

Bargaining Behavior, Demographics and Nationality:

What Can the Experimental Evidence Show?

by

Anabela Botelho, Glenn W. Harrison, Marc A. Hirsch & Elisabet E. Rutström †

August 2003

ABSTRACT

Field experiments have raised important issues of interpretation of bargaining behavior. There is evidence that bargaining behavior appears to vary across groups of populations, such as nationality, ethnicity and sex. Differences have been observed with respect to initial behavior and with respect to the learning pattern over time. Often, such behavioral differences are referred to as cultural, although the delineation of the cultural group has been confined to one or other observable characteristic in isolation. We show that this way of characterizing cultural differences is overly simplistic: at best, it leads to unreliable claims; at worst, it leads to erroneous conclusions. We reconsider the evidence provided by previous experiments using ultimatum game rules, and undertake new experiments that expand the controls for demographics. The lesson from our demonstration is that the task of designing experiments for the field offers many challenges if one wants to define and control for cultural impacts, but that this challenge is one that also faces those conducting and interpreting laboratory experiments.

† Department of Economics, University of Minho and NIMA, Portugal (Botelho); Department of Business and Economics, Murray State University (Hirsch); and Department of Economics, College of Business Administration, University of Central Florida (Harrison & Rutström). Rutström thanks the U.S. National Science Foundation for research support under grants NSF/IIS 9817518, NSF/MRI 9871019 and NSF/POWRE 9973669. Supporting data and instructions are stored at <http://www.bus.ucf.edu/gharrison/data/ee/bargaining/>. Corresponding author: E. Elisabet Rutström at ERutstrom@bus.ucf.edu.

Table of Contents

1. Introduction	-1-
2. Previous Experiments	-3-
The Ultimatum Game	-3-
Roth, Prasnikar, Okuno-Fujiwara, and Zamir	-3-
Slonim and Roth	-4-
Cameron	-6-
New Experiments	-7-
3. Statistical Issues	-8-
4. Results	-13-
Nation Effects and Learning Over Time	-14-
Do Demographics Affect Behavior?	-16-
5. Implications	-20-
References	-23-

1. Introduction

Field experiments have raised important issues of interpretation of bargaining behavior. There is evidence that bargaining behavior appears to vary across samples drawn from populations defined by nationality, ethnicity and sex. Differences have been observed both with respect to initial behavior, and with respect to the learning pattern over time. Such behavioral differences are often referred to as cultural, although the delineation of the cultural group has been confined to one or other observable characteristic in isolation. We show that this way of characterizing cultural differences is overly simplistic, and lead to unreliable claims at best and erroneous conclusions at worst. The lesson from our demonstration is that the task of designing experiments for the field offers many challenges if one wants to define and control for cultural impacts, but that field experiments also offer potential for providing new insights into these issues.

One example that highlights the issues raised here derive from experimental results that suggest that there is an effect of nationality on bargaining behavior. An alternative hypothesis for the interpretation of those findings is that bargaining behavior varies according to some other observable individual characteristics, and that the nation-effect is just a reflection of differences in samples across nations in terms of those other individual characteristics. For example, if men and women vary in the way that they bargain, and samples differ in the mix of men and women across countries, there is an obvious confound present.

We reconsider the evidence provided by previous experiments using ultimatum game bargaining rules, and undertake some new experiments that expand the controls for demographics. We show that inferences about nationality or other demographic effects are sensitive to the way in which the data are analyzed and the controls are incorporated. They are sensitive in a substantive way (bargaining behavior does appear to be affected by both nationality and other individual

characteristics and in non-separable ways), and they are sensitive in a statistical way (the assumptions required to test hypotheses about individual effects are “delicate”).

In section 2 we review three previous experiments and one new experiment that examine nationality effects directly or that are conducted in different countries. These experiments use similar procedures which makes them comparable. Three of them also included controls for demographics.¹ All are based on the experiments conducted in Japan, Israel, Yugoslavia, and the United States by Roth, Prasnikar, Okuno-Fujiwara, and Zamir [1991] (RPOZ). Although RPOZ did not collect demographic information, we also examine their data since it has been so influential.² One is an experiment conducted in the Slovak Republic by Slonim and Roth [1998]. Another is an experiment conducted in Indonesia by Cameron [1999]. The final experiment is a new one that we conducted in the United States and Russia. In section 3 we review the econometric issues involved in teasing apart the confounding effects of individual characteristics and nation, and in section 4 we examine the data from these four studies using statistical procedures that allow such a separation. In section 5 we draw conclusions regarding the dangers in defining culture narrowly in terms of simple unconditional demographic variables.

¹ We are grateful to Alvin Roth, Robert Slonim and Lisa Cameron for making their data available.

² We acknowledge that other studies have looked at bargaining effects relating to demographic characteristics, but do not analyze them here. Henrich, Boyd, Bowles, Camerer, Fehr, Gintis, and McElreath [2001] examine ultimatum bargaining behavior in 15 “exotic” cultures located in 12 countries, and include information on demographics when statistically analyzing behavior. Croson and Buchan [1999] control for sex in their cross-country analysis of behavior in experimental “trust” games, but no other demographics. They do find that sex and country effects interact, consistent with our conclusion, but this would be more robust if they had additionally controlled for other demographics. Other studies which examine cross-country effects *or* demographic effects in standard experimental settings include Andreoni and Vesterlund [2001], Bolton and Katok [1995], Bolton and Zwick [1995], Brown-Kruse and Hummels [1993], Burlando and Hey [1997], Cadsby and Maynes [1998], Eckel and Grossman [1996a][1996b][1998][2001], Nowell and Tinkler [1994], Saijo and Nakamura [1995], Schweitzer and Solnick [1999], Solnick [2001], Willer and Szmataka [1993], Yamagishi [1998a][1998b] and Yamagishi, Cook and Watabe [1998].

2. Previous Experiments

The Ultimatum Game

In the ultimatum game one of two players proposes a split of a fixed monetary pie, and the other player may either accept or reject the proposed split. If the second player accepts the proposal, the payoffs to each are determined by the proposed split. If the second player rejects the proposal, they each get nothing. The subgame perfect equilibrium prediction is for the first player to propose a split that gives him almost 100% of the pie, and for the second player to accept the proposal since any positive offer beats a zero payoff for a player that is not satiated in money. The experimental data consistently shows that the average offer to the second player is substantially greater than predicted, and that the second player often rejects small offers.³

These stylized observations lead to the popular hypothesis that there exists some uncontrolled element in individual utility, and that individuals care about the payoffs of other players as well as their own payoffs. One motivation behind multinational tests of the ultimatum game is the possibility that such “other-regarding” preferences are culturally determined, and that behavior therefore may vary across nations since we intuitively expect culture to vary across nations.⁴

Roth, Prasnikar, Okuno-Fujiwara, and Zamir

RPOZ ran a series of carefully designed ultimatum games in Japan, Israel, Yugoslavia, and the United States. They claim that their data shows significant behavioral differences between

³ See Güth and Tietz [1990] and Harrison and McCabe [1996] for reviews of the empirical findings.

⁴ We use the expression “other-regarding” to remain agnostic as to whether or not the underlying cause of this behavior is altruistic preferences, reciprocity norms, or some other factor such as simple confusion.

subject pools across these nations. Specifically, they concluded that groups in the United States and Yugoslavia displayed the usual experimental results of a modal 50-50 split, but that the groups from Japan and Israel were closer to a 60-40 split. They also found that the propensity to reject lower offers was significantly lower in Japan and Israel.⁵

Three sessions were conducted in each country. In each of Israel, Japan and Yugoslavia only one experimenter was used for all sessions. Each of these three experimenters conducted one session in the United States. Thus it is possible to identify experimenter effects as well as country effects.

No data were collected on the age, sex or other demographics of the individual participants. However, some sessions did have identified differences in subject pool. Two differences are noted in RPOZ (p.1075): in one of the Israeli sessions, and in one of the Yugoslav sessions.

Each subject participated in a session lasting 10 rounds. Each subject faced a new, randomly selected opponent in each round. Thus the data consist of balanced panels of individuals responding over each of 10 rounds.

Slonim and Roth

The Slonim and Roth [1998] experiments with the ultimatum game were conducted in the Slovak Republic in 1994, using procedures that were identical to those employed in by RPOZ. Bargaining was over a pie worth 60 Slovak Crowns (Sk) in one session, a pie worth 300 Sk in

⁵ Unless otherwise stated, all statements about acceptance or rejection rates are conditional on the level of the offer.

another session, and a pie worth 1500 Sk in a third session.⁶ At exchange rates to the U.S. dollar prevailing at the time, these stakes were \$1.90, \$9.70 and \$48.40, respectively. In terms of average local monthly wages, they were equivalent to approximately 2.5 hours, 12.5 hours and 62.5 hours of work, respectively.

The lowest stake level is extremely low by conventional standards in bargaining experiments, and are close to being non-salient. The medium stake level is virtually identical to the standard pie size in most experiments. Hence one could view the lowest stake level as akin to a “hypothetical” payment condition, the medium stake level as akin to the control with other experiments, and the highest stake level as really interesting treatment. We examine behavior in both of the higher stake conditions.

The Slovak Republic experiments consisted of a “practice round” followed by ten rounds. The subjects were paid for one of the ten rounds, chosen at random. Opponents were determined at random each round. The use of repetition is a factor stressed by Slonim and Roth [1998] as central to their ability to detect differences in behavior:

Consistent with prior results, changes in stakes had only a small effect on play for inexperienced players. But the present experimental design allows us to observe that rejections were less frequent the higher the stakes, and proposals in the high stakes conditions declined slowly as subjects gained experience. The Slovak experiment is the first to detect a lower frequency of rejection when stakes are higher and this can be explained by the added power due to multiple observations per subject in the experimental design.

