mean Market Integration

Society	Mean % of diet purchased
Hadza	0
Au	1
Tsimane	7
Yasawa	21
Shuar	22
Sursurunga	24
Gusii	28
Maragoli	43
Dolgan/Nganasan	63
Samburu	69
Isanga	70
Orma	72
Sanquianga	82
Accra, Ghana	100
Rural Missouri	100

INDIVIDUAL SOCIETIES AND RESEARCH SITES

Our sites span the spectrum of human variation in market integration and degree of incorporation in modern states. Many of our small-scale societies are sufficiently remote that the formal institutions of the states under which they reside have minimal impact. For example, primary education is legally mandated in Kenya, but only about one-third of school age Orma girls attend school; female circumcision is also illegal in Kenya, but it is universally practiced among the Orma, as it is among many other Kenyan ethnic groups. In many of the remote populations we studied, the state does not have either the will or the institutional capacity to enforce its authority. Local norms and institutions still predominate in many of these societies, and these norms are reflected in the variation we see across the behavior we studied in the games. Table S2 provides an overview of the societies in the study.

FORAGING AND HORTICULTURAL SOCIETIES

Not surprisingly, the societies in our study that are most remote from national institutions tend also to be those most reliant on subsistence production; they also tend to live in small communities and represent very small ethnic groups. The Hadza of Tanzania (S*19*) are our only purely foraging society where the population subsists on their own hunting and gathering production for most of their calories. Their camps average about 30 people and are quite nomadic, moving 4-10 times per year; the entire ethnic group of Hadza numbers about 1000 individuals; none practice a world religion.

Two other societies in our sample still rely on foraging, but are dependent upon horticulture for much of their calories, and to varying degrees engage in some household-level cash crop farming. The Au of New Guinea (S*18*) and the Sursurunga (S*7*) of New Ireland (an island province of Papua New Guinea) are our two "gifting" societies, and they both fall into this group. The Au number about 10,000. The Au depend almost exclusively on food garnered from foraging and in their small slash and burn gardens. They are five hours walk from the nearest market. But the Au do devote about fifty percent of their land to cash crop farming and have some wage employment, demonstrating some of the complexities that begin to confound a simple categorization of even highly remote, small-scale societies. The Sursurunga number only 3000. They also subsist primarily on swidden agriculture, including some cash cropping, but unlike the Au, they purchase 24 percent of their foodstuffs in the market. A highway bisects many Sursurunga villages, so this population is more closely tied to a market center. Both the Au and the Sursurunga in our sample are virtually 100 percent Evangelical Protestants.

The Tsimane of Bolivia (S*9*) number about 9000 and are, like the Hadza and the Au, a population that depends upon the market for little of their daily food. The Tsimane still acquire about 30 percent of their food from hunting, fishing and gathering and depend upon swidden farming for almost all of the rest. The Tsimane in our sample are fifty percent Catholic and fifty percent Evangelical Protestants. The Shuar (S*20*) represent our second Amazonian population. The Shuar have historically lived a lifestyle similar to that of the Tsimane, with low population density and household units loosely clustered in villages. However, this has recently changed considerably with the encroachment of roads and towns, and the current economic situation of the Shuar is very much in flux as greater access to markets has changed the local economic and settlement patterns. This process is far from played out, however, as the Shuar still purchase only 22 percent of their food and depend upon their own hunting, gathering and horticultural production for the rest. Our Shuar sample is roughly fifty percent non-Evangelical Protestants, twenty-five percent Catholic, and twenty-five percent practice no religion, making them one of our most religiously diverse populations.

The inhabitants of the island of Yasawa in the northwest corner of the Fijian archipelago complete our sample of forager/horticultural societies (S*5*). Like the Shuar and the Sursurunga, the Yasawa depend on the market for approximately a quarter of their foodstuffs. The rest comes from fishing, horticulture (sometimes requiring slashing and burning), and gathering. The Yasawa are fairly isolated from national spheres of administration, including courts and police, and have their own village level governing institutions (chiefdom). Wage work tends to be scarce, but opportunities for trading and employment in the tourist industry are present for those few with the requisite skills. The Yasawa are roughly two-thirds non-Evangelical Protestants and one-third Evangelical.

Our Siberian sample is drawn from two ethnic populations, the Dolgan and the Nganasan (S*21*), who historically were, respectively, reindeer pastoralists (who also hunted and traded) and reindeer hunters. Today, they live a largely sedentary life in town and all depend upon local hunting, fishing, trapping, and a combination of wage work and state-provided social security pensions. While the Dolgan are historically Russian Orthodox, the Nganasan do not practice a world religion.

Table S2. Ethnographic Summary of Societies

Group	Nation/Region	Language Family	Environment	Economic Base	Residence	Researcher
Accra City	Ghana	Mixed	Urban	Wage Work	Sedentary	Barr
Au	Papua New Guinea/Sepik	Torricelli/ Wapei	Mountainous Tropical Forest	Foraging/ Horticulture	Sedentary	Tracer
Dolgan/ Nganasan	Russia/Siberia	Turkic/Samoyedic	Tundra-Taiga	Hunting/Fishing/ Wages	Semi-sedentary	Ziker
Gusii	Kenya	Ekegusii	Fertile High Plains	Mixed Farming/ Wage Work	Sedentary	Gwako
Hadza	Tanzania	Khoisan/Isolate	Savanna-Woodlands	Foraging	Nomadic	Marlowe
Isanga	Tanzania	Bantu	Mountainous Forest	Agriculture/ Wage work	Sedentary	McElreath
Maragoli	Kenya	Logoli	Fertile Plains	Mixed Farming/ Wage Work	Sedentary	Gwako
Orma	Kenya	Cushitic	Semi-arid Savanna	Pastoralism	Semi-nomadic	Ensminger
Samburu	Kenya	Nilotic	Semi-arid Savanna	Pastoralism	Semi-nomadic	Lesorogol
Sanquianga	Columbia/Pacific Coast	Spanish	Mangrove Forest	Fisheries (Fish, clams, shrimp)	Sedentary	Cardenas
Shuar	Ecuador/Amazonia	Jivaroan	Tropical Forest	Horticulture	Sedentary	Barrett
Sursurunga	Papua New Guinea/ New Ireland	Austronesian	Coastal Tropical Island	Horticulture	Sedentary	Bolyanatz
Tsimane	Bolivia/Amazonia	Macro-Panoan Isolate	Tropical Forest	Foraging/ Horticulture	Semi-nomadic	Gurven
U.S., Rural Missouri	U.S./Rural Missouri	English	Prairie	Wage Work	Sedentary	Ensminger
Yasawan	Fiji/Yasawa Island	Oceanic	Coastal Tropical Island	Horticulture/ Marine Foraging	Sedentary	Henrich

FARMING AND WAGE WORK

Three African societies make up our sample of sedentary, non slash and burn farming populations with limited or no dependence upon foraging. These populations are quite distinct in many other respects from the foraging and horticultural populations discussed above. They depend upon the market to a considerably greater degree than our horticultural societies, and they are more highly educated than the societies discussed above. The Gusii and the Maragoli of Kenya (S*22*) both inhabit productive agricultural zones that lend themselves to cash crop farming. Both populations suffer from severely high population pressure that has forced them to rely upon education as a means toward wage employment, as the land cannot support future generations in agriculture. Because of the highly productive nature of the land, however, the percentage of their diet that is purchased is relatively low (28 and 41 percent, respectively). Our Gusii sample is all Evangelical Protestants and our Maragoli sample is all non-Evangelical Protestants.

The mixed ethnic residents of Isanga village (S*23*) in Tanzania inhabit a peri-urban community. The community from which this sample is drawn has a population of 1500 and is only a mile from the regional capital of Mbeya, a major center of trade and commerce. Despite the peri-urban environment, most of this Isanga population still farm small plots that manage to provide roughly 30 percent of their food needs. The rest is purchased in the market and supported by their significant involvement in wage work and business activities. The Isanga sample contains roughly ten percent Muslims and ninety percent Protestants of mixed denominations.

LIVESTOCK HERDERS

Two populations make up our sample of societies that depend largely upon livestock herding. Both have histories of being nomadic and have similar levels of market dependence today, and it is near the high end of our sample—around 70 percent. Both pastoral societies come from Kenya and their subsistence activities are similar, though their ethno-linguistic origins are completely different. The Samburu of Kenya (S*6*), who are closely related to the better known Maasai, are Nilotic, while the Orma of Kenya are Cushitic (S*24*). Both groups herd cattle, small stock, and small numbers of camels. Among both groups, many families today are largely sedentary, though their herds and young men are not. Families typically supplement their lifestyles by marketing livestock, local and migratory wage employment, and local trading businesses. Like the Shuar, the Samburu have considerable religious diversity: forty-eight percent of our Samburu sample is Catholic, fourteen percent are non-Evangelical Protestants, and thirty-four percent practice their indigenous religion. The Orma made a one hundred percent conversion to Islam in the last century.

EXTRACTION OF NATURAL RESOURCES

One of our societies, the Sanquianga of Columbia (S*25*), depends upon extracting natural resources in the form of logging mangrove poles, shrimp, clams, and fish. The population is heavily involved in trading activities and purchases 82 percent of its daily calories. Despite this, the population is not particularly well educated or much involved in wage labor. This Afro-Columbian population actually resides inside what is now a national park, and as a consequence, there are increasing efforts on the part of the national government to regulate their natural resource extraction activities.

INDUSTRIAL SOCIETIES

We have two industrialized populations: one from an urban site in a developing country, and one from a rural site in a western developed country. Our sample from Accra, Ghana (S*26*) is different from our others in that the sample is drawn from a "work community" centered around small firms in the city of Accra, rather than drawn from a residential community. The experiments were carried out in small firms among largely urban immigrants to the mixed ethnic, bustling city of 2 million inhabitants. They are entirely dependent upon the market for their subsistence and have ready access to public transportation, newspapers, radio, and television. The institutions of the Ghanaian state are very much a part of their lives.

Our second fully integrated market sample comes from rural Missouri in the U.S. (S*27*). They were drawn from a town of 1800 where virtually all families are known to each other, and the population is in the heart of the Bible belt. Despite their location, farming is actually a relatively rare form of livelihood; most depend upon wage or small business employment and a few commute to much larger cities for work.

COLLECTION AND OPERATIONALIZATION OF KEY VARIABLES

At each site we collected 25 different economic and demographic variables using standardized collection protocols and forms. Relevant to our analysis here we discuss our measures of age, sex, education, income, wealth, household size and market integration. A complete list of the variables we collected can be found in our template, and our protocols and worksheets for collecting and calculating wealth and income can be found at http://www.hss.caltech.edu/~jensming/roots-of-sociality/ (scroll down to Phase II).

AGE, SEX, AND EDUCATION

In some of our societies ages are not precisely known, but they can usually be estimated using well-established anthropological techniques, often in consultation with other community members who know relative birth orders and can relate these to known historical events or the births of individuals with known birthdates. We measured age in years. Sex was recorded for all participants by visual observation.

For education, some of our researchers interpreted years of education to mean the level completed, and others to mean the total number of years attended (including repeated grades). Any difference here would only create a small discrepancy between sites based on a small number of individuals who had to repeat grades. Since one year of formal education, measured either way, is not equivalent across sites, we standardized our education measures by subtracting the mean value of education in the population and dividing by the standard deviation in education for each population. This allows us to make the best use of within population variation we have in education for most of our sites.

HOUSEHOLD SIZE, WEALTH, AND INCOME

We defined a household as a group of people who shared in the household estate—that is, a corporate body which might or might not live together (including absent school children, for example), but who shared some household accounts, and whose members were subject to some decision-making authority by the head/s of household. The number includes absent members because such individuals may make future claims (for example, for land, livestock, school fees, or bridewealth) upon the estate. It might also include large extended families. For example, polygynous households are often under the decision-making authority of one person with authority to buy and sell land or livestock. Similarly, married sons might also be under the authority of such a person, as might "retired" mothers and fathers of the husband and/or wife(s). This measure was an integer number of individuals.

In addition to defining our variable "household size (HS)," the definition of the household had a bearing upon our wealth measure, because we measure wealth at the household level. We calculated household wealth as a cash equivalent of all revenue generating assets (other than human capital) owned by the household. Because of their complexity, these data were often collected from heads of households in separate surveys not administered on the day of the economic experiment. In some of our societies, such as the Hadza, there was no revenue generating capital, as there was little private property. In most of our societies the bulk of household wealth was held in either land or livestock. However, in some of our populations there were a broad range of assets, from farm equipment to boats and rental property. We did not include the value of non-productive assets in our computation of wealth. For instance: jewelry, radios, bicycles, watches, houses, and household goods, were not counted in our wealth measure.

In contrast to wealth, income was collected for each individual who played a game, and it was that specific individual's income that we used in the regression analyses. Income was defined as the flow of revenue available to the individual from legal, illegal, formal, and informal sources. Given the likely flux in income seasonally, each researcher attempted to get an estimate of annual income taking into account seasonal fluctuations.

Wealth and income are challenging variables to collect under the best of circumstances. We used the techniques of disaggregation by local categories and relevant time periods to be as inclusive as possible, and to facilitate recall. Surveys were created locally that disaggregated all known sources of income and wealth and requested amounts in easily known time periods. Income sources on a weekly, monthly, or one-off basis were then aggregated by the researcher into an annual figure. To verify accuracy we used a variety of standard ethnographic techniques such as cross-checking informant reports by asking multiple informants the same questions (e.g., independently asking fathers and sons at different times about wealth). The fact that most of our researchers were long-time intermittent residents of these communities greatly facilitated the collection of accurate wealth and income data. Most of the income we did record derived from wage work (casual and professional), trading profits, sale of home production, rental income, and remittances.

WORLD RELIGION

Religion has long been considered a potentially important variable in explaining social behavior and cooperation (S*28, S29*), and recently has been shown to influence experimental measures of fairness and honesty (S*15, S30, S31*). Thus, in designing our study we wanted to consider religion. To pursue this, we created a binary variable by coding those who professed belief in Islam, unspecified Christianity, Protestantism, Evangelical Protestantism, Russian Orthodoxy, or Catholicism, as believers in a world religion (setting this variable to 1). No other world religions appeared in our sample. Those who practiced an indigenous religion, and those who professed no religion, were coded as not being members of a world religion (setting this variable to zero). At the end of this document we summarize recent work that theorizes why one might expect differences between world religions and other traditional or local religions.

In our small-scale societies, it is important to note that even among those who profess adherence to Christianity or Islam, the actual composition of people's supernatural beliefs and ritual practices are syncretic with local traditions, beliefs and practices. Most people believe in witchcraft or sorcery, and have usually only added the Abrahamic gods (Allah, Jesus, Yahweh, etc.) to their pantheon. The beliefs and rituals of world religions are merely layered on top of, or inter-meshed with, local belief systems that include other deities, naturalistic spirits (e.g., forest, tundra, and lake spirits), ghosts, ancestor gods, and a variety of rituals. This is the case for all of our societies, to varying degrees.

Important, and consistent with typical findings for small-scale societies (as summarized in the final section of the document), none of these societies traditionally had high, moralizing gods, re-enforced by regular communal rituals, and underwritten by potent afterlife rewards or punishments. High moralizing gods are powerful deities who are concerned about the moral behavior of humans, or at least of the god's people, and willing to intervene in human affairs. These gods provide incentives in the form of punishments and rewards for "moral" behavior (by whatever local definition is relevant). The closest approximation among our traditional religions is probably Nkai, the monotheistic, creator god of the Samburu—see below.

Because more than 20% of the populations of Hadza, Shuar, Nganasan and Samburu report no adherence to a world religion, we will discuss each at greater length here. The Hadza all practice a traditional religion. While they possess little cosmology, Hadza do recognize a high god, Haine, with two manifestations, the sun (Ishoko, female) and the moon (Seta, male). While this being may be perceived as a high god, he or she is not engaged in moralizing or active in human affairs in any way. Haine does not punish or reward behavior of any kind. There are no frequent rituals associated with these beings, or glorious afterlife to be obtained with proper behavior. This is typical of the religions of small-scale foragers. Efforts by Christian missionaries over decades to convert the Hadza have repeatedly failed to make any long term impact.

Interestingly, there is one case in Hadza society of supernatural punishment. Older Hadza men or experienced hunters pass into a social status called *epeme*. *Epeme* males are privileged to eat certain kinds of meat. *Epeme* males cannot eat this meat alone, but only with other *epeme* males. Non-*epeme* individuals cannot even watch *epeme* males eat this meat, let alone partake. Violations of these rules result in illness or death. This punishment applies only to *epeme* violations, and is not associated with any supernatural being.

Traditionally, Shuar religion lacked any gods, at least in the usual sense. Shuar, both traditionally and currently, all believe in a wide range of malevolent forest spirits and ghosts. Illness and death is sometimes caused by the action of one of these beings, or by spiritual attacks made by shamans, but these are not due to improper (immoral or unethical) behavior. There seems to be no connection between the ethical and supernatural realms. There are no traditional regular communal rites of worship, or blissful afterlife beliefs. More recently, Christian beliefs, as well as those of germ theory, have come to predominate in most of the population, though our study reveals that a minority still do not report Christianity as their religion. Even for this minority, however, it is important to realize that evangelical North American missionaries have been hard at work in Amazonia for at least 50 years, so all Shuar have heard something of Christianity.

Many Nganasan reported only adherence to the Nganasan religion, and not the Russian Orthodox faith reported by others in their community. Nganasan do believe in two high gods, an Earth Mother and a Sky Father. However, rather than intervening to sustain prosociality among unrelated individuals, these beings are viewed neutrally, as enforcing their own arbitrary rules disconnected from the realm of human relationships. Perhaps more relevant, this belief system also includes spirits of the tundra, fire, lakes, trees, and reindeer, who do punish and reward inappropriate uses of natural resources, according to local norms and expectations. These incentives, which do seem to impact and organize resource use (S*32*), do not extend beyond particular resources or local context-specific distributional practices.

The indigenous Samburu religion is by far the closest approximation to modern world religions (S*33*). Traditionally, Samburu believe in and pray to Nkai, a potent, monotheistic, creator god who whose presence is greater on mountain tops. However, Christian Samburu, who mix and confuse Nkai with the Abrahamic Gods, have added regular prayer rituals (on Sundays), the construction of special ritual locations (churches), an expansion of the moral circle to include non-Samburu, and the afterlife beliefs common in Christianity.

Table S3: Mean Demographics by Society

Society	Market Integration	Std. Dev (N)	World Religion	Std. Dev (N)	Community Population	Std. Dev (N)	Female Proportion	Std. Dev (N)	Age	Std. Dev (N)	Education	Std. Dev (N)	House-hold Size	Std. Dev (N)	Income USD	Std. Dev (N)	Wealth USD	Std. Dev (N)
Accra	100	0.0	0.97	0.2	44	19.8	0.26	0.4	32	10.9	10.1	3.3	2.6	2.1	529	544	.	.
		590		589		590		589		589		589		588		589		0
Au	1	5.2	1.00	0.0	309	45.7	0.17	0.4	38	11.7	3.3	3.2	5.5	2.1	41	143	89	52.6
		120		145		145		145		78		145		135		120		120
Dolgan/Nganasan	63	29.3	0.59	0.5	612	0.0	0.53	0.5	39	12.7	9.6	2.4	4.7	2.1	1257	1278	.	.
		60		41		60		60		60		58		60		60		0
Gusii	28	5.1	1.00	0.0	4063	727.3	0.47	0.5	45	9.7	11.9	2.5	7.2	1.7	1520	676	6008	1357.7
		140		140		140		140		140		140		140		140		140
Hadza	0	0.0	0.00	0.0	43	24.9	0.43	0.5	37	15.0	1.2	2.0	3.4	2.0	0	0	0	0.0
		116		116		116		116		116		116		114		116		116
Isanga	70	36.9	0.99	0.1	1500	0.0	0.53	0.5	37	12.1	7.6	2.3	5.9	2.1	204	310	153	173.9
		100		100		100		100		100		100		100		100		100
Maragoli	43	8.6	1.00	0.0	3843	1148.5	0.46	0.5	46	8.4	12.5	1.2	7.2	1.7	1193	494	1951	373.4
		140		140		140		140		140		140		140		140		140
Orma	72	3.3	1.00	0.0	125	31.8	0.68	0.5	39	13.0	0.0	0.0	8.7	3.7	106	168	1447	1781.2
		38		43		38		38		37		38		37		38		38
Samburu	69	20.9	0.66	0.5	2000	0.0	0.56	0.5	38	14.6	1.4	2.8	8.7	4.8	359	386	2463	3113.0
		151		117		123		117		117		117		120		117		121
Sanquianga	82	16.3	0.84	0.4	1931	400.1	0.60	0.5	38	15.4	4.0	3.0	6.8	2.9	1853	2419	2400	4727.9
		156		155		156		156		155		156		156		156		156
Shuar	22	0.0	0.76	0.4	498	1410.2	0.41	0.5	39	16.1	6.2	3.7	6.1	2.2	737	956	5962	5876.8
		49		49		49		49		47		47		49		49		49
Sursurunga	24	0.0	1.00	0.0	186	23.5	0.50	0.5	36	13.4	6.6	3.0	5.5	2.3	276	477	5024	5665.9
		125		125		125		125		125		125		125		125		125
Tsimane	7	4.0	1.00	0.0	314	130.5	0.53	0.5	35	15.4	3.6	3.6	7.7	4.0	128	207	454	290.8
		146		146		146		146		146		135		145		71		65
U.S./Rural Missouri	100	0.0	1.00	0.0	1813	0.0	0.59	0.5	47	17.5	13.7	2.1	2.9	1.2	24085	18793	115757	180875.4
		82		82		82		82		82		82		81		82		82
Yasawa	21	0.0	1.00	0.0	109	22.4	0.51	0.5	38	15.3	8.4	2.3	6.9	3.2	1159	1112	424	509.9
		105		105		105		105		104		105		105		103		101
Total	57	39.5	0.89	0.3	1028	1402.9	0.42	0.5	37	13.7	7.5	4.8	5.3	3.4	1646	6065	9129	52022.1
		2118		2093		2115		2108		2036		2093		2095		2006		1353

MARKET INTEGRATION

We had theorized the potential importance of market integration in the initial proposal of the project. At that time, we had in mind operationalization such as the percentage of the diet purchased in the market, but we had no precise measures, only the ethnographers' guesses. Despite this relatively crude measure in Phase 1, we found a remarkably strong statistical relationship between this ranking and UG offers. In Phase 2 we aimed to dramatically improve upon the rigor of our market integration measure. The team agreed that the percentage of the diet purchased in the market was probably the best measure of the degree of market integration, as it provided a clear measure of how much the group depended upon market exchange even for their basic subsistence. It is common for food expenditures in developing nations to make up 60 to 80 percent of a household's monthly expenditures, so measuring food expenditures is central to understanding market reliance. Salt, sugar, cooking oil, rice, and flour are often the first things purchased when cash arrives.

However, to err on the side of caution, we delineated four other measures of market integration (MI 2-5) that might be used to form a stronger composite measure, and each researcher measured these also. In the end, MI2 through MI5 were not used since each had peculiarities or were inapplicable in one or a few sites. All of them captured some aspect of the market phenomena, but none of them were as widely applicable and readily comparable across populations as our MI1 measure (percentage of the diet that is purchased).

We also experimented with a composite index, but it was difficult to design one that did not involve arbitrary judgment calls, and that did not create redundancy with the more general measure captured in MI1. Based upon considerable reflection and analysis of the diverse field sites represented herein, we are confident that MI1 alone captures this concept better than any of the other measures singularly, or collectively.

<u>MI1: The percentage of the diet purchased in the market.</u> This turned out to be the best variable measuring market integration for a variety of reasons noted above, and was the one we used in the analyses.

To create this measure we carried out a consumption survey at all sites. Our locally constructed consumption surveys disaggregated all typical food consumption for the area and recorded quantities consumed by source (purchased versus home production) for the entire family. Our format was a twenty-four hour recall of all food consumed in the household, with follow up questions confirming whether the day was typical, and if not, what a typical day did look like. If seasonal variations existed, multiple surveys were filled to create an average. These quantities were then converted to caloric equivalents and the ratio of purchased versus home grown caloric consumption yielded the percentage of market dependence.

We left it to the discretion of researchers whether they carried out the consumption survey individually with each person playing the game, or whether they carried out approximately 20 random surveys of each community where games were played and created an average market integration measure for participants from that community. Use of an average per site rather than an individual measure of MI is justified on both operational and theoretical grounds. Operationally, households had to be surveyed over time and many sites experience cyclical or erratic short-term fluctuations in market access. This means that an individual household-level measure of MI1 would be affected by this kind of noise. Since the overall sample of households spans across these temporal fluctuations, a community average gives a more accurate picture of people's longer-term MI value. Theoretically, we are aiming for a measure related to norms, which are a group-level phenomena. This measure is a proxy for the institutional or normative environment within which people reside. It is intended to capture the degree of average dependence upon the market, but we also expect that this measure captures many other institutional qualities, such as the strength of rule of law that supports that level of market integration, and which we have not directly measured. We expect that the norms we are ultimately seeking to measure are shared broadly in the community, and we do not predict that these vary within communities to the extent that individuals' dependence upon the market may vary within the same community. It is conceivable, however, that whole communities within an ethnic group might vary significantly in their market dependence, as for instance might be the case if one community was near a road and a market center, and another was remote from both. If researchers drew experimental subjects from different communities, those communities were surveyed separately for market integration and that variation was captured. For the main theoretical findings in the main text and below, one can use MI at the individual, community, or population level, but we used the community average because we deemed it to be theoretically the most appropriate.

