Cooperation and punishment in an economically diverse community in highland Tanzania

Richard McElreath*

Draft 2: June 14, 2004. For inclusion in edited volume detailing second round of cross-cultural economics experiments.

Abstract

Previous cross-cultural variation in economic experiments was unexplained by individual-level variables such as wealth, income, gender and education, even though variation in across societies was explained by market integration and potential returns to cooperation (Henrich *et al.*, 2001; Henrich *et al.*, 2004). However, the previous data contained relatively poor measures of individual income, wealth, and market integration. Additionally, most of the societies sampled contain relatively small amounts of variation in these measures. Thus we do not know if the absence of within-group effects is due to causal unimportance or to insufficient variation. In this chapter, I analyze the game behavior of an economically and ethnically diverse community in southern Tanzania that exhibits substantial variation in these measures. Nevertheless, I find very limited evidence that individual economic variables explain variation in game behavior, echoing the weak and inconsistent results found in other societies in this volume. These results cast doubt on the conjecture that the failure of individual-level variables to explain game behavior results simply from inadequate variation within each site. However, error in measurement might be responsible for the lack of explanatory power at the individual level, while the same variables are explanatory across groups.

1 Introduction

A tremendous amount of data now exists supporting the conclusions that human behavior is poorly predicted by standard economic preferences (Camerer, 2003; Fehr & Gachter, 2002; Hoffman *et al.*, 1998; Roth, 1995). In addition, as more non-Western and non-student populations are sampled, it appears that no single revision to *Homo economicus* will be sufficient (Henrich *et al.*, 2004), or at least that the variables that may eventually explain the tremendous amounts of variation in game behavior will fall outside existing theories.

Adequately addressing the role of individual-level variables such as income and education, however, requires good measures of these variables, and this is complicated by the broad cross-cultural nature of the data. It may not be reasonable to suggest that education or income has the same role in all groups. Nevertheless, it is reasonable to ask if such variables have powerful effects on game behavior in any significant proportion of groups sampled. If so, then it will be more reasonable to infer that across-group variation may be a function of the same variables. Previous attempts

*Department of Anthropology and Graduate Groups in Ecology and Animal Behavior, University of California–Davis, Davis CA 95616. mcelreath@ucdavis.edu

at correlating cross-cultural game behavior with individual-level variables has met with failure, however (Henrich *et al.*, 2001; Henrich *et al.*, 2004). Two problems arise in these analyses. First, many of the groups sampled had inadequately measured variables comparable to those of other groups, making it unclear if education or income or wealth in different samples measured the same econometrically important variable. Second, most of the societies previously sampled contained very little within-group variation in most economic variables. Thus effects, when absent, may be driven by lack of sufficient variation or, when present, by only a few outliers.

In this chapter, I present the results of Ultimatum, Dictator and Third-Party Punishment games in a semi-urban African society in Southwest Tanzania. Instead of interpreting the significance of game behavior itself, I focus on what variables may be able to explain variation within the sample. This society is potentially a good place to look for important individual-level variables, because there is impressive variation within a single community in income, wealth, and market integration. In addition to having adequate variation to detect strong effects of these variables, I perform bootstrapped regressions in order to address the potential problem of measuring highly-significant but spurious effects due to outliers and the non-normal nature of essentially every variable measured—dependent and independent. I find that individual-level variables do a poor job of explaining game behavior, either in terms of individual effects of variables or in overall fit. Effects that are powerful in ordinary regression almost always evaporate after bootstrapping.

2 Ethnographic setting

The village of Isanga is a sprawling community in the Southern Highland region of Tanzania. It is a mile from the regional capital of Mbeya, a center of trade and commerce for the southwestern portion of the country. The total population of Isanga is estimated by the government administration at approximately 8000 individuals, including children, who comprise a little more than half the total.

The physical environment in Isanga is steep and very seasonal. At a little over 4000 feet of elevation, the temperature can reach freezing in early summer nights, but become warm—in the low 80 degrees Fahrenheit—during the day, as at 8 degrees south of the equator, solar radiation is intense. The region receives good rainfall most years, but still experiences occasional dry years.

2.1 Ethnic diversity

Most of the residents of Isanga are ethnically Nyakyusa, a Bantu group originating in a territory just south of Mbeya (Wilson, 1951). The Nyakyusa are traditionally high altitude banana cultivators who keep small numbers of stabled cattle. Since the colonial period, they have been thoroughly Christianized and eager to pursue educational opportunity. Many Nyakyusa are now successful business people, although many are subsistence farmers. The Nyakyusa are well-known among cultural anthropologists for their (now historical) practice of age-villages, in which a generation of men from each village would leave their natal homes and found together a new village, to which they would bring wives when they eventually married. Young men would often return home to sleep when such villages were first formed, but later stayed in the new age-village. Elders suggested to anthroplogists Godfrey and Monica Wilson (Wilson, 1951) that this pattern of settlement was important for building strong ties among men as they matured, leading to a strong village community. Age-villages were possible only under conditions of land availability, however, and the practice has not occurred since the 1890's.