Although not documented in Slonim and Roth [1998], the experimenters collected information from individuals on a range of demographic characteristics. Robert Slonim kindly

⁶ Actually, the subjects bargained over points which were simply converted to currency at different exchange rates. This procedure seems transparent enough, and served to avoid possible focal points defined over differing cardinal ranges of currency.

provided the raw data, which we collated and merged with the data on bargaining behavior responses reported in Slonim and Roth [1998; p.592ff.]. Information was collected on sex, age, field of study enrolled in, whether they had a job, personal income in the previous year, family income range, and the number of family members living with them.

Overall there were 164 subjects, broken down equally of course between the two player types of the Ultimatum game. There were 48, 66 and 50 subjects in the low, medium and high stakes sessions, respectively. If we drop the practice round data, there are 10 observations on virtually all subjects.⁷

Cameron

The experiments conducted by Cameron [1999] offer an opportunity to check conclusions about the effects of stakes on initial behavior. She conducted three sessions with real payoffs, each consisting of two rounds.⁸ In the first round the subjects bargained over 5,000 Indonesian Rupiah (Rp), which was about \$2.50 at prevailing exchange rates to the U.S. dollar in 1994. In the second round the stakes were Rp 5,000, Rp 40,000 and Rp 200,000, respectively, in Games 1, 2 and 3. The highest stake was about \$100, or roughly three times the average monthly expenditure of the student subjects. She also collected information on basic individual demographic characteristics, including sex, religion, cultural background in terms of geography, urban or rural origin, and approximate monthly expenditure level. Lisa Cameron kindly provided these

⁷ Two subjects participated for 11 rounds and 2 subjects for 9 rounds, possibly due to a mix-up in subject ID.

⁸ She also conducted a fourth session with hypothetical payoffs, albeit with a salient show-up fee. The statistical methods we employ to evaluate the data with real payoffs can be extended to evaluate the extent of hypothetical bias in these games. We find evidence of such bias, and prefer to avoid interpreting hypothetical payoffs as just “extremely low real payoffs.”

unpublished data on individual demographics.

New Experiments

We undertook a series of ultimatum game experiments in Russia and the United States in order to test for the effect of demographic variables in addition to country effects. Sixty subjects were recruited from the student population at Moscow Institute of Electronics Technology (MIET). Most of these students were business students at the Zelenograd Business College at MIET. There were two sessions, one in November 1994 and one in March 1995. Each session included 30 subjects. In each session half of the subjects were designated buyers (making offers) and half sellers (accepting or rejecting offers). Subjects made decisions in 5 consecutive bargaining rounds, maintaining their designation as buyers or sellers but playing against different, anonymous opponents in each round. At the end of the experiment one of the rounds was selected at random to determine actual payments. The buyer/seller designation was private information throughout the experiment. Subjects were paid 7000 Rubles for participating and they bargained over 14,000 Rubles in each round during the first session. In the second session subjects were paid 8,000 Rubles for participating and bargained over 16,000 Rubles.⁹

In the United States the same procedures were used in three sessions of 20 subjects each for a total of 60 subjects. These subjects were recruited from the University of South Carolina (USC) and paid \$5 for participating while bargaining over \$10.¹⁰ The same experimenter (Hirsch)

⁹ The amounts were chosen based on comparative purchasing power for a student in either Russia or the United States. The values were meant to be large enough to purchase two reasonable student lunches at a university cafeteria. While the Ruble was devalued significantly over this time period, the price of an average student lunch at the university had not changed as much.

¹⁰ Each session was conducted in a regular classroom where there was plenty of room for subjects to spread out for privacy. Subjects were given a folder which contained all the instructions and the message forms. The language in the instructions used terms like “buyers” and “sellers,” rather than “Senders” and

conducted all experiments, so there should be no experimenter effects across sessions.

We collected those individual characteristics which have been deemed basic for a wide range of general-purpose economic surveys, such as sex, race/ethnicity, age, educational level and income with no further claim that this is either a necessary or sufficient set of controls. One could always add to such lists, but it seemed prudent to ensure that we minimally controlled for these basic demographic characteristics. We accept that these characteristics might just serve as markers for other individual characteristics that could be measured, such as risk aversion, in more elaborate experimental designs.

3. Statistical Issues

Before plunging into the data, it is important to consider the alternative ways to statistically analyze the data. Our goal is to see how much of the observed bargaining behavior is associated with individual demographics, national effects, treatment effects, and their interaction. Limited sample sizes may restrict the ability to consider interaction effects in their full glory, but that is something that depends on the particular data considered.

One statistical issue concerns the specification of time, since one major substantive concern is the extent to which learning paths differ across nations and treatments. We agree here with List and Cherry [2000; p.19/20], who argue that one should use dummy variables for each

“Responders.” Proposals were formulated in terms of number of “tokens,” each of which had the same value to both players. The total number of tokens that could be divided up between the two players in each round was 1000. After the first players had made their proposals, the forms were collected, collated, and handed back to their partners. In order to keep the designation private, we collected and handed back forms to all players every time we went around the room. The player who was not making a decision was asked to report a guess of what decision his partner was making. All players went through a practice round together before starting. The sessions lasted approximately 1¼ hours. The time required for each session varied slightly, based on the subjects’ understanding of the game, the level of difficulty in filling out the required demographic questionnaire, and the size and structure of the classroom in which the experiment was held.

bargaining round in order to remain agnostic about the functional form of the time path of effects. A single trend variable implies that learning behavior is constant in each round, which may not be true.

One drawback of this agnostic stance on the effect of time is that it will be difficult to consider interaction effects with demographics or nationality, due to sample size limitations.

An important statistical issue is the way in which “unobserved individual effects” are characterized. This term can be confusing, but is important to understand in order to sort out the econometric alternatives. An “observed individual effect” is any effect associated with an explanatory variable that varies at the level of the individual *and that is observed* by the experimenter, such as sex or age. Such effects are conventionally defined over variables that are constant for the individual over time.¹¹ An unobserved individual effect is anything else that is correlated with the individual but that is not observed. This individual effect arises because *something is observed*, the individual that generated the observation, and we ought to be able to use that information. Thus the confusion stems from the mixture of observed and unobserved information in an “individual effect.”

One way to handle an unobserved individual effect is to assume *a priori* that it does not exist. Essentially, this approach assumes that everything that one wants to know about the individual is captured in the observed individual effect, or is of no interest (e.g., because theory says it is not). In this case it can be shown that the usual pooled estimator is appropriate, but this is simply because one has assumed *a priori* that the omitted variables implicitly captured by the unobserved individual effect are not statistically relevant. This is a strong assumption, but one

¹¹ Variables that vary over time *and* individuals present little difficulties.

that is not as bad as it might first appear.¹² Some unobserved variable may be relevant in the sense of having an effect on observed behavior, but be correlated with some variable that is observed and hence allowed for, such that it might be *statistically* irrelevant. Moreover, some of the popular ways of handling unobserved individual effects require that the effect be completely uncorrelated with observed variables that are included. This is just an extension of the usual requirement that the error term be uncorrelated with the explanatory variables.

Another way to handle an unobserved individual effect is to use a “fixed effects” specification, such as employed by List and Cherry [2000] in their analysis of some ultimatum experiments.¹³ In this case the unobserved individual effect is assumed to be captured by a fixed individual dummy. This might seem to solve the problem nicely, except for one very unfortunate fact: it is then impossible to include observed individual effects that vary across individuals but are constant for each individual. The reason is that such effects, such as the sex or age of a given individual, are perfectly correlated with the fixed effect dummy for that individual. If no individual observed effects were collected by the experimenter, as in List and Cherry [2000], RPOZ or the published version of Slonim and Roth [1998], such an approach would be appropriate. But it is not appropriate if one has collected observables at the level of the individual and wants to use them, as we do here to be able to identify the source of differences in bargaining behavior across nations.

¹² If the alternative specification is the random effects model, simple tests for the presence of the unobserved individual effect are available in panel probit and panel OLS settings. In the context of panel probit (or logit) a likelihood-ratio test that the panel estimator is the same as the pooled estimator is available (e.g., Wooldridge [2002; p.486]), and in the context of panel OLS the Breusch-Pagan test evaluates the null hypothesis that the variance of the unobserved effect is zero (e.g., Wooldridge [2002; p.264]).

¹³ This is also known as the “within estimator” in panel settings. It is well-known in the case of continuous dependant variables, and is covered in detail by Wooldridge [2000; chapter 10]. In the case of binary dependant variables, the approach was developed by Chamberlin [1980].

The popular alternative in such situations is the “random effects” specification. The good news with such a specification is that one can, under certain assumptions, identify the effects of the individual characteristics while accounting for the fact that each individual might have some distinct unobservable effect on the dependant variable. The bad news is that the certain assumptions might not be correct. Specifically, the unobserved individual effect must be uncorrelated with the observed individual effect (and other explanatory variables). Again, if there are no observed individual effects, due to the information not being collected, there is no major issue here.

Assume, however, that some individual characteristics are observed and included, and are of inferential interest as they are here. How can one test if the additional assumptions of the random effects specification are correct? One possibility is to use the Hausman [1978] test, which compares the results from an estimator that is known to be consistent and the results from an estimator that is efficient under the same assumptions. The null hypothesis that the latter estimator is also consistent can then be tested. In the present case, the fixed effects estimator is a consistent estimator of some variables (e.g., treatment effects) even when the no-correlation assumption of the random effects estimator is violated, and the random effects estimator is efficient and consistent when the no-correlation assumption is valid. The Hausman test examines the estimates of the coefficients for which both estimators should produce consistent estimates under the null.

