

# Chapter 15

---

## Cooperation and Punishment in an Economically Diverse Community in Highland Tanzania

Richard McElreath

Previous cross-cultural variation in economic experiments was unexplained by individual-level variables such as wealth, income, gender, and education, even though variation 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 contained 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.

A tremendous amount of data now exist supporting the conclusions that human behavior is poorly predicted by standard economic preferences (Camerer 2003; Fehr and Gächter 2002; Hoffman et al. 1998; Roth 1995). In addition, as more non-Western and nonstudent 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 the 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 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, because many of the groups sampled had inadequately measured variables comparable to those of other groups, it was unclear whether 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 the game behavior itself, I focus on the variables that 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, and so it is possible to detect strong effects of these variables. I also perform bootstrapped regressions 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, both dependent and independent. I find that individual-level variables do a poor job of explaining game behavior, whether in terms of the individual effects of variables or in overall fit. Effects that are powerful in ordinary regression almost always evaporate after bootstrapping.

## THE 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 eight thousand 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 four thousand feet of elevation, the temperature can reach freezing on early summer nights, but become warm—in the low eighties (Fahrenheit)—during the day, since solar radiation is intense at eight degrees south of the equator. The region receives good rainfall most years, but still experiences occasional dry years.

### 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 businesspeople, 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 would later stay in the new age-village. Elders suggested to anthropologists 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 1890s.

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; see 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 comprised 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, however, such as the Nyasa from the southern border with Malawi, are almost exclusively endogamous. 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 toward other immigrant ethnic groups.

## Economic Diversity

Isanga is an economically diverse community. Most residents farm, in at least 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 on wage labor, as schoolteachers, nurses, doctors, and government employees. Some residents run shops or rely exclusively on the buying and selling of goods. A small fraction provide services such as hairstyling 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.

## 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.

Although 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, or 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 nondyadic 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 households provide the 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.

## DATA COLLECTION

The core of the data collection methodology is described in the summary chapter (chapter 18, available at: <http://www.russellsage.org/Ensminger>). 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 to 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 dictator game (DG) and the strategy method ultimatum game (UG) in three subvillages of Isanga: Isanga proper, a portion of the village lying east across a small stream, and a portion lying 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 by both their statements and their lack of knowledge when tested.

I recruited participants for the games by generating a list of residents and randomly selecting twenty in a given subvillage. These twenty 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 led to very high attendance. When people failed to show up, the administrator recruited replacements immediately by finding available adults. This happened in the case of the last six participants in the third-party punishment game (TPG).

Figure 15.1 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 subdivide, when possible, and told them that card stacks could not be subdivided 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 15.1 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 that none of the individuals felt were 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 the economic and cultural similarities of the different ethnic groups.