There are individuals of many other ethnic groups living and working in Isanga. Most are other groups originating in the southern highlands (Safwa being the group whose ethnic homeland is the region surrounding Isanga) (Shorter, 1972), but there are also substantial numbers of Hehe, originating in the central highlands of Tanzania, as well as small numbers of individuals originating from all over the country. This makes the ethnic picture of Isanga one of a core population of Nyakyusa, who comprise almost 50 percent of my household census, mixed in with members of many other ethnic groups. As is typical in Tanzanian communities, ethnic relations are friendly. Most individuals exhibit a preference for interaction and marriage within their own ethnic group, but friendships and marriages that cross ethnic categories are common. Some groups, such as the Nyasa from the southern border with Malawi, are almost exclusively endogamous, however. Nyasa individuals will often travel to Malawi to find a spouse and then return to live in Tanzania.

The Safwa, who mostly reside in Upper Isanga at higher elevations, are an interesting case, because they are the nominal founders of communities in this area. As the Nyakyusa have increased in number, residents say that the Safwa have retreated up the mountain and now have less influence in the community. Members of other ethnic groups commonly told me that the Safwa were contemptuous of the Nyakyusa and increasingly spiteful towards other immigrant ethnic groups.

2.2 Economic diversity

Isanga is an economically diverse community. Most residents farm, at least in a quarter-acre garden, and many also participate in household production of items (including clothing, charcoal, firewood, woven mats, and furniture) for sale in village or regional markets. Others rely upon wage labor, as schoolteachers, nurses, doctors, and government employees. Others run shops or rely exclusively upon the buying and selling of goods. A small fraction provide services such as hair styling and braiding, both of which are quite profitable. Many individuals also keep domestic animals, including chicken, geese, goats, cattle (both European dairy cattle and Asian/African zebu), pigs, and turkeys. Trade specialization is the norm, and no household is close to being self-sufficient.

2.3 Cooperation

Cooperation within Isanga is principally structured by patrilineal kinship. Most economic activities are carried out by members of nuclear families, and cooperation beyond these units extends first to close patrilineal relatives in other households. This means that sisters who share a father may help one another in planting, harvesting, and market activities.

While the bulk of day-to-day cooperative activity is kinship structured in these ways, there are occasional conspicuous demonstrations of wider sharing and cooperation. Neighbors extending as far as a dozen households outward are able to request small gifts of flour, salt, sugar and labor at any time. These requests, according to my own observations and many statements by community residents, are never refused. "Only a hyena," according to one informant, "would turn away a neighbor."

Larger scale non-dyadic cooperation occurs through church events. Congregations, of which there are several, collectively build and renovate churches and mosques. Collections are taken, as well, to fund clerks and leaders of these congregations. Beyond religious congregations, schools, roads and clinics are usually built by volunteer community labor. The government or an outside company purchases raw materials for the construction and household provide labor. Participation in these volunteer projects is apparently widespread—a majority of households I questioned directly about the most recent school foundation construction reported sending at least one person to participate.

3 Data collection

The core of the data collection methodology is described in the summary chapter. Here I present details relevant only to my field site.

A challenge in conducting economic games in natural communities lies in minimizing contamination as participants leaving the game spread information about it with other members of the community who might play on later days. In a large village like Isanga, it is possible to play in different regions of the village on different days, reducing the likelihood of such contamination. I played the DG and SMUG in three sub-villages of Isanga: (1) Isanga propoer, (2) a portion lying East across a small stream, and (3) a portion north and higher in elevation. In each of these regions, I played one day of games. Participants in each area were almost always ignorant of the game, evidenced both by their statements and their lack of knowledge when tested.

Participants were recruited for the games by generating a list of residents and randomly selecting 20 in a given sub-village. These 20 were then found in person a day ahead of the actual game play and asked to attend the games the next day, in a specified location within the village. A local resident administrator made these requests, which lead to very high attendance. When people failed to show up, the chairperson recruited replacements immediately by finding available adults. This happened in the case of the last six participants in the TPP.

Figure 1 summarizes the final composition of the participants, by ethnic identification and counts in each game. The tree diagram in this figure represents the local conception of the similarities among the different ethnic groups (*makabila*). I derived this tree by a card sort task with a sample of four residents of four different ethnic groups (Nyakyusa, Sangu, Hehe and Nyamwezi). Each individual was given a separate stack of note cards representing the ethnic groups within the sample. Each was instructed to sort the cards into piles of similar ethnic groups, by whatever criteria they thought appropriate. I asked each participant to first sort the cards into the smallest number of groups they could. Then within each group, I asked them to sub-divide, when possible, and card stacks could not be sub-divided any further. This resulted in very similar taxonomies among the four, but with some variation. I then resolved this variation into the tree in Figure 1 by having the four discuss their different taxonomies and arrive at a consensus. They were able to do so very quickly. This resulted in four recognizable families of tribes, as well as two ethnic groups none of the individuals felt was very similar to any of the others. This tree structure is certainly not a phylogeny, either genetic or cultural. However, it does capture the locally perceived degrees of similarity, in economic and cultural aspects, of the different ethnic groups.

I provided cakes and soda to waiting participants, who were separated into two groups in two rooms, those waiting to play and those who had already played. An assistant stayed with each group, ensuring they did not communicate with one another.

I used 100 Shilling coins to illustrate all game examples, as well as to elicit offers and rejection/punishment schedules. The total stake size was 1000 Tanzanian Shillings, a little less than USD 1, at the time of the games, represented as ten 100 Shilling coins. Each game was explained using a prepared sheet of paper with regions representing the different players. I then transferred coins from the stack to illustrate game play and its consequences.