Unfortunately, if the Hausman test leads to a rejection of the random effects specification used, it might not be due to the random effects specification *per se*.¹⁴ It could also be due to some

¹⁴ Another unfortunate feature of applying the Hausman test in practice is that it often fails due to the estimated covariance matrix from finite samples differing from their asymptotic properties under the maintained assumptions of the test. Specifically, if the estimated covariance matrix of the efficient estimator

other mis-specification of the (fixed effect and random effects) models. For example, if one assumed a linear relationship instead of the true non-linear relationship, a Hausman test of the random effects specification might reject it, but this could be due to the mis-specification of functional form.

One constructive solution when the Hausman test rejects the random effects specification is to use the comparison of estimates to guide a re-specification of the analysis (e.g., StataCorp [2003b; p. 202ff.]. The danger with such specification searches is that they are *ad hoc*, and need to be undertaken explicitly so that the reader understands the sequence used. Nonetheless, they can provide useful diagnostics, particularly for the design of new experiments with additional controls.

Another constructive solution when the Hausman test rejects the random effects specification is to use the instrumental variables methods proposed by Hausman and Taylor [1981] and Amemiya and MaCurdy [1986]. In this context, these methods use the exogenous regressors that are not suspected of being correlated with the unobserved individual effect to be instruments for the observed individual effects. Thus one might use treatment effects (e.g., pie size) that are applied exogenously to instrument the variables (e.g., sex, race, age) suspected of being correlated with the unobserved effects. The generic problem with these methods is that the available instruments might be weak, in the sense of not being correlated sufficiently with the variables they are to instrument (Baltagi and Khanti-Akom [1990]). Furthermore, these procedures only apply to continuous dependent variables.¹⁵

does not attain its asymptotic Cramer-Rao bound, the variance of the difference in the estimates can be negative (see Hausman [1976; Corollary 2.6, p.1254]). There exist generalizations of the Hausman test, which employ pooled estimators of the variance of the difference in the original estimators and numerically avoid this problem. However, such asymptotic tests must be applied particularly carefully even if they are numerically feasible, since they are usually only needed when finite sample issues are most important.

¹⁵ Of course, one could treat the discrete dependent variable as if it were continuous. Providing that the average binary response is not too close to 0 and 1, such an approach will often produce reliable

The upshot of this discussion is that one cannot easily avoid random effects specifications if the inferential goal is to say something about the relative contribution of national effects, treatment effects, and observed individual effects on bargaining behavior, while also accounting for possible unobserved individual effects. We adopt these specifications, and report tests of their validity whenever possible.¹⁶

4. Results

We refer to the results of each series of experiments by acronym. These are RPOZ for Roth, Prasnikar, Okuno-Fujiwara, and Zamir [1991], SR for Slonim and Roth [1998], LC for Cameron [1999], and BHHR for our new experiments. Detailed statistical results are collected with the data documentation.¹⁷ We consider data restricted to offers that are 50% or less, since offers in excess of 50% indicate some likely confusion with the game and are very few in number. Any individual or joint effects reported below are statistically significant at the 10% level or better, unless otherwise noted.

Figures 1 through 4 displays the raw behavior observed in these experiments over time, pooled over the country or stake treatments within each design. The left panels show offers and acceptances in round 1, and the right panels show what happened in later rounds. The lines in the right panels are predictions from a simple quadratic fit, and the shaded areas are the associated 95% confidence intervals. These displays provide a useful descriptive context as we examine each

estimates of marginal effects (Wooldridge [2002; p.454ff.]). Unfortunately, in this application the average responses are often close to 0 or 1.

¹⁶ We also encourage further analyses of these data if and when more powerful econometric techniques become available. One by-product of our effort is that the relevant data will be collected (as of mid-September 2003) in one consistent location for replication and extension: the Web-Lab Experimental Social Sciences Digital Archive (WESDA), located at <http://weblab.bus.ucf.edu>.

¹⁷ All statistical results were generated using version 8 of *Stata*, documented in StataCorp [2003].

study in detail.

Nation Effects and Learning Over Time

The experiments of RPOZ exhibit some differences in learning patterns across nations. Table 1 lists the marginal effects of controls on acceptance probabilities based on a random effects logit specification. This specification normalizes acceptances to round 1 in the US, so the country dummies (Israel, Japan, and Yugoslav) show the effect of the change in country in round 1 behavior. The numerical dummies associated with each country (e.g., US_2, Israel_2, etc) show the effect of each period interacted with the country.

Acceptances in Israel start out 15.2 percentage points higher than the US, with a significance level of 7.9%. They generally remain above the US acceptance rates throughout all rounds. The learning pattern in the US is cyclic, in the sense that acceptance rates tend to decline in early rounds, albeit with considerable noise, but increase by 10 percentage points in round 8. The initial decline in acceptances that is observed in the US is not observed in Israel or in Yugoslavia. In Israel, this means that the initial difference in acceptances compared to the US is exacerbated over time. In Yugoslavia the learning pattern diverges in a significant way from the US pattern during rounds 4 and 5, after starting out at a slightly lower level. The learning pattern in Japan does not appear to differ from that in the US.

Learning behavior with respect to offers in the RPOZ experiments are similar to those of acceptances, at least in the sense that we see some different patterns across nations. Table 2 displays the estimates in this case, again normalized in the same manner as in Table 1.¹⁸ In the US

¹⁸ A Hausman test of the random-effects specification fails numerically because the asymptotic assumptions underlying it are violated, but inspection of the common coefficient estimates from the fixed effects and random effects specifications makes it apparent that there are no significant differences in the two

there is a significant decline in percentage offers in rounds 3 and 4, of roughly 5 percentage points, but very little change after that. Thus it appears that the low and declining acceptance rates have a negative effect on offers over time in the US. Israel again displays a difference compared to the US in initial round behavior, with offers 7 percentage points lower than in the US. Contrary to the US, however, there is not much change over time. The regression coefficients displayed in Table 2 simply reflect the difference in the time path for the two nations. Thus, even though acceptances are higher in Israel, this does not appear to cause a decline in offers over time. Offers in Japan start out below those in the US by about 5 percentage points, although this effect is only significant at the 11.3% level. Similar to the learning pattern in the US, the Japanese offers become more generous around the middle of the experiment. In Yugoslavia, the offers are initially not different from those in the US, but the learning pattern is somewhat different since there is a decline (rather than an increase) in offers around the middle of the experiment.

The pattern that emerges here is in one sense quite simple: learning behavior and nationality interact in not-so-simple ways. The behavioral paths differ across countries in terms of initial versus learned effects. No single, *homogeneous* learning pattern would seem to be able to characterize these different paths.

The experiments of BHHR add to the picture of heterogeneous learning paths. Tables 3 and 4 present the statistical analyses of acceptances and offers for these experiments. From Table 3B we observe that Russian acceptances were 16 percentage points more generous in the first round than the US, but this marginal effect is only significant at the 38% level. This initial tendency towards generosity by responders in Russia is immediately offset, however, by 43 and 56

sets of estimates that would indicate violation of the zero-correlation assumption of the random effects specification.

percentage point decreases in their acceptance probability in rounds 2 and 3. These are again in reference to round 1 acceptance rates in the US. In the US acceptance rates declined somewhat later, in rounds 3 and 4, although the effect is only statistically significant in round 4. Thus there appears to a similar qualitative pattern of learning behavior in the US and Russia, but with very different speeds and intensity.

Turning to offers in Table 4B, we detect some late-round learning effects in each country. There was a terminal round effect in the US, with offers being 5 percentage points more generous than in round 1 on average. In Russia offers were also around 5 percentage points higher, but in the penultimate round 4. Although there are differences in timing, in this case the learning patterns do appear to be homogenous across the countries.

Do Demographics Affect Behavior?

The experiments of SR, LC and BHHR collected information on demographic characteristics, allowing a preliminary investigation of behavioral effects of demographics.

The experiments of SR suffer from one “problem” in terms of identifying any demographic effects in the first round, and indeed over time: the vast majority of offers were accepted. In fact, just over 90% of all offers were accepted, making it hard for *any* variation in acceptance rates to be explained by anything other than a constant. This is confirmed in our statistical analysis, shown in Table 5A. There is, however, a statistically significant effect of stake on acceptance rates, although the quantitative effect is only to increase acceptance rates by 2.3 percentage points in each stake increment.¹⁹

¹⁹ These effects are not additive, so the medium stake condition increases acceptance rates relative to the low stake condition by roughly the same amount as the high stake condition increases rates relative to the low stake condition.

There is some room for variation in offers in the SR experiments, however, since they started out a generous 45% on average in round 1. Table 5B shows that there is a decline of rough 2 percentage points in the offers made in rounds 4 through 10, relative to opening offers. Although these effects are statistically significant, this is not a particularly large reduction substantively. There is some slight evidence of demographics affecting offers in this experiment: those that are older, employed, live in larger households, or have Economics training offer less. This effect for household size is significant at the 3.3% level, and the effect for Economics training is significant at the 9.5% level. However, this effect from demographics appears to be an artifact of the invalid assumption underlying the random effects specification. The Hausman-Taylor instrumental variables estimator generates different results, indicating no effect from demographics, when allowance is made for their non-zero correlation with the unobserved individual effects. Thus we conclude that there are no effects from demographics in this experiment, for either acceptances or offers.