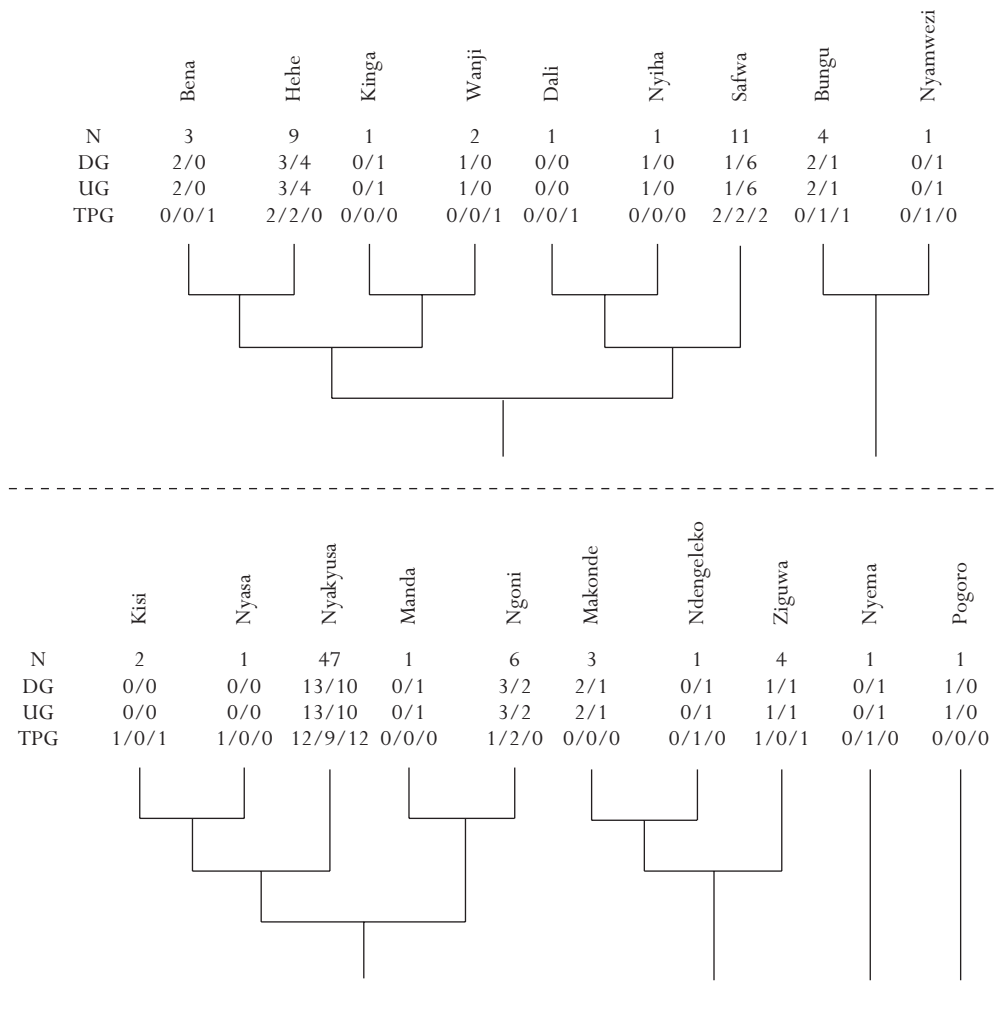
I provided cakes and soda to waiting participants, who were separated into two groups in two rooms, those waiting to play and those who had already played. An assistant stayed with each group, ensuring that they did not communicate with one another.

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 1,000 Tanzanian shillings, a little less than U.S. \$1.00 at the time of the games, represented as ten 100-shilling coins. I explained each game 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.

## RESULTS

In this section, I present the results of the experiments and multivariate analyses of individual player behavior. First I provide 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.

FIGURE 15.1 Game Participants in Isanga, Divided by Ethnic Affiliation



Source: Author's compilation based on author data.

Note: 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 TPG,  $x/y/z$  indicates  $x$  player 1s,  $y$  player 2s, and  $z$  player 3s. Since individuals played multiple games, counts for the games do not sum to the count of participants for each ethnic group.

## Overall Results

Figure 15.2 shows the distributions of offers (as a percentage of total stakes) in the dictator game, ultimatum game, and third-party punishment game. 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 TPG is the lowest, at 32.50 percent.

The error bars in Figure 15.2 plot the 95 percent nonparametric 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, ten thousand times. This generates ten 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 percent tails of the distribution of resampled offer distributions. In all cases, these nonparametric intervals are almost exactly the same as conventional parametric intervals computed from the  $F$ -distribution ( $\mu \pm 1.96\sigma$ ). The nonparametric intervals, however, never fall below zero or above one, since the values to be estimated are proportions.

Although 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 15.3 plots the distributions of responses and punishment schedules for the UG and the TPG. Error bars again show 95 percent nonparametric 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 percent of the stake should also be rejected, resulting in the small probabilities of rejection above 50 percent.