MI2: Individual income from wage labor, rental properties, and trading activities for the game player. This measure included everything captured in our income variable, with the exception of income from home production, which we wished to isolate as a somewhat weaker measure of involvement in the market. We originally identified wage labor and trading as important components of strong market integration, as they involve different market skills and broader networks. As it turned out, MI2 is correlated 0.99 with income, which is already in our regressions. That is, essentially our regressions contain both MI1 and MI2 (which themselves are correlated 0.19).

MI3: Frequency of wage labor in the last month. Each game player was asked about their level of participation in wage labor in the last month. Most individuals in our games reported no wage labor in the previous month, so we had insufficient variation to test the impact of this measure. In designing the project, we thought it might be possible to use it as part of a composite variable, but in hindsight, it is clear that distortions would be created in fully market oriented societies, such as our Missouri sample, when we consider how it would have treated unemployed wage workers, retirees, and non-working spouses. Moreover, comparing MI1 and MI3 (correlated at 0.57) shows that populations become market integrated (as measured by MI1) long before they become wage laborers.

MI4: Trips to market in the last 7 days. This measure was proposed because it seemed plausible that proximity to a market center would be relevant to the level of market integration (S*10, S34*), and that those who frequented such locations might be more attuned to the market. However, while this measure seems to capture what we were going for in some sites, it does not work well to capture much of the phenomena of interest in many other sites. Whether people shop once a week or twice daily may be a function of cash flow problems, the division of labor in the household, or merely individual preferences, as we suspect it is in developed societies. Those in wage work, arguably the most market integrated in our societies, may not be the individuals who do the daily shopping. MI2 is correlated with MI1 at 0.21.

MI5: Frequency of trading goods for purchase and resale in the prior month. Like MI2 and MI3, this was an effort to capture a portion of what is involved in a market economy. The number of people in our sample involved in such activities was relatively small across our sites, and all the more so when one surveyed only the prior month. This yielded a heavily skewed distribution at zero, which made this an ineffective variable on its own, and added to the challenge of creating a viable composite variable that accurately captured the concept in which we were interested. The correlation between MI1 and MI5 is -0.098.

CALCULATING MINIMUM ACCEPTABLE OFFERS (MAO)

For our measures of punishment we collected a vector of rejection/punishment decisions from Player 2s in the UG and Player 3s in TPG. To analyze these vectors we converted to a single number called a minimum acceptable offer (MAO) for each game. A MAO is the lowest offer—between zero and 50%—that a person will accept. For example, if a player stated they would reject (UG) or punish (TPG) an offer of zero, but then accepted 10 through 50, their MAO is set at 10. If an individual accepted all offers (i.e., did not punish in TPG) 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…). In the UG, we converted 387 vectors out of 405 vectors, giving a conversion rate of 96%. In TPG, we were able to convert 305 out of 339 vectors into MAOs (90% conversion rate). A majority of the individuals who could not be converted were those who rejected or punished all offers of 50% and below. Mostly these individuals are scattered around our populations, though for TPG seven come from the Maragoli.

SUPPLEMENTARY ANALYSES

In this section we present and discuss additional analyses that support the primary findings presented in the main text. We discuss analyses for our five experimental measures, DG offers, UG offers, TPG offers, UG MAO and TPG MAO. We also present analyses of all offers compiled together. For each of these six dependent variables we discuss five supplemental analyses. First, building from the baseline model presented in the main text we examine the effects of adding continental control variables as controls for the possibility that populations on the same continent may share a deep cultural phylogeny that reduces the independence of our populations. Note that while we felt it was worth performing these analyses as a check, it seems very unlikely that there is much shared cultural phylogeny given the quite different languages spoken by our populations on the same continent. Second, we explore the robustness of our conclusions by systematically dropping non-significant predictors from the baseline specification. Third, to address the fact that our conversions to U.S. dollars using international exchange rates may distort the real wealth and income difference among individuals, we re-calibrated those variables using conversions to standardized commodities. Fourth, in another check of income and wealth measures, we standardized our income and wealth measures to generate locally relative measures. We also discuss analyses in which we partitioned between population measures of income and wealth within group variation, using both mean income and mean wealth with within-site relative measures. Fifth, we present regressions that make the very conservative assumption that only our populations (and not our individuals) are statistically independent observations. For our analyses of offers only, we present regressions that include each population's mean MAO as a predictor of offers, along with our baseline of predictors. For the MAO analyses, we first discuss our theoretical preference for using the natural logarithm of community size (LNCS) as a predictor of punishment (and not offers), and then in paired tables (20-25 A and B) present our support regressions with community size (Table A) and LNCS (Table B). Note, we choose to use community size in the main text for simplicity of exposition.

Before proceeding to these analyses, however, we first present Table S4, a compilation of summary statistics, and then discuss why the use of population or group fixed effects is not appropriate in this study.

Table S4: Mean Summary Statistics on Offers and Rejections by Society

Society	DG Offer	Std. Dev (N)	UG Offer	Std. Dev (N)	TPG Offer	Std. Dev (N)	UG P2 Minimal Acceptable Offer	Std. Dev (N)	TPG P3 Minimal Offer Not Fined	Std. Dev (N)
Accra	42	16.9	44	15.9	28	16.8	13	17.3	28	17.7
		30		30		39		30		36
Au	41	19.6	44	14.5	33	23.5	20	21.0	31	20.0
		30		30		30		30		30
Dolgan/Nganasan	37	20.8	43	16.2	.	.	17	20.2	.	.
		30		30		.		26		.
Gusii	33	5.4	40	4.5	36	9.4	38	5.8	41	5.5
		25		25		30		25		30
Hadza	26	25.3	26	16.6	26	19.4	17	17.4	8	15.0
		31		31		27		26		24
Isanga	36	18.3	38	12.6	33	17.1	7	10.1	33	14.5
		30		30		20		30		19
Maragoli	35	17.1	25	15.6	34	20.8	30	7.6	33	16.6
		25		25		30		25		23
Orma	42	15.0	.	.	.	.	.	.	.	.
		26		.		.		.		.
Samburu	40	23.2	35	19.1	31	18.0	6	12.3	19	10.9
		31		31		30		31		26
Sanquianga	47	15.6	48	10.1	43	16.0	12	18.1	24	21.6
		30		30		32		30		31
Shuar	35	19.1	37	16.5	37	17.9	7	13.9	19	22.2
		21		21		15		20		15
Sursurunga	41	18.6	51	16.3	37	18.9	25	20.6	10	14.6
		30		30		32		21		25
Tsimane	26	15.5	27	11.1	20	13.3	7	5.4	4	7.8
		38		36		27		33		23
U.S./Rural Missouri	47	10.3	48	10.3	.	.	28	19.5	.	.
		15		26		.		28		.
Yasawa	35	17.9	40	17.5	27	20.4	7	13.8	4	7.8
		35		34		30		32		23
Total	37	18.9	39	16.5	32	18.6	16	18.0	22	19.3
		427		409		342		387		305

WHY NO FIXED EFFECTS FOR POPULATIONS OR GROUPS?

Our central hypotheses are about differences in local norms for dealing with infrequent exchange partners, particularly in monetary transactions. Theoretically, norms can be interpreted as a local stable equilibrium that arises from social interactions, meaning norms are group- or population-level phenomena. Individuals have motivations and beliefs that reflect the local norms to the degree that such motivations foster local coordination, allow individuals to avoid sanctioning or exploit signaling opportunities. The group level nature of norms means that much of the measurable differences in norms, as opposed to individual-level differences created by a lack of willpower or particular individual circumstance, will occur among local groups (e.g., different villages), or among different ethnic populations.

This brings us to the use of fixed effects. Data analysts seeking to assess the predictive power of variables across different populations often use dummy variables for local groups or ethnic populations to control differences among these groups or populations that are not accounted for by the other predictors in the model. These dummy variables absorb all the variation that occurs between groups or populations, meaning the other variables are left *only* to explain the variation *within* the populations or groups. However, since norms are stable equilibria within groups, much of the variation in norms—as opposed to individual deviations from norms—is between groups or populations. Thus, using fixed effects is a mistake if one is studying norms.

Moreover, as mentioned above, in some cases we were only able to obtain a community level average of market integration (MI), while in other cases we sought to average out noise introduced into our household-level MI variable by short-term (theoretically irrelevant) temporal fluctuations in household measures of MI. The absence of individual level measures of MI precludes the possibility of running fixed effects models with our most significant variable included in the regressions. As a check on this, for those sites where we had both, we broke our MI variable up into two parts for each individual, the mean MI for the individual's community and the difference between their own (their household's) MI and their community mean MI. Across a full range of model specifications, the mean MI (used in the main text and below) does all the explanatory work. The coefficient on the relative MI (their MI minus the local mean MI for each community) was always small and non-significant.

We note that our results do not hinge on our choice of using the average MI at the community level. One obtains the same basic findings by using (1) household-level MI values for each player or (2) population level MI values.

REGRESSIONS ON ALL OFFERS (DG, UG AND TPG OFFERS TOGETHER)

Here we present a set of regression models to further support the findings in Table 2 of the main text. We present supporting analyses for each of our dependent variables: (1) all offers together, (2) DG offers, (3) UG offers, and (4) TPG offers. Offers are measured as a percentage of the stake. In each case we begin with our baseline model (used in the main text), which regresses our dependent variables on nine predictor variables: market integration (MI), world religion (WR), age, sex, education, income, wealth, household size (HS) and community size (CS). MI measures market integration as the percentage of the diet in calories purchased in the market, as opposed to grown, hunted, fished, or gathered. As explained above, we use the average MI for each individual's community (village, camp, etc.). WR is a binary variable, with "1" indicating participation in Islam or Christianity, and "0" indicating the practice of a local or traditional religion, or a report of "no religion." Our income and wealth measures are derived from detailed protocols eliciting data disaggregated by source, and have been converted to U.S. dollars and scaled to units of $1000. Income was measured at the individual level, and wealth at the household level. Below, we deal directly with concerns about the buying power equivalence of a U.S.

dollar in different places, and show that this likely has little effect on our findings. Age is measured in years at the time of the experiment. For education, we created standardized values (with mean zero and standard deviation 1) within each population based on self-reports of the number of years a player had spent in formal schooling. We did this because one year of formal schooling is unlikely to be even roughly equivalent across these diverse societies. This approach allows us to get the most from the substantial within population variation in formal education in our samples. Community size (CS) is the number of individuals (in units of 100 people) in the local social group, usually a village (though camps were used for the Hadza, a town in Missouri, and a work place in Ghana).

For our baseline model we estimated the coefficients in this equation for each dependent variable (units are in parentheses, if applicable): Offer (% of stake) = Constant + β_{MI}*MI (%) + β_{WR}*WR + β_I*Income($1000) + β_W*Wealth($1000) + β_H*Household Size (# of people) + β_A*Age (years) + β_S*Sex + β_E*Education +β_{CS}*CS (100 people). Model 2 of Table S5 shows the baseline model for all offers. As explained in the main text, MI and WR are the only large and significant predictors of all offers. In the models in Table S5 we used clustered robust standard errors because the analyses compile all observations across our three experiments, which involved some repeated observations from the same individuals. We clustered on individuals to address the problem of the non-independence of repeated observations from the same person. These standard errors are below the coefficients, with *p*-values beneath these errors.

To examine the effects of including continental-level controls, compare Model 1 (with continental controls) with Model 2 (baseline). For the continental dummies, the reference continent is Eurasia (in which we include the U.S., since it is of European descent) so the coefficient gives effects relative to Eurasia. The coefficient on MI increases from 0.12 to 0.16, and remains highly significant. MI is robust to continental-level controls.

The coefficient on WR, however, drops dramatically to insignificance. This occurs because our individuals with WR=0 are scattered quite unevenly across the continents, residing mostly in Africa. To verify that this is what is driving the evaporation of WR's predictive power with the inclusion of continental controls, we ran our baseline model only on our African populations (a within continent analysis). This analysis gives coefficients for WR and MI as 6.4 (p = 0.026) and 0.10 (p = 0.029), respectively (n = 216, clustered robust standard error, with 117 clusters, R^2 = 0.29).

These analyses indicate that shared cultural phylogenies are unlikely to be responsible for our pattern of results.

Next, to explore robustness of our findings for MI and WR we examined alternative specifications by dropping the terms with the least significant coefficients. In Table S5, Models 2 to 6 show the coefficients from five different specifications. Coefficients on both MI and WR remain large and significant across models. The coefficients on MI range from 0.12 to 0.069. The coefficient on WR ranges from 5.96 to 6.16. This means that moving from a fully subsistence-based society with a local religion to a fully market integrated society with a world religion predicts an increase in percentage offer of between 13 and 18. While the coefficient on income is not significant at conventional levels in the baseline model, it is significant in the other four specifications (Models 3 to 6). Independent of MI and WR, an increase of $4000 in income per year increases the percent offered by between 0.7 and 0.9.

Table S5. Linear regressions for all offers (DG, UG, TPG)

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Model	Model	Model	Model	Model	Model
MI	0.156***	0.120***	0.0842***	0.0830***	0.0695***	0.0691***
	(0.0244)	(0.0231)	(0.0192)	(0.0187)	(0.0182)	(0.0182)
	(3.17e-10)	(2.91e-07)	(1.34e-05)	(1.08e-05)	(0.000148)	(0.000159)
World Religion	0.310	5.956***	5.350***	5.649***	6.158***	6.141***
	(2.164)	(2.038)	(1.844)	(1.807)	(1.819)	(1.817)
	(0.886)	(0.00360)	(0.00384)	(0.00185)	(0.000751)	(0.000766)
Income	0.0967	0.0962	0.174***	0.198***	0.218***	0.227***
(1000 USD)	(0.111)	(0.0887)	(0.0668)	(0.0618)	(0.0610)	(0.0606)
	(0.384)	(0.278)	(0.00940)	(0.00142)	(0.000369)	(0.000196)
Sex	-1.344	-1.525	-1.043	-1.258		
(female = 1)	(1.259)	(1.280)	(1.200)	(1.187)		
	(0.286)	(0.234)	(0.385)	(0.290)		
Age (years)	0.0669	0.0532	0.0695	0.0590		
	(0.0456)	(0.0479)	(0.0441)	(0.0423)		
	(0.143)	(0.266)	(0.115)	(0.164)		
Household Size	-0.257	-0.244	-0.172			
	(0.211)	(0.208)	(0.187)			
	(0.223)	(0.241)	(0.357)			
Education	0.552	0.528	0.423			
(std by pop)	(0.576)	(0.593)	(0.573)			
	(0.338)	(0.374)	(0.461)			
Community Size	0.135***	-0.0250				
(100 people)	(0.0471)	(0.0522)				
	(0.00440)	(0.632)				
Wealth	0.00122	-0.000667				
(1000 USD)	(0.00581)	(0.00588)				
	(0.834)	(0.910)				
UG	0.930	1.294	1.279	1.334	1.353	
	(1.168)	(1.167)	(1.039)	(1.029)	(1.009)	
	(0.426)	(0.268)	(0.219)	(0.195)	(0.181)	
TPG	-4.049**	-2.413	-4.383***	-4.115***	-4.301***	-4.953***
	(1.597)	(1.586)	(1.424)	(1.400)	(1.387)	(1.254)
	(0.0115)	(0.129)	(0.00217)	(0.00341)	(0.00201)	(8.62e-05)
Africa	-0.148					
	(3.594)					
	(0.967)					
South America	4.767					
	(3.617)					
	(0.188)					
Oceania	10.45**					
	(4.123)					

	(0.0115)					
Constant	24.51***	26.89***	27.43***	26.53***	28.79***	29.47***
	(4.557)	(3.291)	(3.049)	(2.810)	(2.021)	(1.902)
	(1.08e-07)	(0)	(0)	(0)	(0)	(0)
Observations	920	920	1065	1071	1120	1120
Number of clusters	596	596	688	691	719	719
R-squared	0.113	0.084	0.079	0.077	0.068	0.067
*p < 0.1 **p< 0.05 ***p<0.01 Clustered robust standard errors and p-value in parentheses below the coefficient Education has been standardized to a mean of zero and standard deviation of one within each population						

Although the TPG dummy is not significant at conventional levels in the baseline model, it is significant in all other specifications. Playing the TPG, as opposed to the DG, predicts a decrease in the percent offered of between 2.5 and 5.

Note that while our measures of wealth, income, age, market integration, and household size are all intercorrelated, the correlations are sufficiently small that they do not create collinearity problems for our regressions. Focusing on MI, the highest correlations in absolute values among our other predictor variables is 0.36 with community size (CS) and -0.34 for household size (HS). World Religion is most highly correlated with MI and CS at just under 0.20. Income and wealth show by far the highest correlation at 0.55, which is one of the reasons why we re-ran everything with wealth dropped. We also ran analyses dropping both wealth and CS to bring the Accra and the Siberian groups into our analyses.

The regressions in Table S5 used our income and wealth variables, which were converted into U.S. dollars based on the international exchange rates at the time of the experiments. This approach could create problems because such rates might distort the real local purchasing power, thereby distorting our measures of income and wealth. Such distortions could result from several factors, including the distance of many of our sites from large market centers (where exchange rates are set) or ephemeral fluctuation in world exchange rates that are unconnected to the material conditions on the ground at our sites. To address this, each researcher compiled a list of the local prices, at the time of the experiments, for twenty-seven different commonly used items, including several staples. From this list we found five items that were present in all field sites and were purchased, at least occasionally. These items are sugar, salt, rice, D-cell batteries, and cooking oil. Using the local prices of these items for each site, we converted our measures of income and wealth from the local currency into quantities of each of these items, giving us five new income and wealth variables now measured in quantities of local consumables. For example, if a person's yearly income is 1000 shillings and sugar was locally priced at 5 shilling per kilogram, we converted the income measure to 200 kilograms of sugar. That is, they could purchase 200kg of sugar locally with their yearly income. This approach avoids the use of international exchange rates and grounds people's income and wealth in the kind of products people commonly purchase in these locations.

Table S6 compares the regression coefficients for our baseline model (Model 2 in Table S5) with those found when wealth and income are re-valued using the local prices of five consumables. The units for wealth and income can be found at the top of each column. The coefficient on MI varies little (increasing if anything with these recalibrations), ranging from 0.12 to 0.13. World religion's coefficient ranges from roughly 5.8 to 6.0, and is significant across the board. No other variables in these models are even marginally significant.

Table S6. Linear regressions for all offers (DG, UG, TPG) with Wealth and Income local calibrations

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	U.S. $	D battery	100kg rice	100kg sugar	100L cooking oil	100kg salt
MI	0.120***	0.121***	0.119***	0.120***	0.129***	0.128***
	(0.0231)	(0.0225)	(0.0235)	(0.0232)	(0.0259)	(0.0265)
	(2.91e-07)	(1.24e-07)	(4.96e-07)	(3.35e-07)	(7.74e-07)	(1.66e-06)
World Religion	5.956***	5.958***	5.968***	5.959***	5.836***	6.040***
	(2.038)	(2.035)	(2.036)	(2.037)	(2.006)	(2.020)
	(0.00360)	(0.00355)	(0.00351)	(0.00357)	(0.00377)	(0.00291)
Community Size	-0.0250	-0.0307	-0.0241	-0.0250	-0.0246	-0.0385
(per 100 people)	(0.0522)	(0.0546)	(0.0520)	(0.0521)	(0.0555)	(0.0648)
	(0.632)	(0.574)	(0.643)	(0.631)	(0.658)	(0.552)
Income	0.0962	4.93e-05	9.19e-05	8.30e-05	-0.000438	-0.000104
	(0.0887)	(0.000138)	(9.62e-05)	(8.67e-05)	(0.000879)	(0.000133)
	(0.278)	(0.721)	(0.340)	(0.339)	(0.618)	(0.433)
Wealth	-0.000667	1.81e-05	-4.59e-08	-4.14e-07	0.000791***	0.000103
	(0.00588)	(1.35e-05)	(6.31e-06)	(5.71e-06)	(0.000239)	(9.42e-05)
	(0.910)	(0.183)	(0.994)	(0.942)	(0.00100)	(0.275)
Sex	-1.525	-1.520	-1.502	-1.511	-1.601	-1.641
(female = 1)	(1.280)	(1.284)	(1.281)	(1.281)	(1.329)	(1.342)
	(0.234)	(0.237)	(0.241)	(0.238)	(0.229)	(0.222)
Age (years)	0.0532	0.0510	0.0538	0.0538	0.0617	0.0603
	(0.0479)	(0.0479)	(0.0478)	(0.0478)	(0.0506)	(0.0509)
	(0.266)	(0.287)	(0.261)	(0.262)	(0.223)	(0.237)
Education	0.528	0.521	0.531	0.533	0.427	0.514
(std by pop)	(0.593)	(0.594)	(0.593)	(0.593)	(0.618)	(0.625)
	(0.374)	(0.381)	(0.371)	(0.369)	(0.490)	(0.411)
Household Size	-0.244	-0.258	-0.249	-0.249	-0.341	-0.301
	(0.208)	(0.205)	(0.207)	(0.208)	(0.222)	(0.225)
	(0.241)	(0.207)	(0.230)	(0.232)	(0.125)	(0.183)
UG	1.294	1.319	1.299	1.304	1.391	1.423
	(1.167)	(1.165)	(1.168)	(1.167)	(1.221)	(1.223)
	(0.268)	(0.258)	(0.266)	(0.265)	(0.255)	(0.245)
TPG	-2.413	-2.462	-2.426	-2.426	-2.938*	-2.508
	(1.586)	(1.582)	(1.585)	(1.586)	(1.604)	(1.603)
	(0.129)	(0.120)	(0.126)	(0.127)	(0.0676)	(0.118)
Constant	26.89***	26.99***	26.86***	26.87***	26.16***	26.54***
	(3.291)	(3.291)	(3.288)	(3.289)	(3.398)	(3.393)
	(0)	(0)	(0)	(0)	(0)	(0)
Observations	920	920	920	920	879	879
R-squared	0.084	0.084	0.084	0.084	0.079	0.068
No. of clusters	596	596	596	596	555	555

Education has been standardized to a mean of zero and standard deviation of one within each population
Clustered robust standard errors and p-value in parentheses below the coefficient
*p < 0.1 **p< 0.05 ***p<0.01

In Table S7, to further address concerns about the comparability of income and wealth across populations, we have also estimated a series of models using a standardized version of wealth and income. For these, we subtracted the mean values for the population divided by the standard deviations for each population. This gives a purely relative measure of income and wealth in each population. We then repeated the procedure used above in Table S5, starting with the base model and dropping the least significant variables. Essentially, for our purpose, nothing changes. MI and WR remain large and highly significant, while nothing else is consistently significant.

Table S7. Linear regressions for all offers (DG, UG, TPG) using standardized measures of Income and Wealth

Variables	(1) Model	(2) Model	(3) Model	(4) Model	(5) Model	(6) Model	(7) Model	(8) Model
MI	0.128***	0.123***	0.122***	0.122***	0.1000***	0.100***	0.0995***	0.101***
	(0.0211)	(0.0200)	(0.0199)	(0.0199)	(0.0181)	(0.0180)	(0.0179)	(0.0173)
	(2.55e-09)	(1.57e-09)	(1.44e-09)	(1.44e-09)	(4.32e-08)	(3.69e-08)	(4.13e-08)	(9.75e-09)
World Religion	5.889***	5.776***	5.809***	5.803***	5.070***	5.341***	5.210***	5.148***
	(2.042)	(2.023)	(2.020)	(2.018)	(1.831)	(1.820)	(1.820)	(1.803)
	(0.00407)	(0.00445)	(0.00418)	(0.00418)	(0.00577)	(0.00344)	(0.00433)	(0.00442)
Age (years)	0.0626	0.0595	0.0586	0.0589	0.0783*	0.0734*	0.0759*	0.0714*
	(0.0473)	(0.0468)	(0.0463)	(0.0462)	(0.0419)	(0.0408)	(0.0406)	(0.0404)
	(0.187)	(0.204)	(0.206)	(0.204)	(0.0622)	(0.0724)	(0.0620)	(0.0775)
Household Size	-0.248	-0.263	-0.255	-0.262	-0.205	-0.213	-0.223	
	(0.203)	(0.200)	(0.197)	(0.197)	(0.179)	(0.178)	(0.177)	
	(0.223)	(0.189)	(0.198)	(0.183)	(0.253)	(0.233)	(0.208)	
Sex	-1.598	-1.628	-1.646	-1.629	-1.097	-1.139		
(female = 1)	(1.299)	(1.298)	(1.283)	(1.283)	(1.176)	(1.166)		
	(0.219)	(0.210)	(0.200)	(0.205)	(0.351)	(0.329)		
Education	0.660	0.649	0.687	0.680	0.433			
(std by pop)	(0.626)	(0.627)	(0.599)	(0.598)	(0.555)			
	(0.292)	(0.300)	(0.252)	(0.256)	(0.436)			
Wealth	-1.042	-1.028	-0.981	-0.972				
(std by pop)	(0.682)	(0.684)	(0.671)	(0.671)				
	(0.127)	(0.133)	(0.144)	(0.148)				
Income	0.262	0.262						
(std by pop)	(0.556)	(0.552)						
	(0.638)	(0.636)						
Community Size	-0.0282							
(100 people)	(0.0509)							
	(0.580)							
UG	1.442	1.405	1.583					
	(1.164)	(1.161)	(1.180)					
	(0.216)	(0.227)	(0.180)					
TPG	-2.424	-2.634*	-2.453	-3.230**	-5.861***	-5.861***	-5.863***	-5.691***
	(1.565)	(1.511)	(1.514)	(1.344)	(1.235)	(1.229)	(1.229)	(1.209)
	(0.122)	(0.0819)	(0.106)	(0.0166)	(2.51e-06)	(2.2e-06)	(2.2e-06)	(3.3e-06)
Constant	26.46***	26.70***	26.59***	27.40***	27.80***	27.79***	27.37***	26.11***
	(3.260)	(3.225)	(3.203)	(3.092)	(2.878)	(2.841)	(2.790)	(2.570)
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Observations	920	920	925	925	1096	1100	1100	1105
Number of clusters	596	596	599	599	716	719	719	722
R-squared	0.086	0.085	0.084	0.083	0.083	0.084	0.083	0.081

Income, Wealth and Education have been standardized to a mean of zero and standard deviation of one within each population
Clustered robust standard errors and p-value in parentheses below the coefficient
*p < 0.1 **p< 0.05 ***p<0.01

Since the effect of participating in a world religion derives from a rather small portion of our sample (11%) that is not diffusely distributed across our populations, we performed some additional analyses to address the concern that our findings may be driven by the few populations in our sample that are dominated by those not associated with a world religion. Of our populations, only five had non-trivial fractions of those who did not report membership in a world religion. We re-ran our baseline model using only these five populations. The coefficient on WR increases from 5.96 to 6.95 (p = 0.017, n = 316 (181 clusters), using robust standard errors, while MI's coefficient remains large and significant (note, the Nganasan drop out because we lack wealth data for them, reducing this to four populations). This suggests the effects are not driven by the difference between a few societies lacking prevalent world religions, but that the phenomena we are capturing is coming also from variation within those communities that maintain diversity in religion.