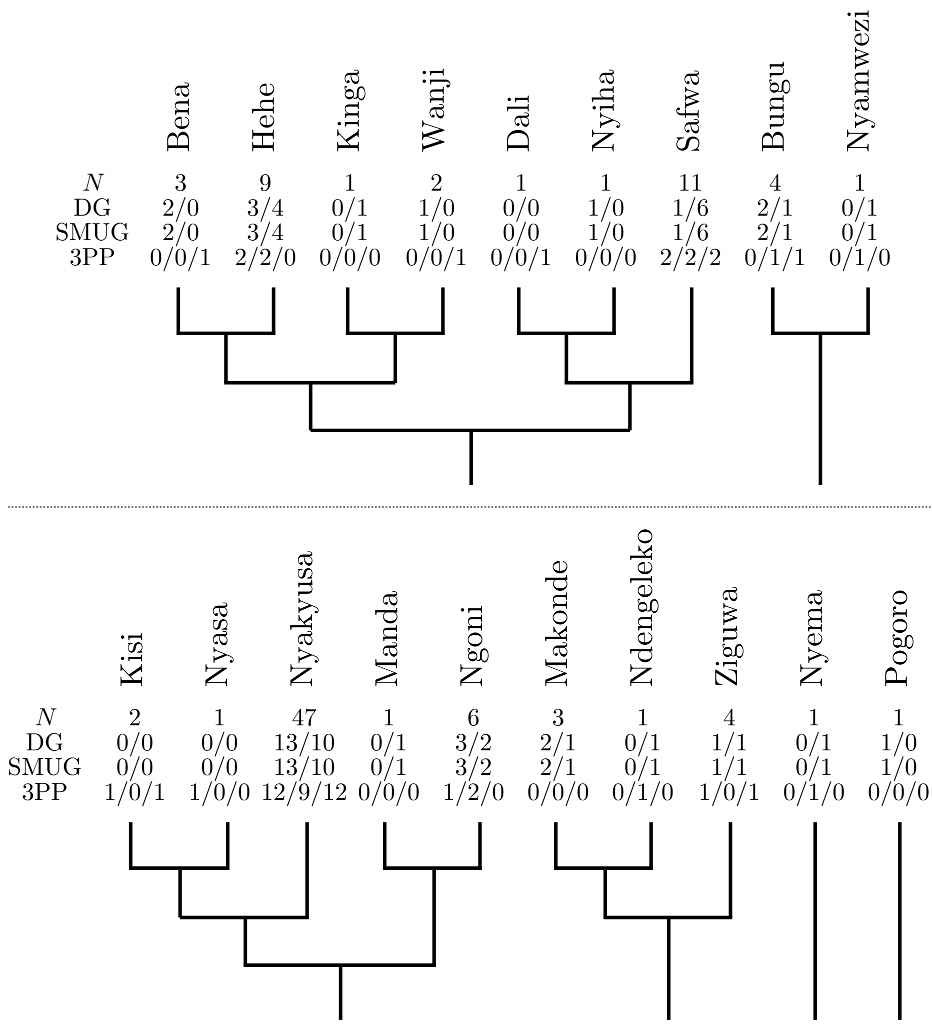

Figure 1: The game participants, divided by ethnic affiliation. The tree diagram shows the local conceptualization of ethnic group similarities. The counts above each branch tip are, in order, the count of participants from each ethnic group and the counts in the dictator game, strategy method ultimatum game, and third-party punishment game. Slashes separate counts of different player types. For example, for the TPP, $x/y/z$ indicates x player ones, y player twos, and z player threes. Since individuals played multiple games, counts for the games do not sum to the count of participants for each ethnic group.

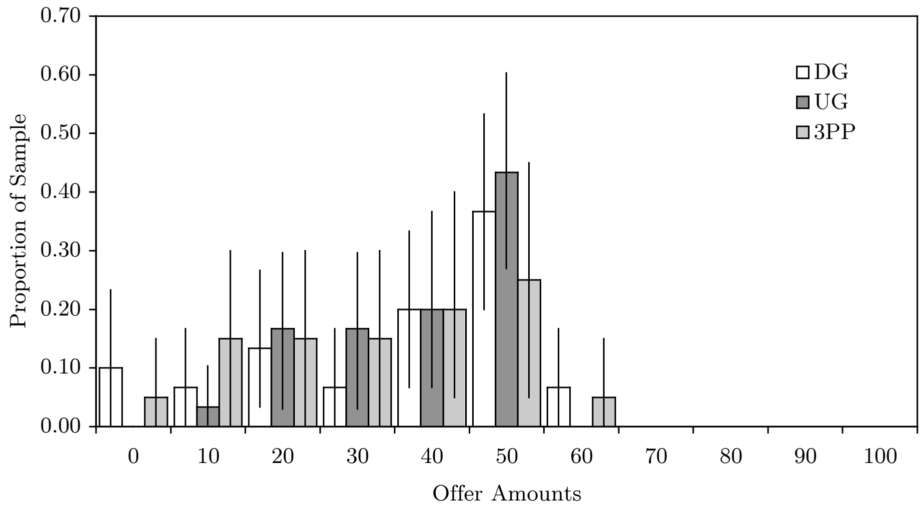

Figure 2: Offer amounts in the Dictator, Ultimatum, and Third-Party Punishment games. Error bars indicate 95% non-parametric bootstrap confidence intervals.

