DISCUSSION PAPER SERIES

IZA DP No. 1341

Comparing Students to Workers: The Effects of Social Framing on Behavior in Distribution Games

Jeffrey Carpenter
Stephen Burks
Eric Verhoogen

October 2004

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Comparing Students to Workers: The Effects of Social Framing on Behavior in Distribution Games

Jeffrey Carpenter
Middlebury College and IZA Bonn

Stephen Burks
University of Minnesota, Morris

Eric Verhoogen
University of California, Berkeley and IZA Bonn

Discussion Paper No. 1341
October 2004

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

IZA Discussion Paper No. 1341
October 2004

ABSTRACT

Comparing Students to Workers: The Effects of Social Framing on Behavior in Distribution Games

To investigate the external validity of Ultimatum and Dictator game behavior we conduct experiments in field settings with naturally occurring variation in "social framing." Our participants are students at Middlebury College, non-traditional students at Kansas City Kansas Community College (KCKCC), and employees at a Kansas City distribution center. Ultimatum game offers are ordered: KCKCC > employee > Middlebury. In the Dictator game employees are more generous than students in either location. This indicates that workers behaved distinctly from both student groups because their allocations do not decrease between games, an effect we attribute to the social framing of the workplace.

JEL Classification: C93, J24, Z13

Keywords: Ultimatum game, Dictator game, fairness, reciprocity, field experiment

Corresponding author:

Jeffrey Carpenter
Department of Economics
Middlebury College
Middlebury, VT 05753
USA
Email: jpc@middlebury.edu

1 Introduction

It is widely acknowledged among experimentalists that the framing of interactions in the laboratory can have significant effects on subjects' behavior. People often follow different norms and rules for behavior in different social contexts, and how they behave in the laboratory may depend on their beliefs about which social context most closely corresponds to the experimental situation. Experimentalists have typically explored the effect of framing by varying the verbal cues given in the descriptions of games, holding constant the underlying payoff structures. But such verbal cues are not the only factor that shapes subjects' beliefs about which set of behavioral rules should be invoked. Subjects' beliefs are also influenced by the real-life social context in which the laboratory is embedded – by their relationship to the people they are playing against and to the experimenter and by the set of norms and habits that dominate the cultural life in the institution in which the experiment is carried out. We refer to this broader set of influences as the *social framing* of the experiment, distinct from the verbal framing given in the experimenter's verbal description of the game.

The vast majority of economic experiments have had one particular social framing: the subjects are college students, playing against other college students, in a laboratory on campus. The ability to randomize subjects into treatment and control groups and to hold the experimental environment constant (or very nearly constant) has enabled experimenters to draw internally valid conclusions about the causal effects of different experimental procedures.[1] But because experiments have tended to be limited to a particular subject pool in a particular social context, the extent to which their results generalize to other groups of people in other social contexts – the external validity of the experiments – remains open to question. One way to explore the external validity of experiments is to examine the extent to which results are robust to variations in changes both in the characteristics of subjects and in the social framing of the experiments.

In this paper, we explore the external validity of experimental results in two simple bargaining games, the Ultimatum Game (UG) and the Dictator Game (DG), by comparing experiments conducted with the standard social framing – among undergraduates at Middlebury College, a small liberal arts college in Vermont – to experiments with identical procedures conducted in the field environment of a workplace – a publishing distribution warehouse in Kansas City, Kansas. We expect the social framing of the workplace to have a quite different effect on subjects' behavior than the social framing of the college campus, controlling for individual characteristics. Workers in the distribution center see each other every day, often work together in teams, and can expect to continue working together for long periods of time. Students, even on a small tight-knit campus like Middlebury, are more likely to be in competition for grades, are likely to have less frequent interactions, and know that their time

[1] We have the most straight-forward definition of internal validity in mind (a la Campbell and Stanley, 1963) – through the proper use of experimental control one can assign causality to independent variables.

together on campus is limited.

A thorny issue in comparing experiments in the two settings is that the experiments may differ along two dimensions: both the social framing and the individual characteristics of subjects may vary. This means that differences in behavior may be attributed to cultural or national differences when they are really, at least partially, attributable to differences in the demographic characteristics of the participant populations (e.g., age or income). To estimate separately the effect of social framing from the effect of differences in individual characteristics, we conducted a third round of experiments at Kansas City Kansas Community College (KCKCC), a junior college near the warehouse. The advantage of KCKCC is that the social framing is similar to that of Middlebury, while the observable demographic characteristics of the participants are similar to those of employees in the distribution center.

Our results indicate that proposers in the UG in the two experiments in Kansas City made more generous offers than proposers in the experiment at Middlebury, even controlling for differences in demographic characteristics. This result is consistent with the hypothesis that regional differences (for example, variations in regional cultural norms) affect behavior; we refer to this as the "Kansas City effect". We also find that our KCKCC students offer significantly more than our KC workers in the UG, while in the Dictator game, the employees allocated more than the students in either location. Perhaps most distinctive is that both groups of students exhibit a large drop in mean allocations between the UG and DG experiments, while the workers offer the same amount, on average, in both games. Together, these facts suggest that social framing matters.

2 Related Work

Interest among economists in framing was stimulated by the work of Daniel Kahneman and Amos Tversky, who noticed that responses to decision problems depended on whether the problem was framed in terms of losses or gains. This recognition later became a component of prospect theory (Kahneman and Tversky, 1979; 2000). Subsequently, this work lead to a standard way of looking at differences in the framing of choice problems in the experimental lab. A common subject pool was presented the same problem, but with distinct frames, and then the results were compared for framing effects.

