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

Joseph Henrich,[1]* Jean Ensminger,[2] Richard McElreath,[3] Abigail Barr,[4] Clark Barrett,[5] Alexander Bolyanatz,[6] Juan Camilo Cardenas,[7] Michael Gurven,[8] Edwins Gwako,[9] Natalie Henrich,[10] Carolyn Lesorogol,[11] Frank Marlowe,[12] David Tracer,[13] John Ziker[14]

Large-scale societies in which strangers regularly engage in mutually beneficial transactions are puzzling. The evolutionary mechanisms associated with kinship and reciprocity, which underpin much of primate sociality, do not readily extend to large unrelated groups. Theory suggests that the evolution of such societies may have required norms and institutions that sustain fairness in ephemeral exchanges. If that is true, then engagement in larger-scale institutions, such as markets and world religions, should be associated with greater fairness, and larger communities should punish unfairness more. Using three behavioral experiments administered across 15 diverse populations, we show that market integration (measured as the percentage of purchased calories) positively covaries with fairness while community size positively covaries with punishment. Participation in a world religion is associated with fairness, although not across all measures. These results suggest that modern prosociality is not solely the product of an innate psychology, but also reflects norms and institutions that have emerged over the course of human history.

At the onset of the Holocene, the stabilization of global climates created possibilities for the emergence of larger-scale sedentary human societies (*1*). Over the next 10 millennia, the scale of some human communities dramatically expanded from kin-based foraging bands to complex, intensely cooperative societies in which strangers frequently engage in mutually beneficial transactions (*2*). Consistent with life in these large-scale societies, behavioral experiments performed with people from these populations reveal fair, trusting, and cooperative behavior among strangers, even in one-shot encounters (*3*).

Two major theoretical approaches have sought to explain both the relatively rapid expansion of human societal scales and the puzzlingly prosocial behavior observed in experiments. The first approach proposes that humans possess an innate social psychology calibrated to life in the small-scale societies of our Paleolithic ancestors. Rooted in the evolutionary logic of kinship and reciprocity, these heuristics were mistakenly extended to nonkin and ephemeral interactants, as societies expanded with the emergence of agriculture. From this view, the prosocial behavior observed in experiments directly reflects the operation of these ancient heuristics (*4*, *5*).

An alternative approach proposes that a crucial ingredient in the rise of more-complex societies was the development of new social norms and informal institutions that are capable of domesticating our innate psychology for life in ever-expanding populations (*6*). Larger and more-complex societies prospered and spread to the degree that their norms and institutions effectively sustained successful interaction in ever-widening socioeconomic spheres, well beyond individuals' local networks of kin and long-term relationships (*7*). It is these particular norms and their gradual internalization as proximate motivations (*8*) that recalibrate our innate psychology for life in small-scale societies in a manner that permits successful larger-scale cooperation and exchange in vast communities.

Much research suggests that norms arise because humans use evolved learning mechanisms to calibrate their behavior, motivations, and beliefs to variable circumstances (*7*, *9*). Modeling work shows that when these learning mechanisms are applied to different kinds of social interactions, such as large-scale cooperation or ephemeral exchange, individually costly behaviors can be sustained by punishment, signaling, and reputational mechanisms (*10–13*). By sustaining such behaviors, norms can facilitate trust, fairness, and cooperation in a diverse array of interactions, thereby allowing the most productive use of unevenly distributed skills, knowledge, and resources, as well as increasing cooperation in exchange, public goods, and warfare. More-effective norms and institutions can spread among societies by a variety of theoretically and empirically grounded mechanisms, including conquest and assimilation, preferential imitation of more-successful societies, or forward-looking decision making by leaders or high-status coalitions (*14*, *15*).

Norms that enhance fairness among strangers are likely causally interconnected with the diffusion of several kinds of institutions. Here we focus on two: (i) the expansion of both the breadth and intensity of market exchange (*16*, *17*), and (ii) the spread of world religions (*18*). The efficiency of market exchange involving infrequent or anonymous interactions improves with an increasingly shared set of motivations and expectations related to trust, fairness, and cooperation. This lowers transaction costs, raises the frequency of successful transactions, and increases long-term rewards (*16*). Although frequent and efficient exchanges among strangers are now commonplace, studies of nonhuman primates and small-scale societies suggest that during most of our evolutionary history, transactions beyond the local group, and certainly beyond the ethnolinguistic unit, were fraught with danger, mistrust, and exploitation (*2*, *16*, *19*). Thus, we propose that such "market norms" may have evolved as part of an overall process of societal evolution to sustain mutually beneficial exchanges in contexts where established social relationships (for example, kin, reciprocity, and status) were insufficient. If our theory is correct, then measures of fairness in situations lacking relationship information (for example, anonymous others) should positively covary with market integration.