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

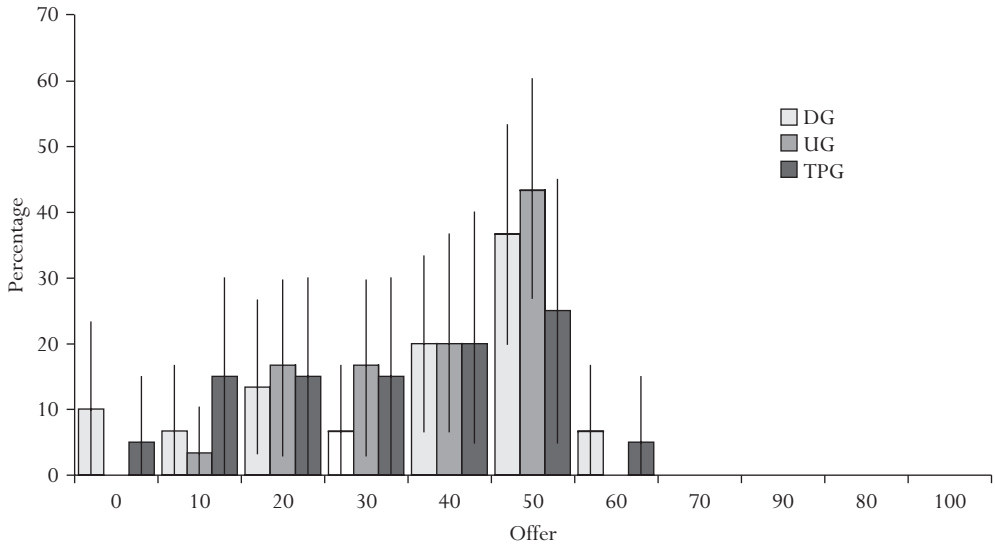
The expected income derived from offers is shown by the trend lines in figure 15.3. The income-maximizing offers (IMOs) in the UG and TPG are 10 percent and 50 percent, respectively. Comparing these values to the modes of each offer distribution, the most common offer in the TPG is the same as the IMO. The expected payoff is very flat, however, from an offer of zero percent to an offer of 50 percent. The UG mode is substantially higher.

## 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 percentage offered to player 2 in the DG, UG, and TPG as a dependent variable. For responses, I used the smallest offer that player 2 accepted—the minimum acceptable offer (MinAO)—in the UG and the smallest offer that 3 player did not fine—the minimum unfined offer (MUO)—in the TPG.

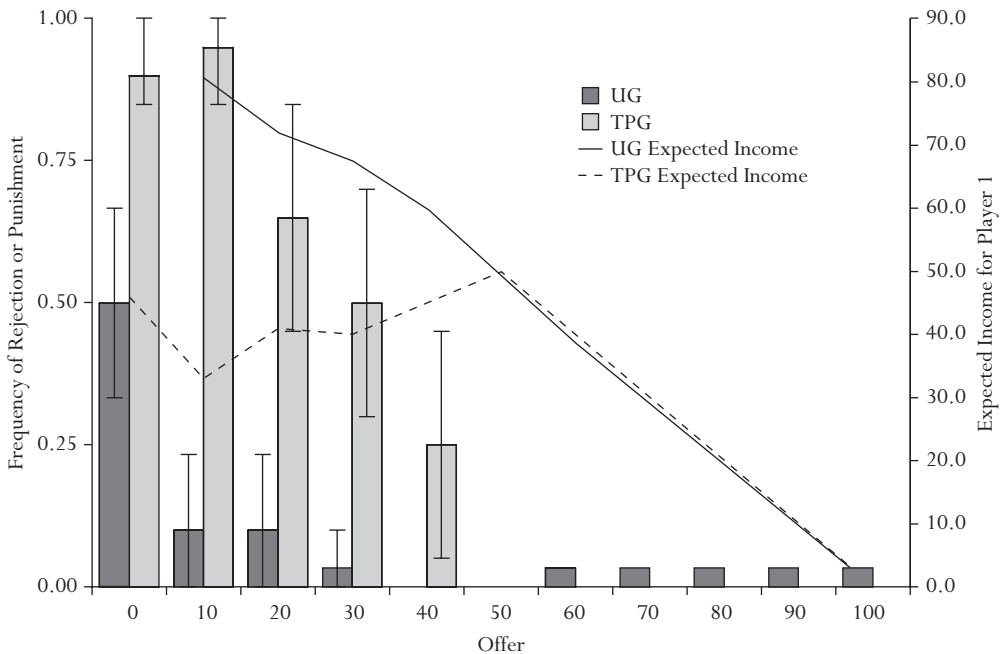
These regressions model both the percentage offered and the MinAO and MUO as continuous measures with normally distributed error. This is certainly not correct: percentage offer is a discrete distribution in these games (eleven levels), and the observed pattern of offers in each game demonstrates that error is not symmetrical. Also, many of the independent variables that I present are highly non-normal (figures 15.4 and 15.5). Transformations cannot help, because