DICTATOR GAME OFFERS

Now we proceed to analyze the offers for each of our three experiments separately. The analytical process is the same as used above except that we will also analyze a model that includes controls both for deep cultural-historical phylogenies (using continental dummies) and assuming *only* our sites are statistically independent. We include this to address the concern that our participants might not represent fully statistically independent observations, since many participants are sampled from the same communities and populations. If true, this would mean that standard errors calculated assuming individuals are independently observed would be wrong. While we think this argument may misunderstand what statistical independence implies or requires, as our participants were alone in making their experimental decisions, we have taken this concern seriously and sought to address it in the analyses of each of our five game measures below. To accomplish this we make the maximally conservative assumption that only our populations (our sites) are independent observations, and we use clustered robust standard errors, clustering on populations (the largest aggregate on which we could cluster). For each of our five dependent experimental measures, we add continental controls to our baseline regression (using nine predictor variables) and then estimate clustered robust standard errors.

Table S8 compares five regression models using the same set of predictor variables deployed in the overall analysis above. The baseline model, Model 2, shows that only MI and WR are large and significant or marginally significant predictors of DG offers. A comparison of Models 1 and 2 in Table S8 shows that when continental controls are included and clustered robust errors are calculated (Model 1), the coefficient on MI increases from 0.17 to 0.20, and remains highly significant. Again, due to the uneven distribution of individuals expressing WR=0, the coefficient on WR drops to about 3.1 and becomes insignificant.

To explore this we again re-ran our baseline regression for only those populations in Africa (to avoid continental differences). As above, the coefficient on WR in this regression jumps up above its value in our baseline regression, going from 6.4 to 7.7, though it is not well estimated due to the smaller sample size (p = 0.18, n = 162, $R2$ = 0.11, using robust standard errors).

Next, we check the robusticity of our findings to alternative model specifications. Models 2 to 5 in Table S8 allow us to examine what happens if we sequentially drop the least significant predictors. While both MI and WR have large coefficients in the baseline model, WR's is only marginally significant at conventional levels. However, as the least significant variables are dropped and the sample size expands (Model 3 to 5), WR becomes significant at conventional levels with coefficients ranging from 6.4 to 7.8. MI's coefficient remains large and highly significant across all models. Together, moving from a fully subsistence-oriented society with a traditional religion to a market integrated economy with a world religion predicts an increase of between 20 and 23 in DG offers, which captures the full range of variation among populations in mean DG offer.[2]

[2] We also ran models to look for any non-linear effects of age, using Age^2, and found none.

Table S8. Linear regressions for Dictator Game offers

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Model	Model	Model	Model	Model
MI (mean)	0.199***	0.171***	0.134***	0.127***	0.121***
(% purchased)	(0.0350)	(0.0352)	(0.0295)	(0.0268)	(0.0263)
	(0.000101)	(1.87e-06)	(6.98e-06)	(3.34e-06)	(5.37e-06)
World Religion	3.105	6.364*	6.315**	7.025**	7.823***
(binary)	(3.513)	(3.608)	(3.078)	(3.033)	(2.908)
	(0.394)	(0.0787)	(0.0409)	(0.0210)	(0.00743)
Sex	-2.692	-2.624	-2.929	-2.967*	-3.367*
(female = 1)	(3.000)	(2.012)	(1.787)	(1.728)	(1.724)
	(0.387)	(0.193)	(0.102)	(0.0867)	(0.0515)
Household Size	-0.0858	-0.130	-0.146	-0.270	
	(0.367)	(0.306)	(0.276)	(0.264)	
	(0.819)	(0.672)	(0.598)	(0.306)	
Education	0.991	0.963	0.618	0.660	
(std by pop)	(0.820)	(1.026)	(0.945)	(0.902)	
	(0.250)	(0.349)	(0.514)	(0.465)	
Income	-0.0535	-0.0124	0.109		
(1000 USD)	(0.161)	(0.150)	(0.134)		
	(0.745)	(0.935)	(0.415)		
Age	-0.0135	-0.0232	0.0132		
(years)	(0.0671)	(0.0780)	(0.0684)		
	(0.844)	(0.766)	(0.847)		
Community Size	0.00691	-0.0728			
(100 people)	(0.0721)	(0.0668)			
	(0.925)	(0.276)			
Wealth	0.00181	0.00129			
(1000 USD)	(0.00695)	(0.00823)			
	(0.799)	(0.876)			
Africa (dummy)	-0.876				
	(3.554)				
	(0.809)				
South America	1.329				
(dummy)	(2.546)				
	(0.611)				
Oceania	5.912				
(dummy)	(4.269)				
	(0.191)				
Constant	26.95***	28.01***	27.36***	28.81***	26.81***
	(5.286)	(5.134)	(4.559)	(3.594)	(3.110)
	(0.000262)	(9.71e-08)	(4.59e-09)	(0)	(0)
Observations	336	336	388	409	416
R-squared	0.115	0.103	0.098	0.089	0.083

*p < 0.1 **p< 0.05 ***p<0.01

Model 1 uses clustered robust standard errors (clustering on population). Other models use robust standard errors; errors and p-values are in parentheses below the coefficient Education has been standardized to a mean of zero and standard deviation of one within each population.

Since the effect of participating in a world religion arises from a small portion of our sample (11%) that is not widely distributed across our populations we performed some additional analyses. We re-ran our baseline model using the four populations that have frequencies of individuals with WR = 1 above 3% and which had data for wealth and community size. The coefficient of WR jumps from 6.4 (baseline model) to 14.9 (p = 0.002, n = 107, using robust standard errors). Then, to add Accra and the Siberian sites, which lack wealth or community size data, but do have non-trivial frequencies of those with WR=1, we re-ran the same regression dropping these variables in order to bring in data from these two populations. Now the coefficient on WR is 11.3 (p = 0.003, n = 157, R^2 = 0.18, robust standard errors). This suggests the WR effects are not driven by the difference between a few societies lacking a prevalent world religion.

The regressions in Table S8 use our income and wealth variables converted into U.S. dollars based on the international exchange rates at the time of the experiments. As explained above, we addressed this by re-valuing income and wealth based on the price of local consumables. Table S9 compares the regression coefficients for our baseline model with those found when wealth and income are re-valued using the local prices of five consumables. The coefficient on MI varies little with these recalibrations, ranging from 0.17 to 0.15. World Religion's coefficient ranges from roughly 6.1 to 6.4, and is marginally significant across the board, as it is in the regression presented in the main text.

Note that when we dropped wealth and CS out of the Models in Table S9, all coefficients of WR across the models with the different valuations for income were significant at conventional levels, except one where p = 0.054. MI's coefficients were robust across the board.

Table S9. Linear regressions for DG offers with local Wealth and Income calibrations

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	U.S. $	D battery	100kg rice	100kg sugar	100L cooking oil	100kg salt
MI	0.171***	0.166***	0.169***	0.170***	0.166***	0.154***
	(0.0352)	(0.0342)	(0.0357)	(0.0354)	(0.0376)	(0.0373)
	(1.87e-06)	(1.85e-06)	(3.28e-06)	(2.48e-06)	(1.49e-05)	(4.66e-05)
World Religion	6.364*	6.299*	6.336*	6.346*	6.064*	6.389*
	(3.608)	(3.600)	(3.603)	(3.606)	(3.565)	(3.555)
	(0.0787)	(0.0811)	(0.0796)	(0.0794)	(0.0899)	(0.0733)
Age (years)	-0.0232	-0.0212	-0.0215	-0.0222	-0.0385	-0.0326
	(0.0780)	(0.0783)	(0.0778)	(0.0779)	(0.0815)	(0.0811)
	(0.766)	(0.787)	(0.782)	(0.776)	(0.638)	(0.687)
Sex	-2.624	-2.377	-2.532	-2.572	-2.701	-2.717
(female = 1)	(2.012)	(2.015)	(2.020)	(2.018)	(2.043)	(2.042)
	(0.193)	(0.239)	(0.211)	(0.203)	(0.187)	(0.184)
Income	-0.0124	0.000192	4.83e-05	2.21e-05	0.00245	0.000316
	(0.150)	(0.000234)	(0.000164)	(0.000146)	(0.00163)	(0.000279)
	(0.935)	(0.411)	(0.768)	(0.880)	(0.134)	(0.259)
Household Size	-0.130	-0.126	-0.124	-0.126	-0.0967	-0.0626
	(0.306)	(0.297)	(0.304)	(0.305)	(0.319)	(0.317)
	(0.672)	(0.672)	(0.685)	(0.681)	(0.762)	(0.843)
Education	0.963	0.989	0.977	0.972	0.851	1.064
(std by pop)	(1.026)	(1.034)	(1.027)	(1.026)	(1.073)	(1.061)
	(0.349)	(0.339)	(0.342)	(0.344)	(0.428)	(0.316)
Community Size	-0.0728	-0.0895	-0.0742	-0.0737	-0.101	-0.0289
(100 people)	(0.0668)	(0.0732)	(0.0666)	(0.0668)	(0.0702)	(0.0747)
	(0.276)	(0.222)	(0.266)	(0.271)	(0.153)	(0.699)
Wealth	0.00129	-1.77e-05	-4.35e-06	-1.98e-06	-0.000649*	-0.000280***
	(0.00823)	(2.45e-05)	(1.10e-05)	(9.01e-06)	(0.000384)	(8.85e-05)
	(0.876)	(0.470)	(0.693)	(0.826)	(0.0923)	(0.00172)
Constant	28.01***	27.91***	27.91***	27.95***	29.06***	28.96***
	(5.134)	(5.121)	(5.129)	(5.131)	(5.217)	(5.174)
	(9.71e-08)	(9.97e-08)	(1.04e-07)	(1.01e-07)	(5.51e-08)	(4.80e-08)
Observations	336	336	336	336	321	321
R-squared	0.103	0.104	0.103	0.103	0.100	0.105

*p < 0.1 **p< 0.05 ***p<0.01
Education has been standardized to a mean of zero and standard deviation of one within each population
Robust standard errors and p-values in parentheses below the coefficient.

We again sought to address concerns about the comparability of income and wealth across populations. As above, we have estimated a series of models using standardized versions of income and wealth to obtain a purely relative measure of income and wealth for each population. We then repeated the procedure used above in Table S8, starting with the baseline model and dropping the least significant variables. Table S10 below shows that both MI and WR have large coefficients. WR is only marginally significant in Model 1 and Model 6, but significant at conventional levels in Models 2 to 5. Here, relative wealth has a large and highly significant negative effect on offers, with relatively wealthier people within a given society offering dramatically less. Income has a positive coefficient, but is only marginally significant in Model 1. MI stands alone, large and highly significant across the board.

Table S10. Linear regressions for Dictator Game offers using standardized Income and Wealth

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Model	Model	Model	Model	Model	Model
MI	0.166***	0.153***	0.146***	0.132***	0.130***	0.123***
	(0.0310)	(0.0295)	(0.0297)	(0.0303)	(0.0305)	(0.0299)
	(1.49e-07)	(3.58e-07)	(1.39e-06)	(1.91e-05)	(2.48e-05)	(4.70e-05)
World Religion	6.171*	7.125**	7.274**	7.100**	6.952**	6.772*
	(3.544)	(3.521)	(3.504)	(3.486)	(3.492)	(3.542)
	(0.0826)	(0.0438)	(0.0386)	(0.0425)	(0.0473)	(0.0567)
Wealth	-3.482***	-3.064***	-3.398***	-3.415***	-3.250***	-3.196***
(std by pop)	(0.978)	(0.917)	(0.950)	(0.949)	(0.920)	(0.912)
	(0.000426)	(0.000931)	(0.000398)	(0.000368)	(0.000468)	(0.000512)
Education	1.249	1.174	1.409	1.416	1.703*	1.844*
(std by pop)	(1.032)	(0.994)	(0.987)	(0.987)	(0.973)	(0.978)
	(0.227)	(0.238)	(0.154)	(0.152)	(0.0809)	(0.0602)
Sex	-2.095	-2.501	-2.853	-2.977	-3.154*	
(female = 1)	(1.989)	(1.905)	(1.899)	(1.889)	(1.861)	
	(0.293)	(0.190)	(0.134)	(0.116)	(0.0911)	
Income	1.552*	1.367	1.255	1.256		
(std by pop)	(0.935)	(0.929)	(0.932)	(0.931)		
	(0.0978)	(0.142)	(0.179)	(0.178)		
Community Size	-0.0696	-0.0745	-0.0819			
(per 100 people)	(0.0588)	(0.0578)	(0.0591)			
	(0.238)	(0.199)	(0.166)			
Household Size	-0.0922	-0.152				
	(0.293)	(0.284)				
	(0.754)	(0.593)				
Age	0.000598					
(years)	(0.0747)					
	(0.994)					
Constant	26.79***	27.56***	27.06***	26.94***	27.07***	26.09***
	(4.978)	(3.983)	(3.620)	(3.612)	(3.606)	(3.499)
	(1.41e-07)	(0)	(0)	(0)	(0)	(0)
Observations	336	354	358	358	360	360
R-squared	0.135	0.121	0.120	0.118	0.111	0.104

*p < 0.1 **p< 0.05 ***p<0.01
Robust standard errors and p-values in parentheses below the coefficient
Income, Wealth and Education has been standardized to a mean of zero and standard deviation of one within each population

Since it is plausible that income and wealth may have different effects within versus between populations, in analyses not reproduced here, we also broke both income and wealth down into population mean measures and within population deviations from the population mean. The population averages allow us to assess if, for example, individuals from populations with higher absolute incomes offer more. Within a population, measures of deviations from a local average allow us to assess whether relative income and wealth predict lower offers. For example, it is theoretically possible that individuals from populations with high mean incomes might offer more, while relatively richer people (locally speaking) might offer less. Further, our regressions above that do not break this down might be non-significant because the coefficients are fitting opposing effects that cancel each other out. Our efforts, however, found no evidence of this in the DG, or elsewhere. None of the coefficients on any of these wealth or income measures approached conventional levels of significance. Meanwhile, the coefficients on MI and WR do not vary much from those shown above. Note that because the variable *mean* wealth and *mean* income are correlated at 0.99, we never entered both into the model at the same time. We did try the complete model with each of these separately.

ULTIMATUM GAME OFFERS

Table S11 repeats the procedure used above for UG offers. Model 2, the baseline model from the main text, shows that MI, WR and age are large, positive and significant predictors of UG offers.

Comparing Models 1 and 2 shows that adding continental controls and using clustered robust standard errors vis-à-vis our baseline regression shows that the coefficient on MI increases from 0.098 to 0.13, and remains significant. WR, however, drops to insignificance due to the uneven distribution of individuals not participating in world religions. As a check we re-ran our regression just in Africa. Now the coefficient on WR jumps back up to 7.6 and is marginally significant (p = 0.098, n = 137, R2 = 0.21, robust standard errors).

Models 2 to 7 in Table S11 show that both MI and WR are robust, positive predictors of higher UG offers, and highly significant at conventional levels across the board. Controlling for a range of economic and demographic variables, a 20 percent increase in calories purchased in the market predicts an average increase of between 1.8 and 2.5 in percent offered in the UG. Changing to a world religion predicts an increase of between 7.7 and 10.4 in percent offered. This suggests that going from a fully subsistence economy with a traditional religion to a fully market dependent economy with a world religion means an increase in offers of between 17 and 21 percent, covering most of the range of variation we observe across societies.

Along with MI and WR, age is also a robust positive predictor of UG offers. Each additional decade of adulthood predicts an increase of between 1.2 and 1.6 in percent offered. Since previous work has suggested a non-linear fit for age (S*35*), we explored adding a squared term for age. We do not have evidence for this non-linear age relationship. Unlike MI and WR, age does not emerge as a significant predictor for offers in the DG and TPG.

To again examine the fact that the effects of participating in a world religion observed in Table S11 arise from a rather small portion of our sample (11%), we performed some additional analyses. We re-ran Model 2 (our baseline) using only the four populations in which individuals had non-trivial frequencies (>1%) of individuals with WR=0 and which had data for wealth and community size. The coefficient of WR is 9.1 (p = 0.032, n = 107, R2 = 0.29, using robust standard errors). Then, since two of these six populations are Accra and the Siberian sites, which lack wealth or community size data, we re-ran the same regression dropping these variables in order to bring in the data from these two populations. Now the coefficient on WR is 7.0 (p = 0.041, n = 157, R^2 = 0.24, robust standard errors). This suggests the effects are not driven by the difference between a few societies lacking world religions.

Table S11. Linear regressions for Ultimatum Game offers

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Model	Model	Model	Model	Model	Model	Model
MI (mean)	0.130**	0.0978***	0.0977***	0.124***	0.122***	0.0857***	0.0904***
(% purchased)	(0.0499)	(0.0349)	(0.0353)	(0.0352)	(0.0349)	(0.0241)	(0.0216)
	(0.0245)	(0.00547)	(0.00599)	(0.000472)	(0.000537)	(0.000421)	(3.66e-05)
World Religion	0.893	10.44***	10.18***	8.678***	8.529***	7.772***	8.447***
(binary)	(5.031)	(2.665)	(2.605)	(2.389)	(2.363)	(2.267)	(2.306)
	(0.862)	(0.000110)	(0.000115)	(0.000325)	(0.000353)	(0.000677)	(0.000283)
Age (years)	0.177*	0.149**	0.144**	0.153**	0.158**	0.121**	
	(0.0842)	(0.0655)	(0.0656)	(0.0630)	(0.0614)	(0.0564)	
	(0.0601)	(0.0239)	(0.0293)	(0.0155)	(0.0104)	(0.0324)	
Income	0.174***	0.157	0.178*	0.127	0.127	0.168**	
(1000 USD)	(0.0483)	(0.104)	(0.100)	(0.0863)	(0.0882)	(0.0743)	
	(0.00414)	(0.134)	(0.0777)	(0.141)	(0.151)	(0.0243)	
Community Size	0.194	-0.135	-0.148	-0.200*	-0.204*		
(100 people)	(0.118)	(0.105)	(0.107)	(0.114)	(0.113)		
	(0.126)	(0.200)	(0.168)	(0.0814)	(0.0732)		
Education	1.085	0.987	1.035	1.017	1.074		
(std by pop)	(0.739)	(0.786)	(0.791)	(0.763)	(0.761)		
	(0.170)	(0.210)	(0.192)	(0.183)	(0.159)		
Sex	-1.681	-1.504	-1.565	-1.272			
(female = 1)	(2.462)	(1.803)	(1.800)	(1.743)			
	(0.509)	(0.405)	(0.385)	(0.466)			
Wealth	-0.00392	-0.00625	-0.00564				
(1000 USD)	(0.00438)	(0.00785)	(0.00762)				
	(0.390)	(0.426)	(0.460)				
Household Size	-0.219	-0.235					
	(0.462)	(0.260)					
	(0.645)	(0.366)					
Africa	-2.011						
(dummy)	(2.621)						
	(0.459)						
South America	5.019						
(dummy)	(3.272)						
	(0.153)						
Oceania	16.78***						
(dummy)	(5.191)						
	(0.00797)						
Constant	19.82***	22.49***	21.50***	22.30***	21.75***	22.95***	27.91***
	(3.575)	(3.728)	(3.487)	(3.404)	(3.247)	(3.162)	(2.206)
	(0.000174)	(4.55e-09)	(2.16e-09)	(2.13e-10)	(8.86e-11)	(0)	(0)
Observations	319	319	320	343	343	374	398
R-squared	0.245	0.140	0.137	0.133	0.132	0.114	0.084

Model 1 uses clustered robust standard errors (clustering on site); other models use robust standard errors; errors and p-values are in parentheses below the coefficient. Education has been standardized to a mean of zero and standard deviation of one within each population

*p < 0.1 **p< 0.05 ***p<0.01

Paralleling our efforts above for the DG to address differences in local purchasing power, we used our income and wealth variables, measured according to the prices of local consumables, to verify that our analyses are not strongly influenced by the use of international currency exchange rates. Table S12 compares the regression coefficients for our baseline model (Model 2 in Table S11) with those found when wealth and income are re-valued using the local prices of the five consumables. The coefficients for MI, WR and age all remain about the same size and all remain significant. Coefficients on MI range between 0.098 and 0.13. World religion’s coefficient ranges from roughly 10.3 to 10.5. The coefficient on age ranges from 0.15 to 0.18. Remember the coefficient for income and wealth cannot be easily compared since they are in different units (which are listed at the top of each column).

Table S12. Linear regressions for UG offers with local Wealth & Income calibrations

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	U.S. $	D battery	100kg rice	100kg sugar	100L cooking oil	100kg salt
MI	0.0978***	0.0981***	0.0966***	0.0970***	0.125***	0.116**
	(0.0349)	(0.0336)	(0.0353)	(0.0351)	(0.0448)	(0.0475)
	(0.00547)	(0.00375)	(0.00653)	(0.00600)	(0.00547)	(0.0153)
World Religion	10.44***	10.42***	10.45***	10.43***	10.32***	10.46***
	(2.665)	(2.655)	(2.662)	(2.664)	(2.538)	(2.621)
	(0.000110)	(0.000107)	(0.000107)	(0.000111)	(6.14e-05)	(8.34e-05)
Age	0.149**	0.145**	0.149**	0.149**	0.176**	0.175**
(years)	(0.0655)	(0.0658)	(0.0654)	(0.0654)	(0.0698)	(0.0715)
	(0.0239)	(0.0280)	(0.0235)	(0.0236)	(0.0124)	(0.0151)
Sex	-1.504	-1.377	-1.456	-1.479	-0.980	-0.915
(female = 1)	(1.803)	(1.799)	(1.802)	(1.802)	(1.865)	(1.925)
	(0.405)	(0.445)	(0.420)	(0.412)	(0.600)	(0.635)
Income	0.157	0.000201	0.000160	0.000153	-0.000560	1.20e-05
	(0.104)	(0.000171)	(0.000111)	(9.90e-05)	(0.00116)	(0.000193)
	(0.134)	(0.241)	(0.149)	(0.124)	(0.631)	(0.950)
Household Size	-0.235	-0.258	-0.240	-0.235	-0.384	-0.318
	(0.260)	(0.257)	(0.258)	(0.259)	(0.286)	(0.282)
	(0.366)	(0.316)	(0.354)	(0.364)	(0.180)	(0.261)
Education	0.987	0.995	0.988	0.987	0.905	1.233
(std by pop)	(0.786)	(0.790)	(0.786)	(0.785)	(0.823)	(0.847)
	(0.210)	(0.209)	(0.210)	(0.210)	(0.272)	(0.146)
Community Size	-0.135	-0.153	-0.134	-0.135	-0.174	-0.186
(100 people)	(0.105)	(0.109)	(0.105)	(0.105)	(0.128)	(0.152)
	(0.200)	(0.160)	(0.201)	(0.198)	(0.174)	(0.221)
Wealth	-0.00625	3.04e-06	-5.68e-06	-6.16e-06	0.00152***	0.000119
	(0.00785)	(1.34e-05)	(8.16e-06)	(7.44e-06)	(0.000544)	(0.000154)
	(0.426)	(0.821)	(0.487)	(0.408)	(0.00550)	(0.443)
Constant	22.49***	22.68***	22.47***	22.48***	20.07***	21.20***
	(3.728)	(3.744)	(3.726)	(3.726)	(3.747)	(3.882)
	(4.55e-09)	(4.00e-09)	(4.63e-09)	(4.58e-09)	(1.76e-07)	(1.04e-07)
Observations	319	319	319	319	293	293
R-squared	0.140	0.139	0.140	0.140	0.152	0.114

Robust standard errors and p-values in parentheses below the coefficient
*p < 0.1 **p< 0.05 ***p<0.01
Education has been standardized to a mean of zero and standard deviation of one within each population

In order to include the data from Accra and our Siberian site we also ran the same analyses shown above with wealth and CS dropped. The coefficients on market integration, world religion, and age show the same kind of robustness seen above.

To further address concerns about the comparability of income and wealth across populations, we have also estimated a series of models using standardized versions of wealth and income. We repeat the procedure used above. Table S13 below shows that both MI and WR have large coefficients and remain highly significant across all specifications. The coefficient on age also remains significant at conventional levels across specifications.