Recent work has also tentatively proposed that certain religious institutions, beliefs, and rituals may have coevolved with the norms that support large-scale societies and broad exchange (*18*, *20*, *21*). Intersocietal competition may have favored those religious systems that galvanize prosocial behavior in broader communities, perhaps using both supernatural incentives (for example, hell) and recurrent rituals that intensify group solidarity (*20*, *22*). Consistent with this view, analyses of ethnographic data show that the emergence of moralizing religions increases with greater societal size and complexity (*18*, *23*). Archaeologically, regular rituals and the con-

[1]Departments of Psychology and Economics, University of British Columbia, 2136 West Mall, Vancouver, British Columbia V6T 1Z4, Canada. [2]Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA. [3]Department of Anthropology, University of California, Davis, CA 95616, USA. [4]Centre for the Study of African Economies, Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, UK. [5]Department of Anthropology, University of California, 341 Haines Hall, Box 951553, Los Angeles, CA 90095, USA. [6]Department of Anthropology, College of DuPage, Glen Ellyn, IL 60137, USA. [7]Facultad de Economia, Universidad de Los Andes, K1 Number 18A-70, Bogotá, Colombia. [8]Department of Anthropology, University of California, Santa Barbara, CA 93106, USA. [9]Department of Sociology and Anthropology, Guilford College, 5800 West Friendly Avenue, Greensboro, NC 27410, USA. [10]Centre for Health Evaluation and Outcome Sciences, Providence Health Care Research Institute, Vancouver, British Columbia V6Z 1Y6, Canada. [11]George Warren Brown School of Social Work, Washington University, 1 Brookings Drive, St. Louis, MO 63130, USA. [12]Department of Anthropology, Florida State University, Tallahassee, FL 32306, USA. [13]Department of Anthropology, University of Colorado, Post Office Box 173364, Campus Box 103, Denver, CO 80217, USA. [14]Department of Anthropology, Boise State University, 1910 University Drive, Boise, ID 83725, USA.

*To whom correspondence should be addressed. E-mail: henrich@psych.ubc.ca

struction of monumental religious architecture co-emerge with societal size and complexity (24). In experiments, unconsciously priming the faithful with religious concepts favors greater fairness toward anonymous others (21). This suggests that, in contrast to the religions that likely dominated our evolutionary history, modern world religions such as Christianity and Islam may be unusual in ways that buttress the norms and institutions that sustain larger-scale interaction [supporting online material (SOM) text]. If this theory is correct, then greater fairness toward anonymous others should be associated with adherence to a world religion.

Evolutionary approaches to norms also afford a prediction about the willingness of individuals across populations to engage in the costly punishment of norm violations. Theoretical modeling reveals at least two different kinds of norm-stabilizing mechanisms: One involves reputational effects in which norm violators are sanctioned in another interaction by, for example, not receiving aid in a dyadic helping situation (10), and a second involves the use of diffuse (25) costly punishment (12). Because the effectiveness of reputational systems in sustaining norms degrades rapidly as communities expand (26, 27), fairness in larger communities must increasingly be maintained by diffuse punishment; that is, larger communities should punish more.

Experiments. The evidence presented below derives from a second round of cross-cultural experiments that were designed to illuminate findings from our first project (28, 29). This round replaces 10 populations from our earlier effort with 10 new ones (swapping several researchers as well), while resampling from five of the same populations used previously (Table 1). This analysis complements prior work from our team, which focused on the distributional patterns of punishment (30), by seeking to explain the variation in our experimental measures of both fairness and punishment. Our analysis converges with recent work comparing diverse industrialized populations (31).

If markets and world religions are linked to the norms that sustain exchange in large-scale societies, we expect that experimental measures of fairness in anonymous interactions will positively covary with measures of involvement in these two institutions. To test this, we used three experiments that were designed to measure individuals' propensities for fairness and their willingness to punish unfairness across 15 populations that vary in their degree of market integration and their participation in world religions. Our three experiments are the Dictator, Ultimatum, and Third-Party Punishment Games (32).

In the Dictator Game (DG), two anonymous players are allotted a sum of money (the stake) in a one-shot interaction (3). Player 1 must decide how to divide this sum between himself or herself and Player 2. Player 2 receives the allocation (offer), and the game ends. Player 1's offer to Player 2 provides a measure of Player 1's behavioral fairness in this context.