This basic method has been applied in many areas of experimental and behavioral economics. Abbink and Hennig-Schmidt (2002) find no difference between a neutrally worded treatment of bribery game and a contextualized treatment of the same game. Many experiments on the effect of framing have been conducted in the context of a voluntary contribution game. Elliott et al. (1998) conduct a two stage experiment in which the first stage frames the free riding problem in terms of autonomous business standards or teamwork and the second stage is a voluntary contribution game. They show that cooperative work frames elicit more cooperation. In the dictator game, Eckel and Grossman (1996) find that subjects behave more generously toward a partner described

as the Red Cross than a partner described as an anonymous student. In the ultimatum game, Hoffman et al. (1994) show that changing the instructions so that participants are called buyers and sellers (i.e. adding a market frame) significantly reduces offers. Other related experiments include Willinger and Ziegelmeyer (1999), Park (2000), and Cookson (2000).

A small number of studies have examined the results of particular games across different subject populations in different real-life social contexts. Murnighan and Saxon (1998) conduct ultimatum games with children of different ages. They find that young children behave more fairly than older children when proposing a distribution, but were less likely to enforce fairness norms when offered a small amount. The authors conclude that small children have a keener sense of fairness and are less competitive than older children and many adults. Carter and Irons (1991) show that economics students offer less and are willing to accept less in the UG; according to the authors, this result may be explained by the fact that more self-interested students study economics. In perhaps the most comprehensive study, Henrich et al. (2001) conducted ultimatum games in fifteen different small-scale communities in developing countries. They found significant variation in behavior across communities, more variation than is typical in cross-population studies in industrialized countries (e.g. Roth et al., 1991).[2] A small related literature has developed on using simple experiments to measure behavioral norms or propensities across cultures or communities (e.g. Camerer and Fehr, 2001 or Carpenter, 2002).

3 Experimental Procedures

Our instructions and survey are available from the authors. What follows is a brief description of our methods. In the Ultimatum Game (UG), first discussed in Gueth et al. (1982), one person is designated as the first-mover or proposer and another as second-mover, or responder. The proposer proposes a split of a sum of money given by the experimenter, and the responder can accept or reject the proposer's offer. If she accepts, the offer is implemented; if she rejects, both players receive nothing. If both proposer and responder were motivated only by monetary payoffs and this were common knowledge, then the proposer would know that the responder would accept any positive offer and hence would offer the smallest possible amount. A series of experiments have shown that results do not conform to this subgame-perfect prediction. Proposers tend to send significantly more than the minimum positive amount, and responders tend to reject low offers (Gueth et al., 1982; Binmore et al., 1985; Gueth and Tietz, 1990). Typically the modal offer in the UG is a 50-50 split.

There are two popular explanations for the fact that proposers offer significantly more than the smallest positive amount. One is that the proposers have non-selfish preferences and are concerned with the outcomes of the responders.

[2]However, it is hard to directly compare Henrich et al. and Roth et al. because of procedural differences. For a critical view of the methodological issues raised by the work reported in Henrich, et al., see Ortmann (this volume).

The other is that the proposers have selfish preferences, but are afraid that responders will spitefully reject low offers. The Dictator Game (DG), developed in Forsythe et al. (1994), is a variant of the UG designed to discriminate between these two explanations. In the DG, the responder does not have veto power over the proposed split; she simply receives whatever she is allocated by the proposer. The subgame-perfect outcome does not change substantially from the UG: the proposer receives all the money instead of nearly all the money. Forsythe et al. (1994) showed that although proposers in the DG typically offer significantly less than proposers in the UG, they still offer non-trivial positive amounts. In terms of the two explanations just mentioned, this suggests a polymorphic population. That is, some subjects (those who might make high offers in the UG but zero in the DG) are risk averse and have selfish preferences, while other subjects (those that might make high offers in both experiments) do indeed have other-regarding preferences, that may be governed by altruistic norms or fairness concerns.

SUMMARY STATISTICS FOR PARTICIPANT DEMOGRAPHICS (FIRSTMOVERS IN BOTH GAMES)

	Middlebury			KCKCC			KC Workers		
Variable	n	mean	std. dev.	n	mean	std. dev.	n	mean	std. dev.
Age	41	19.44	1.34	44	26.91	8.73	67	37.13	10.18
Female	41	0.54	0.55	44	0.66	0.48	68	0.53	0.50
Schooling	41	13.40	1.24	43	13.79	2.04	66	13.08	3.31
Income	41	151,463	97,728	44	36,250	20,349	66	41,287	20,853
Black	41	0	0	44	0.25	0.44	68	0.12	0.32
Hispanic	41	0.07	0.26	44	0.09	0.29	68	0.09	0.29
Non-white	41	0.12	0.34	44	0.41	0.50	68	0.28	0.45
Mach	41	96.31	12.54	44	85.29	13.95	68	87.37	11.56

Table 1 – Demographic summary statistics for first-movers in the ultimatum and dictator games (Note: KC means Kansas City and KCKCC means Kansas City Kansas Community College).

		Social Context: College	Social Context: Work
Demographics	**Younger, More Affluent**	Middlebury 20 UG observations 21 DG observations	
	Older, Less Affluent	KCKCC 18 UG observations 26 DG observations	KC Warehouse 30 UG observations 37 DG observations

Table 2 – Experimental design (Note: KC means Kansas City and KCKCC means Kansas City Kansas Community College).

To assure our participants were highly motivated, the stakes in both games were \$100. Both games were one-shot, to eliminate reputation effects. Table 1 presents a summary of demographic characteristics of our participants in the three contexts, Middlebury College, Kansas City Kansas Community College, and the Kansas City distribution center. Table 2 summarizes our design. The numbers of observations were 20 for the UG and 21 for the DG at Middlebury, 30 for the UG and 37 for the DG at the warehouse, and 18 for the UG and 26 for the DG at KCKCC. The Middlebury students were younger, had dramatically higher family incomes, and were more likely to be white than both the distribution center employees and the KCKCC students ($p<0.01$, $p<0.01$, and $p=0.05$ respectively for Middlebury vs. the distribution center; $p<0.01$, $p<0.01$, and $p<0.01$ respectively for Middlebury vs. KCKCC).