FIGURE 15.2 *Isanga Offer Amounts in the Dictator, Ultimatum, and Third-Party Punishment Games*



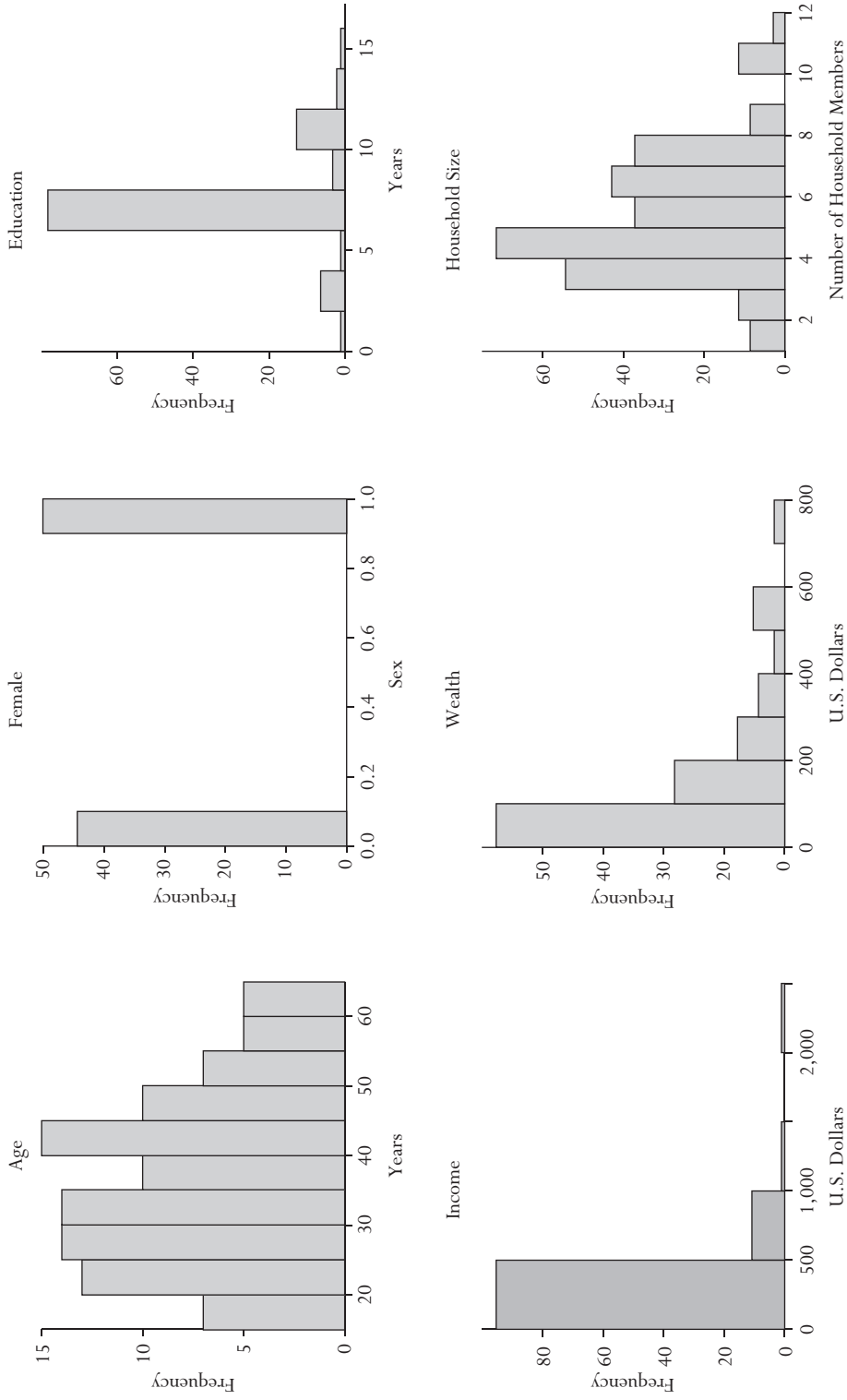
Source: Author's compilation based on author data.  
 Note: Error bars indicate 95 percent nonparametric bootstrap confidence intervals.