In the Ultimatum Game (UG), two anonymous players are again allotted a sum in a one-shot interaction (3). Player 1 can offer a portion of this to Player 2. Player 2, before hearing the actual offer from Player 1, must decide whether to accept or reject each of the possible offers (in 10% increments). Decisions are binding. If Player 2 specified that he or she would accept the amount of the actual offer, then Player 2 receives the offered amount and Player 1 gets the remainder. If Player 2 specified that he or she would reject the offered amount, then both players receive zero. If people are motivated purely by money maximization, Player 2s will always accept any positive offer; realizing this, Player 1 should offer the smallest nonzero amount. Because this is a one-shot anonymous interaction, Player 2's willingness to reject provides a measure of punishment. Player 1's offer measures a combination of social motivations and an assessment of the likelihood of rejection; this gives us a second behavioral measure of fairness.

In the Third-Party Punishment Game (TPG), two players are again allotted a stake, but now a third player also receives the equivalent of one-half of the stake (33). Player 1 must decide how much to allocate to Player 2, who has no choices. Player 3, before hearing the actual amount that Player 1 allocated to Player 2, has to decide whether to pay 20% of his or her allocation to punish Player 1 for each of the possible offers (in 10% increments). If punished, Player 1 loses triple the amount paid by Player 3. Suppose the stake is $100; if Player 1 offers $10 to Player 2 (keeping $90), and Player 3 wants to punish this offer amount, then Player 1 takes home $60 ($90 – $30), Player 2 gets $10, and Player 3 gets $40 ($50 – $10). If Player 3 had instead decided not to punish offers of $10, then the take-home amounts would be $90,

Table 1. Summary information for the populations studied. The column Economic Base classifies the production systems (for example, horticulturalists rely primarily on slash-and-burn agriculture; pastoralists rely on herding). Residence classifies the nature and frequency of each population's movements. Mean MI is the average percentage of total household calories that are purchased in the market. Mean WR gives the percentage of the sample that reported adhering to a world religion. Mean CS gives the average CS for the populations studied. These CSs are for villages, except among the Hadza (who live in camps) and Accra. Com Sam (DG/UG/TPG) gives the number of different communities from which participants were drawn for the DG, UG, and TPG. N gives the number of pairs or trios for each experiment (table S4).

Population	Location	Environs	Economic base	Residence	Mean MI	Mean WR	Mean CS	Com Sam (DG/UG/TPG)	N (DG/UG/TPG)
Accra City	Ghana	Urban	Wage work	Sedentary	100	97	44†	3/3/11	30/30/39
Au*	Papua New Guinea	Mountainous forest	Horticulture and foraging	Sedentary	1	100	309	3/3/1	30/30/30
Dolgan/ Ng.	Siberia	Tundra-taiga	Hunting, fishing, and wages	Semi-sedentary	63	59	612	1/1/0	30/30/0
Gusii	Kenya	Fertile high plains	Farming and wage work	Sedentary	28	100	4063	2/2/1	25/25/30
Hadza*	Tanzania	Savanna-woodlands	Foraging	Nomadic	0	0	43	4/4/3	31/31/27
Isanga village	Tanzania	Mountainous forest	Farming and wage work	Sedentary	70	99	1500	1/1/1	30/30/20
Maragoli	Kenya	Fertile plains	Farming and wage work	Sedentary	43	100	3843	2/2/1	25/25/30
Orma*	Kenya	Semi-arid savanna	Pastoralism	Semi-nomadic	72	100	125	4/0/0	26/0/0
Samburu	Kenya	Semi-arid savanna	Pastoralism	Semi-nomadic	69	66	2000	1/1/1	31/31/30
Sanquianga	Colombia	Mangrove forest	Fisheries	Sedentary	82	84	1931	1/1/2	30/30/32
Shuar	Ecuador	Tropical forest	Horticulture	Sedentary	22	76	498	2/2/1	21/21/15
Sursurunga	Papua New Guinea	Island	Horticulture	Sedentary	24	100	186	3/3/1	30/30/32
Tsimane*	Bolivia	Tropical forest	Horticulture and foraging	Semi-nomadic	7	100	314	2/2/1	38/36/27
United States*	Missouri	Prairie	Wage work	Sedentary	100	100	1813	1/1/0	15/26/0
Yasawa	Fiji	Coastal tropical pacific	Horticulture and marine foraging	Sedentary	21	100	109	2/2/2	35/34/30

*These populations were also sampled from during our first project (28).   †Participants were sampled from small enterprises in the city of Accra. The value 44 is the average number of employees in the enterprises. Because of the anomalous nature of this situation with regard to obtaining a CS comparable with our other populations, we did not include Accra in analyses involving CS.

$10, and $50, respectively. Because a money-maximizing Player 3 would never pay to punish, a similarly motivated Player 1 should always give zero to Player 2. Thus, Player 3's willingness-to-pay provides another measure of punishment. Player 1's offer measures a mixture of his or her social motivations in this context and an assessment of the punishment threat; this provides a third behavioral measure of fairness.