The distribution center employees and the KCKCC students were broadly similar on a number of demographic dimensions. Average family incomes were statistically equal ($p=0.21$). Both subject pools included a significant number of African-American participants (the difference is not significant, $p=0.16$), which was not true of the Middlebury students. In addition, KCKCC is located within a few miles of the distribution center; so if there are any distinctive features of this geographic region (for example, regional cultural norms), it is likely that the KCKCC students and the warehouse workers share them. However, the demographic characteristics of the KCKCC students and distribution center employees were not identical. In particular, the KCKCC students were younger. The mean age of the KCKCC students (26.91) was between that of the Middlebury students (19.44) and that of the distribution center employees (37.13). Although the demographics of our KCKCC participants do not match our warehouse participants perfectly, and although it is of course possible that the two groups differ in other unobservable ways, the demographic similarities make it reasonable to consider the hypothesis that differences in the behavior of these two groups might be due, at least in part, to differences in the social framing of the experiments.

We also had our participants fill out a personality scale called the Mach scale, first developed by Christie and Geis (1970). The Mach scale consists of twenty statements based on Machiavelli's *The Prince* to which subjects are asked to agree or disagree, on a 7-point Likert scale. Their scores are summed over the 20 statements, and a constant of 20 is added, to generate a measure that ranges between 40 and 160, with the neutral score at 100. Those who tend to agree with the Machiavellian statements (i.e., have scores above 100) are termed "high Machs," and those who tend to disagree (i.e., who score less than 100) "low Machs." The Mach scale is designed to capture three components of an individual's behavioral dispositions: (1) the extent to which a subject has a cynical view of human nature, believing that others are not trustworthy; (2) the willingness of a subject to engage in manipulative behaviors; and (3) the extent of the subjects' concern (or lack thereof) with conventional morality (Christie and Geis, 1970; Fehr et al., 1992). The Mach scale is well-established in the social psychology literature (McHoskey et al., 1998). Researchers have found both that the scale is reliable, in that individuals' scores vary little from one administration of the test to another and that it generally accords with other personality assessment tools (Fehr et al., 1992; Wrightsman, 1991; Panitz, 1989; McHoskey et al., 1998).

We included the Mach scale with the goal of controlling for variations in inherent predispositions toward engaging in manipulative or exploitative behaviors. In previous related work, Meyer (1992) found evidence suggesting high Machs are less likely to reject low offers in the ultimatum game, while Gunnthorsdottir, McCabe, and Smith (2000), using a modified trust game, found high Machs reciprocated less.

The procedures we followed for our visit to the distribution center were as follows. Prior to the experiment we posted flyers to recruit participants. On the day of the experiment we walked through the facility to recruit participants in person. We recruited blue-collar workers from the warehouse, white-collar workers from the customer service and accounts receivable departments, and a few supervisors from all three departments.[3] Each session was run at the end of the workday and we designed the protocol to minimize the time commitment of our participants. We gave participants a survey to fill out before the experiment when we recruited them, before the experiment was conducted; most filled out the survey during their afternoon break, approximately two hours before the experiment. This allowed us to keep the experiment to half an hour, on average.[4] At the beginning of the survey we stressed that the responses would be anonymous and not shared with the employer.

At the experiment, participants handed in their surveys, were paid a $10

[3]Approximately 60% of participants were from the warehouse and the remainder from the office. More than 75% of the employees had worked for the company for more than a year at the time we conducted our study. Approximately 45% earned less than $30,000, 45% earned between $30,000 and $50,000, and 10% earned more than $50,000.

[4]Having subjects fill out the survey prior to the experiment is not standard practice, but we followed the same procedure in all treatments and we do not expect this procedure to have had differential effects on the different subject pools.

show-up fee and given a participant number that they were told to keep to themselves. Participants were then given written instructions and told to follow along as one of the experimenters read aloud. After any questions were answered, we flipped a coin to see whether the people with odd or even participant numbers would become proposers. Responders were taken to a different break room and waited silently for the proposers to make their decisions. Proposers were asked to choose between eleven discrete allocations of the hundred dollars: (0, 100), (10, 90), (20, 80), ..., (100, 0). When all the proposal forms were completed, one experimenter brought them to the other room and distributed them, face down, randomly to the responders. In the UG, responders circled either Accept or Reject. When all the responders were finished, the proposal forms were collected and the responders were paid, one at a time. In the DG, recipients were allowed to see what had been allocated to them by the dictator, the forms were collected, and then each recipient was paid, one at a time. Each second-mover was then free to go. After paying the second-movers, the proposal forms were given back to each first-mover. First-movers were then paid one at a time and allowed to leave.

The procedures for the student sessions (both at Middlebury and at KCKCC) were similar, except for the following minor variations. Because it was not obvious what convenient times for sessions would be at KCKCC, the students there were recruited by posters on bulletin boards which asked students to return a response card indicating interest at a choice of particular dates and times. Letters or phone calls were used to confirm participation. The Middlebury students were recruited by email rather than by flyers. However all recruiting materials contained the same information (the dates and anticipated length of the experiment, the amount of the show-up fee, etc.). Second, all students filled out their surveys once they arrived at the experiment (before making decisions), rather than a few hours prior to the experiment as in the warehouse.

4 Comparing Distributions Across Locations

In this section, we compare the distributions of responses across locations. The comparison of the Middlebury distribution with the KCKCC distribution gives us a rough estimate of the effect of demographic differences, holding social framing constant. The comparison of the KCKCC distribution with the workplace distribution gives us a rough estimate of the effect of social framing, holding individual characteristics constant. In the next section we will augment this analysis by adding demographic controls.

Summary Statistics for Ultimatum Games			
	Middlebury	**KCKCC**	**KC Workers**
Observations	20	18	30
Mean Offer	0.41	0.50	0.45
Median Offer	0.45	0.50	0.50
Minimum Offer	0.10	0.50	0.00
Maximum Offer	0.60	0.50	0.70
Standard Deviation	0.13	0.00	0.15
Rejection Rate	1 of 20	0 of 18	2 of 30
Highest Rejected Offer	0.10	NA	0.10

Table 3 – Data comparisons for the ultimatum game (Note: KC means Kansas City and KCKCC means Kansas City Kansas Community College).