Table S13. Linear regressions for UG offers using standardized Income and Wealth

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Model	Model	Model	Model	Model	Model	Model
MI	0.114***	0.119***	0.139***	0.137***	0.141***	0.142***	0.102***
	(0.0321)	(0.0328)	(0.0328)	(0.0325)	(0.0315)	(0.0312)	(0.0223)
	(0.000424)	(0.000349)	(2.75e-05)	(3.02e-05)	(9.79e-06)	(7.18e-06)	(6.29e-06)
WR	10.67***	10.60***	9.005***	8.829***	8.693***	8.833***	7.940***
	(2.648)	(2.636)	(2.450)	(2.412)	(2.337)	(2.303)	(2.244)
	(7.07e-05)	(7.33e-05)	(0.000277)	(0.000292)	(0.000234)	(0.000150)	(0.000455)
Age	0.150**	0.168**	0.179***	0.184***	0.183***	0.163***	0.150***
(years)	(0.0655)	(0.0667)	(0.0631)	(0.0619)	(0.0617)	(0.0606)	(0.0567)
	(0.0227)	(0.0124)	(0.00495)	(0.00321)	(0.00332)	(0.00761)	(0.00859)
Comm. Size	-0.145	-0.149	-0.186	-0.191*	-0.216*	-0.214*	
(100 people)	(0.106)	(0.110)	(0.114)	(0.114)	(0.116)	(0.115)	
	(0.173)	(0.176)	(0.104)	(0.0939)	(0.0627)	(0.0636)	
Education	0.960	1.039	1.033	1.091	1.182		
(std by pop)	(0.826)	(0.793)	(0.741)	(0.740)	(0.744)		
	(0.246)	(0.191)	(0.164)	(0.141)	(0.113)		
Household	-0.325	-0.256	-0.282	-0.276			
Size	(0.253)	(0.258)	(0.255)	(0.254)			
	(0.199)	(0.323)	(0.270)	(0.276)			
Sex	-1.305	-1.615	-1.336				
(female = 1)	(1.807)	(1.822)	(1.747)				
	(0.471)	(0.376)	(0.445)				
Wealth	0.471	0.417					
(std by pop)	(1.002)	(0.977)					
	(0.639)	(0.670)					
Income	0.272						
(std by pop)	(0.971)						
	(0.780)						
Constant	22.59***	21.58***	22.52***	21.93***	20.39***	20.98***	21.39***
	(3.753)	(3.815)	(3.683)	(3.585)	(3.193)	(3.171)	(3.119)
	(4.97e-09)	(3.51e-08)	(2.67e-09)	(2.65e-09)	(5.62e-10)	(1.43e-10)	(0)
Observations	319	321	344	344	346	347	377
R-squared	0.136	0.139	0.137	0.135	0.132	0.127	0.111

Robust standard errors and p-values in parentheses below the coefficient
Income, Wealth, and Education have been standardized to a mean of zero and standard deviation of one within each population. *p < 0.1 **p< 0.05 ***p<0.01.

As above, to address the possibility that income and wealth may have different effects within versus between populations, we broke both income and wealth down into population mean measures and within population deviations from the population mean (see above for detailed description). These analyses find no evidence of income or wealth effects in the UG. None of the coefficients on any of these wealth or income measures approached conventional levels of significance. Meanwhile, the coefficients on MI and WR do not vary much from those shown above.

THIRD PARTY PUNISHMENT OFFERS

For offers in the Third Party Punishment Game our analyses do again show the predictive importance of market integration, but do not show parallel findings for World Religion. In addition to these two theoretically important variables, we also find that higher incomes predict lower TPG offers, more wealth predicts higher offers, and larger households predict lower offers (Model 2 in Table S14). We return to these results at the end of our discussion of offer variation.

To address concerns about both historically shared cultural phylogeny and the non-independence of individuals from the same groups, compare Model 1 (with continental controls and clustered robust standard errors) with Model 2 (baseline) in Table S14. Since we lack data from the Siberian sites and from the U.S. for the TPG, the regression uses Africa as the reference for the continental dummies. These modifications have no quantitative effect on our main results in the TPG. The coefficients on income, wealth and household size remain large and significant with the addition of continental controls and use of clustered robust errors. Interestingly, the coefficient on community size does increase and becomes highly significant in Model 1.

Models 2 to 7 allow comparison of the coefficients on MI, as the least significant predictors are sequentially dropped from the model. MI is again significant and meaningful. An increase of 20 percent in calories purchased in the market is associated with an increase in percent offered of roughly 2.3. For income, every $1000 increase predicts a decrease in percent offered of between 1.9 and 2.4. For wealth, each increment of $1000 predicts an increase in percent offered of 1.2. For household size, each additional member of a player’s household decreases offers by about 1 percentage point.

Table S14 Linear regressions for Third Party Game offers[3]

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Model	Model	Model	Model	Model	Model	Model
MI (mean)	0.103***	0.111**	0.111**	0.112**	0.109**	0.101**	0.118***
(% purchased)	(0.0173)	(0.0436)	(0.0435)	(0.0435)	(0.0433)	(0.0421)	(0.0416)
	(0.000212)	(0.0116)	(0.0113)	(0.0107)	(0.0123)	(0.0168)	(0.00494)
Wealth	0.993***	1.203***	1.207***	1.210***	1.212***	1.158***	1.184***
(1000 USD)	(0.112)	(0.248)	(0.246)	(0.243)	(0.242)	(0.242)	(0.237)
	(9.61e-06)	(2.09e-06)	(1.65e-06)	(1.17e-06)	(1.03e-06)	(2.89e-06)	(1.04e-06)
Income	-3.177***	-2.250**	-2.259**	-2.244**	-2.149**	-2.238**	-1.873**
(1000 USD)	(0.633)	(0.938)	(0.930)	(0.915)	(0.889)	(0.889)	(0.906)
	(0.000716)	(0.0171)	(0.0158)	(0.0149)	(0.0163)	(0.0124)	(0.0396)
Household Size	-0.953**	-1.005**	-1.013**	-1.015**	-1.026**	-0.996**	-0.914**
	(0.358)	(0.425)	(0.420)	(0.419)	(0.419)	(0.420)	(0.413)
	(0.0259)	(0.0188)	(0.0167)	(0.0162)	(0.0149)	(0.0184)	(0.0277)
Community Size	0.294***	0.0910	0.0910	0.0931	0.0917	0.0971	
(100 people)	(0.0419)	(0.0682)	(0.0681)	(0.0664)	(0.0660)	(0.0644)	
	(6.17e-05)	(0.183)	(0.183)	(0.162)	(0.166)	(0.133)	
Age (years)	0.0254	0.0509	0.0564	0.0574	0.0580		
	(0.0757)	(0.0856)	(0.0812)	(0.0810)	(0.0810)		
	(0.745)	(0.553)	(0.488)	(0.480)	(0.474)		
Sex	-0.410	-0.763	-0.715	-0.690			
(female = 1)	(2.046)	(2.308)	(2.317)	(2.292)			
	(0.846)	(0.741)	(0.758)	(0.764)			
World Religion	-3.939**	0.651	0.515				
(binary)	(1.304)	(2.995)	(2.948)				
	(0.0144)	(0.828)	(0.861)				
Education	-0.702	-0.273					
(std by pop)	(1.785)	(1.305)					
	(0.703)	(0.834)					
South America	10.59***						
(dummy)	(1.497)						
	(5.84e-05)						
Oceania	9.587***						
(dummy)	(1.568)						
	(0.000176)						
Constant	27.85***	30.27***	30.15***	30.46***	30.21***	33.01***	33.18***
	(4.448)	(5.228)	(5.195)	(4.938)	(4.788)	(3.000)	(2.998)
	(0.000148)	(2.06e-08)	(1.92e-08)	(2.66e-09)	(1.21e-09)	(0)	(0)
Observations	265	265	265	265	265	272	272
R-squared	0.130	0.097	0.097	0.096	0.096	0.086	0.080
Model 1 uses clustered robust standard errors (clustering on site); other models use robust standard errors; errors and p-values are in parentheses below the coefficient. Education has been standardized to a mean of zero and standard deviation of one within each population. *p < 0.1 **p< 0.05 ***p<0.01							

[3] In the TPG, and only the TPG, we lack data on income and wealth among the Tsimane.

We again examined the effect of using international exchange rates to convert our measures of income and wealth into U.S. dollars by deploying five different models in which these two variables were re-valued using local prices of consumables. The coefficient on MI remains large, varying from 0.091 to 0.11, and is significant across all models. Wealth also remains significant across all six models, as does household size. Income is significant in all models except Model 5, where the price of cooking oil is used to recalibrate our measures. Note that the sizes of the coefficients on income and wealth cannot be easily compared across models since units are different depending on the consumable used.

Table S15. Linear regressions for TPG offers with local Wealth and Income calibrations

Variables	U.S. $	D battery	100kg rice	100kg sugar	100L cooking oil	100kg salt
MI	0.111**	0.108**	0.0913**	0.0990**	0.110**	0.0976**
(% purchased)	(0.0436)	(0.0444)	(0.0440)	(0.0435)	(0.0441)	(0.0444)
	(0.0116)	(0.0157)	(0.0390)	(0.0236)	(0.0133)	(0.0286)
World Religion	0.651	-0.210	1.411	1.557	0.0926	0.921
	(2.995)	(3.036)	(2.984)	(2.950)	(3.009)	(2.938)
	(0.828)	(0.945)	(0.637)	(0.598)	(0.975)	(0.754)
Income	-2.250**	-0.000580**	-0.00122**	-0.00144***	-0.00247*	-0.000555**
	(0.938)	(0.000263)	(0.000570)	(0.000555)	(0.00135)	(0.000221)
	(0.0171)	(0.0284)	(0.0332)	(0.00985)	(0.0689)	(0.0126)
Wealth	1.203***	0.000407***	0.000843***	0.000982***	0.00130***	0.000463***
	(0.248)	(9.70e-05)	(0.000193)	(0.000215)	(0.000271)	(0.000118)
	(2.2e-06)	(3.75e-05)	(1.83e-05)	(7.41e-06)	(2.61e-06)	(0.000112)
Household Size	-1.005**	-0.971**	-1.195***	-1.114**	-0.936**	-1.047**
	(0.425)	(0.428)	(0.439)	(0.432)	(0.424)	(0.430)
	(0.0188)	(0.0240)	(0.00689)	(0.0105)	(0.0279)	(0.0156)
Community Size	0.0910	0.0747	0.104	0.0816	0.115*	0.0520
(100 people)	(0.0682)	(0.0693)	(0.0660)	(0.0677)	(0.0656)	(0.0721)
	(0.183)	(0.282)	(0.116)	(0.229)	(0.0800)	(0.471)
Age (years)	0.0509	0.0584	0.0542	0.0489	0.0459	0.0523
	(0.0856)	(0.0861)	(0.0868)	(0.0861)	(0.0858)	(0.0863)
	(0.553)	(0.498)	(0.533)	(0.571)	(0.594)	(0.545)
Education	-0.273	-0.177	-0.244	-0.371	-0.319	-0.354
(std by pop)	(1.305)	(1.318)	(1.318)	(1.312)	(1.313)	(1.330)
	(0.834)	(0.893)	(0.853)	(0.778)	(0.808)	(0.791)
Sex	-0.763	-0.235	-0.359	-0.579	-0.451	-0.182
(female = 1)	(2.308)	(2.302)	(2.313)	(2.311)	(2.295)	(2.313)
	(0.741)	(0.919)	(0.877)	(0.802)	(0.844)	(0.937)
Constant	30.27***	30.20***	30.72***	30.58***	29.99***	30.50***
	(5.228)	(5.193)	(5.218)	(5.202)	(5.240)	(5.156)
	(2.1e-08)	(1.81e-08)	(1.23e-08)	(1.29e-08)	(2.93e-08)	(1.06e-08)
Observations	265	265	265	265	265	265
R-squared	0.097	0.086	0.089	0.097	0.092	0.093

Education has been standardized to a mean of zero and standard deviation of one within each population
*p < 0.1 **p< 0.05 ***p<0.01
Robust standard errors and p-values in parentheses below the coefficient

To further address concerns about the comparability of income and wealth across populations, we have also estimated a series of models using a standardized version of wealth and income, as described above. While these models show no effects of wealth (standardized), income (standardized), or household size, MI's coefficient remains large across all seven specifications. MI is significant at conventional levels in three of the models and has *p*-values just above 0.05 in four of the specifications. Compared to the models using either U.S. dollars or local consumable prices for wealth and income, these models explain much less of the variation, with the percentage of variance explained dropping from about 9% to 4%.

Table S16. Linear regressions for TPG offers using standardized Income and Wealth

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Model	Model	Model	Model	Model	Model	Model
MI	0.0875**	0.0869**	0.0816*	0.0833*	0.0832*	0.0840**	0.0750*
	(0.0436)	(0.0435)	(0.0426)	(0.0424)	(0.0424)	(0.0421)	(0.0420)
	(0.0460)	(0.0470)	(0.0563)	(0.0505)	(0.0505)	(0.0471)	(0.0757)
Income	-1.037	-1.017	-1.097	-1.083	-1.114	-1.187	-1.232
(std by pop)	(0.943)	(0.929)	(0.937)	(0.935)	(0.896)	(0.888)	(0.890)
	(0.273)	(0.275)	(0.243)	(0.248)	(0.215)	(0.183)	(0.167)
Community Size	0.0823	0.0825	0.0797	0.0885	0.0821	0.0819	0.0675
(100 people)	(0.0663)	(0.0659)	(0.0658)	(0.0630)	(0.0625)	(0.0625)	(0.0616)
	(0.216)	(0.212)	(0.227)	(0.161)	(0.190)	(0.191)	(0.274)
Household Size	-0.590	-0.595	-0.593	-0.596	-0.553	-0.572	
	(0.455)	(0.458)	(0.457)	(0.456)	(0.438)	(0.435)	
	(0.196)	(0.195)	(0.196)	(0.192)	(0.208)	(0.189)	
Education	-0.567	-0.556	-1.085	-0.952	-0.738		
(std by pop)	(1.368)	(1.365)	(1.330)	(1.289)	(1.273)		
	(0.679)	(0.684)	(0.415)	(0.461)	(0.563)		
Wealth	0.730	0.747	0.984	0.944			
(std by pop)	(1.866)	(1.856)	(1.805)	(1.798)			
	(0.696)	(0.688)	(0.586)	(0.600)			
World Religion	1.184	1.167	1.876				
	(3.144)	(3.117)	(3.141)				
	(0.707)	(0.708)	(0.551)				
Age	0.0251	0.0256					
(years)	(0.0872)	(0.0867)					
	(0.774)	(0.768)					
Sex	-0.216						
(female = 1)	(2.319)						
	(0.926)						
Constant	30.58***	30.52***	31.46***	32.79***	32.68***	32.75***	29.91***
	(5.350)	(5.263)	(3.847)	(3.217)	(3.066)	(3.052)	(2.163)
	(3.04e-08)	(1.96e-08)	(0)	(0)	(0)	(0)	(0)
Observations	265	265	271	271	274	275	275
R-squared	0.037	0.037	0.037	0.036	0.034	0.034	0.028

Robust standard errors and p-values in parentheses below the coefficient
Income, Wealth, and Education has been standardized to a mean of zero and standard deviation of one within each population.
*p < 0.1 **p< 0.05 ***p<0.01

OFFER REGRESSIONS USING MAO AS A PREDICTOR

To address concerns that our use of UG and TPG offers as measures of “fairness” might not be appropriate, since as we explain the main text, these measures may combine a variety of social preferences or motivations as well as beliefs about the likelihood of punishment or rejection, we introduce an MAO control.

In this paper we are primarily interested in *behavioral fairness* (what people do), and not with arguing for the importance of any one internal motivation, preference (e.g., equity aversion) or belief (e.g., punishment is likely). However, building on our DG findings (which lack any threat of punishment), we do think our data suggest that internalized motivations are part of the observed phenomena. To address this concern, we re-ran our baseline regression with each population's mean MAO for the UG and TPG, respectively. The idea is that if individuals are responding to the local chances of rejection or punishment in making their offers, the MAO of an individual's population should capture something of this response. Comparing Models 1 and 2 for the UG and Models 3 and 4 for the TPG indicates little impact on our key theoretical variables (Table S17). The coefficient on MI in UG increases from 0.098 to 0.13 with the addition of mean MAO (in the UG) as a predictor, bringing it more into line with the coefficient of MI on DG offers. Both the coefficients on MAO mean are in the expected direction, with more punishment predicting greater offers, though neither is significant at conventional levels. Little else changes in the analysis, suggesting that even when the local threat of punishment is controlled for, MI and WR (in the UG) still predict offers.

Table S17. Linear regressions for offers in UG and TPG with and without using the mean MAO as a predictor

Variables	UG Offers		TPG Offers	
	Model 1	Model 2	Model 3	Model 4
MI (community mean)	0.0978***	0.132**	0.111**	0.110**
(% calories purchased)	(0.0349)	(0.0597)	(0.0436)	(0.0439)
	(0.00547)	(0.0272)	(0.0116)	(0.0132)
World Religion (binary)	10.44***	9.886***	0.651	0.404
	(2.665)	(2.676)	(2.995)	(3.113)
	(0.000110)	(0.000261)	(0.828)	(0.897)
MAO mean		0.201		0.0461
		(0.154)		(0.165)
		(0.192)		(0.779)
Age (years)	0.149**	0.134**	0.0509	0.0503
	(0.0655)	(0.0649)	(0.0856)	(0.0860)
	(0.0239)	(0.0396)	(0.553)	(0.559)
Income (1000 USD)	0.157	0.0943	-2.250**	-2.169**
	(0.104)	(0.121)	(0.938)	(0.913)
	(0.134)	(0.438)	(0.0171)	(0.0182)
Community Size	-0.135	-0.249	0.0910	0.0689
(100 people)	(0.105)	(0.208)	(0.0682)	(0.115)
	(0.200)	(0.232)	(0.183)	(0.548)
Education (std by pop)	0.987	1.096	-0.273	-0.293
	(0.786)	(0.789)	(1.305)	(1.302)
	(0.210)	(0.166)	(0.834)	(0.822)
Sex (female = 1)	-1.504	-1.729	-0.763	-0.590
	(1.803)	(1.814)	(2.308)	(2.306)
	(0.405)	(0.341)	(0.741)	(0.798)
Wealth (1000 USD)	-0.00625	-0.00940	1.203***	1.202***
	(0.00785)	(0.00891)	(0.248)	(0.248)
	(0.426)	(0.292)	(2.09e-06)	(2.10e-06)
Household Size	-0.235	-0.121	-1.005**	-1.002**
	(0.260)	(0.309)	(0.425)	(0.428)
	(0.366)	(0.695)	(0.0188)	(0.0198)
Constant	22.49***	19.84***	30.27***	29.74***
	(3.728)	(4.514)	(5.228)	(5.605)
	(4.55e-09)	(1.53e-05)	(2.06e-08)	(2.44e-07)
Observations	319	319	265	265
R-squared	0.140	0.150	0.097	0.097
Robust standard errors and p-values in parentheses below the coefficient. *p < 0.1 **p< 0.05 ***p<0.01				

WHY DOES THE EFFECT OF WORLD RELIGION DISAPPEAR IN THE THIRD PARTY PUNISHMENT GAME?

Offers in the TPG differ from offers in the DG and UG in two interesting respects. First, TPG offers were generally lower than offers in the other two games. Overall mean offers across these three games were

32 (TPG), 37 (DG) and 39 (UG) percent of the stake. Of our 12 populations, only 3 had TPG mean offers greater than their DG mean (greater by only 1.7 in percent offered, on average), while 9 had mean offers in the TPG less than in the DG (less by 6.3 on average). Both parametric and non-parametric comparisons of means show the TPG means are lower than both the DG and UG means (all $p < 0.02$). The DG and UG mean offers, however, cannot be distinguished at conventional levels of confidence (see Table S5). Second, while WR is a potent predictor of offers in both the DG and UG, it is not for the TPG. Interestingly, while WR's effects disappear for TPG offers, the coefficients on income, wealth and household size all emerge as potent predictors.

Here we explore one hypothesis that may help account for both of these phenomena: the general drop in TPG offers (vs. DG and UG offers), and the reversal of relative predictive importance of WR and the economic variables (income, wealth, etc.). We acknowledge at the outset that this is post-hoc theorizing. Our hypothesis emerges from a long-standing and well-documented empirical phenomenon termed "crowding out", which suggests that external incentives (e.g., third party punishments and rewards) can under some circumstances reduce non-self-regarding intrinsic motivations (S*36, S37*). We suspect that adding a third party who can impose fines—as in our TPG—may drive out some of people's intrinsic motivation toward equal divisions, whether those be based on altruism, inequity aversion, norm-adherence, or some other motivation. This phenomenon may have emerged as particularly salient in our version of the TPG because our setup prevented third parties from punishing sufficiently to discourage purely self-interested players from offering zero.

To explore this hypothesis we reasoned that any differences between DG offers or UG offers (see below) and TPG offers among populations created by the crowding out effect should be manifest in the predictive effects of World Religion, income and/or wealth (or wealth per household member). Participation in a world religion may imbue individuals with ethical principles or prosocial motivations (toward those beyond close durable relationships) that they seek to demonstrate to themselves, God, or others. World religions, through ritualized reminders, may make us judge ourselves in ways more critical, explicit, and harsh than other systems of belief. The threat of a fine might destroy either the signaling content of an individuals' generosity, or it might diminish the intrinsic pleasure or satisfaction derived from it (or both). For example, one might take pleasure in personal altruism towards one's fellow humans, or experience the pride and approval of the All Mighty by taking the time to give blood; however, if one gets paid $10 for the blood, that good feeling and sense of divine approval may disappear, while the cash might not provide sufficient compensation for the cost in time and discomfort.

With regard to our experiments, this reduction in non-pecuniary intrinsic motivations may be moderated by a player's wealth or income, because these would impact the degree to which a game decision influences one's material self-interest. For example, when richer individuals tradeoff their—now reduced—fairness motivations against their material self-interest, fairness motivations will be relatively more important (compared to poorer individuals) because the same amount of money will have a smaller impact on the material self-interest of richer individuals (concave utility functions). Alternatively, less well off people may be looking for a rationalization to be less fair, and they find it once the responsibility for enforcing fairness is deflected to a third party. From this, the prediction is straightforward: populations with greater participation in world religions ought to show a larger difference between their DG and TPG offers, once economic variables are controlled for. Greater income or wealth ought to favor smaller differences, since higher values mean the decision has a smaller effect on material self-interest.

To analyze this we use the population means because different players were used in the DG and TPG. We regressed the mean population differences between DG and TPG on mean values of WR, income, and wealth (absolute and per household member). However, as noted previously, at the population aggregate level income and wealth (and wealth per household member) are all highly intercorrelated (all > 0.9), so we paired each with WR in Table S18 (Models 1, 2, and 3). The effects of WR and each of our economic variables goes in the direction predicted by our hypothesis. The coefficients on WR are large in all three models, though only significant at conventional levels when income is used. The coefficients on income, wealth and wealth per household member are all negative and large, though only the two wealth measures are significant at 0.05. Models (1) to (3) show that either WR is significant or the economic variable is significant. To include all 12 of our data points, we use income in Models (4) and (5).

Model 4 adds the mean MAO in the TPG to Model 1, to control for the effects of the threat of punishment. The coefficient on WR increases, achieving conventional levels of significance. Model 5 adds MI to this, and the coefficient on WR remains significant. Societies with more participation in world religions show larger drops in offers from the DG to TPG. Model 5 captures 57% of the variation in the difference between DG and TPG mean offers.

Table S18. Linear regressions on the difference between mean DG and TPG offers

VARIABLES	(1) Model	(2) Model	(3) Model	(4) Model	(5) Model
WR	6.350**	4.931	4.956	7.466**	6.926**
	(2.696)	(2.992)	(2.763)	(3.064)	(2.826)
	(0.0429)	(0.138)	(0.111)	(0.0408)	(0.0440)
Income	-2.973			-2.462	-3.482
(per $1000)	(2.038)			(1.946)	(2.215)
	(0.179)			(0.241)	(0.160)
Wealth		-1.135**			
(per $1000)		(0.344)			
		(0.0109)			
Wealth per HH member			-6.980**		
($1000 per HH member)			(2.377)		
			(0.0188)		
mean MAO TPG				-0.0987	-0.178
				(0.102)	(0.122)
				(0.363)	(0.188)
MI					0.0927*
					(0.0450)
					(0.0783)
Constant	0.846	1.852	1.667	1.644	0.849
	(2.178)	(2.839)	(2.636)	(2.489)	(1.551)
	(0.707)	(0.532)	(0.545)	(0.527)	(0.601)
Observations	12	11	11	12	12
R-squared	0.202	0.468	0.459	0.254	0.572

Robust standard errors and p-values in parentheses below the coefficient. *p < 0.1 **p< 0.05 ***p<0.01. Wealth per HH member is household wealth divided by household size.

Table S19 provides analyses parallel to those observed in Table S18, now for the difference between mean UG and TPG offers. Model 1 shows that both WR and income have effects in the predicted directions, with large coefficients, though only WR's coefficient is significant at conventional levels. Adding the difference in the mean MAOs in the UG and TPG as a predictor in Model (2), to control for differences in the expectation of punishment, shows the coefficients on both income and WR change little. Model 3 shows that adding MI to this estimation changes little, though now the coefficient on the difference in MAO is larger and marginally significant.