FIGURE 15.3 *Frequency of Rejection (UG) and Punishment (TPG) in Isanga Sample and Expected Income for Player 1s*



Source: Author's compilation based on author data.  
 Note: Error bars indicate 95 percent nonparametric bootstrap confidence intervals.

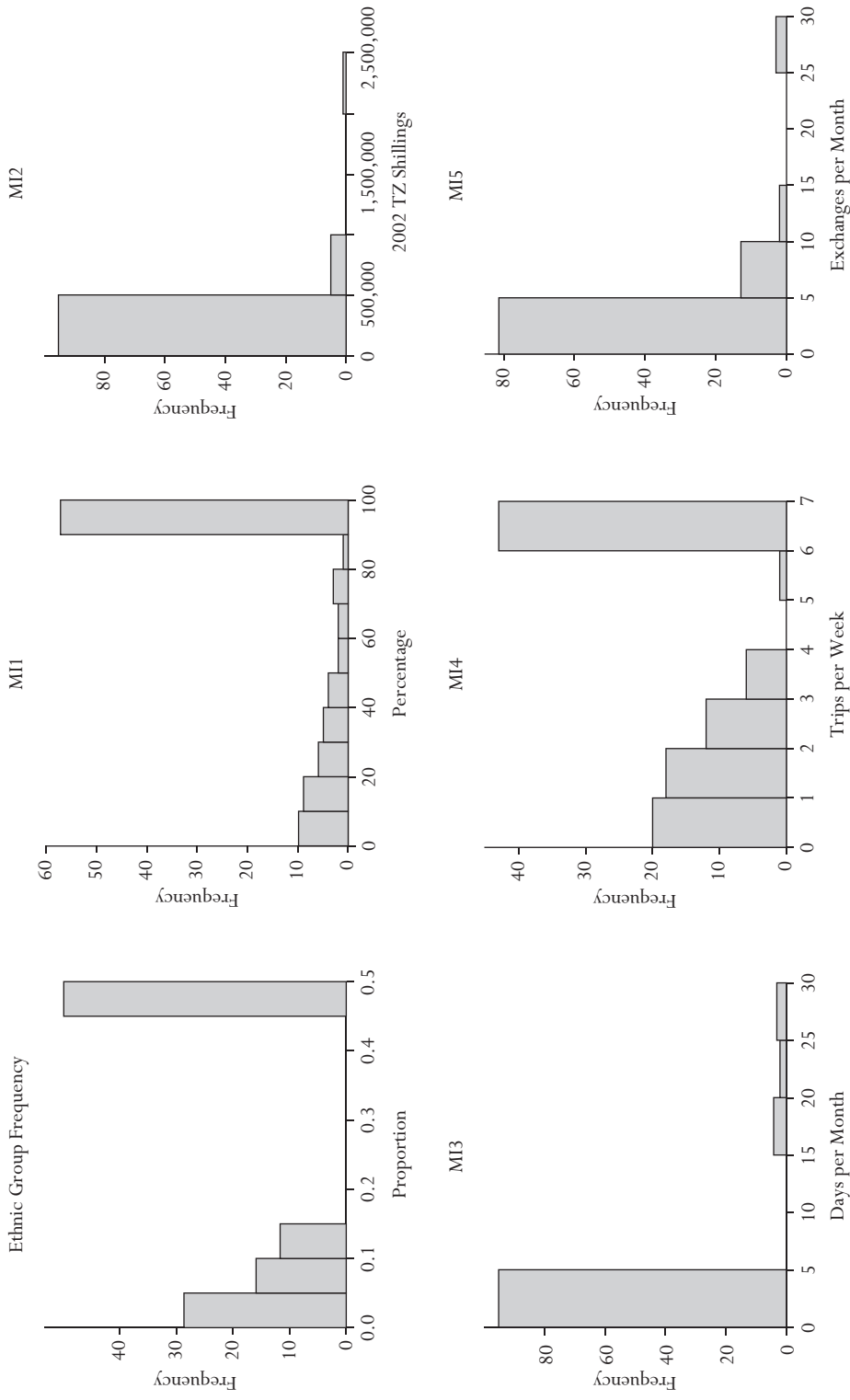
FIGURE 15.4 Distributions of Five Basic Independent Variables Used in the Regressions