4 Results

In this section, I present the results of the experiments and multivariate analyses of individual player behavior. First I show an overall description of offers and responses in each game. Then I present a series of regressions of these data against individual level explanatory variables.

4.1 Overall results

Figure 2 shows the distributions of offers (as a percent of total stakes) in the Dictator (DG), Ultimatum (UG) and Third-Party Punishment (TPP) games. DG and UG offer distributions are very similar. The mean offer in the DG is 35.67 percent, and the mean UG offer is 38.33 percent. The UG mean is slightly higher, as is commonly observed in other experimental populations. The mean offer in the TPP is the lowest, at 32.50 percent.

The error bars in Figure 2 plot the 95% non-parametric bootstrap confidence intervals for each offer amount, in each game. These intervals are derived by resampling with replacement from the offer distribution in each game, 10-thousand times. This generates 10-thousand offer distributions. The interval for each offer amount is then the range of frequencies for that offer that were not in the 2.5% tails of the distribution of resampled offer distributions. In all cases, these non-parametric intervals are almost exactly the same as conventional parametric intervals computed from the F-distribution ($\mu \pm 1.96 \times \sigma$). The non-parametric intervals, however, never fall below 0 or above 1, since the values to be estimated are proportions.

While there is substantial overlap between the games, note that the observed means in each case are still the best estimates of the population offer distributions. The point of these confidence intervals is to quantify the precision of the estimates, given the precise nature of the sample. They are not intended to provide significance tests, since null-hypotheses that the games have the same offer distributions or that different offers are equally likely within each game are not reasonable, given the existing experimental literature.

Figure 3 plots the distributions of responses, punishment schedules (UG) and punishment schedules (TPP). Error bars again show 95% non-parametric bootstrap confidence intervals. The rejection profiles in the UG show high probabilities of rejection for the lowest offers, but the probabilities quickly approach zero. One individual insisted that offers above 50% of the pie should also be rejected, resulting in the small probabilities of rejection above 50%.

The punishment profile for the TPP shows much higher probabilities, across all offer amounts less than 50%. The best estimate of the probability of being punished when offering 40% of the pie is 0.25. The probabilities within the 95% confidence interval for all offers below this are all above 0.25, and in each case the best estimate is at least 0.50.

The expected income derived from offers are shown by the trend lines in Figure 3. The income maximizing offers in the UG and TPP are 10% and 50%, respectively. Comparing these values to the modes of each offer distribution, the TPP most common offer is the same as the income maximizing offer. The expected payoff is very flat from an offer of zero to an offer of 50%, however. The UG mode is substantially higher.

4.2 Regression analyses

I performed a series of regressions to estimate the importance of a series of individual-level variables on both offer and response (rejection and punishment) behavior. For the offers, I used percent offered to the second player in the DG, UG and TPP as a dependent variable. For responses, I used the smallest offer the second player accepted (the Minimum Acceptable Offer, MAO) in the UG and the smallest offer the third player did not fine (the Minimum Unfined Offer, MUO) in the TPP.

These regressions model both the percents offered and the MAO and MUO as a continuous measures with normally distributed error. This is certainly not correct: percent offer is a discrete distribution in these games (11 levels), and the observed pattern of offers in each game demonstrate that error is not symmetrical. Also, many of the independent variables I will present are highly non-normal (Figures 4 and 5). Transformations cannot help, because the mode in several important cases is at the minimum or maximum value (for example, wealth in Figure 4). Still, while the precise parameter estimates and p-values will be suspect, the direction and rough magnitude of effects may still be useful. In order to estimate how robust any measured effect is, I bootstrapped [using the `boot` library in `R` (R Development Core Team, 2004)] each regression ten-thousand times to generate non-parametric confidence intervals for each parameter in each model. An effect that remains strong in the bootstrap generalization is robust to variations in the exact data used in the regression. In the result tables for each regression, I indicate the bootstrap p-values by daggers trailing the standard error of each original estimate.

I begin by presenting a basic model in which the dependent variables are functions of age (years), sex (male = 0, female = 1), education (years schooling), individual annual income (in USD), household wealth (in USD), and household size (adults and children) (Table 1). This model is the foundation of the regressions to come.

The majority of model effects are small and imprecise. Only age in the UG remains significant (large and precisely estimated) after bootstrapping. This effect is actually quite modest in explaining variation: the adjusted R^2 for the UG model is only 5%. The moderate positive effect of income on the TPP MUO vanishes after bootstrapping, and the stronger negative effects of education and household wealth also both vanish in the bootstrap.