To understand the lack of substantial learning behavior in these experiments, consider the behavior over time shown in Figure 2. Essentially, these subjects started out making relatively generous offers, experienced high rates of acceptance as one might expect, and changed this pattern very little over time. At the risk of imputing a rationale for this behavior, it is as if the subjects making offers were extremely risk averse, not wanting to risk a rejection by reducing their offer. Of course, many other explanations are possible. But the main point of Figure 2 is that one would not expect to see much evidence of learning behavior or heterogeneity in this experiment, and that is what the statistical analysis shows.²⁰ There is evidence of a significant drop in offers

²⁰ The experiments of List and Cherry [2000] were designed to examine the same stakes issues as SR, but with a design that allowed more rejections to be observed. The raw behavior in this experiment is displayed in Figure 5. The most significant difference seems to be the starting point in round 1, where offers

after round 3 of between 1 and 2 percentage points, and that is about all.²¹

The experiments of CR also exhibit little variation in acceptance rates in the first round, with 81% of offers being accepted. No demographics were significant, individually or jointly, in terms of initial acceptance rates or offers, consistent with this general absence in acceptance rates.²² One feature of these data is that a large number of subjects gave the same response in the two non-practice rounds, so there is very little variation in the data for the “random effect” to explain. There are some significant stake effects on acceptances (roughly 6 percentage points higher with medium and high stake conditions) and offers (5.5 percentage points higher with the medium stake condition).

In the new experiments we conducted, individual demographic characteristics appear to play a role in both the US and Russia. For example, in the US males were significantly more likely to accept a given offer, as were those who reported higher incomes (either for themselves or for their parents), those whose mother was privately employed, and those living in urban areas. In Russia males are also more likely to accept, as are those from urban areas. These results were obtained from statistical models estimated using the sample within each country, to implicitly allow for country interactions with all demographics.²³ On the offer side, students and younger subjects tended to offer much more in Russia, but there are no significant effects in the US.

were relatively “tough” and the acceptance rate much less than observed in round 1 of SR. In any event, all of the learning in the List and Cherry [2000] experiments appears to have been done by responders, who steadily increased their acceptance rates over time. We do not expand on the analysis of List and Cherry since they did not collect demographic variables, and their data does not identify the individual proposer.

²¹ These time effects were obtained from a fixed-effects estimator, which provides consistent estimates. The random-effects estimator fails a Hausman test for the assumption that the unobserved individual effects are exogenous. The Hausman-Taylor instrumental variables estimator generated results that were consistent with the fixed-effects estimator in terms of the time effects, after allowing for the possibility that some of the demographics are correlated with the unobserved individual effects.

²² Detailed statistical results are not presented here, but are available in the data documentation.

²³ Detailed statistical results are not presented here, but are available in the data documentation.

Our data also indicates that there are interesting interaction effects between several of the key demographic variables, making our concern about unobserved individual variation substantive.

First, adding additional demographic controls, over those of sex and nationality, significantly lowers the acceptance probabilities for males in the US, but not in Russia. When sex is the only demographic characteristic estimated with the US sample, apart from the percent offer and time-effects, it appears to have a significant effect on acceptances: on average, males accept offers at a rate that is 44 percentage points higher than the rate for females, and this effect is significant at the 0.1% level. But when the same regression is undertaken with the addition of all of the demographics we collected, the pure effect of males drops to just 26 percentage points. Although this is still a significant effect, statistically and substantively, it indicates how one could over-estimate the effect of sex if other demographic controls are left out.²⁴

Second, the fact that there is such a confounding effect in the US and not in Russia shows that the effect interacts with nationality (and perhaps other unobserved demographics).

Third, allowing for additional demographics also increases the offers among males, particularly in Russia.

The results from the new experiments can also be used to explore the effect of adding controls for demographics and national effects when studying time effects. In Panel A of Table 3 these interactions are not used when evaluating the acceptance probabilities. Different conclusions emerge compared to the analysis with more controls in Panel B of Table 3. In addition

²⁴ Of course, one should distinguish between the partial effect of varying *only* sex while counterfactually holding all other demographics constant (which is what we refer to here) from the total effect of varying sex along with the characteristics that are correlated with it. Harrison, Lau and Williams [2002] illustrate the difference in demographic effects when partial and total effects are calculated for a field experiment eliciting individual rates of time preference.

to those discussed above (e.g., the differences in speed of adjustment), one could erroneously conclude that there is no nation-effect from inspect of Panel A, whereas this is just masked by the nation-time interaction. The same point applies to offers if one compares Panels A and B in Table 4.

Thus we offer further evidence in favor of the view that learning behavior is heterogeneous across nations, at least to the extent that we have built in controls to explain the heterogeneity. The fact that there appears to be no effects associated with individual demographics in Indonesia, some slight effects in the Slovak Republic, and significant effects in the US and Russia, says that one cannot claim that the demographics considered here explain the variation in behavior. The alternative is that there is some interaction between nationality and demographics, *or* there are some demographics that have not been measured and that vary across these national samples, *or* both. The only way to tease these hypotheses apart will be to conduct more experiments with larger samples that control for a wider array of demographics.

We do not claim that our analysis, or experimental design, is exhaustive. Nor need it be to make our central point: that omitting basic controls in the analysis of experimental data from different countries and samples can lead to fragile conclusions.

5. Implications

It is tempting, but incorrect, to equate the effect of culture on bargaining behavior with the effect of a simple country dummy variable, or perhaps with a demographic variable such as sex. The word “culture” connotes systematic beliefs and modes of behavior that are associated with a group of individuals, where the group can be defined along many different characteristics boundaries. One can have a Swedish culture, and even an Australian culture, but one can also

have a “geek culture” or a “gay culture.” In general there are many characteristics of individuals that can be used to identify systematic beliefs and patterns of behavior, and nationality and sex are just two examples of such characteristics.

Moreover, it is completely plausible that some of these characteristics might interact. Thus the effect of sex in one country could be very different from the effect of sex in another country. In other words, the differential effect of sex could be a reflection of the effects of “national culture.” RPOZ (p.1092) note that there were differences in the age and sex mix of their subject pool in different countries, reflecting differences in national cultures with respect to attendance at higher-education and the necessity of military duty. The possibility of interactions makes it even more difficult to claim that culture can be reduced to one simple dummy variable, or identified by unconditioned bilateral comparisons of distributions of behavior between two groups. These considerations become particularly important in field experiments, where the subject pools become much more heterogeneous and the cultural group delineations more difficult to discern.

One specific conclusion that can be drawn from the new experiments presented here is that the effect of nationality is not robust with respect to the addition of modest controls for other characteristics of the subject pool. This finding is consistent with previous literature that has demonstrated that behavior is correlated with demographic characteristics such as sex, ethnicity and race. Statistically, testing for the effect of one effect without controlling for the others will lead to unreliable claims at best, and erroneous conclusions at worst. We suspect that these type of confounding factors are not limited to nationality, sex and ethnicity, although in our initial approach to these issues here we are limited to these.

We do not intend to offer a methodological solution to the conduct of studies of culture, in the sense of proposing a definitive and sufficient list of controls that have to be included, but

simply to demonstrate some of the problems. We conclude that cross-national and other cross-cultural experimental projects have a serious design problem to solve before the questions of nationality, sex and cultural effects on behavior can be properly studied. We believe, however, that field experiments can contribute substantial knowledge about these issues due to the increased heterogeneity in subject pools that these experiments offer.

References

- Amemeyi, T., and MaCurdy, Thomas, "Instrumental-Variable Estimation of an Error-Components Model," *Econometrica*, 54(4), 1986, 869-880.
- Andreoni J. and Vesterlund, L., "Which is the fair sex? Gender differences in altruism," *Quarterly Journal of Economics*, February 2001, 293-312.
- Baltagi, Badi H., and Khanti-Akom, S., "On Efficient Estimation With Panel Data: An Empirical Comparison of Instrumental Variables Estimators," *Journal of Applied Econometrics*, 5, 1990, 401-406.
- Bolton, Gary and Katok, Elena. "An Experimental Test for Gender Differences in Beneficent Behavior," *Economics Letters*, 48(3-4), June 1995, 287-92.
- Bolton, Gary E., and Zwick, Rami, "Anonymity versus punishment in ultimatum bargaining," *Games and Economic Behavior*, 10, 1995, 95-121.
- Brown-Kruse, J. and Hummels D., "Gender effects in laboratory public goods contributions: do individuals put their money where their mouth is?" *Journal of Economic Behavior & Organization*, 22, 1993, 255-268.
- Burlando, R., and Hey, J.D., "Do Anglo-Saxons Free-ride More?" *Journal of Public Economics*, 64, 1997, 41-60.
- Cadsby, C. B. and Maynes E., "Gender and free riding in a threshold public goods game: experimental evidence," *Journal of Economic Behavior & Organization*, 34, 1998, 603-620.
- Cameron, Lisa A., "Raising The Stakes in the Ultimatum Game: Experimental Evidence from Indonesia," *Economic Inquiry*, 37(1), January 1999, 47-59.
- Croson, Rachel, and Buchan, Nancy, "Gender and Culture: International Experimental Evidence from Trust Games," *American Economic Review (Papers & Proceedings)*, 89(2), May 1999, 386-391.
- Eckel, Catherine C., and Grossman, Philip J., "Altruism in Anonymous Dictator Games," *Games and Economic Behavior*, 16, 1996a, 181-191.
- Eckel, Catherine C., and Grossman, Philip, "The Relative Price of Fairness: Gender Differences in a Punishment Game," *Journal of Economic Behavior and Organization*, 30, 1996b, 143-158.
- Eckel, Catherine C., and Grossman, Philip, "Are Women Less Selfish than Men? Evidence from Dictator Experiments," *Economic Journal*, 108(448), May 1998, 726-735.
- Eckel, Catherine and Grossman, Philip, "Chivalry and Solidarity in Ultimatum Games," *Economic Inquiry*, 39(2), April 2001, 171-188.