Table S19. Linear regressions for the difference in UG and TPG offers

VARIABLES	(1) Model	(2) Model	(3) Model
WR	11.02***	11.88**	11.86**
	(3.237)	(3.669)	(3.668)
	(0.00781)	(0.0119)	(0.0144)
Income	-4.322	-4.295	-5.585
(per $1000)	(2.718)	(2.950)	(3.075)
	(0.146)	(0.184)	(0.112)
Diff (UGMAO –TPGMAO)		0.100	0.255*
		(0.130)	(0.128)
		(0.462)	(0.0859)
MI			0.0864
			(0.0618)
			(0.205)
Constant	-0.654	-0.854	-2.499
	(1.862)	(1.810)	(2.074)
	(0.733)	(0.650)	(0.267)
Observations	12	12	12
R-squared	0.255	0.279	0.369
Robust standard errors and p-values in parentheses below the coefficient. *p < 0.1 **p< 0.05 ***p<0.01			

Lest the reader be concerned that this drop in offers found in moving from the DG to the TPG only occurs in the unusual contexts of our diverse small-scale societies, where for example, anti-social punishment may occur (S*4, S38*), our protocols have also been administered to university students in the U.S. (S*4, S5*). The undergraduate means offers are 32% (DG), 39% (UG) and 27% (TPG). Undergraduates show a five percent drop in moving from the DG to the TPG, making them about average in our sample of societies. This suggests that alternative explanations based on anti-social punishment (S*38, S39*) are not likely.

Note, we have not included these student samples in our analyses because, unlike all our other populations, these participants are not random samples from an overall community, but instead represent a narrow age range of high SES individuals (Emory University), most of whom lack any real economic responsibilities. It is well established that student behavior in experiments like ours has not reached its adult plateau (S*40, S41, S42*). Student prosociality in experiments hugs the lower bound of prosociality observed in older adults (S*43, S44*), and student prosociality continues to increase over the university years (S*45*). Entering these samples into our database for this analysis would mix adult variation among populations with developmental variation.

While necessarily tentative, the analysis presented in this section provides some very preliminary support for our religion-crowding-out hypothesis described above, as world religion and the economic variables seem to account for a non-trivial fraction of the differences between the TPG offers and the other two experiments. Our small sample of twelve populations makes stronger conclusions impossible.

WHY DOESN'T THE PUNISHMENT THREAT REDUCE OFFERS IN THE UG TOO?

When we performed analyses parallel to those above on the differences between DG and UG offers, nothing was ever significant at conventional levels and the R^2 never went above 0.06. However, given our tentative results above, suggesting that crowding out effects may be important, we must ask why the threat of punishment present in the UG does not cue the same loss of intrinsic motivation that we observe for the TPG. We consider two possibilities and evaluate their consistency with the data.

Hypothesis 1: the strength of punishment is structurally stronger in the UG compared to the TPG. For low UG offers, punishment is cheap for Player 2 and costly for Player 1. For higher offers in the UG, nearing 50/50, punishment becomes quite expensive for Player 2 and the costs inflicted on Player 1 are lower. In contrast, in our particular design of the TPG, punishing low and high offers costs exactly the same amount, and the damage to Player 1 is the same. Consequently, the threat of punishment in the TPG may be structurally too weak to fully compensate for the loss of intrinsic motivation created by the threat of punishment. An income maximizer in our TPG will still give zero even when he believes punishment is likely (this is not the case in the UG). This *compensatory hypothesis* suggests that the same motivational loss occurs for both the UG and TPG, but that only the UG has sufficiently potent punishment to compensate.

The compensatory hypothesis does not hold up. The above mentioned analyses of the difference between mean DG and UG offers find only small and non-significant coefficients on MAO in the UG. Populations with greater punishment in the UG do not show greater differences in mean offers. We also substituted other statistics for the mean, including the 80^{th} and 90^{th} percentile MAO (this is the offer that gives an 80% or 90% chance of acceptance). These measures of the threat of punishment also showed no relationship to the difference in mean DG and UG offers. In short, the difference between DG and UG does not seem related to any differences in the actual threats of punishment, which are substantial. This remains true even when mean DG offer is controlled for, thereby addressing differences in intrinsic motivation.

Hypothesis 2: There may be an important psychological difference between the threat of second party punishment, which can be motivated by revenge, and that of third party punishment. Perhaps the possibility of second party punishment is perceived as endemic to any interaction, while third party punishment is cognitively perceived and encoded as an external source of rewards and punishments. We do not have a direct test for this hypothesis. However, our findings above are broadly consistent with it in showing the same patterns for explaining the difference between mean DG offer and mean TPG offers, and between mean UG mean offers and mean TPG offers.

Finally, we want to emphasize that this section has been a post-hoc attempt to grapple with something that unexpectedly emerged from our findings. Our project was not designed to address these questions, so these unexpected patterns may be merely an artifact of our design. The DG and UG were played in rapid succession, with the same people as Player 1 in both games (without feedback). The TPG was either played with different participants, three weeks or more later in the same community, or was played in a different community within the same population. Comparing the DG and TPG raises the fewest concerns, since the DG always came first and the TPG was played much later with different people (so it was first for them). However, comparisons of the difference between the DG and UG on one hand and the DG and TPG on the other raise a variety of methodological concerns that cannot be adequately addressed in our design. Overall, we think the only thing this analysis strongly favors is further investigation.

SUMMARY OF OFFER ANALYSIS

Overall, the above analyses of offers in all three experiments increase our confidence in the robustness of the effects observed for MI and WR. *MI is completely robust to all checks.* In contrast, while WR is fairly robust in the UG and DG, we should be less confident in the WR finding, both because of the rather small proportion of our sample with WR=0 and because of weaknesses that emerge in some of the analyses and checks. It was concerns about the historical non-independence of our sites that prompted the use of such robustness checks in the first place, and given the outcome of our additional checks, we think there is reason to believe this is not a problem.

PUNISHMENT IN THE THIRD PARTY PUNISHMENT GAME

Community size predicts more punishment in both the UG and TPG. People from larger communities engage in more costly punishment. As discussed in the main text, theoretical work examining various mechanisms capable of sustaining costly norms, including those associated with fairness among strangers, suggests that smaller groups can sustain costly norms with reputational systems that, for example (S*46*), allow individuals to withdraw help from norm violators (instead of punishing them at a personal cost). Larger cooperative groups require costly punishment because reputational systems rapidly breakdown as group size increases (S*47*). At the same time reputational systems are collapsing, the anonymity of larger groups mitigates the threat of counter-punishment, thereby increasing the range of conditions in which costly punishment can sustain larger scale cooperation and prosocial norms. That is, the possibility of counter-punishment, punishing someone back who punished you, threatens the effectiveness of punishment in maintaining cooperation in smaller groups. This problem declines in larger populations because anonymity increases, so "pushing back" is less more costly or difficult. This line of theory predicts that larger coherent communities must have diffuse punishment—otherwise they will break down and not remain large communities anymore. Since some theoretical work suggests that reputational breakdown is roughly proportional to the natural logarithm of the group size (S*48*), we will use both community size and the natural logarithm of community size (LNSC) as the key theoretical predictor of variables below (S*48*). We decided to use community size in the main text both because it was difficult to provide the necessary theoretical justification for using the natural logarithm in the available space, and because our theoretical justifications may be controversial. We also discuss the use of the square of community size.

As described above, we reduce our vector of punishment decisions for both the TPG and UG to a single number called a *minimum acceptable offer* (MAO). This is the lowest offer below 50 for which an individual would not reject (in the UG) or pay to punish (in the TPG). To analyze the MAOs in both the

UG and TPG we used an ordered logistic regression instead of a least squares regression because our dependent variables (MAOs) are both discrete and bimodal. The diagnostics for our initial linear regressions indicated that basic assumptions were being dramatically violated. The ordered logistic regression (OLR) assumes that the dependent variables are discrete and rank ordered, but the distance between discrete ranks is not meaningful. It is worth noting that all of the important results highlighted below can also be found using linear regression analyses that assume MAO is a continuous normally distributed variable.

All coefficients shown and discussed below are reported as odds ratios, for ease of interpretation.

The TPG provides the most straightforward measure of an individual's willingness to engage in diffuse norm-enforcing costly punishment, because unlike the UG, the motivation behind this punishment cannot be simply revenge for a direct personal slight.

Model 2 in both Tables S20A and S20B show that CS and LNCS (logarithm of community size, respectively, are large and significant predictors of MAOs in the TPG. Focusing first on Table S20A: comparing Model 1 (continental controls plus clustered robust standard errors) and Model 2 (baseline) shows that adding both continental controls and using clustered robust standard errors results in little change to our findings. Models 2 through 9 test the robusticity of community size (CS) as a predictor, following the procedure used above. Instead of the coefficients, we have provided odds ratios. For an ordered logistic regression, the odds ratio tells us how much an increase of magnitude 100 people in community size influences an individual's chances of punishing in the next higher MAO category (e.g., the increased chance of delivering an MAO of 20 instead of 10). CS remains large and highly significant across all specifications. Figure S4 illustrates the effect size.

The reader should note that WR is positive and significant, or marginally so, across all these specifications. Participating in a world religion predicts punishment more in the TPG. We think this is worth noting, but have not made much of it because if we use the natural logarithm of community size (Table S20B), our preferred theoretical predictor, the WR effect evaporates.

Table S20A. Ordered Logistic Regression for MAO in TPG using Community Size

VARS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Model	Model	Model	Model	Model	Model	Model	Model	Model
CS	1.054***	1.053***	1.049***	1.049***	1.049***	1.049***	1.055***	1.054***	1.056***
(per 100)	(0.0170)	(0.0089)	(0.008)	(0.00798)	(0.00790)	(0.008)	(0.0073)	(0.0073)	(0.0072)
	(0.0010)	(1.1e-09)	(2.7e-10)	(2.25e-10)	(2.3e-10)	(2.8e-10)	(0)	(0)	(0)
World	2.430*	2.143**	2.408***	2.411***	2.340**	2.338**	1.772*	1.772*	
Religion	(1.113)	(0.750)	(0.812)	(0.821)	(0.773)	(0.772)	(0.547)	(0.538)	
(binary)	(0.0526)	(0.0294)	(0.0092)	(0.00972)	(0.0101)	(0.0101)	(0.0639)	(0.0596)	
Sex	0.619	0.632*	0.676	0.677	0.701	0.703	0.761		
(f = 1)	(0.204)	(0.163)	(0.161)	(0.161)	(0.163)	(0.165)	(0.167)		
	(0.146)	(0.0746)	(0.100)	(0.101)	(0.126)	(0.133)	(0.213)		
Wealth	0.981	0.987	0.976	0.976	0.972	0.971			
(1K USD)	(0.0480)	(0.0473)	(0.0446)	(0.0444)	(0.0417)	(0.0410)			
	(0.688)	(0.789)	(0.594)	(0.586)	(0.501)	(0.481)			
MI	1.000	1.005	1.001	1.001	1.001				

(% purch)	(0.0067)	(0.0058)	(0.0051)	(0.00476)	(0.00472)				
	(0.975)	(0.423)	(0.776)	(0.778)	(0.829)				
Income	0.916	0.952	0.955	0.956					
(1K USD)	(0.0965)	(0.0950)	(0.0924)	(0.0908)					
	(0.406)	(0.619)	(0.634)	(0.639)					
Household	0.990	0.986	0.994						
Size	(0.0426)	(0.0399)	(0.0370)						
	(0.824)	(0.722)	(0.872)						
Age	0.998	1.000							
(years)	(0.0130)	(0.0113)							
	(0.869)	(0.977)							
Education	0.936	0.958							
(std pop)	(0.121)	(0.138)							
	(0.610)	(0.766)							
S. Am.	1.566								
(dummy)	(0.736)								
	(0.340)								
Oceania	0.815								
(dummy)	(0.619)								
	(0.788)								
Constant	1.167	1.379	1.356	1.389	1.374	1.332	1.500	1.736*	1.111
Cut1	(0.525)	(0.909)	(0.555)	(0.517)	(0.509)	(0.420)	(0.459)	(0.488)	(0.184)
	(0.731)	(0.626)	(0.458)	(0.377)	(0.390)	(0.364)	(0.185)	(0.0501)	(0.526)
Constant	2.201	2.602	2.413**	2.522**	2.487**	2.407***	2.727***	3.153***	2.005***
cut2	(1.258)	(1.756)	(1.024)	(0.987)	(0.966)	(0.789)	(0.874)	(0.941)	(0.354)
	(0.168)	(0.157)	(0.0380)	(0.0181)	(0.0190)	(0.0074)	(0.0017)	(0.0001)	(8.4e-05)
Constant	3.805**	4.474**	3.934***	4.106***	4.125***	3.991***	4.434***	5.118***	3.237***
cut3	(2.310)	(3.073)	(1.721)	(1.661)	(1.657)	(1.369)	(1.483)	(1.602)	(0.634)
	(0.0277)	(0.0291)	(0.0017)	(0.00048)	(0.00042)	(5.5e-05)	(8.5e-06)	(1.8e-07)	(2e-09)
Constant	7.562***	8.800***	7.735***	8.067***	8.091***	7.838***	8.668***	9.985***	6.396***
cut4	(5.508)	(6.139)	(3.474)	(3.369)	(3.361)	(2.824)	(3.072)	(3.333)	(1.503)
	(0.00548)	(0.0018)	(5.3e-06)	(5.8e-07)	(4.84e-07)	(1.1e-08)	(1.1e-09)	(0)	(0)
Constant	22.16***	25.49***	21.42***	22.32***	22.37***	21.71***	23.64***	26.97***	17.14***
cut5	(20.34)	(18.47)	(10.43)	(10.31)	(10.29)	(8.986)	(9.737)	(10.63)	(5.319)
	(0.000740)	(7.9e-06)	(3.1e-10)	(0)	(0)	(0)	(0)	(0)	(0)
Obs	227	227	242	243	244	244	268	268	269
pseudo R2	0.0860	0.0837	0.0733	0.0735	0.0733	0.0732	0.0776	0.0759	0.0725
Model 1 uses clustered robust standard errors (clustering on site); other models use robust standard errors; p-values are in parentheses below the coefficient; Education has been standardized to a mean of zero and standard deviation of one within each population. Coefficients' reported as odds ratios. *p < 0.1 **p< 0.05 ***p<0.01									

Table S20B parallels the analysis in Table S20A, now using LNCS, the natural logarithm of community size, in place of CS. Comparing Models 1 and 2 shows that the odds ratio for the coefficient of LNSC increases from 2.4 to 2.7 and remains highly significance when phylogenetic controls are added and clustered robust standard errors are used. Little else changes in this specification. If anything, adding continental level controls increases the effects of LNCS.

In contrast to the results shown above for CS, these models also indicate that MI has a negative effect on MAO. This odds ratio indicates that more market integrated societies punish less. However, if LNCS is dropped from the specification, the direction and effect of MI switches to positive and significant. In fact, without LNSC in the model, the effect of MI changes direction and is significant across all of these specifications. MI and LNCS are correlated 0.6. The converse is not true. If LNSC is kept in the specification, but MI is dropped, the effect of LNCS remains large (odds ratio about 2) and highly significant.

We also tried using combinations of community size, community size squared, and the logarithm of Community size. The goodness of fit measures are somewhat better for the LNCS vis-à-vis CS and CS^2 (together), as we would expect from theory. When CS is entered along with CS^2, both are significant. The effect of CS is large and positive (or odds ratio > 1) and CS^2 is smaller and negative, suggesting a declining effect of CS as populations get large. Goodness of fit measures (both pseudo R^2 and AIC), however, still indicate that entering LNCS alone is superior to entering both CS and CS^2.

Table S20B. Ordered Logistic Regression for MAO in TPG using LNSC

VARS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Model	Model	Model	Model	Model	Model	Model	Model	Model	Model
LNCS	2.718***	2.434***	2.431***	2.396***	2.385***	2.232***	2.272***	2.272***	2.256***	1.981***
	(0.585)	(0.294)	(0.290)	(0.269)	(0.253)	(0.215)	(0.211)	(0.211)	(0.208)	(0.169)
	(3.38e-06)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
MI (mean)	0.974**	0.982**	0.983**	0.981***	0.982***	0.991*	0.990*	0.989**	0.989**	
(% purchased)	(0.0102)	(0.00699)	(0.00678)	(0.00614)	(0.00583)	(0.00548)	(0.00529)	(0.00526)	(0.00514)	
	(0.0132)	(0.0125)	(0.0111)	(0.00218)	(0.00205)	(0.0845)	(0.0543)	(0.0413)	(0.0284)	
Household Size	0.980	0.978	0.978	0.979	0.978	0.951	0.951	0.952		
	(0.0414)	(0.0399)	(0.0399)	(0.0366)	(0.0367)	(0.0340)	(0.0344)	(0.0347)		
	(0.640)	(0.592)	(0.590)	(0.572)	(0.545)	(0.157)	(0.168)	(0.179)		
Sex (female = 1)	0.615	0.638*	0.643*	0.671*	0.659*	0.739	0.738			
	(0.193)	(0.167)	(0.165)	(0.161)	(0.159)	(0.165)	(0.165)			
	(0.122)	(0.0853)	(0.0844)	(0.0959)	(0.0839)	(0.176)	(0.175)			
World Religion	1.407	1.393	1.386	1.521	1.448	1.225				
(binary)	(0.575)	(0.509)	(0.511)	(0.533)	(0.507)	(0.417)				
	(0.403)	(0.365)	(0.376)	(0.232)	(0.291)	(0.552)				
Income	0.901	0.966	0.965	0.964	0.939					
(1000 USD)	(0.107)	(0.101)	(0.0990)	(0.0980)	(0.0870)					
	(0.377)	(0.742)	(0.727)	(0.717)	(0.494)					
Wealth	0.965	0.978	0.978	0.972						
(1000 USD)	(0.0404)	(0.0474)	(0.0472)	(0.0456)						
	(0.400)	(0.642)	(0.642)	(0.546)						
Age (years)	0.997	1.001	1.002							
	(0.0123)	(0.0113)	(0.0100)							
	(0.810)	(0.946)	(0.880)							
Education	0.949	0.975								
(std by pop)	(0.127)	(0.138)								
	(0.694)	(0.858)								
Africa (dummy)	1.006									
	(0.536)									
	(0.992)									
South America	2.383**									

(dummy)	(0.902)									
	(0.0218)									
Constant	69.10***	51.21***	52.61***	48.69***	49.12***	45.41***	41.88***	48.71***	61.44***	39.80***
cut1	(59.02)	(37.45)	(37.47)	(29.47)	(29.39)	(27.27)	(23.41)	(27.01)	(34.12)	(22.53)
	(7.08e-07)	(7.36e-08)	(2.64e-08)	(1.36e-10)	(7.53e-11)	(2.08e-10)	(0)	(0)	(0)	(7.60e-11)
Constant	138.5***	102.7***	105.5***	91.92***	92.26***	84.24***	77.66***	90.09***	115.3***	74.82***
cut2	(145.6)	(78.55)	(78.90)	(58.47)	(57.88)	(52.87)	(45.18)	(52.12)	(67.05)	(44.54)
	(2.74e-06)	(1.39e-09)	(4.75e-10)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Constant	249.2***	183.4***	188.3***	156.4***	156.3***	139.2***	128.4***	148.4***	189.1***	122.1***
cut3	(267.4)	(145.7)	(147.1)	(103.9)	(102.2)	(90.72)	(78.01)	(89.59)	(114.8)	(75.74)
	(2.73e-07)	(5.36e-11)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Constant	520.7***	376.4***	386.4***	323.8***	323.1***	272.9***	251.6***	295.9***	375.7***	237.0***
cut4	(636.4)	(312.2)	(316.3)	(227.1)	(222.7)	(186.2)	(161.5)	(189.2)	(241.7)	(154.8)
	(3.08e-07)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Constant	1602***	1133***	1163***	936.4***	934.4***	741.7***	683.5***	795.0***	1002***	616.5***
cut5	(2315)	(1005)	(1022)	(710.3)	(696.0)	(544.8)	(475.9)	(551.0)	(700.4)	(434.3)
	(3.25e-07)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Observations	227	227	227	242	243	267	267	268	269	269
pseudo R2	0.118	0.112	0.112	0.104	0.105	0.0934	0.0930	0.0913	0.0900	0.0845

Model 1 used clustered robust standard errors (clustering on site); other models use robust standard errors; p-values are in parentheses below the coefficients. Education has been standardized to a mean of zero and standard deviation of one within each population. Coefficients reported as odds ratios

*p < 0.1 **p< 0.05 ***p<0.01

Following the same procedure we used above for offers, we ran six models with our wealth and income revalued by local prices in batteries, rice, sugar, cooking oil and salt. Table S21A shows all qualitative conclusions from our baseline model remain robust to this check.

Table S21A. OLR for MAO in TPG with local Wealth and Income calibrations using Community Size

VARS	$US	Oil	Sugar	Salt	battery	rice
Community Size	1.053***	1.051***	1.052***	1.054***	1.053***	1.051***
(100 people)	(0.00890)	(0.00825)	(0.00862)	(0.00957)	(0.00922)	(0.00816)
	(1.09e-09)	(1.78e-10)	(4.59e-10)	(6.73e-09)	(3.96e-09)	(1.25e-10)
MI (mean)	1.005	1.004	1.005	1.004	1.004	1.005
(% purchased)	(0.00579)	(0.00573)	(0.00561)	(0.00573)	(0.00587)	(0.00557)
	(0.423)	(0.434)	(0.383)	(0.483)	(0.495)	(0.375)
World Religion	2.143**	2.119**	2.138**	2.145**	2.144**	2.128**
(binary)	(0.750)	(0.748)	(0.738)	(0.740)	(0.748)	(0.732)
	(0.0294)	(0.0334)	(0.0277)	(0.0269)	(0.0288)	(0.0282)
Income	0.952	1.000	1.000	1.000	1.000	1.000
(1000 USD)	(0.0950)	(0.000133)	(6.11e-05)	(2.45e-05)	(2.90e-05)	(5.85e-05)
	(0.619)	(0.852)	(0.690)	(0.884)	(0.992)	(0.793)
Wealth)	0.987	1.000	1.000	1.000	1.000	1.000
(1000 USD	(0.0473)	(5.28e-05)	(3.80e-05)	(1.92e-05)	(1.77e-05)	(3.21e-05)
	(0.789)	(0.787)	(0.694)	(0.505)	(0.622)	(0.625)
Sex	0.632*	0.649*	0.632*	0.649*	0.654*	0.642*
(female = 1)	(0.163)	(0.166)	(0.163)	(0.167)	(0.167)	(0.165)
	(0.0746)	(0.0912)	(0.0749)	(0.0925)	(0.0958)	(0.0845)
Age (years)	1.000	1.001	1.000	1.001	1.000	1.000
	(0.0113)	(0.0112)	(0.0113)	(0.0113)	(0.0113)	(0.0113)
	(0.977)	(0.959)	(0.973)	(0.951)	(0.967)	(0.970)
Education	0.958	0.954	0.957	0.957	0.953	0.954
(std by pop)	(0.138)	(0.138)	(0.138)	(0.137)	(0.138)	(0.138)
	(0.766)	(0.747)	(0.762)	(0.758)	(0.740)	(0.747)
Household Size	0.986	0.986	0.987	0.990	0.988	0.989
	(0.0399)	(0.0398)	(0.0402)	(0.0405)	(0.0404)	(0.0411)
	(0.722)	(0.727)	(0.754)	(0.796)	(0.762)	(0.794)
Constant	1.379	1.404	1.391	1.429	1.402	1.401
cut1	(0.909)	(0.924)	(0.911)	(0.932)	(0.922)	(0.917)
	(0.626)	(0.606)	(0.614)	(0.585)	(0.608)	(0.606)
Constant	2.602	2.649	2.626	2.698	2.646	2.645
cut2	(1.756)	(1.785)	(1.762)	(1.806)	(1.782)	(1.775)
	(0.157)	(0.148)	(0.150)	(0.138)	(0.149)	(0.147)
Constant	4.474**	4.552**	4.517**	4.640**	4.552**	4.551**
cut3	(3.073)	(3.121)	(3.087)	(3.163)	(3.118)	(3.111)
	(0.0291)	(0.0271)	(0.0274)	(0.0244)	(0.0270)	(0.0266)
Constant	8.800***	8.948***	8.887***	9.121***	8.952***	8.959***
cut4	(6.139)	(6.235)	(6.174)	(6.325)	(6.232)	(6.225)
	(0.00182)	(0.00166)	(0.00166)	(0.00143)	(0.00164)	(0.00160)
Constant	25.49***	25.92***	25.76***	26.45***	25.95***	25.98***
Cut5	(18.47)	(18.77)	(18.61)	(19.09)	(18.77)	(18.78)
	(7.88e-06)	(6.96e-06)	(6.94e-06)	(5.70e-06)	(6.79e-06)	(6.58e-06)
Observations	227	227	227	227	227	227
pseudo R2	0.0837	0.0834	0.0839	0.0841	0.0838	0.0838

Table S21B re-runs the above income and wealth recalibrations using LNCS as our key predictor variable. The odds ratio for LNCS ranged from 2.4 to 2.5 (all highly significant), placing our U.S. dollar estimate in the middle of the range. The size and significance of MI remains roughly the same as in our baseline model.