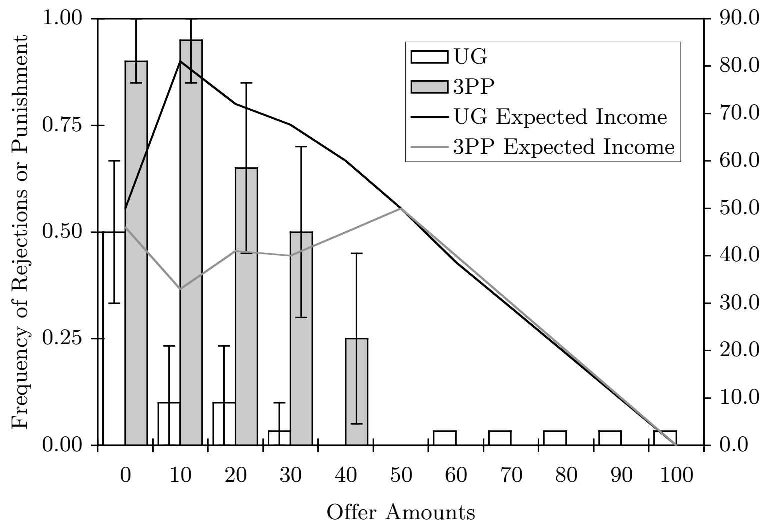

Figure 3: Frequency of rejection (UG) and punishment (TPP), plotted on the left axis, as well as expected profit for each offer amount, plotted on the right axis. Error bars indicate 95% non-parametric bootstrap confidence intervals.

Table 1: Regressions of six standard variables against DG, UG and TPP offers, as well as UG Minimum Acceptable offer (MinAO) and TPP Minimum Unpunished offer (MinUO). All coefficients are standardized. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Daggers (†) after the standard error indicate an estimate was significant in a bootstrap test of robustness (see main text).

Variable	DG offer Std β (Std Err)	UG offer Std β (Std Err)	TPP offer Std β (Std Err)	UG MinAO Std β (Std Err)	TPP MinUO Std β (Std Err)
age	0.1081	0.3587*	0.1610	−0.0306	0.1967
	(0.2109)	(0.1986)†	(0.2569)	(0.2469)	(0.2096)
female	0.3577	0.3049	−0.2316	−0.0376	−0.0485
	(0.2163)	(0.2038)	(0.2793)	(0.2164)	(0.1748)
education	0.1080	0.2954	−0.3270	0.0946	−0.6271**
	(0.2136)	(0.2012)	(0.3001)	(0.2146)	(0.2289)
income (USD)	0.1867	0.0582	0.1292	−0.0407	0.3568*
	(0.2147)	(0.2022)	(0.2591)	(0.2059)	(0.1666)
wealth (USD)	0.0730	−0.0861	−0.1690	−0.0742	−0.4953**
	(0.2562)	(0.2414)	(0.3305)	(0.2187)	(0.2041)
household size	−0.0523	0.2465	−0.1492	−0.3217	−0.2417
	(0.2044)	(0.1925)	(0.2666)	(0.2051)	(0.2366)
Observations	30	30	20	30	20
Adjusted R^2	−0.0653	0.0547	−0.0404	−0.0711	0.5661

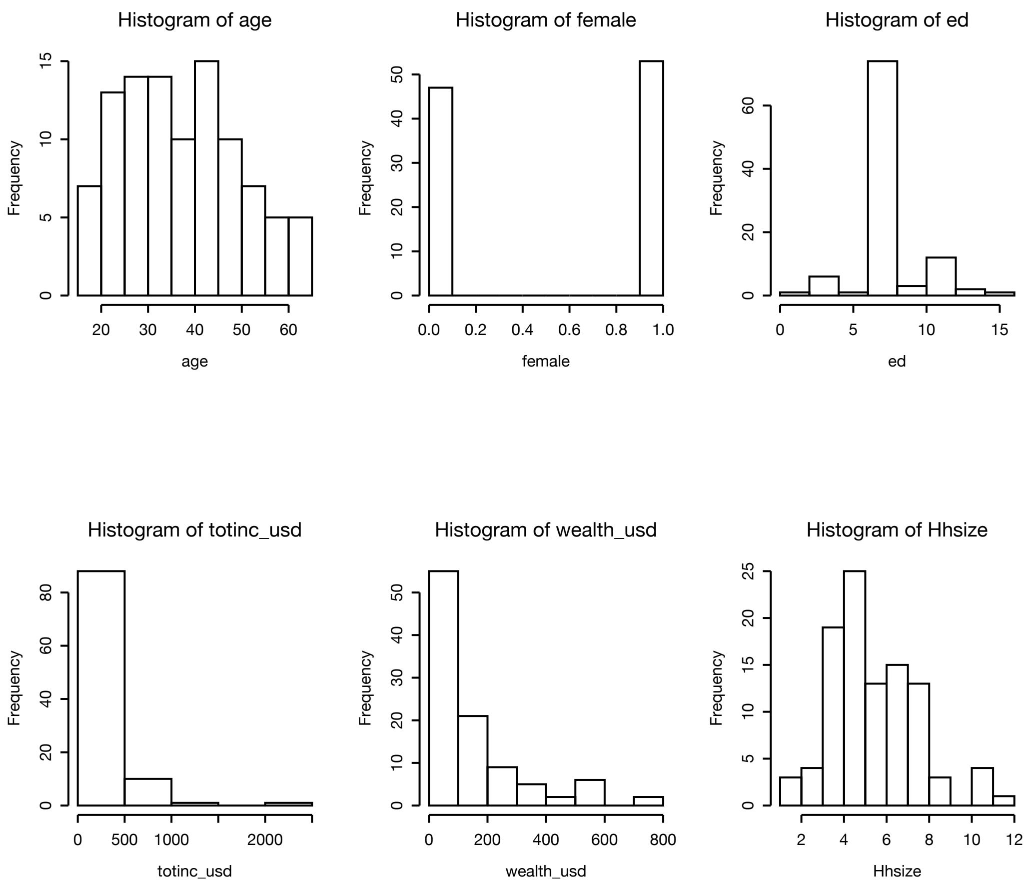

Figure 4: Distributions of five basic independent variables used in the regressions.