Source: Author's compilation based on author data.



FIGURE 15.5 Distributions of Independent Variables, Ethnic Group Frequency, and the Five Market Integration Variables



Source: Author's compilation based on author data.  
Notes: All coefficients are standardized.

TABLE 15.1 *Regressions of Six Standard Variables Against Dictator Game, Ultimatum Game, and Third-Party Punishment Game Offers, UG Minimum Acceptable Offers, and TPG Minimum Unpunished Offers*

Variable	DG Offer Standard $\beta$ (Standard Error)	UG Offer Standard $\beta$ (Standard Error)	TPG Offer Standard $\beta$ (Standard Error)	UG MinAO Standard $\beta$ (Standard Error)	TPG MUO Standard $\beta$ (Standard Error)
Age	0.1081 (0.2109)	0.3587* (0.1986)†	0.1610 (0.2569)	-0.0306 (0.2469)	0.1967 (0.2096)
Female	0.3577 (0.2163)	0.3049 (0.2038)	-0.2316 (0.2793)	-0.0376 (0.2164)	-0.0485 (0.1748)
Education	0.1080 (0.2136)	0.2954 (0.2012)	-0.3270 (0.3001)	0.0946 (0.2146)	-0.6271** (0.2289)
Income (U.S. dollars)	0.1867 (0.2147)	0.0582 (0.2022)	0.1292 (0.2591)	-0.0407 (0.2059)	0.3568* (0.1666)
Wealth (U.S. dollars)	0.0730 (0.2562)	-0.0861 (0.2414)	-0.1690 (0.3305)	-0.0742 (0.2187)	-0.4953** (0.2041)
Household Size	-0.0523 (0.2044)	0.2465 (0.1925)	-0.1492 (0.2666)	-0.3217 (0.2051)	-0.2417 (0.2366)
Observations	30	30	20	30	20
Adjusted R-squared	-0.0653	0.0547	-0.0404	-0.0711	0.5661

Source: Author's compilation based on author data.

Notes: All coefficients are standardized.

\*  $p < 0.10$ ; \*\*  $p < 0.05$ ; † after the standard error, indicates that an estimate was significant in a bootstrap test of robustness

the mode in several important cases is at the minimum or maximum value (for example, wealth in figure 15.4). Still, while the precise parameter estimates and  $p$ -values are suspect, the direction and rough magnitude of effects may still be useful. To estimate how robust any measured effect is, I bootstrapped (using the boot library in R; see R Development Core Team 2011) each regression ten thousand times to generate nonparametric 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 results 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 of schooling), individual annual income (in U.S. dollars), household wealth (in U.S. dollars), and household size (number of adults and children) (see table 15.1). 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-squared for the UG model is only 5 percent. The moderate positive effect of income on the TPG MUO vanishes after bootstrapping, and the stronger negative effects of education and household wealth also both vanish in the bootstrap.

## Ethnicity Effects

In an attempt to estimate the role of ethnic variation (see figure 15.1), I added to the basic models in table 15.1 a set of model effects coding the ethnicity of the participants (table 15.2). There

TABLE 15.2 *Regressions of Six Standard and Five Ethnicity Variables Against Dictator Game, Ultimatum Game, and Third-Party Punishment Game Offers, UG Minimum Acceptable Offers, and TPG Minimum Unpunished Offers*