- Güth, Werner, and Tietz, Reinhard, "Ultimatum Bargaining Behavior: A Survey and Comparison of Experimental Results," *Journal of Economic Psychology*, 11, September 1990, 417-449.
- Harrison, Glenn W.; Lau, Morten Igel, and Williams, Melonie B., "Estimating Individual Discount Rates for Denmark: A Field Experiment," *American Economic Review*, 92(5), December 2002, 1606-1617.
- Harrison, Glenn W., and McCabe, Kevin A., "Expectations and Fairness in a Simple Bargaining Experiment," *International Journal of Game Theory*, 25(3), 1996, 303-327.
- Hausman, Jerry A., "Specification Tests in Econometrics," *Econometrica*, 46(6), November 1978, 1251-1271.
- Hausman, Jerry A., and Taylor, William E., "Panel Data and Unobservable Individual Effects," *Econometrica*, 49(6), November 1981, 1377-1398.
- Henrich, Joseph; Boyd, Robert; Bowles, Samuel; Camerer, Colin; Fehr, Ernst; Gintis, Herbert; and McElreath, Richard, "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies," *American Economic Review (Papers & Proceedings)*, May 2001, 73-78.
- List, John, and Cherry, Todd, "Learning to Accept in Ultimatum Games: Evidence from an Experimental Design that Generates Low Offers," *Experimental Economics*, 3(1), 2000, 11-29.
- Nowell, C. and Tinkler, S., "The influence of gender on the provision of a public good," *Journal of Economic Behavior and Organization*, 25, 1994, 25-36.
- Roth, Alvin E.; Prasnikar, Vesna; Okuno-Fujiware, Msahiro, and Zamir, Shmuel, "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *American Economic Review*, 81(5), December 1991, 1068-1095.
- Saijo, T., and Nakamura, H., "The 'Spite' Dilemma in Voluntary Contribution Mechanism Experiments," *Journal of Conflict Resolution*, 39, 1995, 535-560.
- Schweitzer, Maurice, and Solnick, Sara, "The Influence of Physical Attractiveness and Gender on Ultimatum Game Decisions," *Organizational Behavior and Human Decision Processes*, 79(3), September 1999, 199-215.
- Slonim, Robert, and Roth, Alvin E., "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic," *Econometrica*, 66(3), May 1998, 569-596.
- Solnick, Sara, "Gender differences in the ultimatum game," *Economic Inquiry*, 39(2), April 2001, 189-200.
- StataCorp, *Stata Statistical Software: Release 8.0* (College Station, TX: Stata Corporation, 2003a).

StataCorp, *Stata Cross-Sectional Time-Series Reference Manual: Release 8.0* (College Station, TX: Stata Corporation, 2003b).

Willer, David and Szmataka, Jacek, "Cross-National Experimental Investigations of Elementary Theory: Implications for the Generality of the Theory and the Autonomy of Social Structures," *Advances in Group Processes*, vol. 10, 1993, 37-81.

Yamagishi, T., "Exit From the Group as an Individualistic Solution to the Public Good Problem in the United States and Japan," *Journal of Experimental Social Psychology*, 24, 1998a, 530-542.

Yamagishi, T., "The Provision of a Sanctioning System in the United States and Japan," *Social Psychology Quarterly*, 51, 1998b, 265-271.

Yamagishi, T.; Cook, K.; and Watabe, M., "Uncertainty, Trust, and Commitment Formation in the United States and Japan," *American Journal of Sociology*, 1998.

Figure 1: Behavior in RPOZ Experiments

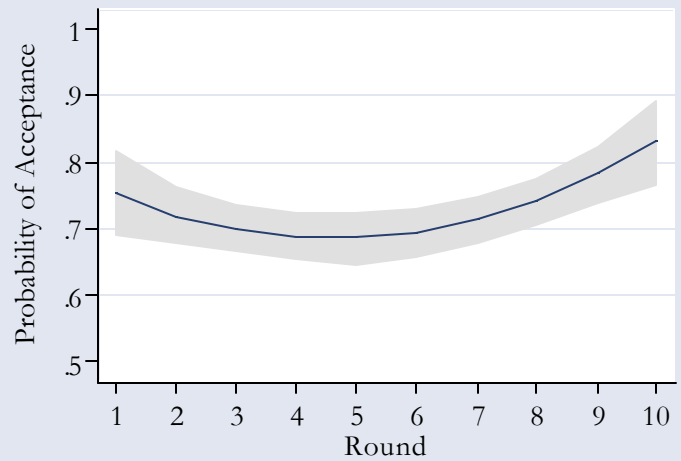
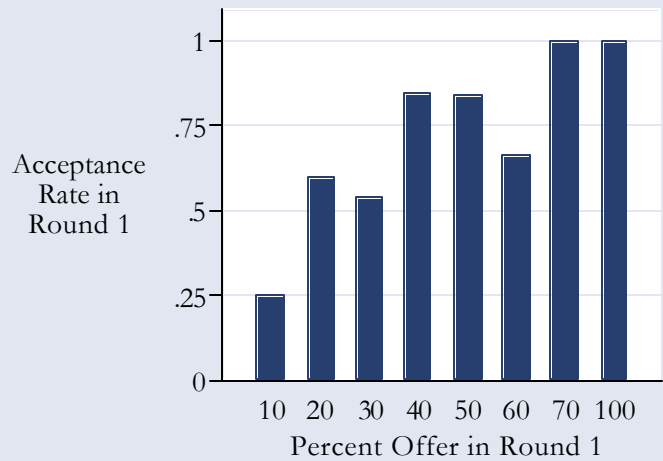
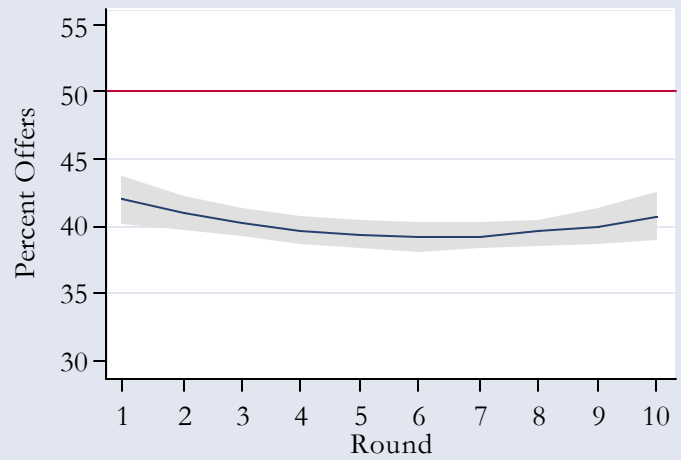
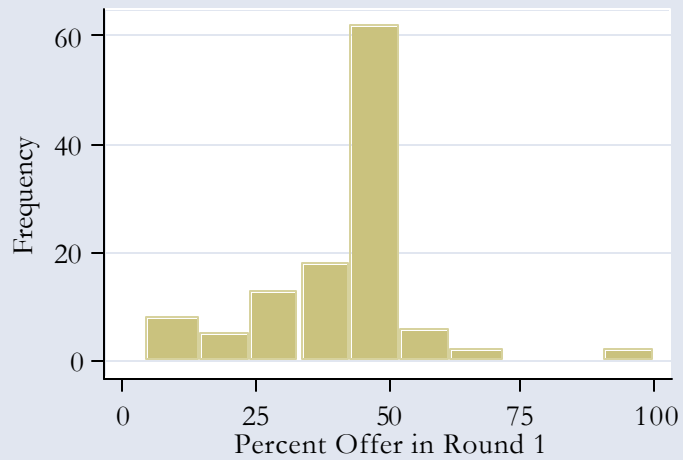