Consider first the results for the UG. Table 3 presents summary statistics and Figure 1 presents histograms for the distribution of offers in each location, with the fraction of the initial $100 offered by the proposer to the responder on the horizontal axis, and the fraction of proposers making the offer on the vertical axis. It appears that proposers at both KCKCC and the distribution center made higher offers overall than the Middlebury students. All 18 offers at KCKCC were for 50-50 splits. There were a few less generous offers at the distribution center, but over 70% of proposers offered the 50-50 split. At Middlebury, by contrast, although the 50-50 split was the mode, fewer than half of proposers made this offer. Table 4 reports statistical tests of these differences. We employ two tests: the Wilcoxon test of differences in central tendencies and the Kolmogorov-Smirnov test for differences in cumulative distributions. The tests indicate that the Middlebury distribution is significantly different from the KCKCC distribution. The difference between the Middlebury distribution and the workplace distribution, however, is only marginally significant according to the Wilcoxon test, and insignificant according to the Kolmogorov-Smirnov test. The KCKCC and distribution center results are not significantly different from each other.

Difference Tests for Ultimatum Games		
	KCKCC	**KC Workers**
Middlebury	Z=-2.94, p<0.01	Z=-1.82, p=0.07
	KS=0.50, p=0.01	KS=0.30, p=0.20
KCKCC		Z=1.16, p=0.24
		KS=0.20, p=0.70

Table 4 – Wilcoxon (Z) and Kolmogorov-Smirnov (KS) tests for differences in the ultimatum game (Note: KC means Kansas City and KCKCC means Kansas City Kansas Community College).

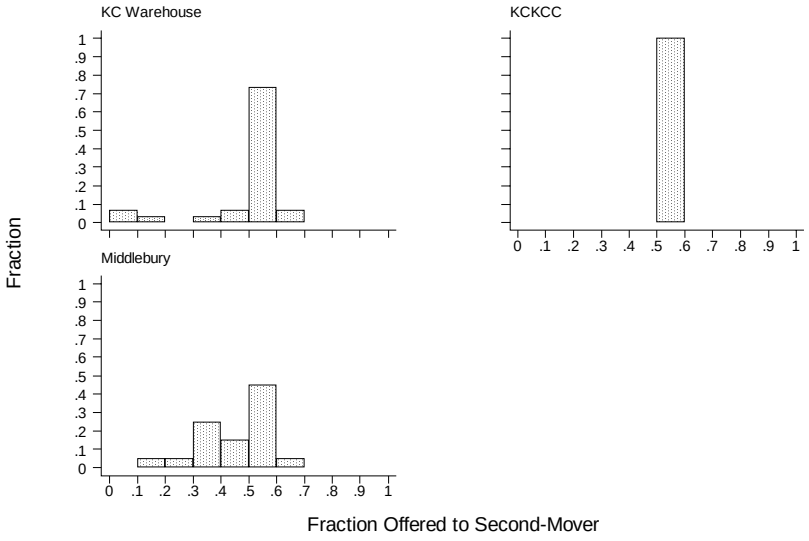

Figure 1 – The effect of social framing on offers in the 100 dollar ultimatum game.

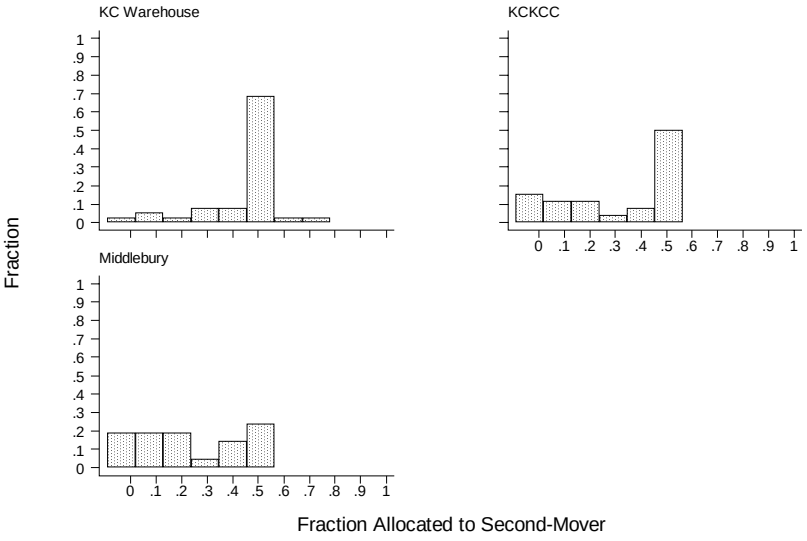

Figure 2 – The effect of social framing on allocations in the 100 dollar dictator game.

Consider next the results for the DG. Table 5 presents summary statistics and Figure 2 presents histograms of the distributions of offers. In this case, the distribution of KCKCC offers appears to be intermediate between that of the Middlebury students and the Kansas City workers. The mean and median offers, for instance, fall between those of the other locations. Table 6 presents statistical tests of the differences in distributions. In this case, the distribution center results are significantly different from both the KCKCC and the Middlebury results, while the KCKCC and Middlebury results are not significantly different from each other.

Summary Statistics for Dictator Games			
	Middlebury	**KCKCC**	**KC Workers**
Observations	21	26	37
Mean Allocation	0.25	0.33	0.45
Median Allocation	0.20	0.45	0.50
Minimum Allocation	0.00	0.00	0.10
Maximum Allocation	0.50	0.50	0.70
Standard Deviation	0.19	0.20	0.12

Table 5 – Data comparisons for the dictator game (Note: KC means Kansas City and KCKCC means Kansas City Kansas Community College).