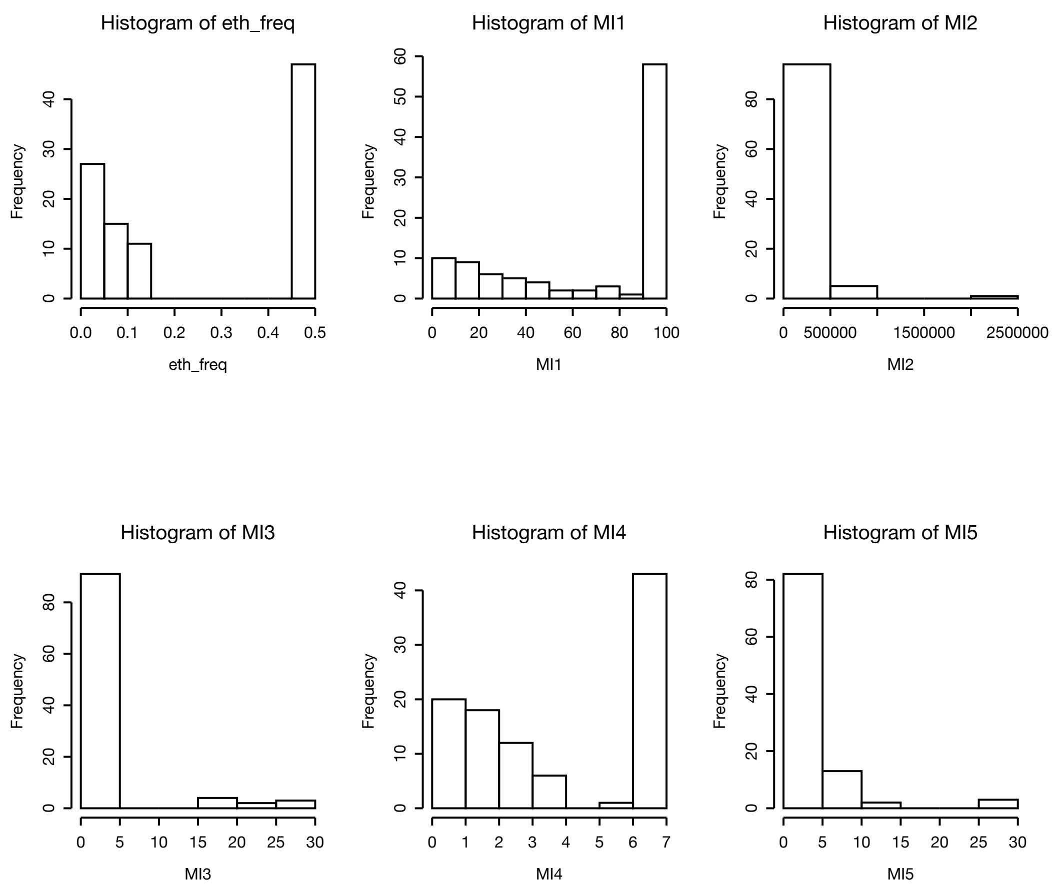

Figure 5: Distributions of independent variables ethnic group frequency and the five market integration variables, MI1-MI5.

4.3 Ethnicity effects

In an attempt to estimate the role of ethnic variation (see Figure 1), I added to the basic models in Table 1 a set of model effects coding ethnicity of the participants. There are too many different individual ethnic groups to hope to estimate reliable parameters for each. Instead, I grouped ethnic labels by the major groups in Figure 1. There are four major groups: the Southern Highlands (SH) groups (Nyakyusa etc.), the neighboring Central Highlands (CH) groups (Hehe, Safwa, etc.), a Nymawezi (NY) culture group, and a Makonde (MA) culture group. The remaining ethnic groups (Pogoro and Nyema) were omitted from this coding scheme. I also included ethnic identity not as a categorical variable, but as the frequency of individuals of each ethnicity present in the total sample. These frequencies are estimates of the overall community frequencies of each ethnic group. One might predict that more common ethnic groups, placed in an apparently anonymous game with a random member of their community, would offer more, owing to favoritism of co-ethnics and an increased chance of interacting with a co-ethnic when one is a member of a common ethnic group. The widespread ethnographic observation of in-group biases in cooperation (Levine & Campbell, 1972) and the theoretical plausibility of the stability of such behavior (McElreath *et al.*, 2003) suggests that it is worthwhile to look for ethnic group effects of these kinds.

Controlling for ethnic identity, several of the control variables become more predictive. For UG offers, female, education and household size are now sizable positive effects that survive the bootstrap. Ethnic frequency is also a very powerful effect for UG offers, but this effect depends upon a small number of individuals from rare ethnic groups, and so the effect flattens in the bootstrap regressions. Effects elsewhere in the table are not robust in bootstrap, except for household size (UG MAO, negative) and household wealth (TPP MUO, negative).