Table S21B. OLR for MAO in TPG with local Wealth and Income calibrations using LNCS

Variables	U.S. $	100L cooking oil	100kg sugar	100kg Salt	1 D cell battery	100K rice
LNCS	2.434***	2.383***	2.415***	2.511***	2.461***	2.382***
	(0.294)	(0.264)	(0.282)	(0.331)	(0.312)	(0.262)
	(0)	(0)	(0)	(0)	(0)	(0)
MI	0.982**	0.983**	0.983**	0.981***	0.981**	0.983**
	(0.00699)	(0.00667)	(0.00670)	(0.00713)	(0.00720)	(0.00653)
	(0.0125)	(0.0108)	(0.0106)	(0.00746)	(0.0100)	(0.0101)
World Religion	1.393	1.369	1.376	1.382	1.399	1.367
	(0.509)	(0.506)	(0.500)	(0.502)	(0.509)	(0.496)
	(0.365)	(0.395)	(0.380)	(0.373)	(0.356)	(0.389)
Income	0.966	1.000	1.000		1.000	1.000
	(0.101)	(0.000138)	(6.31e-05)	1.000	(3.08e-05)	(6.06e-05)
	(0.742)	(0.943)	(0.919)	(2.59e-05)	(0.839)	(0.983)
Wealth	0.978	1.000	1.000	(0.627)	1.000	1.000
	(0.0474)	(5.34e-05)	(3.84e-05)	1.000	(1.87e-05)	(3.23e-05)
	(0.642)	(0.750)	(0.577)	(1.96e-05)	(0.437)	(0.567)
Sex	0.638*	0.653	0.645*	(0.299)	0.661	0.654
(female = 1)	(0.167)	(0.170)	(0.169)	(0.173)	(0.171)	(0.171)
	(0.0853)	(0.102)	(0.0932)	(0.115)	(0.110)	(0.104)
Age (years)	1.001	1.001	1.001	1.001	1.001	1.001
	(0.0113)	(0.0113)	(0.0114)	(0.0114)	(0.0114)	(0.0114)
	(0.946)	(0.943)	(0.941)	(0.911)	(0.943)	(0.947)
Education	0.975	0.970	0.973	0.976	0.970	0.970
(std by pop)	(0.138)	(0.139)	(0.137)	(0.137)	(0.138)	(0.137)
	(0.858)	(0.832)	(0.848)	(0.863)	(0.832)	(0.828)
Household Size	0.978	0.979	0.981	0.984	0.981	0.983
	(0.0399)	(0.0400)	(0.0403)	(0.0406)	(0.0406)	(0.0413)
	(0.592)	(0.595)	(0.640)	(0.696)	(0.643)	(0.682)
cut1	51.21***	47.84***	50.72***	61.24***	54.15***	48.23***
Constant	(37.45)	(35.03)	(37.53)	(47.17)	(39.88)	(35.42)
	(7.4e-08)	(1.3e-07)	(1.1e-07)	(9.2e-08)	(6.0e-08)	(1.3e-07)
cut2	102.7***	95.74***	101.6***	123.2***	108.9***	96.59***
Constant	(78.55)	(73.09)	(78.65)	(99.28)	(83.98)	(74.09)
	(1.39e-09)	(2.30e-09)	(2.34e-09)	(2.34e-09)	(1.19e-09)	(2.55e-09)
cut3	183.4***	170.7***	181.4***	220.2***	194.9***	172.5***
Constant	(145.7)	(135.2)	(145.9)	(184.2)	(156.2)	(137.5)
	(5.36e-11)	(8.57e-11)	(1.00e-10)	(1.12e-10)	(0)	(1.04e-10)
cut4	376.4***	350.2***	372.3***	451.8***	400.5***	354.3***
Constant	(312.2)	(289.0)	(312.4)	(393.4)	(335.0)	(294.5)

	(0)	(0)	(0)	(0)	(0)	(0)
cut5	1133***	1055***	1121***	1362***	1208***	1068***
Constant	(1005)	(927.5)	(1004)	(1264)	(1082)	(946.4)
	(0)	(0)	(0)	(0)	(0)	(0)
Observations	227	227	227	227	227	227
pseudo R2	0.112	0.112	0.112	0.113	0.113	0.112
Robust standard errors in parentheses; Education standardized to mean zero and standard deviation 1 for each population. *** p<0.01, ** p<0.05, * p<0.1						

As above, we also explored the effects of standardizing our income and wealth variables. As shown in Table S22A these results vary little from what is observed above. Both CS and WR are large and highly significant across the board.

Table S22A. OLR for MAO in TPG using standardized Income and Wealth and Community Size

VARIABLES	(1)	(2)	(3)	(4)
	Model	Model	Model	Model
Community Size	1.050***	1.047***	1.047***	1.047***
(100 people)	(0.00823)	(0.00723)	(0.00746)	(0.00751)
	(3.67e-10)	(0)	(1.42e-10)	(9.63e-11)
World Religion	2.053**	2.194**	2.193**	2.135**
(binary)	(0.712)	(0.735)	(0.733)	(0.710)
	(0.0380)	(0.0190)	(0.0188)	(0.0225)
Sex	0.666	0.714	0.715	0.719
(female = 1)	(0.178)	(0.172)	(0.173)	(0.169)
	(0.127)	(0.162)	(0.165)	(0.162)
Income	1.221	1.205	1.206	1.150
(std by pop)	(0.207)	(0.194)	(0.194)	(0.170)
	(0.239)	(0.247)	(0.244)	(0.344)
Education	0.939	0.866	0.865	0.864
(std by pop)	(0.137)	(0.106)	(0.106)	(0.105)
	(0.667)	(0.241)	(0.235)	(0.230)
Wealth	0.777	0.883	0.882	
(std by pop)	(0.168)	(0.185)	(0.185)	
	(0.244)	(0.552)	(0.548)	
MI (mean)	1.003	1.001		
(% purchased)	(0.00565)	(0.00463)		
	(0.557)	(0.911)		
Age (years)	1.000			
	(0.0116)			
	(0.993)			
Household Size	1.005			
	(0.0438)			
	(0.911)			
Constant	1.498	1.359	1.339	1.342
cut1	(0.982)	(0.505)	(0.435)	(0.433)
	(0.537)	(0.408)	(0.369)	(0.362)

Constant	2.845	2.481**	2.442***	2.432***
cut2	(1.912)	(0.971)	(0.829)	(0.820)
	(0.120)	(0.0202)	(0.00856)	(0.00838)
Constant	4.910**	4.050***	3.986***	3.955***
cut3	(3.347)	(1.642)	(1.419)	(1.398)
	(0.0196)	(0.000559)	(0.000103)	(9.98e-05)
Constant	9.700***	7.983***	7.862***	7.781***
cut4	(6.719)	(3.351)	(2.934)	(2.886)
	(0.00104)	(7.43e-07)	(3.30e-08)	(3.17e-08)
Constant	28.22***	22.17***	21.85***	21.62***
cut5	(20.37)	(10.35)	(9.338)	(9.185)
	(3.72e-06)	(0)	(0)	(0)
Observations	227	242	242	243
pseudo R2	0.0865	0.0740	0.0740	0.0739
Robust standard errors in parentheses; Education, Income, and Wealth are standardized to mean zero and standard deviation 1 for each population. *** p<0.01, ** p<0.05, * p<0.1				

Table S22B presents the same analyses as in Table S22A, except now using LNCS in place of CS. The coefficient on LNCS is large and highly significant across the board. The significant effects of WR drop while those of MI emerge. In summary, the results of the baseline regression hold when income and wealth are standardized.

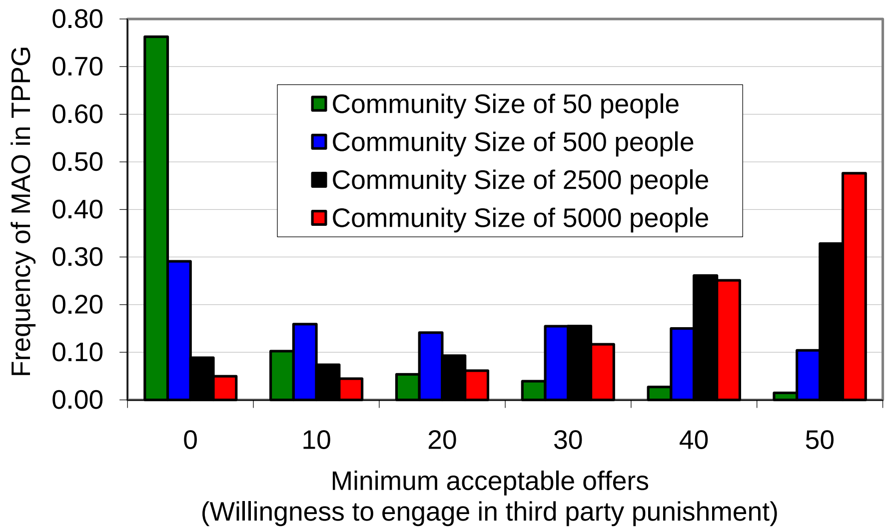
Table S22B. OLR for MAO in TPG using standardized Income and Wealth and LNCS

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Model	Model	Model	Model	Model	Model	Model	Model
LNCS	2.359***	2.354***	2.311***	2.333***	2.232***	2.272***	2.272***	2.256***
	(0.261)	(0.257)	(0.234)	(0.236)	(0.215)	(0.211)	(0.211)	(0.208)
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
MI	0.982***	0.983***	0.981***	0.982***	0.991*	0.990*	0.989**	0.989**
	(0.00657)	(0.00638)	(0.00585)	(0.00590)	(0.00548)	(0.00529)	(0.00526)	(0.00514)
	(0.00761)	(0.00712)	(0.00164)	(0.00192)	(0.0845)	(0.0543)	(0.0413)	(0.0284)
Household	0.996	0.995	0.990	0.986	0.951	0.951	0.952	
Size	(0.0445)	(0.0443)	(0.0399)	(0.0378)	(0.0340)	(0.0344)	(0.0347)	
	(0.924)	(0.916)	(0.808)	(0.719)	(0.157)	(0.168)	(0.179)	
Sex	0.662	0.669	0.720	0.720	0.739	0.738		
(female = 1)	(0.179)	(0.176)	(0.175)	(0.172)	(0.165)	(0.165)		
	(0.128)	(0.128)	(0.178)	(0.170)	(0.176)	(0.175)		
World	1.325	1.315	1.418	1.388	1.225			
Religion	(0.491)	(0.492)	(0.495)	(0.483)	(0.417)			
	(0.447)	(0.463)	(0.317)	(0.347)	(0.552)			
Income	1.194	1.184	1.136	1.095				
(std by pop)	(0.206)	(0.198)	(0.186)	(0.158)				
	(0.304)	(0.312)	(0.436)	(0.529)				
Wealth	0.792	0.792	0.916					

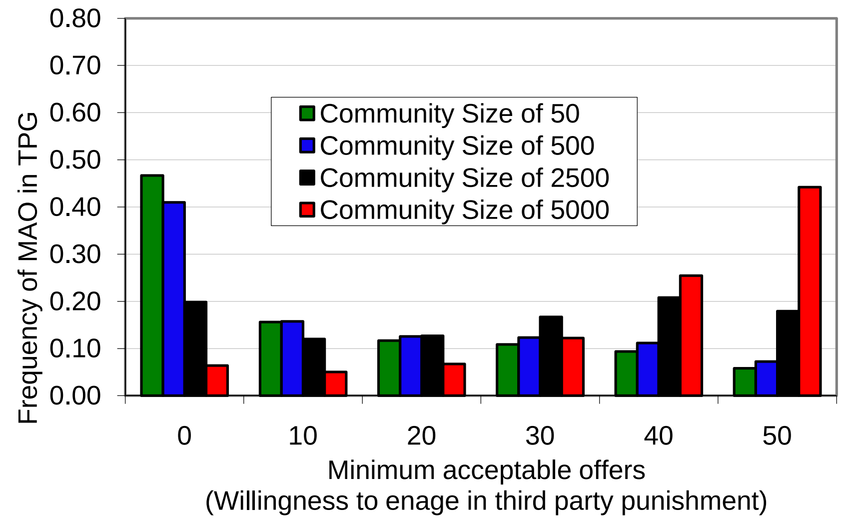
(std by pop)	(0.174)	(0.174)	(0.196)					
	(0.288)	(0.290)	(0.682)					
Age (years)	1.000	1.001						
	(0.0117)	(0.0103)						
	(0.997)	(0.899)						
Education	0.956							
(std by pop)	(0.138)							
	(0.757)							
cut1	48.79***	51.10***	46.03***	47.44***	45.41***	41.88***	48.71***	61.44***
Constant	(36.01)	(36.66)	(27.47)	(28.23)	(27.27)	(23.41)	(27.01)	(34.12)
	(1.4e-07)	(4.2e-08)	(1.4e-10)	(8.9e-11)	(2.1e-10)	(0)	(0)	(0)
cut2	98.25***	102.8***	86.67***	89.02***	84.24***	77.66***	90.09***	115.3***
Constant	(75.79)	(77.46)	(54.31)	(55.54)	(52.87)	(45.18)	(52.12)	(67.05)
	(2.7e-09)	(7.8e-10)	(0)	(0)	(0)	(0)	(0)	(0)
cut3	175.8***	183.9***	146.9***	150.6***	139.2***	128.4***	148.4***	189.1***
Constant	(140.6)	(144.5)	(96.26)	(98.20)	(90.72)	(78.01)	(89.59)	(114.8)
	(1.0e-10)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
cut4	362.3***	378.8***	303.7***	311.0***	272.9***	251.6***	295.9***	375.7***
Constant	(301.9)	(311.6)	(209.8)	(213.6)	(186.2)	(161.5)	(189.2)	(241.7)
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
cut5	1094***	1144***	878.9***	901.0***	741.7***	683.5***	795.0***	1002***
Constant	(972.0)	(1007)	(655.0)	(667.5)	(544.8)	(475.9)	(551.0)	(700.4)
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Observations	227	227	242	243	267	267	268	269
pseudo R2	0.114	0.114	0.104	0.105	0.0934	0.0930	0.0913	0.0900
Robust standard errors in parentheses; Education standardized to mean zero and standard deviation 1 for each population. *** p<0.01, ** p<0.05, * p<0.1								

In our sample, local community sizes are highly correlated with overall ethnic group size. In a few cases the boundaries of the ethnic group are somewhat unclear, so several different demarcations could be drawn, making it somewhat difficult to nail a precise value for this relationship. However, using a sensible set of demarcations, we have estimated the correlation at 0.97 (S*49*). This relationship makes good sense from our theoretical perspective: those populations with more third party punishment can stabilize more fairness and cooperation in larger populations, leading to larger more stable ethnic groups, and success in competition with other ethnic groups. Since we get similar findings whether we use ethnic group size or local community size, these analyses cannot tease out which is the best predictor. From our theoretical perspective these are causally interconnected in any case. Such findings are consistent with recent work suggesting that cultural group selection, driven by differences in political complexity, can help explain the size and diversity of languages globally (S*50*).

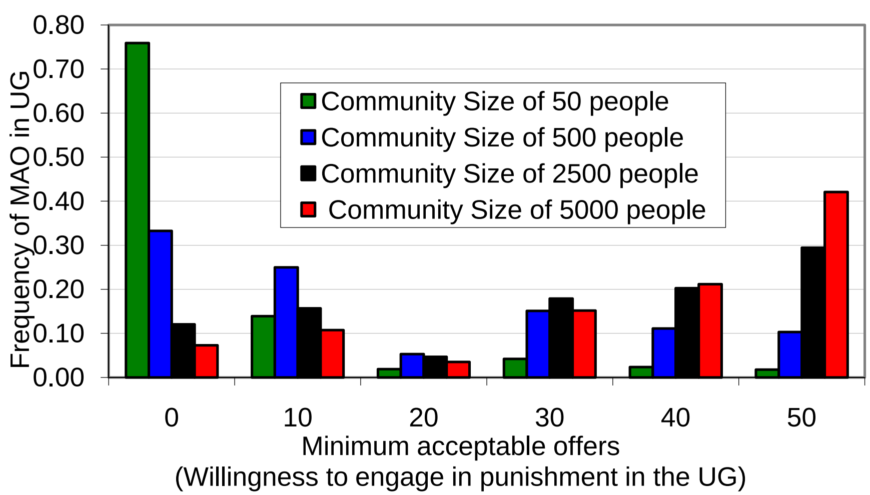
Panel A: LNCS predicting Third Party Punishment MAO



Panel B: Community size predicting Third Party Punishment MAO



Panel C: LNCS predicting Ultimatum MAO



Panel D: Community size predicting Ultimatum MAO

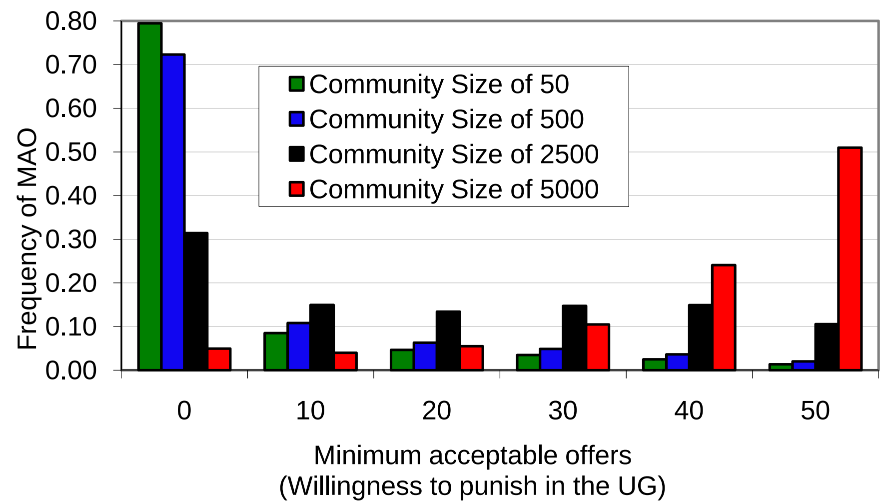


Figure S4. Shows the difference between using LNCS vs. CS as the predictor of MAO for TPG and the UG.

Each set of colored bars shows the distribution of MAOs, ranging discretely from 0 to 50, for different sizes of communities. The coefficients used to create the plots are from an ordered logistic regression containing all eight of our other variables.

Willingness to engage in costly punishment as measured by rejecting in the UG is also positively related to both community size and LNCS. Model 2 in Tables S23A and S23B shows the baseline models discussed in the main text. As with our regressions for the TPG, both the coefficients on LNCS and CS are large and highly significant. Individuals from large communities tend to punish more. The coefficient on MI is also a potent predictor in both Tables S23A and S23B. The coefficient on MI shows that—*ceteris paribus*—individuals from less market integrated communities punish less (more on these relationships below). Unlike the above TPG MAO analysis, income and household size are also significant at conventional levels, with individuals with higher incomes punishing more and those from larger households punishing less. However, note that income switches direction (above or below 1) depending on the regression model. This occurs both when LNCS and CS are in the model. The effects of household size and wealth are consistent.

To check for the effects of cultural phylogeny and the use of clustered robust standard errors, compare Models 1 and 2 (baseline). When continental controls are added and clustered robust errors are used, the odds ratio for both CS (Table S23A) and LNCS (Table S23B) increases (a larger effect) and remains highly significant. The odds ratio for MI also withstands these checks in both tables. In contrast, the odds ratio for income drops below 1, indicating a change in direction, while the odds ratio for wealth becomes significant. The odds ratio for household size increases towards 1, and becomes non-significant.

Models 2 to 7 in Tables S23A and S23B examine the effect of alternative specifications for CS and LNCS. The odds ratios for both CS and LNCS remain large and highly significant across these models. Figure S4 provides an illustration of the size of the estimated coefficients and odds ratios. The effects of MI and household size are also robust across these specifications, with little variation in the size of the effect. The effect for income, which is significant in both Tables S23A and S23B, switches direction depending on whether one uses LNCS or CS. It also switches direction depending on the presence of the continental controls.

Table S23A Ordered Logistic Regressions for MAO in UG using Community Size

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Com. Size	1.119***	1.091***	1.091***	1.090***	1.090***	1.091***	1.081***	
(100 people)	(0.0156)	(0.0120)	(0.0119)	(0.0116)	(0.0114)	(0.0109)	(0.0100)	
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	
MI (mean)	0.962***	0.975***	0.975***	0.975***	0.975***	0.976***	0.980***	0.996
(% purch.)	(0.00742)	(0.00488)	(0.00489)	(0.00480)	(0.00482)	(0.00461)	(0.00418)	(0.00291)
	(6.75e-07)	(4.2e-07)	(4.4e-07)	(3.5e-07)	(3.6e-07)	(3.1e-07)	(3.3e-06)	(0.13)
Household	0.983	0.922**	0.922**	0.922**	0.927**	0.924**	0.918***	0.981
Size	(0.0269)	(0.0294)	(0.0294)	(0.0292)	(0.0287)	(0.0285)	(0.0273)	(0.0246)
	(0.530)	(0.0105)	(0.0105)	(0.0105)	(0.0146)	(0.0105)	(0.00413)	(0.451)
Income	0.959***	1.060***	1.060***	1.060***	1.061***	1.062***	1.043***	1.036***
(1000 USD)	(0.0137)	(0.0193)	(0.0190)	(0.0189)	(0.0189)	(0.0185)	(0.0149)	(0.0133)
	(0.00366)	(0.00131)	(0.00108)	(0.00118)	(0.000893)	(0.000564)	(0.00308)	(0.00560)
Wealth	0.998***	0.998	0.998	0.998	0.998	0.998		
(1000 USD)	(0.000421)	(0.00226)	(0.00225)	(0.00224)	(0.00225)	(0.00223)		
	(1.49e-05)	(0.303)	(0.303)	(0.308)	(0.309)	(0.297)		
Age (years)	1.004	1.006	1.006	1.006	1.007			
	(0.0131)	(0.00956)	(0.00914)	(0.00895)	(0.00889)			
	(0.740)	(0.508)	(0.479)	(0.489)	(0.458)			
World	0.814	1.156	1.154	1.169				
Religion	(0.207)	(0.355)	(0.351)	(0.348)				
(binary)	(0.418)	(0.638)	(0.638)	(0.599)				
Sex	1.060	1.055	1.056					
(female = 1)	(0.243)	(0.228)	(0.226)					
	(0.799)	(0.806)	(0.798)					
Education	1.058	0.993						
(std by pop)	(0.147)	(0.118)						
	(0.686)	(0.950)						
Africa	0.0128***							
(dummy)	(0.0102)							
	(4.95e-08)							
S. America	0.00926***							
(dummy)	(0.00916)							
	(2.21e-06)							
Oceania	0.0249***							
(dummy)	(0.0323)							
	(0.00444)							
Constant	0.00704***	0.577	0.579	0.568	0.531	0.445***	0.415***	0.523***
cut1	(0.00679)	(0.267)	(0.266)	(0.253)	(0.220)	(0.123)	(0.111)	(0.130)
	(2.74e-07)	(0.235)	(0.234)	(0.204)	(0.127)	(0.00350)	(0.00102)	(0.00892)
Constant	0.0238***	1.737	1.744	1.709	1.590	1.305	1.228	1.324
cut2	(0.0199)	(0.817)	(0.814)	(0.773)	(0.677)	(0.365)	(0.331)	(0.328)
	(7.45e-06)	(0.240)	(0.234)	(0.236)	(0.276)	(0.342)	(0.445)	(0.257)
Constant	0.0317***	2.229*	2.238*	2.193*	2.038	1.701*	1.604*	1.698**
cut3	(0.0260)	(1.066)	(1.063)	(1.009)	(0.883)	(0.486)	(0.439)	(0.423)
	(2.68e-05)	(0.0937)	(0.0900)	(0.0879)	(0.100)	(0.0628)	(0.0838)	(0.0336)

Constant	0.0811***	5.133***	5.154***	5.050***	4.681***	3.691***	3.259***	3.001***
cut4	(0.0603)	(2.663)	(2.653)	(2.517)	(2.225)	(1.181)	(0.981)	(0.788)
	(0.000724)	(0.00162)	(0.00145)	(0.00116)	(0.00117)	(4.51e-05)	(8.6e-05)	(2.8e-05)
Constant	0.217**	12.62***	12.67***	12.41***	11.49***	8.401***	6.718***	5.900***
cut5	(0.160)	(7.219)	(7.207)	(6.851)	(6.135)	(3.198)	(2.348)	(1.723)
	(0.0380)	(9.37e-06)	(8.01e-06)	(5.07e-06)	(4.83e-06)	(2.26e-08)	(5.02e-08)	(1.22e-09)
Observations	297	297	297	297	298	318	349	379
pseudo R2	0.106	0.0625	0.0625	0.0624	0.0621	0.0576	0.0470	0.00589
Model 1 used clustered robust standard errors (clustering on site); other models use robust standard errors; p-values are in parentheses below the coefficient. Education has been standardized to a mean of zero and standard deviation of one within each population. Coefficients reported as odds ratios. *p < 0.1 **p< 0.05 ***p<0.01								

Model 8 in Table S23A and Models 8 and 9 in Table S23B explore how dropping CS or LNCS, respectively, influences what happens to the coefficients on MI, income, and household size. The result is that both the odds ratios for MI and HS move toward 1 and become non-significant. The odds ratio for income also moves toward 1, but remains a significant predictor. Model 9 in Table 23B drops everything except MI and shows the odds ratio moves even closer to 1. These models show that the significant negative relationship for MI and household size depends on keeping CS or LNCS in the regression (and only on this). In contrast, if CS or LNCS are the only predictors in the model, both remain highly significant, though the odds ratio for their coefficients do decrease.