Difference Tests for Dictator Games		
	KCKCC	**KC Workers**
Middlebury	Z=-1.43, p=0.15	Z=-4.17, p<0.01
	KS=0.26, p=0.33	KS=0.52, p<0.01
KCKCC		Z=-2.63, p<0.01
		KS=0.30, p=0.09

Table 6 – Wilcoxon (Z) and Kolmogorov-Smirnov (KS) tests for differences in the dictator game (Note: KC means Kansas City and KCKCC means Kansas City Kansas Community College).

What can we take away from these comparisons? First, the fact that Middlebury proposers appear to have made lower offers in both games than proposers in the other locations – in particular, lower than proposers at KCKCC, with similar social framing – suggests that there may indeed be an effect of individual characteristics. The older subjects in Kansas City with less experience with higher education appear to make higher offers than the elite college students in Vermont, although we should keep in mind that the difference between Middlebury and KCKCC is only significant in the UG. Second, the fact that KCKCC proposers made less generous offers than the distribution center workers in the DG suggests that social framing may be important as well.

Tests for Differences Between Ultimatum and Dictator Behavior		
	Wilcoxon	**Kolmogorov-Smirnov**
Middlebury	Z=2.66, p<0.01	KS=0.47, p<0.01
KCKCC	Z=3.47, p<0.01	KS=0.50, p<0.01
KC Workers	Z=0.34, p=0.73	KS=0.07, p=1.00

Table 7 – Wilcoxon (Z) and Kolmogorov-Smirnov (KS) tests for differences between the ultimatum and dictator games (Note: KC means Kansas City and KCKCC means Kansas City Kansas Community College).

As in Forsythe et al. (1994), we can also compare behavior in the UG to behavior in the DG within each subject population. In Table 7 we see that the workers behave differently from both groups of students on this dimension, because their allocations do not drop between the UG and the DG. That is, once the threat of veto by the second-mover is taken away, and choices solely reflect the generosity of the proposers, the workplace framing appears to lead subjects to allocate more to the recipient. It is important to note that, because

the demographics between KCKCC and the distribution center do not match exactly and because there may be subtle differences in social framing between Middlebury and KCKCC, these results comparing overall distributions remain suggestive. To better tease apart the effects of individual characteristics and social framing we now turn to regression analyses.

5 Regression Results

As mentioned above, the advantage of having run the same experiment at KCKCC as well as at Middlebury College and the distribution center is that we can use the variation in subject pools between KCKCC and Middlebury to estimate the effect of individual characteristics separately from the effect of social framing. There are a variety of ways in which the relationship between the individual characteristics and the social framing could be modeled econometrically. In our baseline estimates, we take the simplest, most straightforward approach, and assume that the effects of observable individual characteristics and social framing are additively separable. That is, we estimate a model of the following form:

$$f_i = \beta_0 + T_{1,i} \cdot \beta_1 + T_{2,i} \cdot \beta_2 + x' \cdot \beta_3 + \in_i$$

where i indexes individuals, f is the fraction offered to the responder by the proposer, T_1 is an indicator for KCKCC, T_2 is an indicator for the KC distribution center, and x is a vector of demographic characteristics (age, sex, years of schooling, family income, dummy variable for African-American and a dummy variable for non-white, non-African-American) and $\in$ is an error term.

Note that we do not explicitly include a term for geographic region. If we were to include, for instance, a dummy for Kansas City, it would be exactly collinear with T_1 and T_2. Rather, if we are correct in seeing the social framing of KCKCC as similar to the social framing of Middlebury, and if our observed demographic variables adequately capture the remaining variations in individual characteristics, then the coefficient T_1 can be interpreted as the regional "Kansas City effect", and the difference $T_2 - T_1$ can be thought of as the difference between the "college student" frame and the "warehouse employee" frame.

We think it is important to be careful when interpreting differences between subject groups. Implicit in our formulation are two key assumptions. The first is that the coefficients on the demographic terms do not vary across locations.[5] The second is that the treatment variables, T_1 and T_2, are uncorrelated with the error term. This amounts to an assumption that conditional on observable characteristics (and unobservable characteristics exactly collinear with T_1 and

[5]Botelho et al. (this volume) provide an insightful discussion of the pitfalls of this assumption. To examine the validity of our assumptions about the additive separability of the effects of demographics and social framing, we also estimated an OLS model with a complete set of interaction terms of KCKCC and KC Worker with all other independent variables. We then tested the restriction that all the interactions are jointly zero, and could not reject this hypothesis at conventional levels of significance.

T_2 , as discussed above) the unobservable characteristics of individuals are not related in a systematic way to the location of the experiment. This assumption is admittedly restrictive. Ideally, we would be able to conduct an experiment in which we could randomly assign subjects to different locations and social contexts, but since that experiment is infeasible, we feel that the assumption that subjects' unobservable characteristics are ignorable, conditional on differences in their observable characteristics, is a reasonable first step.

Analysis of Ultimatum Game Offers (dependent variable equals fraction of endowment offered to the second mover)		
	(1)	(2)
	Tobit	Interval
T_1, KCKCC	0.14	0.14
	(0.02)	(0.00)
T_2, KC Worker	0.10	0.10
	(0.11)	(0.00)
Mach Score	-0.0002	-0.0002
	(0.88)	(0.77)
Age	-0.003	-0.003
	(0.21)	(0.00)
Female	-0.04	-0.04
	(0.31)	(0.00)
Schooling (years)	-0.009	-0.009
	(0.20)	(0.02)
Family Income	7.6e-8	8.0e-8
	(0.78)	(0.13)
African American	0.05	0.05
	(0.48)	(0.40)
Not African American,	-0.03	-0.03
Not White	(0.56)	(0.66)
Intercept	0.60	0.63
	(0.00)	(0.00)
N	65	65

Table 8 – The determinants of offers in the ultimatum game (Notes: p-values in parentheses; we report marginal effects conditional on being uncensored for the Tobit regression; we assume that decision-makers always choose the allocation at the bottom of whatever interval in which their "true" value lies for the interval regression; errors for the interval regression are clustered by location).