Figure 2: Behavior in Slonim and Roth Experiments

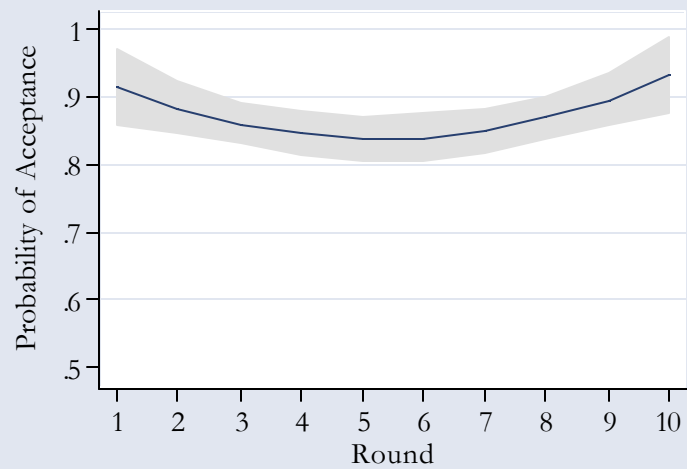
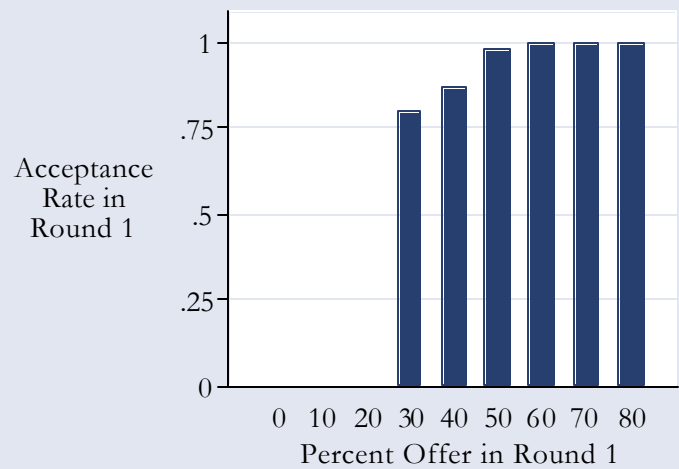
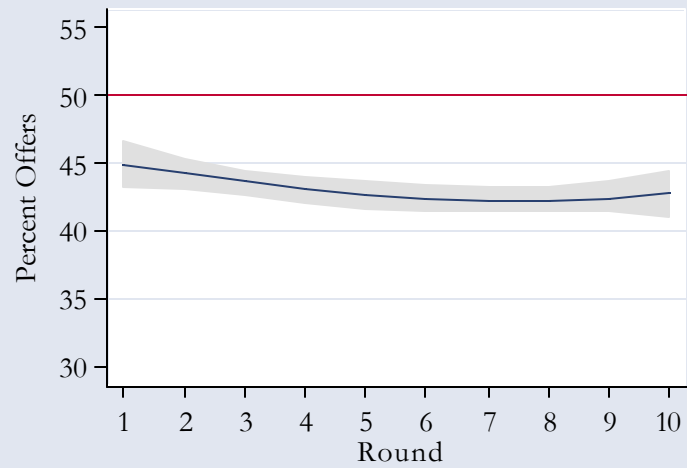
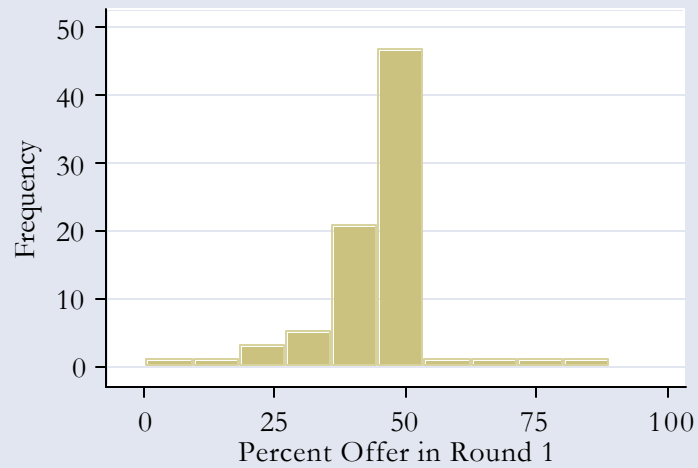


Figure 3: Behavior in Cameron Experiments

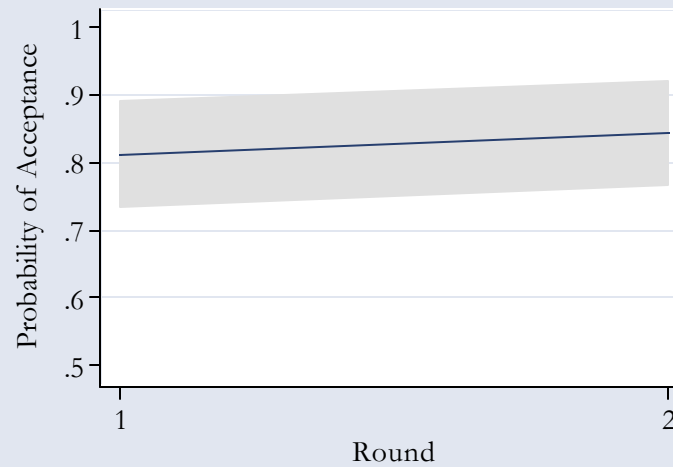
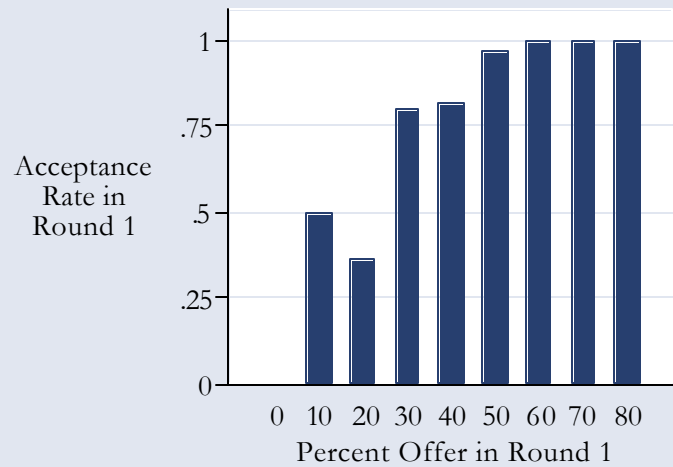
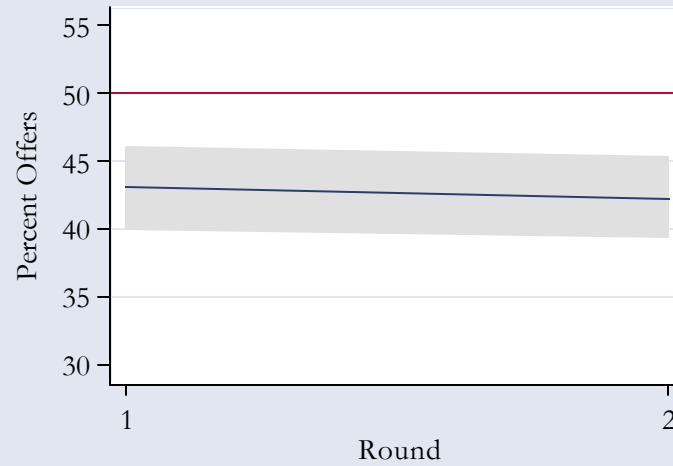
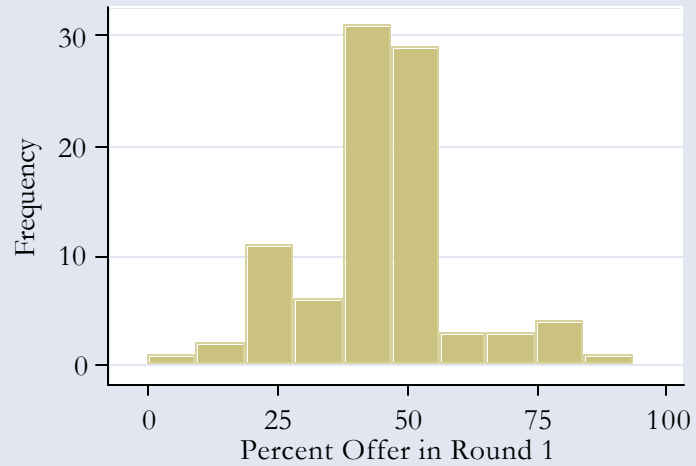


Figure 4: Behavior in New BHHR Experiments

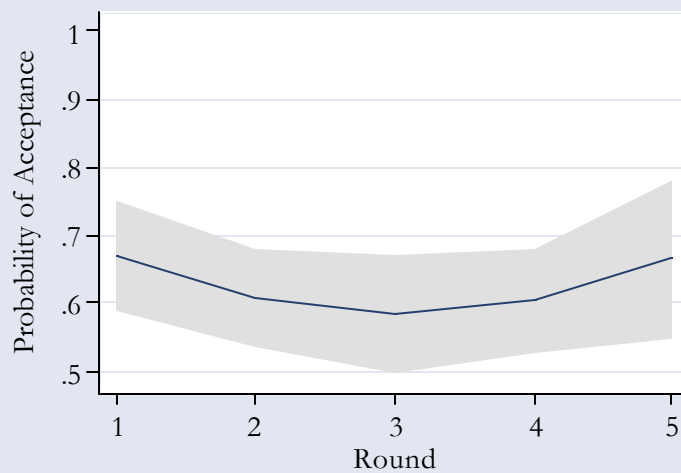
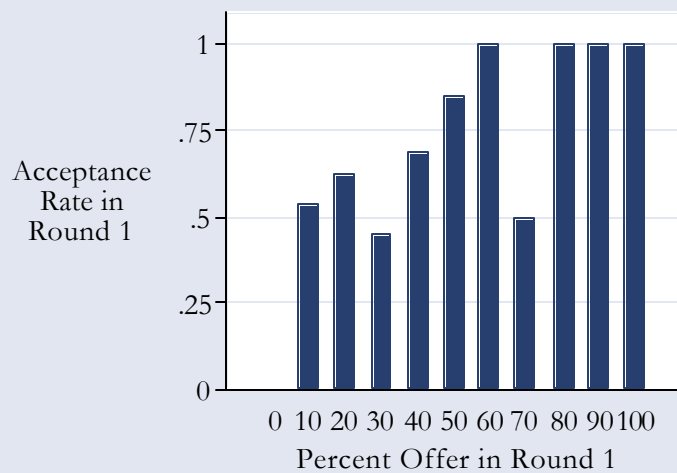
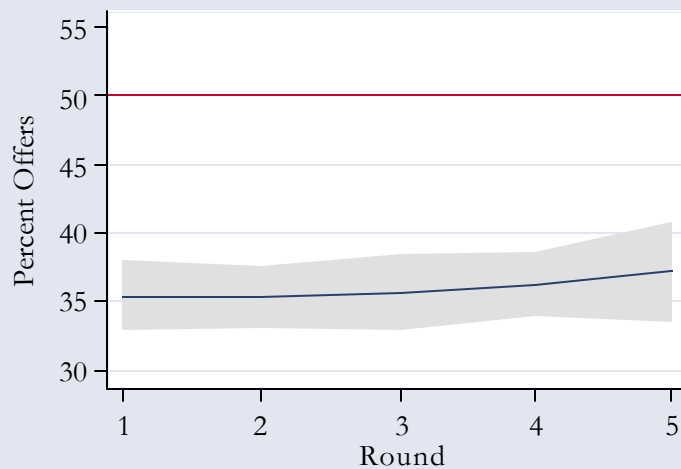
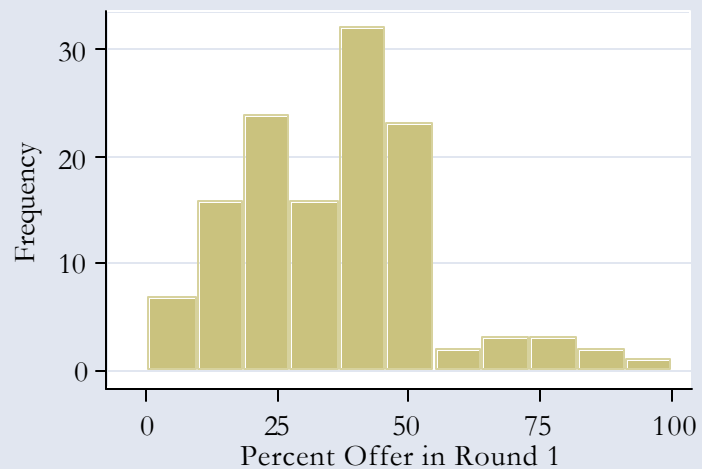