These experiments, with their salient contextual cues of cash and anonymity, seem well suited to tapping the particular norms that support cooperation and exchange among ephemeral interactants in market transactions. Cash is closely associated with market transactions (34) and often signals a desire to avoid a longer-term nonmarket relationship (35). Anonymity in our games means that players lack the cues or information necessary to apply the expectations and motivations associated with other kinds of relationships, such as those based on kinship, reciprocity, or status differences, forcing players to default to local norms for dealing with people outside durable relationships. However, an important concern in interpreting such experiments involves the degree to which participants accept the experimental situation, believe the anonymity, and worry about the experimenter's judgments of them (SOM text).

Our standardized protocols and methods included the following: (i) random adult samples from our communities (with little attrition in most sites), (ii) stakes set at 1 day's local wage, (iii) a show-up fee of 20% of the stake, (iv) back-translation of scripts, (v) one-on-one instruction and pregame testing for comprehension, (vi) steps to preclude collusion and contamination, and (vii) no deception (32). Our pregame tests combined with analyses of the relationships between experimental decisions and measures of comprehension, indicate no measurable impact of differences in comprehension on behavior (30).

We performed these experiments with 2148 individuals across populations from Africa, North and South America, Oceania, New Guinea, and Asia that included small-scale societies of hunter-gatherers, marine foragers, pastoralists, horticulturalists, and wage laborers.

Table 1 provides the location (mapped in fig. S3), environment, economic base, residence pattern, and sampling information for each population, as well as averages for three key variables.

Results. The theory outlined above predicts a positive relationship between our three measures of fairness—offers in each game—with market integration and adherence to a world religion. Individuals' offers are measured as a percentage of the total stake (1 day's local wage). Market Integration (MI) is measured at the household level by calculating the percentage of a household's total calories that were purchased from the market, as opposed to home-grown, hunted, or fished, and then averaged to obtain a community-level measure. We use the community average for MI both to remain consistent with our definition of norms (as local equilibria) and to remove day-to-day stochastic variation (32). Table 1 shows that the population means for MI range from 0 to 100%, with a mean of 57.3%. World Religion (WR) was assessed by asking participants what religion they practiced, and coding these as a binary variable, with "1" indicating Islam or Christianity, and "0" marking the practice of a tribal religion or "no religion." Table 1 provides the percentage of each population that adheres to a world religion. The mean value of WR is 89%.

To analyze these data, we regressed offers on MI and WR, as well as seven control variables: age, sex, education, income, wealth, household size, and community size (CS). Except for wealth and household size, which are both measured at the household level, and CS, these control variables are individual-level measures. We used 30 (DG), 26 (UG), and 16 (TPG) community-level means for MI to predict 416 (DG), 398 (UG), and 272 (TPG) individual offers with a minimal set of controls, or 336 (DG), 319 (UG), and 265 (TPG) individual offers with the full set of controls.

Table 2 shows four regression models using MI and WR to predict all offers together and offers from the DG, UG, and TPG, separately. Independent of the other sociodemographic variables, the coefficients on MI are large and significant at conventional levels across all four models. A 20–percentage point increase in MI is

associated with an increase in percentage offered ranging from roughly 2 to 3.4. The same qualitative results are obtained for MI whether one uses household measures, community averages, or population averages.

The coefficients on WR are also large across all offers and in both the UG and DG, though not well estimated in the DG ($P < 0.10$). However, WR's coefficients are significant ($P < 0.05$) across all other specifications (tables S5, S8, and S11). Participating in a world religion is associated with an increase in percentage offered of between about 6 and 10 (36).

Taken together, these data indicate that going from a fully subsistence-based society (MI = 0) with a local religion to a fully market-incorporated society (MI = 100%) with a world religion predicts an increase in percentage offered of roughly 23, 20, and 11 in the DG, UG, and TPG, respectively. This spans most of the range of variation across our populations: DG means range from 26 to 47%, UG from 25 to 51%, and TPG from 20 to 43%.

For the seven other socioeconomic variables, none of their coefficients is significant ($P < 0.05$) across all offers or in the DG (tables S5 and S8). For UG offers, the coefficient on age is also positive ($P < 0.05$, table S11). For TPG offers, Table 2 shows that the coefficients on income, wealth, and household size are significant ($P < 0.05$, table S14).

Figure 1 plots mean offers versus mean MI values for our 15 populations. Population mean MI values account for 52% of the variation in mean DG offers [correlation $\rho = 0.72$, 95% bootstrap confidence interval (CI) = 0.4 to 1.0 , $P < 0.01$, $n = 15$ populations]. In designing this second round of experiments, we sought out an additional New Guinea population (Sursurunga), because in the first round, the Au of New Guinea revealed highly unusual behavior, including relatively high offers with little market integration. The Au pattern replicated, and now extends to the Sursurunga. However, because we targeted a second population that skews our world sample unrepresentatively toward New Guinea, we also examined this relationship with either the Au or the Sursurunga dropped. Dropping the Sursurunga

Table 2. Linear regression models for offers. These ordinary least-squares models include four additional control variables (sex, age, community size, and education). Coefficients are followed by standard errors, indicated with ±; P values are given in parentheses.