An additional word of caution about the "Kansas City effect". There have been many recent economics experiments that seek to explain variations in subject behavior in terms of "culture". However, most such studies use a definition

of culture that is quite loose, and ours is no exception.[6] While we think our interpretation of the difference $T_2 - T_1$ as a social framing effect can be straight forwardly linked to existing experimental work, we have no developed theory about why Kansas City should be regionally distinctive, and so we are essentially using the idea of regional cultural differences as black box in our interpretation of T_1.

We first consider results for the UG. Column 1 of Table 8 presents Tobit results of our baseline model. We use the Tobit procedure to account for the fact that our dependent variable (the fraction of the pie offered) is bounded between 0 and 1. The coefficient on T_1 is significant at the 98% level, and indicates that, conditional on demographic characteristics and being uncensored, proposers at KCKCC on average offered 14% more of the initial sum to responders than did proposers at Middlebury, the omitted category. The coefficient on T_2 is also positive, and indicates that on average proposers at the warehouse offered 10% more than proposers at Middlebury, although the p-value of 0.11, while suggestive, is just below the 90% conventional significance level. More importantly, the coefficients on T_1 and T_2 are not statistically different from each other (p=0.38) which suggests that location differences matter in the UG.

Note that the Tobit estimator treats the fraction sent as continuous within the unit interval. In fact, proposers were constrained to choose among eleven discrete offers, between \$0 and \$100. Given the discrete and cardinal nature of the dependent variable, we think that the interval regression estimator is more reasonable. Column 2 of Table 8 presents interval regression results for the same model. To create the intervals for each participant's choice we assumed that decision-makers always choose an allocation that is at the bottom of the interval in which their true choice lies. For example, if a participant really wants to allocate 25% to the second-mover, we assume they will pick 20% instead of 30%. Therefore, the interval assigned to a 20% allocation is [0.20, 0.29].[7] Switching to the interval estimator also allows us to better deal with heteroskedasticity by clustering our errors by location. The results are stronger than the Tobit results and the interval regression, in general, is a better fit. Both the coefficient on T_1 and the coefficient on T_2 are now significant at better than the 99% level and a number of other demographic effects become significant. We see that offers are decreasing in age and years of schooling and that women offer less than men. Our more precise interval regression results now suggest a significant difference between the coefficients on T_1 and T_2 (p<0.01). KCKCC students offered more than the warehouse workers, and the warehouse workers offered more than the Middlebury students. Theses results suggest that behavior is not dominated by location differences in the UG. There appear to be countervailing forces at

[6] For an example of an experimental study which sets a higher standard, see Nisbett and Cohen (1996).

[7] Two referees suggested that this model of choice was consistent with expected utility theory given the discrete set of allocations. Our first instinct was to allow decision-makers to move in both directions. Specifically, we simply assumed that people picked whichever allocation was closest to their true preference. In this case, an observed allocation of 20% was assigned the interval [0.15, 0.25]. As one would expect, the difference in the results is miniscule.

work. Location increases offers, but the social frame of the workplace partially reduces them.

Analysis of Dictator Game Allocations (dependent variable equals fraction of endowment allocated to the second player)		
	(1)	(2)
	Tobit	Interval
T_1, KCKCC	0.03	0.02
	(0.67)	(0.23)
T_2, KC Worker	0.14	0.12
	(0.05)	(0.00)
Mach Score	-0.003	-0.003
	(0.02)	(0.03)
Age	0.003	0.003
	(0.20)	(0.11)
Female	-0.008	-0.01
	(0.83)	(0.74)
Schooling (years)	-0.007	-0.007
	(0.38)	(0.01)
Family Income	-1.5e-7	-1.9e-7
	(0.70)	(0.13)
African American	-0.14	-0.12
	(0.01)	(0.22)
Not African American,	0.10	0.09
Not White	(0.11)	(0.00)
Intercept	0.64	0.65
	(0.00)	(0.00)
N	81	81

Table 9 – The determinants of allocations in the dictator game (Notes: p-values in parentheses; we report marginal effects conditional on being uncensored for the Tobit regression; we assume that decision-makers always choose the allocation at the bottom of whatever interval in which their "true" value lies for the interval regression; errors for the interval regression are clustered by location).

We now turn to the DG results. Column 1 of Table 9 reports the Tobit results for the DG. The coefficient on the KCKCC dummy is no longer significant, suggesting that in the DG there is no "Kansas City" effect. The coefficient of the warehouse treatment is quite a bit larger than the coefficient on the KCKCC treatment and significant at the 95% level. Furthermore, we can reject the hypothesis that the coefficient on the KCKCC and warehouse treatments are equal (p=0.03). As in the UG, our DG interval specification (column 2) fits the data better. The coefficient on the warehouse treatment is significantly different from both the Middlebury and the KCKCC treatments at better than

the 99% level. While these results should not be overstated, they provide some evidence that the social framing of the workplace is important in the DG game.

Analysis of Dictator Game Allocations (dependent variable equals fraction of endowment allocated to the second player)			
	(1)	(2)	(3)
	Middlebury College	KCKCC	KC Workers
DG indicator	-0.16	-0.13	0.005
	(0.001)	(0.00)	(0.87)
Mach Score	-0.003	-0.004	-0.001
	(0.16)	(0.01)	(0.26)
Age	-0.08	0.006	0.0004
	(0.07)	(0.004)	(0.83)
Female	-0.02	0.10	-0.01
	(0.66)	(0.02)	(0.69)
Schooling (years)	0.04	0.001	-0.01
	(0.31)	(0.88)	(0.14)
Family Income	-9.2e-08	-3.5e-07	-1.1e-06
	(0.72)	(0.73)	(0.25)
African American		-0.27	0.01
		(0.00)	(0.72)
Not African American,	0.11	0.08	-0.04
Not White	(0.12)	(0.03)	(0.59)
Intercept	1.66	0.67	0.75
	(0.00)	(0.00)	(0.00)
N	41	43	62

Table 10 – Controlled tests for the difference in ultimatum and dictator behavior (Notes: p-values in parentheses; we assume that decision-makers always choose the allocation at the bottom of whatever interval in which their "true" value lies for the interval regression; errors are robust).