Table S23B. Order Logistic Regressions for MAO in UG using LNSC

VARS	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
LNCS	3.292***	2.229***	2.218***	2.128***	2.135***	2.075***	1.963***		
	(0.802)	(0.277)	(0.273)	(0.245)	(0.242)	(0.228)	(0.194)		
	(1.00e-06)	(1.12e-10)	(1.01e-10)	(5.13e-11)	(0)	(0)	(0)		
MI (mean)	0.945***	0.966***	0.966***	0.967***	0.967***	0.970***	0.973***	0.996	0.998
(% purchased)	(0.0137)	(0.00630)	(0.00620)	(0.00608)	(0.00603)	(0.00568)	(0.00517)	(0.00291)	(0.00271)
	(0.000104)	(1.19e-07)	(9.95e-08)	(7.82e-08)	(5.72e-08)	(3.03e-07)	(3.36e-07)	(0.130)	(0.479)
Household Size	0.987	0.922**	0.922**	0.918***	0.917***	0.920***	0.916***	0.981	
	(0.0246)	(0.0295)	(0.0294)	(0.0290)	(0.0290)	(0.0285)	(0.0274)	(0.0246)	
	(0.600)	(0.0107)	(0.0107)	(0.00651)	(0.00631)	(0.00739)	(0.00329)	(0.451)	
Income	0.963***	1.068***	1.067***	1.064***	1.065***	1.064***	1.048***	1.036***	
(1000 USD)	(0.0132)	(0.0209)	(0.0208)	(0.0202)	(0.0197)	(0.0191)	(0.0152)	(0.0133)	
	(0.00587)	(0.000794)	(0.000897)	(0.000962)	(0.000676)	(0.000566)	(0.00122)	(0.00560)	
Wealth	0.998***	0.998	0.998	0.998	0.998	0.998			
(1000 USD)	(0.000323)	(0.00226)	(0.00226)	(0.00225)	(0.00224)	(0.00222)			
	(7.51e-10)	(0.287)	(0.292)	(0.301)	(0.294)	(0.283)			
Age (years)	1.003	1.008	1.007	1.008	1.008				
	(0.0129)	(0.00970)	(0.00941)	(0.00946)	(0.00911)				
	(0.794)	(0.436)	(0.446)	(0.392)	(0.392)				
Education	1.090	1.025	1.022	1.018					
(std by pop)	(0.154)	(0.121)	(0.120)	(0.119)					
	(0.545)	(0.833)	(0.855)	(0.881)					
World Religion	0.332***	0.679	0.692						
(binary)	(0.0887)	(0.231)	(0.230)						
	(3.65e-05)	(0.256)	(0.267)						
Sex	1.113	1.067							
(female = 1)	(0.250)	(0.232)							
	(0.633)	(0.764)							
Africa	0.00767***								
(dummy)	(0.00847)								
	(1.04e-05)								

South America	0.00431***								
(dummy)	(0.00664)								
	(0.000408)								
Oceania	0.0191***								
(dummy)	(0.0268)								
	(0.00476)								
Constant	0.472	15.09***	14.44***	15.50***	15.58***	12.52***	8.908***	0.523***	0.627***
cut1	(0.609)	(9.607)	(8.762)	(9.550)	(9.596)	(7.024)	(4.446)	(0.130)	(0.0943)
	(0.561)	(2.01e-05)	(1.09e-05)	(8.67e-06)	(8.31e-06)	(6.57e-06)	(1.18e-05)	(0.00892)	(0.00190)
Constant	1.658	45.09***	43.12***	45.96***	46.19***	35.87***	26.01***	1.324	1.568***
cut2	(1.892)	(30.38)	(27.79)	(30.00)	(30.13)	(21.19)	(13.60)	(0.328)	(0.237)
	(0.658)	(1.57e-08)	(5.23e-09)	(4.52e-09)	(4.22e-09)	(1.36e-09)	(4.66e-10)	(0.257)	(0.00287)
Constant	2.204	57.31***	54.81***	58.37***	58.66***	46.11***	33.64***	1.698**	2.020***
cut3	(2.560)	(39.60)	(36.29)	(39.13)	(39.29)	(28.06)	(18.13)	(0.423)	(0.317)
	(0.496)	(4.67e-09)	(1.48e-09)	(1.30e-09)	(1.21e-09)	(3.09e-10)	(6.91e-11)	(0.0336)	(7.51e-06)
Constant	5.608	128.0***	122.3***	129.8***	130.3***	95.83***	66.52***	3.001***	3.524***
cut4	(6.665)	(95.24)	(87.50)	(93.82)	(94.19)	(62.07)	(38.20)	(0.788)	(0.588)
	(0.147)	(7.03e-11)	(0)	(0)	(0)	(0)	(0)	(2.82e-05)	(0)
Constant	15.10**	311.0***	297.2***	314.8***	316.1***	215.3***	136.3***	5.900***	6.804***
cut5	(20.60)	(257.4)	(237.6)	(254.3)	(255.2)	(155.2)	(87.06)	(1.723)	(1.327)
	(0.0466)	(0)	(0)	(0)	(0)	(0)	(0)	(1.22e-09)	(0)
Observations	297	297	297	298	298	318	349	379	387
pseudo R2	0.113	0.0583	0.0582	0.0571	0.0570	0.0492	0.0422	0.00589	0.000433

Model 1 used clustered robust standard errors (clustering on site); other models use robust standard errors; p-values are in parentheses below the coefficient. Education has been standardized to a mean of zero and standard deviation of one within each population. Coefficients reported as odds ratios. *p < 0.1 **p< 0.05 ***p<0.01.

As above, we checked our use of U.S. dollars to measure wealth and income by replacing these variables with income and wealth variables derived from the prices of local consumables at the site. Table S24A and S24B compare our full model with five other models using our same set of consumables. In Table S24A both CS and MI remain highly significant across different recalibrations of income and wealth. Neither the coefficients on household size nor income are significant in Models 4 (salt) and 5 (cooking oil). The reason for this has nothing to do with revaluing of income and wealth in oil and salt, but instead results entirely from the fact we lack oil and salt prices for our rural U.S. sample, so these data drop out of the regression. When U.S. dollars are used for income and wealth, but the U.S. data is removed from the analysis, neither income nor household size remain even remotely significant. Since Americans have dramatically higher incomes and smaller households than the other populations in our sample, this is not surprising. This also demonstrates that, unlike income or household size, the effects of MI and LNSC are robust and not contingent on the U.S. data (we have no comparative U.S. data for TPG).

Table S24A. OLR for MAO in UG with local Wealth and Income calibrations using Community Size

VARIABLES	U.S. $	D battery	100kg rice	100kg sugar	100L cooking oil	100kg salt
Community Size	1.091***	1.078***	1.092***	1.091***	1.111***	1.104***
(100 people)	(0.0120)	(0.0114)	(0.0122)	(0.0121)	(0.0160)	(0.0176)
	(0)	(0)	(0)	(0)	(0)	(4.80e-10)
MI (comm. mean)	0.975***	0.977***	0.975***	0.975***	0.961***	0.962***
(% purchased)	(0.00488)	(0.00462)	(0.00497)	(0.00490)	(0.00530)	(0.00533)
	(4.21e-07)	(1.20e-06)	(5.57e-07)	(5.05e-07)	(0)	(0)
Household Size	0.922**	0.917***	0.919***	0.920***	0.954	0.958
	(0.0294)	(0.0297)	(0.0293)	(0.0295)	(0.0308)	(0.0309)
	(0.0105)	(0.00726)	(0.00805)	(0.00932)	(0.142)	(0.182)
World Religion	1.156	1.184	1.171	1.156	0.981	0.970
(binary)	(0.355)	(0.359)	(0.359)	(0.355)	(0.325)	(0.322)
	(0.638)	(0.578)	(0.607)	(0.636)	(0.953)	(0.927)
Income	1.060***	1.000**	1.000***	1.000***	1.000	1.000
	(0.0193)	(3.05e-05)	(1.87e-05)	(1.70e-05)	(0.000295)	(5.91e-05)
	(0.00131)	(0.0120)	(0.00178)	(0.00167)	(0.776)	(0.736)
Wealth	0.998	1.000	1.000	1.000	1.000**	1.000***
	(0.00226)	(5.67e-06)	(2.40e-06)	(2.20e-06)	(3.42e-05)	(4.07e-06)
	(0.303)	(0.830)	(0.353)	(0.356)	(0.0394)	(0.00787)
Sex (female = 1)	1.055	1.040	1.057	1.055	1.099	1.089
	(0.228)	(0.223)	(0.228)	(0.228)	(0.248)	(0.246)
	(0.806)	(0.855)	(0.797)	(0.805)	(0.678)	(0.707)
Age (years)	1.006	1.007	1.007	1.007	1.007	1.007
	(0.00956)	(0.00956)	(0.00954)	(0.00955)	(0.0102)	(0.0102)
	(0.508)	(0.437)	(0.474)	(0.492)	(0.482)	(0.522)
Education	0.993	0.997	0.996	0.993	1.045	1.042
(std by pop)	(0.118)	(0.118)	(0.118)	(0.118)	(0.142)	(0.141)
	(0.950)	(0.977)	(0.972)	(0.953)	(0.746)	(0.760)
Constant	0.577	0.607	0.591	0.583	0.542	0.525
	(0.267)	(0.279)	(0.273)	(0.270)	(0.263)	(0.255)
	(0.235)	(0.277)	(0.255)	(0.245)	(0.208)	(0.184)

Constant	1.737	1.804	1.771	1.751	1.928	1.869
	(0.817)	(0.844)	(0.833)	(0.824)	(0.960)	(0.930)
	(0.240)	(0.208)	(0.224)	(0.234)	(0.187)	(0.209)
Constant	2.229*	2.310*	2.270*	2.246*	2.342*	2.272
	(1.066)	(1.100)	(1.086)	(1.074)	(1.185)	(1.148)
	(0.0937)	(0.0785)	(0.0865)	(0.0907)	(0.0926)	(0.104)
Constant	5.133***	5.302***	5.218***	5.166***	6.688***	6.503***
	(2.663)	(2.740)	(2.708)	(2.681)	(3.854)	(3.753)
	(0.00162)	(0.00125)	(0.00145)	(0.00155)	(0.000973)	(0.00118)
Constant	12.62***	13.00***	12.81***	12.69***	18.48***	17.94***
	(7.219)	(7.403)	(7.331)	(7.256)	(12.44)	(12.10)
	(9.37e-06)	(6.70e-06)	(8.26e-06)	(8.89e-06)	(1.46e-05)	(1.87e-05)
Observations	297	297	297	297	270	270
pseudo R2	0.0625	0.0598	0.0615	0.0617	0.0969	0.0968
Models use robust standard errors; p-values are in parentheses below the coefficient. Education has been standardized to a mean of zero and standard deviation of one within each population. Coefficients reported as odds ratios. *p < 0.1 **p< 0.05 ***p<0.01.						

In Table S24B LNCS and MI remain significant predictors in all these models. The odds ratio of the coefficient on LNSC ranges from 2.0 (batteries) to 2.7 (cooking oil), putting our U.S. dollar estimate near the center of the range. The coefficient on MI does not vary much across the models. Again, neither the coefficients on household size nor income are significant in Models 4 (salt) and 5 (cooking oil). The same reasoning as above applies.

Table S24B. OLR for MAO in UG with local Wealth and Income calibrations using LNSC

Variables	U.S. $	D battery	100kg rice	100kg sugar	100L cooking oil	100kg salt
LNCS	2.229***	1.979***	2.240***	2.235***	2.710***	2.536***
	(0.277)	(0.231)	(0.282)	(0.279)	(0.410)	(0.400)
	(1.12e-10)	(5.02e-09)	(1.54e-10)	(1.24e-10)	(0)	(3.60e-09)
MI	0.966***	0.969***	0.966***	0.966***	0.949***	0.950***
	(0.00630)	(0.00581)	(0.00643)	(0.00634)	(0.00683)	(0.00682)
	(1.19e-07)	(2.00e-07)	(1.81e-07)	(1.38e-07)	(0)	(0)
Household Size	0.922**	0.917***	0.919***	0.920***	0.953	0.954
	(0.0295)	(0.0299)	(0.0294)	(0.0296)	(0.0315)	(0.0319)
	(0.0107)	(0.00811)	(0.00805)	(0.00949)	(0.146)	(0.158)
World Religion	0.679	0.747	0.688	0.677	0.482**	0.497*
	(0.231)	(0.248)	(0.235)	(0.231)	(0.176)	(0.180)
	(0.256)	(0.378)	(0.274)	(0.254)	(0.0455)	(0.0532)
Income	1.068***	1.000***	1.000***	1.000***	1.000	1.000
	(0.0209)	(3.09e-05)	(1.98e-05)	(1.83e-05)	(0.000305)	(5.74e-05)
	(0.000794)	(0.00227)	(0.000937)	(0.000860)	(0.799)	(0.305)
Wealth	0.998	1.000	1.000	1.000	1.000	1.000**
	(0.00226)	(5.87e-06)	(2.37e-06)	(2.19e-06)	(4.67e-05)	(4.82e-06)
	(0.287)	(0.807)	(0.321)	(0.322)	(0.196)	(0.0262)
Sex	1.067	1.060	1.070	1.069	1.134	1.134
(female = 1)	(0.232)	(0.228)	(0.232)	(0.232)	(0.258)	(0.259)

	(0.764)	(0.787)	(0.755)	(0.757)	(0.579)	(0.581)
Age (years)	1.008	1.008	1.008	1.008	1.008	1.007
	(0.00970)	(0.00970)	(0.00968)	(0.00971)	(0.0105)	(0.0105)
	(0.436)	(0.405)	(0.400)	(0.426)	(0.472)	(0.522)
Education	1.025	1.019	1.030	1.025	1.079	1.065
(std by pop)	(0.121)	(0.120)	(0.121)	(0.121)	(0.146)	(0.144)
	(0.833)	(0.874)	(0.805)	(0.835)	(0.575)	(0.643)
cut1	15.09***	9.647***	15.87***	15.42***	29.96***	21.68***
Constant	(9.607)	(6.051)	(10.20)	(9.849)	(22.06)	(16.51)
	(2.01e-05)	(0.000302)	(1.70e-05)	(1.85e-05)	(3.87e-06)	(5.34e-05)
cut2	45.09***	28.59***	47.20***	45.96***	106.7***	78.07***
Constant	(30.38)	(18.83)	(32.08)	(31.05)	(84.41)	(63.56)
	(1.57e-08)	(3.59e-07)	(1.42e-08)	(1.47e-08)	(3.50e-09)	(8.68e-08)
cut3	57.31***	36.35***	59.92***	58.40***	128.6***	94.40***
Constant	(39.60)	(24.56)	(41.76)	(40.47)	(103.8)	(78.35)
	(4.67e-09)	(1.05e-07)	(4.26e-09)	(4.38e-09)	(1.77e-09)	(4.28e-08)
cut4	128.0***	81.60***	133.5***	130.3***	350.8***	262.0***
Constant	(95.24)	(59.28)	(100.1)	(97.26)	(308.7)	(236.6)
	(7.03e-11)	(1.37e-09)	(6.79e-11)	(6.76e-11)	(0)	(7.00e-10)
cut5	311.0***	198.6***	323.9***	316.5***	954.9***	716.1***
Constant	(257.4)	(160.3)	(269.8)	(262.6)	(952.2)	(730.9)
	(0)	(5.64e-11)	(0)	(0)	(0)	(1.18e-10)
Observations	297	297	297	297	270	270
pseudo R2	0.0583	0.0574	0.0569	0.0576	0.0957	0.0989

Models use robust standard errors; p-values are in parentheses below the coefficient. Education has been standardized to a mean of zero and standard deviation of one within each population. Coefficients reported as odds ratios. *p < 0.1 **p< 0.05 ***p<0.01.

As above, we also explored the effects of standardizing our income and wealth variables on the robustness of our theoretical findings. Tables S25A and S25B show that these results vary little from what is observed above, without standardization, except that now income is not an important predictor of MAO.

Table S25A. OLR for MAO in UG using standardized Income and Wealth and Community Size

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5
Community Size	1.084***	1.084***	1.080***	1.081***	1.077***
(100 people)	(0.0117)	(0.0114)	(0.0108)	(0.0106)	(0.0103)
	(0)	(0)	(0)	(0)	(0)
MI (mean)	0.981***	0.981***	0.982***	0.982***	0.983***
(% purchased)	(0.00446)	(0.00434)	(0.00408)	(0.00409)	(0.00399)
	(2.18e-05)	(1.19e-05)	(1.80e-05)	(1.76e-05)	(3.95e-05)
Household Size	0.908***	0.907***	0.908***	0.908***	0.909***
	(0.0304)	(0.0305)	(0.0284)	(0.0283)	(0.0277)
	(0.00382)	(0.00375)	(0.00211)	(0.00196)	(0.00167)
Age (years)	1.011	1.012	1.009	1.009	1.010
	(0.00915)	(0.00888)	(0.00844)	(0.00821)	(0.00783)

	(0.215)	(0.193)	(0.291)	(0.294)	(0.214)
Income	0.956	0.959	0.915	0.917	0.949
(std by pop)	(0.122)	(0.122)	(0.0891)	(0.0895)	(0.0977)
	(0.725)	(0.742)	(0.359)	(0.373)	(0.610)
World Religion	1.249	1.240	1.172	1.174	
(binary)	(0.383)	(0.373)	(0.336)	(0.336)	
	(0.469)	(0.475)	(0.579)	(0.575)	
Education	1.033	1.035	1.021		
(std by pop)	(0.123)	(0.123)	(0.114)		
	(0.785)	(0.774)	(0.850)		
Wealth	0.972	0.971			
(std by pop)	(0.136)	(0.136)			
	(0.838)	(0.834)			
Sex	0.966				
(female = 1)	(0.211)				
	(0.875)				
Constant	0.688	0.698	0.595	0.589	0.544
	(0.315)	(0.308)	(0.253)	(0.250)	(0.204)
	(0.413)	(0.415)	(0.221)	(0.213)	(0.105)
Constant	2.005	2.034	1.812	1.795	1.610
	(0.936)	(0.916)	(0.781)	(0.773)	(0.618)
	(0.136)	(0.115)	(0.168)	(0.174)	(0.215)
Constant	2.557**	2.594**	2.309*	2.288*	2.062*
	(1.220)	(1.194)	(1.016)	(1.005)	(0.805)
	(0.0490)	(0.0383)	(0.0572)	(0.0597)	(0.0638)
Constant	5.774***	5.858***	5.043***	4.995***	4.339***
	(2.986)	(2.924)	(2.386)	(2.358)	(1.840)
	(0.000698)	(0.000397)	(0.000625)	(0.000655)	(0.000538)
Constant	13.99***	14.20***	11.50***	11.38***	9.398***
	(7.951)	(7.809)	(5.968)	(5.897)	(4.409)
	(3.41e-06)	(1.39e-06)	(2.55e-06)	(2.69e-06)	(1.79e-06)
Observations	297	297	319	319	329
pseudo R2	0.0533	0.0533	0.0485	0.0484	0.0443
Coefficients reported as odds ratios. Robust standard errors and p-values in parentheses below the coefficient. Education, Income and Wealth have been standardized to a mean of zero and standard deviation of one within each population. *p < 0.1 **p< 0.05 ***p<0.01					

Table S25B. OLR for MAO in UG using standardized Income and Wealth and LNCS

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
LNCS	2.015***	1.969***	1.970***	1.915***	1.920***	1.931***	1.904***	1.672***
	(0.238)	(0.216)	(0.216)	(0.190)	(0.190)	(0.189)	(0.185)	(0.149)
	(2.92e-09)	(7.14e-10)	(6.22e-10)	(5.66e-11)	(0)	(0)	(0)	(8.07e-09)
MI	0.974***	0.974***	0.974***	0.976***	0.975***	0.975***	0.979***	0.983***
	(0.0057)	(0.0056)	(0.0056)	(0.00508)	(0.00503)	(0.00501)	(0.00476)	(0.00469)
	(7.8e-06)	(5.6e-06)	(5.3e-06)	(2.35e-06)	(1.47e-06)	(1.21e-06)	(8.37e-06)	(0.00027)
Household	0.907***	0.905***	0.905***	0.904***	0.903***	0.901***	0.904***	
Size	(0.0304)	(0.0301)	(0.0299)	(0.0278)	(0.0278)	(0.0277)	(0.0273)	

	(0.00351)	(0.00262)	(0.00253)	(0.000997)	(0.000922)	(0.000695)	(0.000864)	
Age (years)	1.013	1.014	1.014	1.011	1.011	1.011		
	(0.00924)	(0.00924)	(0.00926)	(0.00853)	(0.00835)	(0.00806)		
	(0.152)	(0.134)	(0.135)	(0.203)	(0.176)	(0.178)		
Education	1.068	1.060	1.057	1.017	1.019			
(std by pop)	(0.128)	(0.125)	(0.124)	(0.108)	(0.108)			
	(0.580)	(0.621)	(0.636)	(0.876)	(0.860)			
Sex	0.960	0.956	0.966	0.932				
(female = 1)	(0.211)	(0.205)	(0.206)	(0.188)				
	(0.853)	(0.835)	(0.870)	(0.727)				
Wealth	1.000	1.004	0.992					
(std by pop)	(0.137)	(0.136)	(0.117)					
	(0.998)	(0.974)	(0.945)					
Income	0.953	0.963						
(std by pop)	(0.119)	(0.118)						
	(0.701)	(0.760)						
World	0.808							
Religion	(0.271)							
	(0.525)							
cut1	12.35***	12.93***	13.02***	9.789***	10.27***	10.26***	7.767***	8.253***
Constant	(7.897)	(8.357)	(8.388)	(5.494)	(5.544)	(5.505)	(3.812)	(3.954)
	(8.41e-05)	(7.52e-05)	(6.83e-05)	(4.81e-05)	(1.59e-05)	(1.45e-05)	(2.95e-05)	(1.06e-05)
cut2	35.43***	36.88***	37.11***	28.74***	30.14***	29.96***	22.04***	22.83***
Constant	(23.72)	(24.93)	(25.02)	(16.84)	(17.07)	(16.85)	(11.26)	(11.46)
	(9.94e-08)	(9.52e-08)	(8.26e-08)	(9.89e-09)	(1.82e-09)	(1.49e-09)	(1.42e-09)	(4.59e-10)
cut3	44.70***	46.50***	46.78***	36.01***	37.76***	38.10***	28.32***	29.36***
Constant	(30.63)	(32.16)	(32.27)	(21.58)	(21.91)	(21.97)	(14.87)	(15.14)
	(2.9e-08)	(2.9e-08)	(2.5e-08)	(2.21e-09)	(3.88e-10)	(2.75e-10)	(1.90e-10)	(5.5e-11)
cut4	97.61***	101.3***	102.0***	74.26***	77.88***	78.48***	55.25***	55.97***
Constant	(71.77)	(75.14)	(75.43)	(47.69)	(48.69)	(48.80)	(30.85)	(30.73)
	(4.6e-10)	(4.7e-10)	(4.1e-10)	(0)	(0)	(0)	(0)	(0)
cut5	233.4***	242.1***	243.6***	159.9***	167.7***	168.9***	112.3***	111.2***
Constant	(190.0)	(198.7)	(199.5)	(113.5)	(116.6)	(116.8)	(69.72)	(67.86)
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Observations	297	298	298	329	329	330	350	357
pseudo R2	0.0467	0.0461	0.0460	0.0412	0.0411	0.0410	0.0351	0.0248
Robust standard errors and p-values in parentheses below the coefficient; Education has been standardized to a mean of zero and standard deviation of one within each population; coefficients reported as odds ratios. *p < 0.1 **p< 0.05 ***p<0.01.								

SUMMARY OF MAO ANALYSIS

Both community size and the natural logarithm of community size are robust and potent predictors of MAO in both the Third Party Punishment and Ultimatum Games. Larger communities punish more, controlling for our eight other variables. Both measures of pseudo R^2 and AIC indicate that LNCS is superior to CS as a predictor. Market Integration predicts less punishment in most of our analyses above, though never when CS is used to predict MAO in the TPG. MI's effect requires that CS or LNSC be in the regression equation. If CS or LNCS is dropped, MI's effect dramatically drops or reverses directions. Household size and wealth sometimes emerge as important predictors of the MAO in the UG. The direction of the effect is such that more wealth or bigger household lead to lower MAOs in the UG. Income also shows significant effects for the MAO in the UG, but the direction of this effect depends on the specification. Finally, participation in a world religion had the effect of predicting greater MAOs in the TPG, but only if CS is used.

ANONYMITY HYPOTHESES

Some researchers have proposed that the prosociality observed in behavioural experiments results from purely selfish motivations combined with an uncertainty about the anonymity in the experiments. There are at least three versions of this idea. The first version proposes that individuals' beliefs regarding the reality of the anonymity in our experiments may be influenced by either their experience in their daily lives or by calibrations that reflect some ancestral world (and associated mechanisms for figuring out repeated interactions) or both. If daily life is an important influence, people from societies with more actual ephemeral interactions and real anonymity (Western societies, for example) ought to be the best able to handle the anonymity and behave purely selfishly.

Our findings reveal precisely the opposite pattern from that suggested by this line of reasoning. The smaller-scale, least anonymous and most face-to-face societies are generally less prosocial (New Guinea is the singular exception), while larger, more complex and anonymous societies are more prosocial. In the larger societies, people actually do have lots of interactions with strangers and in contexts lacking expectations of future interactions. If people's experienced-derived beliefs about anonymity were in fact driving prosociality in these experiments, we would expect the smaller-scale societies to be MORE prosocial in the experiments, not less. That is, the societies that live more like our ancestors in those small groups are actually less prosocial in these experiments. In our first project we showed that anonymity, market integration, and societal complexity are all highly correlated (S*51*). From this project we learn that MI predicts higher offers in all three games. If we use Community Size as a proxy for anonymity and face-to-face interactions in populations, then our work shows that CS is not negatively associated with prosocial behaviour, as predicted by this anonymity hypothesis. Our findings show that larger societies engage in more prosocial punishment. For offers, CS is never significant. However, if MI is removed from the regression, CS always has a positive coefficient and it sometimes approaches significance. This is true with or without the full host of control variables. Overall, our findings do not support this version of the anonymity hypothesis.