Figure 5: Behavior in List and Cherry Experiments

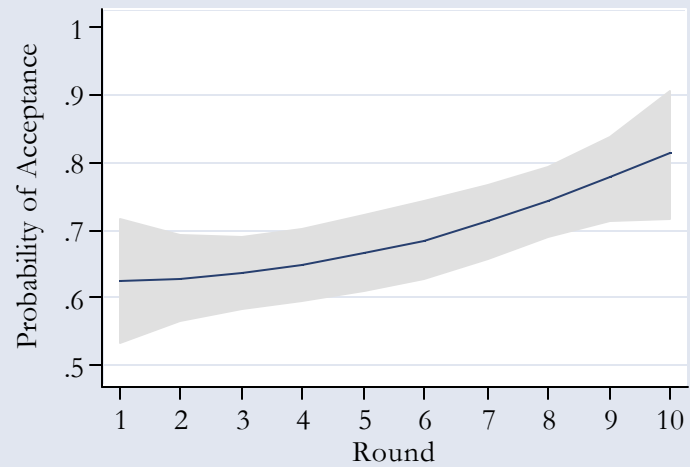
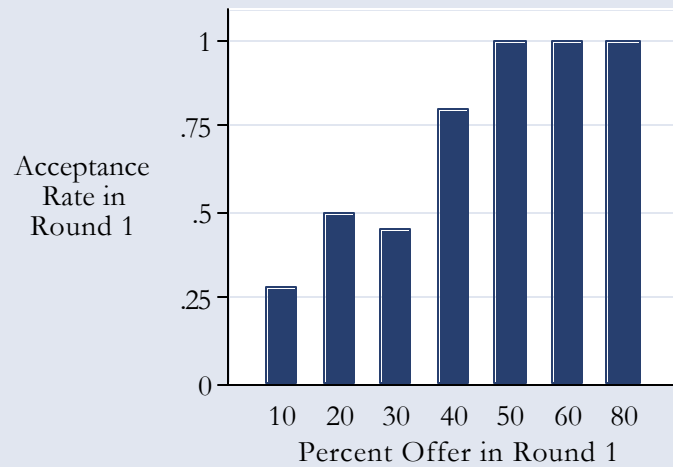
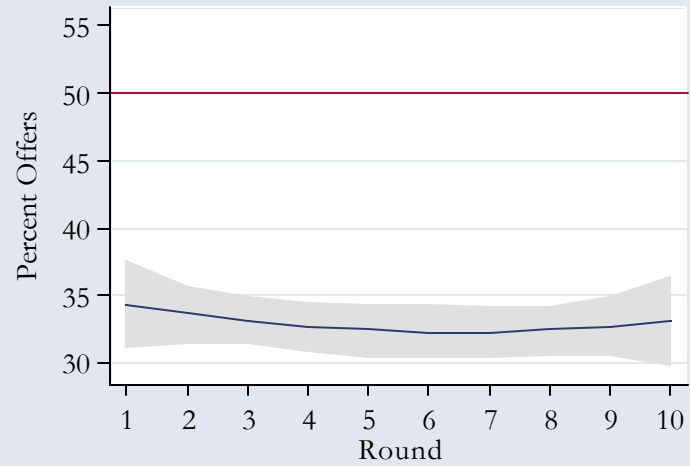
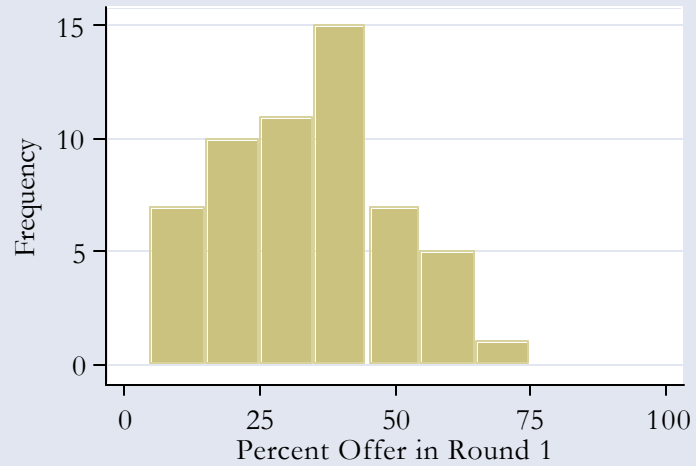


Table 1: Marginal Effects on Acceptance Probability in RPOZ Experiments

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
offer	.0243225	.00397	6.12	0.000	.016535	.03211	38.9317	
US_2	-.0913897	.15509	-0.59	0.556	-.395358	.212578	.020796	
US_3	-.0710585	.14627	-0.49	0.627	-.357751	.215634	.023508	
US_4	-.1780045	.18841	-0.94	0.345	-.547286	.191277	.024412	
US_5	-.2587144	.21034	-1.23	0.219	-.670967	.153538	.023508	
US_6	-.0634496	.14305	-0.44	0.657	-.343829	.21693	.023508	
US_7	.00245	.10706	0.02	0.982	-.207382	.212282	.023508	
US_8	.0928186	.04075	2.28	0.023	.012942	.172696	.024412	
US_9	-.0138002	.10816	-0.13	0.898	-.225791	.19819	.024412	
US_10	.0395336	.07752	0.51	0.610	-.112405	.191472	.024412	
Israel	.1528925	.08666	1.76	0.078	-.016948	.322733	.267631	
Israel_2	.0883512	.0417	2.12	0.034	.006625	.170077	.027125	
Israel_3	-.015762	.11329	-0.14	0.889	-.237802	.206278	.026221	
Israel_4	.0209402	.08836	0.24	0.813	-.152251	.194131	.027125	
Israel_5	.0700895	.05246	1.34	0.182	-.032738	.172917	.027125	
Israel_6	.0609966	.05711	1.07	0.286	-.050941	.172934	.026221	
Israel_7	.0163483	.08894	0.18	0.854	-.15797	.190667	.026221	
Israel_8	.079365	.04511	1.76	0.079	-.009057	.167787	.027125	
Israel_9	.122358	.02714	4.51	0.000	.069156	.17556	.027125	
Israel_10	.1132184	.02851	3.97	0.000	.057346	.169091	.027125	
Japan	.0930439	.11775	0.79	0.429	-.137738	.323826	.230561	
Japan_2	-.1253788	.19046	-0.66	0.510	-.498666	.247909	.0217	
Japan_3	-.1863459	.23559	-0.79	0.429	-.6481	.275409	.018987	
Japan_4	-.1796466	.20775	-0.86	0.387	-.586822	.227529	.022604	
Japan_5	-.1503167	.19721	-0.76	0.446	-.536843	.236209	.022604	
Japan_6	-.2777681	.24856	-1.12	0.264	-.764943	.209406	.022604	
Japan_7	.0475814	.07803	0.61	0.542	-.105358	.200521	.025316	
Japan_8	-.1041263	.17228	-0.60	0.546	-.441781	.233529	.024412	
Japan_9	-.0972221	.16713	-0.58	0.561	-.424791	.230347	.025316	
Japan_10	.0161805	.09891	0.16	0.870	-.177677	.210038	.026221	
Yugoslavia	-.205074	.18525	-1.11	0.268	-.568151	.158003	.266727	
Yugoslavia_2	.0594432	.05012	1.19	0.236	-.03879	.157676	.025316	
Yugoslavia_3	.0534076	.05362	1.00	0.319	-.051687	.158502	.026221	
Yugoslavia_4	.0540901	.05432	1.00	0.319	-.05237	.160551	.027125	
Yugoslavia_5	.0432096	.05746	0.75	0.452	-.069419	.155838	.026221	
Yugoslavia_6	.0504546	.05323	0.95	0.343	-.053881	.15479	.027125	
Yugoslavia_7	-.0240187	.09235	-0.26	0.795	-.205016	.156979	.027125	
Yugoslavia_8	-.0013145	.07896	-0.02	0.987	-.156073	.153444	.027125	
Yugoslavia_9	-.007982	.08124	-0.10	0.922	-.16721	.151246	.027125	
Yugoslavia_10	.0449911	.05626	0.80	0.424	-.065277	.15526	.027125	
Shmuel	-.0959244	.12047	-0.80	0.426	-.332042	.140193	.341772	
Masahiro	-.0172904	.14269	-0.12	0.904	-.296949	.262368	.320976	
IsDiff	.0567176	.06421	0.88	0.377	-.069133	.182568	.090416	
YuDiff	.0781105	.0368	2.12	0.034	.005982	.150239	.087703	