Variable	DG Offer Standard $\beta$ (Standard Error)	UG Offer Standard $\beta$ (Standard Error)	TPG Offer Standard $\beta$ (Standard Error)	UG MinAO Standard $\beta$ (Standard Error)	TPG MUO Standard $\beta$ (Standard Error)
Age	0.1700 (0.2441)	0.3439 (0.2059)	0.2266 (0.3007)	-0.0093 (0.2647)	-0.0624 (0.2494)
Female	0.1199 (0.2676)	0.4697* (0.2257)†	-0.2993 (0.3352)	-0.1382 (0.2401)	-0.2887 (0.2095)
Education	0.0559 (0.2392)	0.5071** (0.2018)†	-0.1603 (0.4086)	-0.0894 (0.2367)	-0.5094* (0.2411)
Income (U.S. dollars)	0.1154 (0.2248)	0.0032 (0.1896)	0.0518 (0.4880)	-0.1139 (0.2099)	0.3075* (0.1644)
Wealth (U.S. dollars)	0.0472 (0.2753)	-0.2196 (0.2322)	-0.2485 (0.4592)	-0.2654 (0.2490)	-0.7019** (0.2505)†
Household Size	-0.3298 (0.2801)	0.5368** (0.2362)†	-0.0837 (0.3251)	-0.5983** 0.2439†	-0.0371 (0.2672)
Ethnic group frequency	-0.7142 (0.4380)	0.8401** (0.3694)	-0.1428 (0.6555)	-0.3671 (0.3381)	0.2262 (0.3632)
Ethnic branch: CH	0.6438 (0.5778)	0.2924 (0.4873)	-0.2189 (0.5177)	1.4420** (0.6608)	-0.0436 (0.3023)
Ethnic branch: MA	0.4626 (0.3956)	0.5546 (0.3336)	0.2062 (0.4176)	0.8446* (0.4342)	-0.1867 (0.1969)
Ethnic branch: NY	0.6539 (0.4229)	0.2032 (0.3567)		0.5080 (0.3607)	0.2836 (0.2232)
Ethnic branch: SH	1.3350 (0.8302)	-0.4637 (0.7002)		1.6040** (0.7571)	
Observations	30	30	20	30	20
Adjusted R-squared	-0.07533	0.2351	-0.1988	-0.06169	0.6062

Source: Author's compilation based on author data.

Notes: All coefficients are standardized.

\*  $p < 0.10$ ; \*\*  $p < 0.05$ ; † after the standard error, indicates that an estimate was significant in a bootstrap test of robustness

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 15.1. There are four major groups: the southern highlands (SH) groups (Nyakyusa, and so on), the neighboring central highlands (CH) groups (Hehe, Safwa, and so on), 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 the favoritism of coethnics and an increased chance of interacting with a coethnic when one is a member of a common ethnic group. The widespread ethnographic observation of in-group biases in cooperation (Levine and Campbell 1972) and the theoretical plausibility of the stability of such behavior (McElreath, Boyd, and Richerson 2003) suggest that it is worthwhile to look for ethnic group effects of these kinds.

TABLE 15.3 *Pairwise Correlations Between Five Measures of Market Integration*

	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	—

Source: Author's calculations based on author data.

Note: MI1: percentage of diet (calories) purchased; MI2: Income from wage labor, rental, or trade (Tanzanian 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.

When we control 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 on 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 (negative effect on UG MinAO) and household wealth (negative effect on TPG MUO).

### Market Integration Effects

Since market integration (MI) 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 15.3 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 on the market for their food (higher MI1), have more income from wages, and so on (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 15.4). 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 TPG offers, these effects are not at all robust, vanishing in the bootstrap.

### Results Summary

No variable explains 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 integration variables are inconsistent in direction and in their effects, and none are powerful explainers of variation in the bootstrap.