A second version of the anonymity hypothesis proposes that individuals from smaller communities may have avoided punishing because they feared that their actions might be found out and interpreted as an aggressive move against Player 1. The idea is that the likelihood of being found out increases in smaller communities, thus yielding the observed relationship between community size and MAO. When seen in the light of the offer data we think this interpretation is largely consistent with our theory. To see this, first, recall that community size was not an important predictor of offers in any of the experiments. This means that, somehow, this anonymity concern was only a factor on the punishment side, and did not influence offer decisions. Individuals from some smaller communities (though not all), such as in Fiji (S*5*), entered the experiment and gave fairly, but were rarely willing to punish. Following our theoretical approach, this likely reflects local prosocial norms stabilized by reputational mechanisms, such as being dropped from dyadic helping networks (S*46*), and does not involve costly punishment. Costly punishment is, in fact, often frowned upon in these small communities because it can generate cycles of reprisals. Individuals who live in communities with norms maintained by such reputational mechanisms ought to show both fair behavior and an unwillingness to engage in costly punishment. As is the case for many norms (S*52*), habitually performed and frequently observed behaviors are partially internalized so that actions in daily life reflect some combination of internalized motivations (not to punish, in this case) and beliefs about the consequences of punishing, for example. Thus, in our view, the unwillingness to punish in our experiments reflects the rules of daily life, and arises from some combination of beliefs and/or motivations imported into the experiment from routine practices and interactions.

The fear that might deter a potential punisher from inflicting costs in the experiment can only arise if a certain community would judge this action as "aggressive" (and inappropriate). This depends entirely on local rules about appropriateness. For example, in many Western communities, it is perfectly fine, encouraged actually, to punch someone as hard as you can in the face, as long as you are in a boxing ring. In this light, one can just as easily run the anonymity hypothesis the other way. If individuals do not believe the anonymity, or believe it less in smaller communities, they should take the opportunity to demonstrate their commitment to fairness, their toughness, or their "hard bargaining" to their fellow community members by dishing out large punishments. Community size should negatively predict MAOs. In short, this anonymity hypothesis depends on local norms that influence how "punishing" is judged. Thus, it falls broadly within our explanatory framework based on norms.

Relevant to assessing this anonymity hypothesis, we performed these experiments among university students in the U.S., as mentioned above (S*5*). In an effort to approximate the small communities from which we drew our subjects we randomly selected students from the same small freshman dormitory. If the size of the pool from which players are drawn influences their assessment of future potential anonymity, and thereby causes them to punish less, then these students should have punished less than students typically observed, and potentially a lot less, since the pool sizes vary from one freshman dormitory to the entire university. While a carefully controlled comparison is not possible, the results for punishment among these student subjects do not appear much different from those observed in student samples drawn from larger more anonymous populations (S*3*). This suggests that merely manipulating the size of the pool of frequently interacting potential subjects is not driving the impact of the community size variable.

Yet a third version of the anonymity hypothesis is that prosocial behavior in our experiments results from efforts to manage one's reputation with the experimenter (S*53*). Inspiration for this effect comes from work using the Dictator Game (S*54, S55, S56*) involving experimenter-blind treatments and work manipulating anonymity cues (S*57*). Most experiments are single-blind, meaning that a player's behavior will never be known by the other players (and everyone is told this). In double-blind experiments, neither the other players nor the experimenters can figure out what a specific player did. Among students, protocols that make the double-blind transparent (giving maximum confidence in anonymity vis-à-vis the experimenter) causes Dictator Game offers to decline. The approach could explain our variation if different populations varied in how much they wanted to impress the experimenter, in a manner that happened to strongly correlate with MI, WR and CS.

To put such experiments in a broader context, it is important to recognize two things. First, most of these effects have been limited to students in Dictator Games. Both students (S*58*) and Dictator Games are notorious for being relatively easy to manipulate using framing effects. The same manipulations show little or no effects in the UG, or other games like the Trust Game; there is no data on the TPG. Manipulations of anonymity cues using "eye-spots," for example, do not work in the Trust Game (Fehr, personal communication).

Second, a norm-learning approach offers a ready interpretation to these findings. Some student players find the DG is ambiguous in terms of which norms apply, since it lacks the strategic conflict of other games. This ambiguity causes them to seize on otherwise minor framing effects to figure out which norms to apply to the game. The double-blind procedures provide strong cues—the game equivalent of neon-signs—proclaiming: "I am not looking, so feel free to behave self-interestedly." Psychologists call this well-known phenomenon the experimenter demand effect. Attempting to remove concerns about the impacts of experimenter knowledge may actually dramatically raise experimenter demand effects by signaling to the subject how they are "supposed" to behave (cuing them as to what the norms are). Interestingly, as we discuss below, unconsciously priming subjects with "God" before playing a double-blind Dictator Game increases fairness, switching the mode from offers of zero to offers of 50%. God is way more important than the experimenter, among non-atheists, and there is no experimental treatment that blinds him/her/it.

Our project sought to address these concerns by running double-blind experiments among non-student subjects in four populations (S*24, S59*). Among the Orma and Samburu we find no measureable impact of the double-blind treatment (S*24*). Moreover, in the U.S., the standard finding using the DG with students, which reveals a large drop in mean offers, does not readily extend to non-student populations. Among U.S. non-student adults we found a substantially diminished effect in the double-blind treatment compared to typical student findings. Among the Sanquianga (S*25*) in Columbia, we did find some impact of the double-blind treatments. Across our four comparative experiments, the two with no detectable effects for the double-blind treatments were conducted by non-co-ethnics of the participants. The U.S. experiment and the Sanquianga experiment were done, not by locals, but by co-nationals. Perhaps people do not care what out-group others think of them. This is good for us, since with a couple of exceptions, all experiments were administered by out-group, non-co-ethnics. This could suggest that the large effects of double-blind procedures is an artifact of relying on Western student subjects with western experimenters whose approval they seek (S*58*).

Our previous work has also addressed the issue of experimenter anonymity by reasoning that if the variation we observe among populations results from different subject populations caring differing degrees about what the experimenter might think, believing different things about what he might do with the information, or possessing different beliefs about what the experimenter might want, we sought to control for this by including the number of months each experimenter had spent in the community prior to doing the experiments. The coefficient on this predictor was close to zero, and non-significant. Its inclusion had no influence on our findings (S*51*).

Finally, in this current project, as in our previous work, the anonymity hypotheses generally suffer because they cannot explain the strong relationship we find for MI, WR and CS in our data.

EVOLUTION OF MORALIZING (WORLD) RELIGIONS

Recent work has suggested that certain religious institutions, beliefs, and rituals may have coevolved with the norms that support complex societies (S*30, S60, S61*). Competition among societies or religions may have favoured the spread of beliefs, institutions, and rituals—including those associated with faith in potent, moralizing gods—because these emerging religious systems helped promote or sustain prosocial behaviour towards co-religionists (and the exploitation of non-co-religionists). The culturally-evolved psychological tools used by these religions to promote in-group prosociality and successful economic exchange may include the threat of supernaturally supplied rewards or punishments (blissful afterlives) and recurrent community rituals that build group solidarity (S*62*).

MODERN RELIGIONS ARE UNUSUAL RELIGIONS

Comparative ethno-historical analyses show that the emergence of potent moralizing religions and gods increases with greater societal size and complexity. Cross-cultural analysis of 186 societies found that larger, more complex, societies were much more likely to subscribe to potent deities directly concerned with morality and willing to punish norm violators (S*63, S64, S65*). The gods of the smallest-scale societies are typically weak, whimsical, and morally ambiguous (S*60, S64*). In fact, anthropologists have long pointed out that in the religions of the smallest scale societies there is little or no overlap between the domains of morality and the supernatural (S*66, S67*). This is absolutely not to suggest that small-scale societies lack morality, but merely that supernatural agents were not involved in policing behavior. Among the anthropologically well-known San foragers in the Kalahari, for example, Marshall wrote (S*68*), "Man's wrong-doing against man is not left to ≠Gao!na's [the relevant god] punishment nor is it considered to be his concern. Man corrects or avenges such wrong-doings himself in his social context" (p. 245, the "[]" are ours). The connection between morality and religion has evolved culturally over human history, probably rather recently.

While the smallest scale societies lack a connection between supernatural beliefs and incentives regarding anti- and prosocial behaviour, the emergence of larger-scale chiefdoms, after the origins of sedentary food production, seems tightly associated with changes in religious beliefs, rituals and institutions. The ancestor gods of even the simplest chiefdoms, with all the faults of humans, would sometimes punish errant individuals for violations like theft, murder, and adultery using the instruments of illness, accidents (e.g. shark attacks in Polynesia and Fiji), and generally bad luck (S*69, S70*). Ancestors would also (1) punish people for whimsy, (2) demand payments in the form of sacrifice, and (3) remain absent during critical times. Moreover, the religions of Chiefdoms notoriously favoured political stability by endowing chiefs with divine wisdom and power, or in super-chiefdoms like Egypt, divinity itself (S*71*).

While far from omnipotent, omniscient, and benevolent, these supernatural overseers reveal the first footprints of the track created by the competition among religions.

In a recent synthesis, Wright (S*60*) summarizes the historical evidence revealing the gradual evolution of the Abrahamic god from a rather limited, whimsical, tribal war god, a subordinate in the Pantheon, to the unitary, supreme, moralizing deity of the world's largest religious communities. By the time that gods enter the historical records, after the fifth millennia B.P. in Mesopotamia, they look a lot like the gods observed ethnographically in Chiefdoms. They were still relatively small, local, city, or tribal gods. Even in the cases in which they would seem to call for prosociality towards others, these calls were confined to co-ethnics. By 3790 B.P a variety of gods provide the supernatural support and justification for the Code of Hammurabi. It is not until around 500 B.C.E. that ideas of salvation emerge, giving religion for the first time power to amply reward and punish beyond the confines of earthly life (S*72*).

Archaeological research indicates that both the expansion of regular rituals and the construction of religiously significant monumental architecture co-emerges with increasing societal size and political complexity (S*73*). Research on the coevolution of ritual and society indicate that rituals became much more formal, elaborate and costly as societies developed from foraging bands into chiefdoms and states (S*73*). In Mexico before 4000 B.P., for example, foraging bands relied on informal, unscheduled rituals just as ethnographically known foragers do (S*74*). With the establishment of multi-village chiefdoms (4000-3000 B.P.), rituals expanded and distinct religious specialists emerged. This also applies to pre-dynastic Egypt (6000-5000 B.P.) and China (4500-3500 B.P.), as well as to North American chiefdoms. After state formation in Mexico (2500 B.P.), key rituals were performed by a class of full-time priests, subsidized by the populace, using religious calendars and occupying temples built at immense costs. This is also true for the earliest state-level societies of Mesopotamia (after 5500 B.P.) and India (after 4500 B.P.). Combining this with comparative ethnography suggests that high moralizing gods likely coevolved with costly regularized rituals, creating a mutually re-enforcing package capable of enhancing internal cooperation and harmony, while providing a justification to exploit out-groups.

BELIEF IN MODERN RELIGIONS AND RITUAL PARTICIPATION IMPACTS EXPERIMENTAL MEASURES OF ADHERENCE TO PROSOCIAL NORMS

A slowly rising tide of experimental evidence supports the linkage between modern religious faith, ritual participation and prosocial behavior (S*30*). Behavioral experiments reveal that unconsciously activating religious concepts leads to increased generosity toward strangers (S*15, S75, S76*), except among atheists. For example, using the Dictator Game, Shariff and Norenzayan (S*15*) show that priming God among those practicing a world religion increases offers rather dramatically, even in double-blind conditions. Atheists are not impacted by this unconscious priming. Using parallel techniques, other authors have shown that unconsciously priming religious concepts decreases cheating (S*31*).

The linkage of rituals to prosocial behavior towards in-group members has also been demonstrated in a variety of ways. Among Israeli kibbutzim, individuals from religious kibbutzim cooperated more in a behavioral experiment than those from non-religious kibbutzim, where the increased cooperativeness of religious members can be attributed to greater ritual participation (S*77*). Religious kibbutzim also economically outperform secular kibbutzim (S*78, S79*).

Ritual on its own seems sufficient to increase prosociality. Recent studies found that strangers acting in synchrony —marching, singing and dancing — cooperated more in subsequent group exercises, even in situations requiring personal costs. Synchronous action (rhythmically moving together) increases cooperation, even when no positive emotion was attached to the actions (S*80*). The ability of music, rhythm, and synchrony to instill commitment and trust is no doubt why military drill routines and practices developed over the centuries to train soldiers and bond armies (S*81*). Religion, not the military, may have been the first human institution to evolve to tap this useful piece of psychology.

FIELD EVIDENCE SUPPORTS NOTION THAT RELIGION EVOLVES VIA COMPETITION

There is a wide range of field evidence on how the religions and rituals of small-scale societies change over time in a manner illustrating how competitive forces can drive the evolution of religion. While a comprehensive review is impossible here (S*62, S82*), we supply a few examples. In their ethnohistorical study of the Enga, Wiessner and Tumu (S*83*) lay out the evolution of various belief-ritual complexes in the highlands of New Guinea. Central to the emergence of these ritually galvanized ideological systems, which the authors describe as promoting "identity, welfare, and unity" within larger and larger groups over time, is the cultural transmission of belief-ritual complexes, or elements of them, both within and across linguistic boundaries. The authors write, "cults were readily transmitted across linguistic boundaries when (1) donors and recipients faced comparable problems, so that underlying beliefs and overt procedures were meaningful, and (2) the owners of the cults were perceived as being successful…Cults were imported in order to acquire new and more effective ways to communicate with the spirit world, as well as to emulate those who appeared more successful" (S*83: 195-196*). While the data are lacking for a precise calculation, it appears clear that the frequency and intensity of costly rituals increases along with sociopolitical complexity in New Guinea, in the face of increasing military and economic competition. Sosis and his collaborators show similar effects in their quantitative analysis, linking costly ritual practices and warfare (S*84*).

Elsewhere in New Guinea, Tuzin has examined the historical co-emergence of a strong group ideology, an intricate form of social organization, a complex ritual system, and a high degree of cooperation and solidarity. In a region where villages often breakdown when they grow above approximately 300 people, this study of the Ilahita Arapesh reveals how an interlocking segmented moiety system, galvanized by the rehearsal of a secret ritual system called Tambaran, allows 1,500 people to live together in harmony, sustain high levels of cooperation and solidarity, and survive in a very competitive regional environment that has long included both military and economic threats (S*85, S86*). The basic elements of the institutional-ritual complex, which the Ilahita Arapesh elaborated and improved upon, were first imitated from a highly successful and aggressively expanding group called the Abelam around the 1870's. Their acquisition and modification (and improvement) of the Abelam systems probably permitted Ilahita to resist being driven out by the Abelam, and has since permitted both military and economic success.[4]

In discussing one means by which Islam spread into Africa, Ensminger (S*87*) argues that Islamic beliefs—galvanized by costly displays involving abstaining from alcohol, blood and pork, limiting polygyny, Pilgrimages, and fasting—permitted greater trust, shared rules of exchange and the use of credit institutions among converted Muslims. This facilitates more trade and greater economic success. The Orma, and presumably other African groups, observed successful Muslim models performing costly displays and began adopting the religious beliefs along with the associated institutions and rituals. Ensminger describes an empirical pattern when she writes (p. 26-27), “such groups [Islamic groups] may have attracted followers at a greater rate than others, thus increasing the ranks of the converted.”

This last example is particularly crucial since it is about the connection between religion and reliable, mutually beneficial, exchange. The Orma had a local religion, but it did not allow them to do what Islam now facilitates. The Orma are one of the populations in our sample.

REFERENCES

S1. E. Hoffman, V. Smith, K. McCabe, *J. Econ. Perspect.* **9**, 236.
S2. W. Guth, R. Schmittberger, B. Schwarze, *Journal of Economic Behavior & Organization* **3**, 367.
S3. E. Fehr, U. Fischbacher, *Evolution and Human Behavior* **25**, 63.
S4. J. Henrich *et al.*, *Science* **312**, 1767.
S5. J. Henrich, N. Henrich, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Russell Sage Press, New York, Submitted).
S6. C. K. Lesorogo, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Submitted).
S7. A. H. Bolyanatz, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Submitted).
S8. A. Barr, J. Ensminger, J. C. Johnson, in *Who Can We Trust: How Groups, Networks, and Institutions make Trust Possible* K. S. Cook, M. Levi, R. S. Hardin, Eds. (Russell Sage Foundation, New York, forthcoming) pp. 65-90.
S9. M. Gurven, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Submitted).
S10. M. Gurven, in *Foundations of Human Sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies* J. Henrich *et al.*, Eds. (Oxford University Press, Oxford, 2004) pp. 194-231.
S11. M. Gurven, A. Zanolini, E. Schniter, *Journal of Economic Behavior & Organization* **67**, 587.

[4] Tuzin argues that the Ilahita Arapesh also acquired their elaborate garden technology for growing yams from the Abelam. He also points out that the transmission went one way, from more successful to less successful. The rich mythology and elaborate hunting magic of the Ilahita Arapesh did not transmit to the Abelam, or anyone else (S*86: 79*).

S12. J. Henrich, N. Smith, in *Foundations of Human Sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies* J. Henrich *et al.*, Eds. (Oxford University Press, Oxford, 2004) pp. 125-167.
S13. J. Henrich, *Amer. Econ. Rev.* **90**, 973.
S14. E. Ostrom, R. Gardner, J. Walker, *Rules, Games and Common-Pool Resource Problems* (University of Michigan Press, 1994).
S15. A. F. Shariff, A. Norenzayan, *Psychological Science* **18**, 803.
S16. D. Tracer, *Curr Anthrop* **44**, 432.
S17. D. Tracer, in *Foundations of Human Sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies* J. Henrich *et al.*, Eds. (Oxford University Press, Oxford, 2004) pp. 232-259.
S18. D. P. Tracer, I. Mueller, J. Morse, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Submitted).
S19. F. Marlowe, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Submitted).
S20. H. C. Barrett , K. J. Haley, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Submitted).
S21. J. P. Ziker, in *Experimenting with Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Submitted).
S22. E. L. Gwako, J. Ensminger, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Submitted).
S23. R. McElreath, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Submitted).
S24. C. Lesorogol, J. Ensminger, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Submitted).
S25. J.-C. Cardenas, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Submitted).
S26. A. Barr, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Henrich, J. Ensminger, Eds. (Submitted).
S27. J. Ensminger, K. Cook, in *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* J. Ensminger, J. Henrich, Eds. (Russell Sage Foundation Press, New York, Submitted).
S28. M. Weber, *The Protestant ethic and the spirit of capitalism* (Scribner, New York, 1958).
S29. E. Durkheim, *Elementary Forms of Religious Life* (Free Press, New York, 1995).
S30. A. Norenzayan, A. F. Shariff, *Science* **322**, 58.
S31. B. Randolph-Seng, M. E. Nielsen, *Int J Psych Rel* **17**, 303.
S32. J. P. Ziker, *Hum. Ecol.* **31**, 331.
S33. P. Spencer, *The Samburu* (Routledge & Kegan Paul, London, 1965).
S34. J. Henrich, *Hum. Ecol.* **25**, 319.
S35. D. L. Bahry, R. K. Wilson, *Journal of Economic Behavior & Organization* **60**, 37.
S36. S. Bowles, *Science* **320**, 1605.
S37. B. Frey, R. Jegen, *J. Econ. Surveys* **15**, 589.
S38. B. Herrmann, C. Thöni, S. Gächter, *Science* **319**, 1362.
S39. E. Fehr, K. Hoff, M. Kshetramade, *Amer. Econ. Rev.* **98**, 494.
S40. M. Sutter, M. Kocher, *Games Econ. Behav.* **59**, 364.
S41. W. T. Harbaugh, K. Krause, S. G. Liday, http://economics.uoregon.edu/papers/UO-2002-4_Harbaugh_Kid_Bargaining.pdf

S42. J. Carpenter, S. Burks, E. Verhoogen, in *Field Experiments in Economics* J. Carpenter, G. W. Harrison, J. A. List, Eds. (JAI, Greenwich, Connecticut, 2005).
S43. C. Bellemare, S. Kroeger, A. Van Soest, *Econometrica* **76**, 815.
S44. C. Bellemare, S. Kroger, *Europ. Econ. Rev.* **51**, 183.
S45. J. R. Carter, M. D. Irons, *J. Econ. Perspect.* **5**, 171.
S46. K. Panchanathan, R. Boyd, *Nature* **432**, 499.
S47. K. Panchanathan, R. Boyd, *J. Theor. Biol.* **224**, 115.
S48. R. F. I. Cancho, R. V. Sole, R. Kohler, *Phys. Rev. E* **69**.
S49. F. Marlowe *et al.*, *Proc. R. Soc. London Ser. B* **275**, 587.
S50. T. E. Currie, R. Mace, *Proceedings of the National Academy of Sciences* **106**, 7339.
S51. J. Henrich *et al.*, Eds., *Foundations of Human Sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies* (Oxford University Press, Oxford, 2004).
S52. N. Henrich, J. Henrich, *Why Humans Cooperate: A Cultural and Evolutionary Explanation* (Oxford University Press, Oxford, 2007).
S53. S. D. Levitt, J. A. List, *J. Econ. Perspect.* **21**, 153.
S54. T. L. Cherry, P. Frykblom, J. F. Shogren, *Amer. Econ. Rev.* **92**, 1218.
S55. J. A. List, T. L. Cherry, *Journal of Economic Behavior & Organization* **65**, 1.
S56. E. Hoffman, K. McCabe, K. Shachat, V. Smith, *Game and Economic Behavior* **7**, 346.
S57. K. Haley, D. M. T. Fessler, *Evolution and Human Behavior* **26**, 245.
S58. J. Henrich, S. J. Heine, A. Norenzayan, *Behavior and Brain Science*.
S59. J. C. Cardenas. (Roots of Human Sociality Conference, California Institute of Technology, Pasadena, 2005).
S60. R. Wright, *The Evolution of God* (Little Brown and Company, Boston, 2009).
S61. A. Shariff, A. Norenzayan, J. Henrich, in *Evolution, Culture and the Human Mind* M. Schaller, A. Norenzayan, S. Heine, T. Yamaguishi, T. Kameda, Eds. (Lawrence Erlbaum Associates, 2009) pp. 119-136.
S62. J. Henrich, *Evolution and Human Behavior* **30**, 244.
S63. D. D. P. Johnson, *Hum. Nature* **16**, 410.
S64. F. L. Roes, M. Raymond, *Evolution and Human Behavior* **24**, 126.
S65. F. L. Roes, *Politics and the Life Sciences* **14**, 73.
S66. E. B. Tylor, *Primitive culture : researchs into the development of mythology, philosophy, religion, language, art and custom* (John Murray, London, 1920).
S67. G. E. Swanson, *The birth of the gods; the origin of primitive beliefs* (University of Michigan Press, Ann Arbor, 1960).
S68. L. Marshall, *Africa: Journal of the International African Institute* **32**, 221.
S69. E. S. C. Handy, *Polynesian religion* (The Museum, Honolulu, Hawaii, 1927).
S70. R. H. Lowie, *Primitive religion* (Liveright Pub. Corp., New York, 1948).
S71. P. Nolan, G. Lenski, *Human Societies: An Introduction to Macrosociology* (Paradigm Publishers, Boulder, ed. 9, 2004).
S72. W. H. McNeil, *The Rise of the West: A History of the Human Community* (University of Chicago Press, Chicago, 1991).
S73. J. Marcus, K. V. Flannery, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 18257.
S74. R. Lee, *The !Kung San: Men, women and work in a foraging society* (Cambridge University Press, New York, 1979).
S75. J. A. Bargh, T. L. Chartrand, *American Psychologist* **54**, 462.
S76. N. Mazar, D. Ariely, *J Public Policy Mark* **25**, 117.
S77. R. Sosis, B. J. Ruffle, *Curr Anthrop* **44**, 713.
S78. B. J. Ruffle, R. Sosis, *Journal of Economic Behavior & Organization* **60**, 147.

S79. A. Fishman, Y. Goldschmidt, *Journal for the Scientific Study of Religion* **29**, 505.
S80. S. S. Wiltermuth, C. Heath, *Psychological Science* **20**, 1.
S81. W. H. McNeil, *Pursuit of Power: Technology, Armed Force, and Society since A.D. 1000* (University of Chicago Press, Chicago, 1982).
S82. D. S. Wilson, *Darwin's Cathedral: Evolution, Religion, and the Nature of Society* (University of Chicago, Chicago, 2002).
S83. P. Wiessner, A. Tumu, *Historical Vines*. W. Merrill, I. Karp, Eds., Smithsonian Series in Ethnographic Inquiry (Smithsonian Institution, Washington D.C, 1998).
S84. R. Sosis, H. Kress, J. Boster, *Evolution and Human Behavior* **28**, 234.
S85. D. Tuzin, *Social Complexity in the Making: A case study among the Arapesh of New Guinea* (Routledge, London, 2001).
S86. D. Tuzin, *The Ilahita Arapesh* (University of California, Berkeley, 1976).
S87. J. Ensminger, *J. Inst. Theoretical Econ.* **153**.