Table 3: Marginal Effects on Acceptance Probability in BHR Experiments

A. No Nation-Time Interactions

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
offer	.0303516	.00453	6.70	0.000	.021467 .039237	35.5828
Round2	-.2234373	.13496	-1.66	0.098	-.487956 .041082	.203448
Round3	-.4193737	.1253	-3.35	0.001	-.664966 -.173781	.2
Round4	-.2930531	.13339	-2.20	0.028	-.554492 -.031614	.203448
Round5	-.2231247	.13486	-1.65	0.098	-.487438 .041188	.203448
Russia	-.0244756	.09121	-0.27	0.788	-.203238 .154287	.506897

B. Adding Nation-Time and Nation-Gender Interactions

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
offer	.028863	.00415	6.96	0.000	.020732 .036994	35.5828
US_2	.0081222	.15605	0.05	0.958	-.297739 .313984	.1
US_3	-.2094879	.17675	-1.19	0.236	-.55592 .136944	.096552
US_4	-.313298	.18698	-1.68	0.094	-.679767 .053171	.1
US_5	-.1327786	.17793	-0.75	0.456	-.481516 .215959	.1
Russia	.1595116	.18237	0.87	0.382	-.197918 .516941	.506897
Russia_2	-.4262789	.18508	-2.30	0.021	-.789034 -.063524	.103448
Russia_3	-.5616082	.14691	-3.82	0.000	-.849537 -.273679	.103448
Russia_4	-.2640974	.20312	-1.30	0.194	-.662198 .134003	.103448
Russia_5	-.24871	.20164	-1.23	0.217	-.643916 .146496	.103448
US_male	.318247	.05947	5.35	0.000	.201682 .434812	.203448
R_male	.2892508	.06359	4.55	0.000	.164622 .41388	.224138

Note: US_male and R_male are interactions of nation and sex.

Table 4: Effects of Controls on Offers in BHHR Experiments

A. No Nation-Time Interactions

```

Random-effects GLS regression           Number of obs   =       289
Group variable (i): idS                 Number of groups =        59

R-sq:  within = 0.0248                  Obs per group:  min =         4
        between = 0.0093                  avg =           4.9
        overall = 0.0132                  max =           5

Random effects u_i ~ Gaussian           Wald chi2(5)     =        6.22
corr(u_i, X) = 0 (assumed)              Prob > chi2      =       0.2851
    
```

Offer	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Round2	1.679608	1.610063	1.04	0.297	-1.476058	4.835274
Round3	2.010175	1.618339	1.24	0.214	-1.161711	5.18206
Round4	2.916896	1.610063	1.81	0.070	-.2387697	6.072562
Round5	3.688083	1.610063	2.29	0.022	.5324167	6.843749
Russia	1.303663	2.741866	0.48	0.634	-4.070296	6.677621
_cons	32.68294	2.214275	14.76	0.000	28.34304	37.02284
sigma_u	9.8101752					
sigma_e	8.5197146					
rho	.57005476	(fraction of variance due to u_i)				

Note: sigma_u is the estimate of the panel-level standard deviation (the random effect), sigma_e is the estimate of the overall standard deviation (other than the random effect), and rho is the fraction of the total variance due to the random effect.

B. Adding Nation-Time and Nation-Gender Interactions

```

Random-effects GLS regression           Number of obs   =       289
Group variable (i): idS                 Number of groups =        59

R-sq:  within = 0.0454                  Obs per group:  min =         4
        between = 0.1909                  avg =           4.9
        overall = 0.1396                  max =           5

Random effects u_i ~ Gaussian           Wald chi2(11)    =       23.31
corr(u_i, X) = 0 (assumed)              Prob > chi2      =       0.0160
    
```

Offer	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
US_2	1.52019	2.251432	0.68	0.500	-2.892536	5.932916
US_3	.7914856	2.275327	0.35	0.728	-3.668073	5.251044
US_4	1.227087	2.251432	0.55	0.586	-3.185639	5.639813
US_5	5.209846	2.251432	2.31	0.021	.7971195	9.622572
Russia	3.662378	4.499392	0.81	0.416	-5.156268	12.48102
Russia_2	1.882063	2.289357	0.82	0.411	-2.604995	6.369121
Russia_3	3.182063	2.289357	1.39	0.165	-1.304995	7.669121
Russia_4	4.59873	2.289357	2.01	0.045	.1116719	9.085788
Russia_5	2.265397	2.289357	0.99	0.322	-2.221661	6.752455
R_male	-11.04675	3.632963	-3.04	0.002	-18.16723	-3.926278
US_male	-6.204525	3.66902	-1.69	0.091	-13.39567	.9866225
_cons	36.61695	3.151915	11.62	0.000	30.43931	42.79458
sigma_u	8.9751577					
sigma_e	8.5046668					
rho	.52689674	(fraction of variance due to u_i)				

Note: US_male and R_male are interactions of nation and sex; sigma_u is the estimate of the panel-level standard deviation (the random effect), sigma_e is the estimate of the overall standard deviation (other than the random effect), and rho is the fraction of the total variance due to the random effect.

Table 5: Marginal Effects of Controls on Behavior in SR Experiments

A. Acceptances

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
offer	.0021404	.00137	1.57	0.117	-.000539	.00482	43.0555	
Round2	-.0008298	.00709	-0.12	0.907	-.014717	.013057	.1	
Round3	.0002211	.00671	0.03	0.974	-.012937	.013379	.1	
Round4	-.0036009	.00894	-0.40	0.687	-.021122	.01392	.1	
Round5	-.0076567	.012	-0.64	0.524	-.031184	.015871	.1	
Round6	-.0094692	.01362	-0.70	0.487	-.036162	.017223	.1	
Round7	-.0051986	.01012	-0.51	0.607	-.025033	.014636	.1	
Round8	.0019051	.00567	0.34	0.737	-.009199	.013009	.1	
Round9	-.0004938	.00655	-0.08	0.940	-.013339	.012351	.102439	
Round10	.0059465	.0053	1.12	0.261	-.004432	.016325	.097561	
mid	.0235733	.01113	2.12	0.034	.001764	.045382	.402439	
hi	.0230285	.01374	1.68	0.094	-.003893	.04995	.304878	
Female	.000467	.0033	0.14	0.887	-.005998	.006932	.292683	
Age	-.000152	.00117	-0.13	0.897	-.002443	.002139	20.8049	
Employed	.000786	.00637	0.12	0.902	-.011691	.013263	.097561	
Nhhd	-.0007982	.00154	-0.52	0.605	-.003823	.002227	3.28049	
Pincome	-.0083622	.00708	-1.18	0.237	-.022235	.005511	.268293	
Hincome	1.57e-07	.00000	0.37	0.710	-6.7e-07	9.8e-07	9390.24	
Econ	-.0010914	.00383	-0.29	0.776	-.008596	.006413	.219512	

B. Offers

Random-effects GLS regression	Number of obs	=	765
Group variable (i): ProID	Number of groups	=	81
R-sq: within = 0.0235	Obs per group: min	=	2
between = 0.1540	avg	=	9.4
overall = 0.1298	max	=	10
Random effects u_i ~ Gaussian	Wald chi2(18)	=	29.36
corr(u_i, X) = 0 (assumed)	Prob > chi2	=	0.0441

offer	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Round2	-.1656309	.7775514	-0.21	0.831	-1.689604 1.358342
Round3	-1.23669	.7739109	-1.60	0.110	-2.753527 .2801477
Round4	-1.777967	.7782267	-2.28	0.022	-3.303263 -.2526709
Round5	-1.861203	.7648388	-2.43	0.015	-3.360259 -.3621466
Round6	-1.978497	.7694166	-2.57	0.010	-3.486526 -.4704681
Round7	-1.678499	.7686237	-2.18	0.029	-3.184974 -.1720242
Round8	-1.397034	.7695454	-1.82	0.069	-2.905315 .1112471
Round9	-2.069989	.7609298	-2.72	0.007	-3.561384 -.5785941
Round10	-1.694451	.7773497	-2.18	0.029	-3.218028 -.1708732
mid	-1.286465	2.12756	-0.60	0.545	-5.456406 2.883475
hi	-2.557485	2.259274	-1.13	0.258	-6.98558 1.870611
Female	-1.3873	1.962302	-0.71	0.480	-5.233342 2.458741
Age	-.9841636	.6819567	-1.44	0.149	-2.320774 .352447
Employed	-5.065536	3.365	-1.51	0.132	-11.66082 1.529743
Nhhd	-1.769091	.8274864	-2.14	0.033	-3.390935 -.1472478
Pincome	-1.833938	2.043437	-0.90	0.369	-5.839001 2.171125
Hincome	.000225	.0002202	1.02	0.307	-.0002066 .0006566
Econ	-3.702302	2.215516	-1.67	0.095	-8.044633 .6400299
_cons	71.10788	14.0886	5.05	0.000	43.49474 98.72103
sigma_u	7.4073822				
sigma_e	4.7614075				
rho	.70762303	(fraction of variance due to u_i)			

Note: sigma_u is the estimate of the panel-level standard deviation (the random effect), sigma_e is the estimate of the overall standard deviation (other than the random effect), and rho is the fraction of the total variance due to the random effect.