4.4 Market integration effects

Since market integration turned out to be explanatory across societies in the first round of games, we developed five measures of individual market integration that we might use to explain variation in offers within societies. Table 3 shows the pairwise correlations among these measures, as well as their definitions. In some cases, the pairwise correlations are moderate, but never above 0.50. Individuals who often buy and sell for profit (high MI5) also rely upon the market for their food (higher MI1), have more income from wages etc. (higher MI2), and make more trips to the market (higher MI4). Predictably, individuals who do more wage labor (high MI3) have more income from such sources (higher MI2).

To investigate the effect of these variables on offers in the three games, I regressed the five MI variables against offers and responses (Table 4). Overall, market integration variables do a poor job of explaining variation in offers and responses, whether alone or controlling for demographic and wealth effects. While MI4 and MI have strong estimated effects for TPP offers, these effects are not at all robust, vanishing in the bootstrap.

4.5 Results summary

To summarize these regressions, no variable offers to explain a large portion of the variance in offers. Several have consistent but weak effects across the games, and there is weak evidence that families of ethnic groups differ slightly in their offers. Individual wealth and income are inconsistently related to offers, and even have different estimated directions of effect in the same game. Market

Table 2: Regressions of six standard and five ethnicity variables against DG, UG and TPP offers, as well as UG Minimum Acceptable offer (MinAO) and TPP Minimum Unpunished offer (MinUO). All coefficients are standardized. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Daggers (†) after the standard error indicate an estimate was significant in a bootstrap test of robustness (see main text).

Variable	DG offer Std β (Std Err)	UG offer Std β (Std Err)	TPP offer Std β (Std Err)	UG MinAO Std β (Std Err)	TPP MinUO Std β (Std Err)
age	0.1700	0.3439	0.2266	−0.0093	−0.0624
	(0.2441)	(0.2059)	(0.3007)	(0.2647)	(0.2494)
female	0.1199	0.4697*	−0.2993	−0.1382	−0.2887
	(0.2676)	(0.2257)†	(0.3352)	(0.2401)	(0.2095)
education	0.0559	0.5071**	−0.1603	−0.0894	−0.5094*
	(0.2392)	(0.2018)†	(0.4086)	(0.2367)	(0.2411)
income (USD)	0.1154	0.0032	0.0518	−0.1139	0.3075*
	(0.2248)	(0.1896)	(0.4880)	(0.2099)	(0.1644)
wealth (USD)	0.0472	−0.2196	−0.2485	−0.2654	−0.7019**
	(0.2753)	(0.2322)	(0.4592)	(0.2490)	(0.2505)†
household size	−0.3298	0.5368**	−0.0837	−0.5983**	−0.0371
	(0.2801)	(0.2362)†	(0.3251)	0.2439†	(0.2672)
eth−freq	−0.7142	0.8401**	−0.1428	−0.3671	0.2262
	(0.4380)	(0.3694)	(0.6555)	(0.3381)	(0.3632)
eth−branchCH	0.6438	0.2924	−0.2189	1.4420**	−0.0436
	(0.5778)	(0.4873)	(0.5177)	(0.6608)	(0.3023)
eth−branchMA	0.4626	0.5546	0.2062	0.8446*	−0.1867
	(0.3956)	(0.3336)	(0.4176)	(0.4342)	(0.1969)
eth−branchNY	0.6539	0.2032		0.5080	0.2836
	(0.4229)	(0.3567)		(0.3607)	(0.2232)
eth−branchSH	1.3350	−0.4637		1.6040**	
	(0.8302)	(0.7002)		(0.7571)	
Observations	30	30	20	30	20
Adj R−squared	−0.07533	0.2351	−0.1988	−0.06169	0.6062

Table 3: Pairwise correlations between the different measures of market integration. MI1 ≡ percent of diet (calories) purchased; MI2 ≡ income from wage labor, rental, or trade (Shillings); MI3 ≡ frequency (days) of wage labor in previous month; MI4 = trips to any market in the last week; MI5 ≡ frequency of purchasing goods expressly for resale in the last month.

	MI1	MI2	MI3	MI4	MI5
MI1	–	0.0974	-0.0020	0.1891	0.2481
MI2	0.0974	–	0.3233	0.1542	0.4658
MI3	-0.0020	0.3233	–	-0.0985	-0.0790
MI4	0.1891	0.1542	-0.0985	–	0.3040
MI5	0.2481	0.4658	-0.0790	0.3040	–

Table 4: Regressions using market integration variables. All coefficients are standardized. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Daggers (†) after the standard error indicate an estimate was significant in a bootstrap test of robustness (see main text).