Among the demographic factors in our interval regression, both the Mach score (p=0.03) and the years of schooling (p=0.01) variables are associated with a lower fraction offered and being non-white nor African American (p<0.01) is associated with being more generous.[8] In addition, the positive effect of age on allocations is on the boundary of conventional significance (p=0.11). The

[8] A referee hypothesized that our years of schooling variable might have been better modeled as an exposure to college indicator variable. The idea was that exposure to college might affect behavior more than simply adding another year of schooling. Because some of our warehouse participants have been exposed to college the indicator is not collinear with our treatments. However, adding this variable or replacing the years of schooling variable does not improve our estimates. In the UG, the variable is significant but it's coefficient is similar in magnitude to the years of schooling regressor in the original specification. In the DG, the college variable is not significant (either with the years of schooling variable or on it's own). Further, the log likelihoods are worse in the new regressions. Based on this evidence we think the current specification is appropriate.

result for the Mach score is particularly noteworthy, since it corresponds to our theoretical expectation: high Machs may offer a fair split in the UG, even if they have selfish preferences, because they believe responders will reject fair offers, but once they no longer have to worry about the veto power of responders, they will reduce their offers.[9]

As a final exercise we examine how robust our comparisons of the UG and DG are when we control for demographic factors. In Table 10 we regress the fraction of the $100 endowment sent on an indicator variable for the DG and the same personal characteristics as in tables 8 and 9.[10] We organize our analysis by location. We see that, controlling for demographic factors, Middlebury college students allocate 15% less in the DG than in the UG ($p<0.01$), KCKCC students allocate 13% less ($p<0.01$), but workers in Kansas City offer the same amount, roughly half the pie, in both games ($p=0.87$). Considering demographic determinants *within a population*, we see that few factors matter in Middlebury and at the warehouse, while among KCKCC students a number of our regressors are significant. At KCKCC, controlling for the effect of the rules of the game, higher Machs and African Americans allocate less and older students, women, and people who describe their ethnicity as neither white or African American all allocate more to the second-mover.

6 Concluding Remarks

What do our results suggest about the external validity of results in the Ultimatum and Dictator Games? In the UG, we have two results: we find a "Kansas City" effect, a label we give to the fact that differences across regions (which could be cultural in origin) appear to affect behavior in the UG, and we find a social framing effect in which warehouse workers offer more than college students in Vermont, but less than college students in Kansas city. Combined, and controlling for demographic differences, we can order offers in the UG from highest to lowest KCKCC>KC Warehouse>Middlebury. In the DG, we find a highly significant effect of social framing: dictators are more likely to choose an equal allocation in the warehouse, even controlling for observable demographic characteristics. In addition, the mean offers of students drop significantly from the UG to the DG, while those of workers do not.

What is the economic significance of these results? We offer two answers, a narrow one and a broader one. Although the range of variation in observed behavior across our subject groups and social framing treatments is much smaller than that found across fifteen small societies by Henrich et al. (2001), a narrow conclusion would be that, while our results qualitatively suggest the external va-

[9] These results, consistent with prior expectations about Machiavellian behavior, contrast with our results in a trust game reported in Burks et al. (2003), in which high Machs were not less trustworthy than others, although we would have expected high Machs to behave opportunistically and not reward other players who had trusted them.

[10] We continue to use the interval regression procedure. Notice that the African American regressor has been dropped in the Middlebury regression because none of the participants at the college fell into this category.

lidity of standard UG results, they also show some limitations in the precision of external extrapolation: call this a "limitation in calibration". We observe enough variation in UG behavior to suggest that, even within an advanced industrial society, the specific patterns observed in trials with young, four-year, full-time college students, under an intra-collegiate social framing, should not be automatically assumed to translate precisely into the patterns of UG behavior to be expected among other subject groups or with other frames. However, we feel less comfortable explaining our DG differences in terms of calibration. By comparing students to workers in the DG where normative behavior is unconfounded by strategic considerations we see that in interactions with a more economically significant frame (e.g., within the workplace), altruistic norms affect behavior to a greater degree than in the classroom.

More broadly, our results may be of some interest to those (like us) who find other-regarding, or "social preference," explanations for UG and DG behavior attractive. Placed in this interpretive framework, our results suggest an interesting and consistent story. High offers in the UG are here taken to be a mixture of strategic avoidance of rejection by selfish but risk-averse subjects, along with fair-mindedness by subjects with social preferences. The DG then provides a check on the extent to which these two different motivations are at work. In this regard the two student subject groups are essentially similar—there is an extremely sharp drop in offers from the UG to the DG. This shows that few high offers in the UG are made by fair-minded student subjects; most are made by selfish subjects worried about rejection. (In this context, the fact that the KCKCC students offer more in the UG than do the Middlebury students would be most parsimoniously explained by higher risk aversion among the KCKCC student group.)

However, the KC warehouse workers are quite distinctive in comparison, because their offers do not change from UG to DG. Conditional on the social preference interpretation of subject behavior in these experiments, this suggests that something about the social framing of the warehouse has shifted the behavior of worker subjects sharply towards fair-mindedness: many more of the high offers by workers in the UG are due to an intrinsic preference for sharing gains with their co-workers. Because the overlap in demographic characteristics across our subject pools is imperfect (in particular, with respect to age between KCKCC and the KC warehouse), as well as because of the always present potential for significant unobservable differences, this evidence is only suggestive, but it is nonetheless quite interesting.

Our findings suggest a few directions for future research. We should continue experimenting in the field to get a better sense of the size of the variations in external validity "calibration" mentioned above. At this point we have only one observation of a 10% difference in the UG (and a 13% difference in the DG). We have no idea how robust this estimate is. Second, we might well ask what is it about the nature of social interactions in workplaces that reinforces prosocial behavior in these experiments, presumably through reinforcing prosocial norms? Does this happen in all workplaces, or is there something distinctive about our particular warehouse? Do all groups of workers behave similarly, or do

boundaries within the workplace, such as between blue collar and white collar, or between labor and managers, ever matter? There is substantial field and experimental evidence that norms against free-riding and in favor of cooperation are particularly strong among work groups (e.g., Acheson, 1988; Ostrom, 1990). It would be interesting to investigate whether this is especially true in cases where workers produce in teams and their individual contributions to group productivity are difficult to distinguish, as suggested by Tyler and Blader (2000).