Variables	All offers*	DG offers†	UG offers†	TPG offers†
MI	0.12 ± 0.023 (<0.001)	0.17 ± 0.035 (<0.001)	0.098 ± 0.035 (0.005)	0.11 ± 0.044 (0.044)
WR‡	5.96 ± 2.04 (0.0036)	6.4 ± 3.61 (0.079)	10.4 ± 2.67 (<0.001)	0.45 ± 3.06 (0.879)
Income (per U.S. $1000)	0.096 ± 0.089 (0.28)	−0.012 ± 0.15 (0.93)	0.16 ± 0.10 (0.13)	−2.25 ± 0.94 (0.017)
Wealth (per U.S. $1000)	0.0012 ± 0.006 (0.83)	0.0013 ± 0.008 (0.88)	−0.0056 ± 0.008 (0.43)	1.2 ± 0.25 (<0.001)
Household size	−0.24 ± 0.21 (0.24)	−0.13 ± 0.31 (0.67)	−0.24 ± 0.2 (0.37)	−1.0 ± 0.43 (0.019)
Observations; R^2	920; 0.084	336; 0.10	319; 0.14	265; 0.10

*Clustered robust standard errors adjust for repeated observations of the same people, 596 individuals (clusters). Indicator variables for each experiment (UG, TPG) were included to control for differences among offers in the three games (table S5).   †Robust standard errors used. See (32) for models with clustered robust standard errors [clustering on populations (Table 1) to control for the nonindependence of individuals], continental-level controls to address any shared culture history (Model 1 in tables S8, S11, and S14), and a variety of other robustness checks, which are summarized in the text.   ‡WR is an individual-level dummy variable with "1" indicating Islam or Christianity, and "0" marking the practice of a tribal religion or "no religion." We have few "no religion" responses, and ethnography suggests that in the wake of missionary activities people associate "religion" with "Christianity." Thus, responses of "no religion" probably imply belief in the local or traditional religion.

leads to mean MI accounting for 58% of the variation ($\rho = 0.76$, 95% bootstrap CI = 0.44 to 0.95, $P < 0.001$). Dropping the Au instead leads to mean MI capturing 71% of the variation ($\rho = 0.84$, 95% bootstrap CI = 0.59 to 0.96, $P < 0.001$).

On the punishment side, we looked for a relationship between our individual-level measures of punishment, from the UG and TPG, and our theoretically important variable, CS, while controlling for our other eight variables. For both the UG and TPG, we reduced the vector of "punish or not punish" responses to a single number called a minimum acceptable offer (MAO). This is the lowest offer ($\leq 50\%$) for which an individual no longer punishes. Thus, if a player accepts all offers, his or her MAO is zero. If the player punishes offers of 0 and 10% but accepts all higher offers, his or her MAO is set at 20. MAO measures an individual's willingness to punish low offers (32).

To analyze MAOs, we used an ordered logistic regression (OLR) because MAOs are both discrete and bimodally distributed. This model assumes that the dependent variable is discrete and rank ordered, but that the distance between ranks is not meaningful. Because the output of an OLR is nonintuitive (tables S20A and S23A), we have captured the effects of the highly significant ($P < 0.001$) coefficients for CS in Fig. 2, A and B. Holding the other eight variables at their mean values, the figure shows how the distribution of MAOs for the TPG and UG shift with increases in CS. Small communities, the size of foraging bands, are the least willing to punish. As CS increases from 50 to 5000, there is a dramatic shift from a modal MAO of 0 to 50. Our communities' sizes range from 20 to 4600 people.

Theoretical arguments suggest that punishment (MAO) should be related more directly to the natural logarithm of CS, because the effectiveness of reputational systems decays in rough proportion to this variable (26, 37). Consistent with this, the natural logarithm of CS is a better predictor of MAO in both experiments than CS itself (tables S20B to S25B), and this amplifies the effects illustrated in Fig. 2 (fig. S4).