## DISCUSSION AND CONCLUSIONS

Despite a considerable effort to collect 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

TABLE 15.4 *Regressions Using Market Integration Variables Against Dictator Game, Ultimatum Game, and Third-Party Punishment Game Offers, UG Minimum Acceptable Offers, and TPG Minimum Unpunished Offers*

Variable	DG Offer Standard $\beta$ (Standard Error)	UG Offer Standard $\beta$ (Standard Error)	TPG Offer Standard $\beta$ (Standard Error)	UG MinAO Standard $\beta$ (Standard Error)	TPG MUO Standard $\beta$ (Standard Error)
Age	0.1872 (0.2527)	0.2841 (0.2266)	-0.0041 (0.3240)	-0.1848 (0.2623)	0.4812 (0.2839)
Female	0.2828 (0.2956)	0.4006 (0.2651)	-0.4267 (0.2389)	-0.1098 (0.2123)	0.6301 (0.3740)
Education	0.1504 (0.4437)	0.5089 (0.3980)	-0.5173 (0.2856)	-0.0479 (0.2114)	-0.7932** (0.3045)
Income (U.S. dollars)	0.0646 (0.2745)	0.2103 (0.2462)	1.6640 (1.6730)	-0.3856 (0.4075)	0.5232** (0.2043)
Wealth (U.S. dollars)	-0.0708 (0.3707)	-0.2700 (0.3325)	-0.4908 (0.3812)	0.2964 (0.2897)	-0.2607 (0.2619)
Household Size	-0.1721 (0.3677)	0.4917 (0.3297)	-0.2014 (0.2615)	-0.1611 (0.2047)	-0.5805 (0.3943)
MI1	-0.3886 (0.2936)	0.1871 (0.2633)	-0.5161 (0.3471)	0.4547 (0.2842)	0.0508 (0.2258)
MI2	0.1367 (0.3090)	0.0513 (0.2771)	-1.7180 (1.7700)	1.0730* (0.5718)	0.0590 (0.3093)
MI3	-0.0725 (0.5002)	-0.3030 (0.4486)	0.0516 (0.3020)	-0.4915 (0.4769)	-0.2967 (0.2619)
MI4	0.1812 (0.3177)	-0.1752 (0.2849)	-0.6239** (0.2403)	-0.0678 (0.2824)	-0.5110 (0.3260)
MI5	0.0139 (0.2496)	-0.3555 (0.2239)	0.8936** (0.3060)	-0.5423* (0.2918)	-0.5774 (0.3117)
Observations	30	30	20	30	20
Adjusted R-squared	-0.1958	0.03823	0.3414	0.1002	0.5627

Source: Author's calculations based on author data.

Notes: All coefficients are standardized.

\*  $p < 0.10$ ; \*\*  $p < 0.05$

games, and in only one case (TPG MUO) did wealth have a strong effect in all of the regressions. Similarly, market integration measures, across games, either do a poor job of explaining variation or are inconsistently related to offers.

In light of these results, it seems unlikely that the standard explanatory variables (age, sex, income, and so on) and market integration measures can explain much of the variation in Isanga 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 support for this finding demands that the direction and magnitude of the effects be replicated 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 thirty 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 intercorrelated, 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 offer or response schedules, without any reliable causal relationship.

Thus, it is important to compare the direction and magnitude of model effect estimates across study sites if the explanatory robustness of these estimates is to be validated. 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 the Roots of Human Sociality Project is developing. It is important, however, to beware 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, Colin. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, N.J.: Princeton University Press.
- Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415: 137–40.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis, eds. 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "In Search of Homo Economicus: Behavioral Experiments in Fifteen Small-Scale Societies." *American Economic Review* 91(2): 73–78.
- Hoffman, Elizabeth, Kevin McCabe, and Vernon Smith. 1998. "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology." *Economic Inquiry* 36(3): 335–52.
- Levine, Robert A., and Donald T. Campbell. 1972. *Ethnocentrism: Theories of Conflict, Ethnic Attitudes, and Group Behavior*. New York: John Wiley & Sons.
- McElreath, Richard, Robert Boyd, and Peter J. Richerson. 2003. "Shared Norms and the Evolution of Ethnic Markers." *Current Anthropology* 44(1): 122–29.
- R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Roth, Alvin E. 1995. "Bargaining Experiments." In *The Handbook of Experimental Economics*. Princeton, N.J.: Princeton University Press.
- Shorter, Aylward. 1972. *Chieftship in Western Tanzania*. Oxford: Clarendon Press.
- Wilson, Monica. 1951. *Good Company: A Study of Nyakyusa Age-Villages*. Oxford: Oxford University Press.