Acknowledgements: We thank Julia Assael, Gary Carpenter, Pamela Carpenter, and Marla Weinstein for research assistance, Laura Burks for help with the data, Mary Grunke and Cathy Plaster for facilitating our experiments at KCKCC, three referees for their thoughtful comments, and the MacArthur Foundation's norms and preferences working group for generous financial support. In addition Carpenter acknowledges the National Science Foundation (SES-CAREER 0092953) and Verhoogen acknowledges the MacArthur Foundation's cost of inequality network.

7 Bibliography

Abbink, K., Hennig-Schmidt, H., 2002. Neutral versus loaded instructions in a bribery experiment. Bonn Econ Discussion Paper 23/2002.

Acheson, J., 1988. The lobster gangs of Maine. University Press of New England, Hanover.

Binmore, K., Shaked, A., Sutton, J., 1985. Testing noncooperative bargaining theory: A preliminary study. American Economic Review 75(5), 1178-1180.

Botelho, A., Harrison, G., Hirsch, M., Rutstrom, E., this volume. Bargaining behavior, demographics and nationality: What can the experimental evidence show?

Burks, S., Carpenter, J., Verhoogen, E., 2003. Playing both roles in the trust game. Journal of Economic Behavior and Organization 51(2), 195-216.

Camerer, C., Fehr, E., forthcoming. Measuring social norms and preferences using experimental games: A guide for social scientists. In: Henrich, J., Boyd, R., Bowles, S., Gintis, H., Fehr, E., Camerer, C. (Eds.), Foundations of human sociality: Experimental and ethnographic evidence from 15 small-scale societies. Oxford University Press, Oxford.

Campbell, D., Stanley, J., 1963. Experimental and quasi-experimental designs for research. Rand McNally, Chicago.

Carpenter, J., 2002. Measuring social capital: Adding field experimental methods to the analytical toolbox. In: Isham, J., Kelly, T., Ramaswamy, S., (Eds.), Social capital and economic development: Well-being in developing countries. Edward Elgar, Northampton, pp.119-137.

Carter, J., Irons, M., 1991. Are economists different, and if so, why? Journal of Economic Perspectives 5(2), 171-177.

Christie, R., Geis, F., 1970. Studies in Machiavellianism. Academic Press, New York.

Cookson, R., 2000. Framing effects in public goods experiments. Experimental Economics 3(1), 55-79.

Eckel, C., Grossman, P., 1996. Altruism in anonymous dictator games. Games and Economic Behavior 16, 181-191.

Elliott, C., Hayward, D., Canon, S., 1998. Institutional framing: Some experimental evidence. Journal of Economic Behavior & Organization 35(4), 455-464.

Fehr, B., Samson, D., Paulhus, D., 1992. The construct of Machivellianism: Twenty years later. In: Spielberger, C., Butcher, J., (Eds.), Advances in personality assessment. Lawrence Erlbaum Associates, Hillsdale, NJ, 77-116.

Forsythe, R., Horowitz, J., Savin, N., Sefton, M., 1994. Fairness in simple bargaining experiments. Games and Economic Behavior 6, 347-369.

Gueth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. Journal of Economic Behavior and Organization 3, 367-388.

Gueth, W., Tietz, R., 1990. Ultimatum bargaining behavior: A survey and comparison of experimental results. Journal of Economic Psychology 11, 417-449.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., 2001. In search of homo economics: Behavioral experiments in 15 small-scale societies. American Economic Review 91(2), 73-78.

Hoffman, E., McCabe, K., Shachat, J., Smith, V., 1994. Preferences, property rights, and anonymity in bargaining games. Games and Economic Behavior 7, 346-380.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of choice under risk. Econometrica 47, 263-291.

Kahneman, D., Tversky, A., 2000. Choices, values, and frames. Cambridge University Press, Cambridge.

McHoskey, J., Worzel, W., Szyarto, C., 1998. Machiavellianism and psychopathy. Journal of Personality and Social Psychology 74(1), 192-210.

Meyer, H.-D., 1992. Norms and self-interest in ultimatum bargaining: The prince's prudence. Journal of Economic Psychology 13, 215-232.

Murnighan, K., Saxon, M., 1998. Ultimatum bargaining by children and adults. Journal of Economic Psychology 19, 415-445.

Nisbett, R., Cohen, D., 1996. Culture of honor: The psychology of violence in the south. Westview Press, Denver.

Ortmann, A, this volume. Field experiments in economics: Some methodological caveats.

Ostrom, E., 1990. Governing the commons: The evolution of institutions for collective action. Cambridge University Press, Cambridge.

Panitz, E., 1989. Psychometirc investigation of the mach iv scale measuring Machiavellianism. Psychological Reports 64, 963-968.

Park, E.-S., 2000. Warm-glow versus cold-prickle: A further experimental study of framing effects on free-riding. Journal of Economic Behavior & Organization 43(4), 405-421.

Roth, A., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An experimental study. American Economic Review 81(5), 1068-1095.

Tyler, T., Blader, S., 2000. Cooperation in groups: Procedural justice, social identity, and behavioral engagement. Psychology Press, Philadelphia.

Willinger, M., Ziegelmeyer, A., 1999. Framing and cooperation in public goods games: An experiment with an interior solution. Economics Letters 65(3), 323-328.

Wrightsman, L., 1991. Interpersonal trust and attitudes towards human nature. In: Robinson, J., Shaver, P., Wrightsman, L., (Eds.), Measures of personality and social psychological attitudes. Academic Press, San Diego, 373-412.