Variable	DG offer Std β (Std Err)	UG offer Std β (Std Err)	TPP offer Std β (Std Err)	UG MinAO Std β (Std Err)	TPP MinUO Std β (Std Err)
age	0.1872	0.2841	−0.0041	−0.1848	0.4812
	(0.2527)	(0.2266)	(0.3240)	(0.2623)	(0.2839)
female	0.2828	0.4006	−0.4267	−0.1098	0.6301
	(0.2956)	(0.2651)	(0.2389)	(0.2123)	(0.3740)
education	0.1504	0.5089	−0.5173	−0.0479	−0.7932**
	(0.4437)	(0.3980)	(0.2856)	(0.2114)	(0.3045)
income (USD)	0.0646	0.2103	1.6640	−0.3856	0.5232**
	(0.2745)	(0.2462)	(1.6730)	(0.4075)	(0.2043)
wealth (USD)	−0.0708	−0.2700	−0.4908	0.2964	−0.2607
	(0.3707)	(0.3325)	(0.3812)	(0.2897)	(0.2619)
household size	−0.1721	0.4917	−0.2014	−0.1611	−0.5805
	(0.3677)	(0.3297)	(0.2615)	(0.2047)	(0.3943)
MI1	−0.3886	0.1871	−0.5161	0.4547	0.0508
	(0.2936)	(0.2633)	(0.3471)	(0.2842)	(0.2258)
MI2	0.1367	0.0513	−1.7180	1.0730*	0.0590
	(0.3090)	(0.2771)	(1.7700)	(0.5718)	(0.3093)
MI3	−0.0725	−0.3030	0.0516	−0.4915	−0.2967
	(0.5002)	(0.4486)	(0.3020)	(0.4769)	(0.2619)
MI4	0.1812	−0.1752	−0.6239**	−0.0678	−0.5110
	(0.3177)	(0.2849)	(0.2403)	(0.2824)	(0.3260)
MI5	0.0139	−0.3555	0.8936**	−0.5423*	−0.5774
	(0.2496)	(0.2239)	(0.3060)	(0.2918)	(0.3117)
Observations	30	30	20	30	20
Adjusted R^2	−0.1958	0.03823	0.3414	0.1002	0.5627

integration variables are inconsistent in the direction and their effects, and none are powerful explainers of variation in the bootstrap.

5 Discussion and conclusion

Despite considerable effort in collecting plausible explanatory variables, few consistent and powerful effects explaining variation in either offers or responses (rejection and fining schedules) were found in these data. Income and wealth have inconsistent effects in their direction across games, and in only one case (TPP MNO) did wealth have a strong effect in all of the regressions. Market integration measures similarly either do a poor job of explaining variation or are inconsistently related to offers, across games.

In light if these results, it seems unlikely that the standard explanatory variables (age, sex, income, etc.) and market integration measures can explain much of the variation in game behavior. Given the likely imprecision in measuring market integration, it is of course possible that measurement error in the independent variables is masking their effects. It is also possible that MI measures could have different effects in different games, but supporting this finding demands replicating the direction and magnitude of the effects in other data sets. A basic problem in these data is that there are many interesting and intuitively plausible explanatory variables to include in the models. With 30 observations per game, the number of interesting explanatory variables can in fact approach the number of observations. Many of the interesting explanatory variables are also inter-correlated, sometimes substantially. These two facts together make it dangerous to simply try out every plausible effect and report significant or sizable effects. By chance alone, a number of these variables will correlate strongly with offers or response schedules, without any reliable casual relationship.

Thus it is important to compare the direction and magnitude of model effect estimates across study sites, in an effort to validate the explanatory robustness of these estimates. It is always tempting, especially for a field anthropologist, to devise a clever and appealing explanation of any significant result found in a regression. Given the usual small number of observations and potentially vast number of explanatory variables, significant effects are guaranteed. Finding that the effects exist in other places talks to the explanatory power of those variables, while post hoc explanations of their causal effects in a specific population have comparatively little value to general understandings of behavior. This is not to say that unique effects do not exist. They most certainly do, and understanding the cross-population interactions among variables is a necessary component of the larger understanding of cooperative behavior that this group is developing. However, it is important to be cautious of developing special explanations of the results in each population as an alternative to building general frameworks to explain experimental game behavior cross-culturally.

References

Camerer, C. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton: Princeton University Press.

Fehr, E., & Gachter, S. 2002. Altruistic Punishment in Humans. *Nature*, **415**, 137–140.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (eds). 2004. *Foundations of*

Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies. Oxford: Oxford University Press.

Henrich, Joseph, Boyd, Robert, Bowles, Samuel, Camerer, Colin, Fehr, Ernst, Gintis, Herb, & McElreath, Richard. 2001. In search of Homo oeconomicus: Behavioral experiments in 15 small scale societies. *American Economic Review*, **91**, 73–78.

Hoffman, E., McCabe, K., & Smith, V. 1998. Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology. *Economic Inquiry*, **36**, 335–352.

Levine, Robert A., & Campbell, Donald T. 1972. *Ethnocentrism: Theories of Conflict, Ethnic Attitudes, and Group Behavior.* New York: John Wiley & Sons, Inc.

McElreath, Richard, Boyd, Robert, & Richerson, Peter J. 2003. Shared norms and the evolution of ethnic markers. *Current Anthropology*, **44**(1), 122–129.

R Development Core Team. 2004. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.

Roth, A. E. 1995. *The handbook of experimental economics*. Princeton: Princeton University Press. Chap. Bargaining Experiments, pages 253–248.

Shorter, A. 1972. *Chiefship in Western Tanzania*. Oxford: Clarendon Press.

Wilson, Monica. 1951. *Good Company: A study of Nyakyusa age-villages.* Oxford: Oxford University Press.