Our theoretical approach makes no predictions about the relationship between punishment

and MI, because, for small communities, there are numerous reputational mechanisms (not involving costly punishment) that can sustain equality norms. Similarly, because religion should galvanize the existing mechanisms for norm stabilization, whether participants in a world religion punish more depends entirely on the local stabilization mechanisms, which themselves depend on CS (38). Nevertheless, for MAOs in the TPG, we find that WR is associated with significantly ($P < 0.05$) more punishment, although MI reveals no such relationship. None of the other predictors is significant ($P < 0.05$, table S20A). For MAO in the UG (table S23A), income, wealth, and MI all significantly predict lower MAOs ($P < 0.05$). The effect for MI, however, is contingent on having CS in the equation. If CS is dropped, the effect of MI becomes nonsignificant. Thus, unlike CS, none of these other effects is consistent for both MAO measures or across alternative specifications.

Our analyses for both offers and MAOs are robust to a variety of checks, including alternative model specifications and adjustments to our wealth and income variables to account for local differences in purchasing power (tables S6, S9, S12, S15, S21, and S24). We also included continental-level dummies (Africa, Oceania, South America, and Eurasia) to address concerns about shared cultural phylogenies, and we used clustered robust standard errors (clustering on site) to control for the potential nonindependence of individual observations within our sites (tables S5 to S25). The findings for MI and CS are robust to these checks. However, when continental controls are applied, the effects of WR on offers disappears because of the highly uneven distribution of populations containing individuals with WR = 0 (tables S5, S8, and S11, Model 1). We reran our models for Africa only and generally reconfirmed the above findings (32). Nevertheless, because of the rather small sample of individuals with WR = 0, any conclusions about the effect of WR must remain quite tentative (39).

Discussion. Our results on the relationship between MI and offers extend the findings of our previous project (29) in several ways, including using fresh samples, better mea-

sures, new experiments, and additional controls. Despite swapping in 10 new sites and using a different protocol, we have replicated our earlier UG findings at both the level of individual sites (in four sites) and in obtaining a positive relationship between UG offers and market integration with a better measure. This relationship is now demonstrated for DG and TPG offers, each of which reveals similarly large effects and is robust to the inclusion of a range of socio-economic variables not previously measured and to a suite of statistical checks (tables S5 to S25).

These findings also delineate two additional lines of research. First, our inclusion of participation in a world religion converges with other recent findings (21) and tentatively supports the notion that religion may have coevolved with complex societies to facilitate larger-scale interactions. Second, our analyses (tables S18 and S19) open up the question of why, when a third-party punisher is added to a DG to create a TPG, do mean offers usually decrease, the predictive effects of WR disappear (Table 2), and economic variables emerge as potent correlates of offers (tables S14 to S16)?

On the punishment side, our analyses of MAOs are consistent with the idea that as reputational systems break down in larger populations, increasing levels of diffuse costly punishment are required to sustain large harmonious communities. This extends earlier bivariate analyses using population means in the TPG (40) and converges with ethnographic data, suggesting that large communities, lacking sufficient and effective punishing mechanisms, fragment into smaller groups as their community sizes reach about 300 people (41).

Although our primary interest here is behavioral fairness, which includes whatever combination of motivations and expectations yields more equal divisions, two features of these results suggest that they may stand when considered as measures of intrinsic (internalized) motivations. First, because no punishment opportunities exist in the DG, this experiment likely provides a purer measure of intrinsic motivations for equal offers. Second, we reran the above offer regressions for UG and TPG offers including each population's mean MAO as an additional predictor variable. If individuals are well calibrated to the threat of punishment posed by their local communities, then adding this variable should (at least partially) remove the impact of any threat of punishment on offers. The inclusion of mean MAO as a predictor variable does not change the coefficients on MI or WR (table S17). Finally, anonymity concerns regarding our experiments and results are taken up in the SOM text, where we discuss our double-blind treatments and related analyses, which indicate that differences in perceived anonymity are unlikely to explain our major findings.

These findings indicate that people living in small communities lacking market integration or world religions—absences that likely character-

Fig. 1. Mean DG offers for each population plotted against mean value of MI. Error bars are bootstrapped standard errors (bias corrected and accelerated) on the population mean.
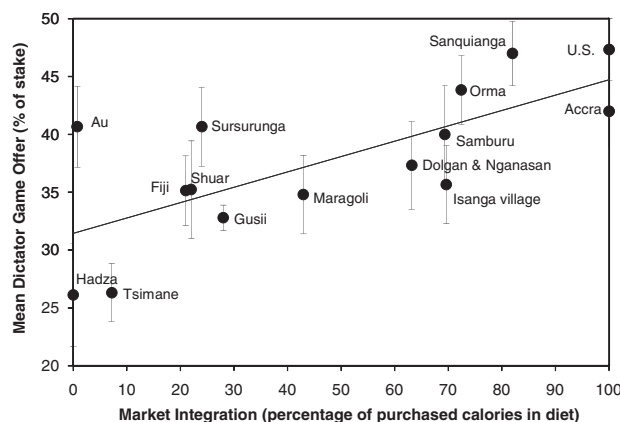
Fig. 2. Effect of CS on MAOs. Each set of bars shows the distribution of MAOs, ranging discretely from 0 to 50, for different CSs. The coefficients used to create the plot are for the variable CS in an ordered logistic regression containing all eight of our other variables. All variables except CS are held constant at their mean values. The coefficients on CS represented graphically here are 0.052 ± 0.0085 (robust standard error, $P < 0.001$, $n = 227$) for the TPG and 0.087 ± 0.011 (robust standard error, $P < 0.001$, $n = 297$) for the UG. (**A**) TPG (see table S20A for full regression). (**B**) UG (see table S23A).
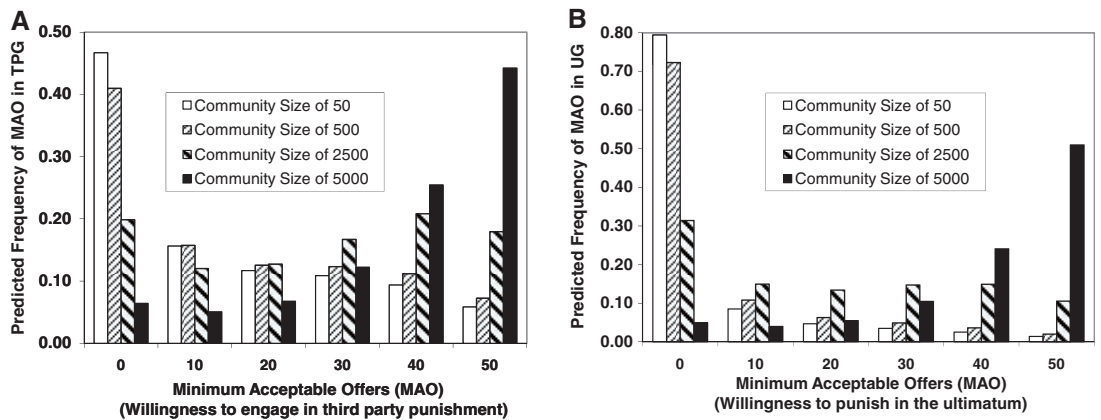
ized all societies until the Holocene—display relatively little concern with fairness or punishing unfairness in transactions involving strangers or anonymous others. This result challenges the hypothesis that successful social interaction in large-scale societies—and the corresponding experimental findings—arise directly from an evolved psychology that mistakenly applies kin and reciprocity-based heuristics to strangers in vast populations (4, 5), without any of the "psychological workarounds" (42) that are created by norms and institutions. Moreover, it is not clear how this hypothesis can explain why we find so much variation among populations in our experimental measures and why this variation is so strongly related to MI, WR, and CS. The mere fact that the largest and most anonymous communities engage in substantially greater punishment relative to the smallest-scale societies, who punish very little, challenges this interpretation.

Methodologically, our findings suggest caution in interpreting behavioral experiments from industrialized populations as providing direct insights into human nature. Combining our findings with work on the links between behavioral experiments and real life (31, 43, 44) suggests that such experiments elicit norms, or reflect institutions, that have evolved to facilitate interactions among individuals not engaged in durable relationships. Given this, much current work using behavioral games appears to be studying the interaction between a particular set of norms and our evolved psychology, not tapping this psychology directly.

Overall, these findings lend support to the idea that the evolution of societal complexity, especially as it has occurred over the last 10 millennia, involved the selective spread of those norms and institutions that best facilitated the successful exchange and interaction in socioeconomic spheres well beyond local networks of durable kin and reciprocity-based relationships. Although differences in environmental affordances probably had a profound impact on the emergence of complex societies across the globe (2), the rate-determining step in societal evolution may have involved the assembly of the norms and institutions that are

capable of harnessing and extending our evolved social psychology to accommodate life in large, intensely cooperative communities. Recent experimental work among diverse industrialized populations suggests that the gradual honing of these norms and institutions continues in modern societies (31).

References and Notes

1. P. J. Richerson, R. Boyd, R. L. Bettinger, *Am. Antiq.* **66**, 387 (2001).
2. J. M. Diamond, *Guns, Germs, and Steel: The Fates of Human Societies* (Norton, New York, 1997).
3. C. Camerer, *Behavior Game Theory: Experiments in Strategic Interaction* (Princeton Univ. Press, Princeton, NJ, 2003).
4. T. C. Burnham, D. D. Johnson, *Anal. Kritik* **27**, 113 (2005).
5. M. A. Nowak, K. M. Page, K. Sigmund, *Science* **289**, 1773 (2000).
6. D. C. North, *Institutions, Institutional Change, and Economic Performance* (Cambridge Univ. Press, New York, 1990).
7. S. Bowles, *Microeconomics: Behavior, Institutions, and Evolution* (Princeton Univ. Press, Princeton, NJ, 2004).
8. E. Fehr, C. F. Camerer, *Trends Cogn. Sci.* **11**, 419 (2007).
9. P. Richerson, R. Boyd, *Not by Genes Alone: How Culture Transformed Human Evolution* (Univ. of Chicago Press, Chicago, 2005).
10. K. Panchanathan, R. Boyd, *Nature* **432**, 499 (2004).
11. H. Gintis, E. A. Smith, S. Bowles, *J. Theor. Biol.* **213**, 103 (2001).
12. J. Henrich, R. Boyd, *J. Theor. Biol.* **208**, 79 (2001).
13. K. Sigmund, C. Hauert, M. A. Nowak, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10757 (2001).
14. R. Boyd, P. J. Richerson, *J. Theor. Biol.* **215**, 287 (2002).
15. J. Ensminger, J. Knight, *Curr. Anthropol.* **38**, 1 (1997).
16. S. Bowles, *J. Econ. Lit.* **36**, 75 (1998).
17. A. Smith, *The Theory of Moral Sentiments* (Prometheus, New York, 2000).
18. R. Wright, *The Evolution of God* (Little, Brown, Boston, 2009).
19. E. Fehr, J. Henrich, in *Genetic and Cultural Evolution of Cooperation*, P. Hammerstein, Ed. (MIT Press, Cambridge, MA, 2003).
20. A. Shariff, A. Norenzayan, J. Henrich, in *Evolution, Culture and the Human Mind*, M. Schaller, A. Norenzayan, S. Heine, T. Yamaguishi, T. Kameda, Eds. (Psychology Press, New York, 2009), pp. 119–136.
21. A. Norenzayan, A. F. Shariff, *Science* **322**, 58 (2008).
22. J. Henrich, *Evol. Hum. Behav.* **30**, 244 (2009).
23. F. L. Roes, M. Raymond, *Evol. Hum. Behav.* **24**, 126 (2003).
24. J. Marcus, K. V. Flannery, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 18257 (2004).
25. Diffuse punishment implies that the responsibility for punishing is spread over a large segment of the population.
26. K. Panchanathan, R. Boyd, *J. Theor. Biol.* **224**, 115 (2003).
27. A. K. Dixit, *Lawlessness and Economics: Alternative Modes of Governance* (Princeton Univ. Press, Princeton, NJ, 2004).
28. J. Henrich et al., Eds., *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (Oxford Univ. Press, Oxford, 2004).
29. J. Henrich et al., *Behav. Brain Sci.* **28**, 795, discussion 815 (2005).
30. J. Henrich et al., *Science* **312**, 1767 (2006).
31. B. Herrmann, C. Thöni, S. Gächter, *Science* **319**, 1362 (2008).
32. Background, methods, and supplemental analyses are available as supporting material on *Science* Online.
33. E. Fehr, U. Fischbacher, *Evol. Hum. Behav.* **25**, 63 (2004).
34. Cash has value in all our societies. Everyone wants steel, salt, and sugar, if nothing else.
35. K. D. Vohs, N. L. Mead, M. R. Goode, *Science* **314**, 1154 (2006).
36. We found no robust differences among different forms of Christianity (Catholics versus Protestants), or between Christianity and Islam.
37. R. F. I. Cancho, R. V. Sole, R. Kohler, *Phys. Rev. E* **69**, 051915 (2004).
38. Traditional religions may sometimes galvanize local norms (SOM text), but they do not galvanize "market norms," which are for interacting with strangers or anonymous others. When religions do underwrite norms, they must operate in concert with the local norm-stabilizing mechanisms.
39. We did not use fixed effects for communities or populations in any of our regressions, because our theoretical focus is on differences in social norms. Because norms are group-level properties (different locally stable equilibria), much of the relevant norm-related variation occurs between, not within, communities or populations (SOM text).
40. F. W. Marlowe et al., *Proc. Biol. Sci.* **275**, 587 (2008).
41. A. Forge, in *Man, Settlement and Urbanisation*, P. Ucko, R. Tringham, G. Dimbelby, Eds. (Duckworth, London, 1972), pp. 363–376.
42. P. Richerson, R. Boyd, *Hum. Nat.* **10**, 253 (1999).
43. M. Poppe, *J. Econ. Psychol.* **26**, 431 (2005).
44. M. Gurven, J. Winking, *Am. Anthropol.* **110**, 179 (2008).
45. We especially thank the communities that participated in our research. We also thank the National Science Foundation, the MacArthur Norms and Preferences Network, the Max Planck Institute for Social Anthropology, and the Russell Sage Foundation for funding this project. Thanks also to the many audiences and readers who contributed to improving our efforts.