FRIEND OR FOE? A NATURAL EXPERIMENT OF THE PRISONER'S DILEMMA

John A. List*

Abstract—This study examines data drawn from the game show *Friend or Foe?* which is similar to the classic prisoner's dilemma tale: partnerships are endogenously determined, and players work together to earn money, after which they play a one-shot prisoner's dilemma game over large stakes: varying from $200 to (potentially) more than $22,000. The data reveal several interesting insights; perhaps most provocatively, they suggest that even though the game is played in front of an audience of millions of viewers, some of the evidence is consistent with a model of discrimination. The observed patterns of social discrimination are unanticipated, however.

I. Introduction

THE prisoner's dilemma has become *the* classic example of the theory of strategy and its implications for predicting the behavior of players. The resulting equilibrium concept has been used in many disparate strategic applications, including economic, social, political, and biological competitions. Indeed, its influence has extended well beyond scholarly circles, as people who have never been formally introduced to game theory typically are familiar with the basic story underlying the game: Two people form a team and rob a bank. They are subsequently apprehended by the police and brought to the precinct under the suspicion that they were conspirators in the bank robbery. Detectives explain to each suspect that even if neither confesses they are both looking at jail time (say, 2 years) for the robbery. If one confesses, however, the confessor goes free and the other serves substantial jail time (say, 10 years). If both confess, then jail terms will be negotiated down (say, 5 years). A simple transformation of this story produces many realistic economic circumstances, including oligopoly pricing, international trade negotiations, and labor arbitration.

Ever since the initial theoretical conjectures of Nash, economists have used the prisoner's dilemma game as a basis for testing the predictions of noncooperative game theory. Yet almost all of this evidence comes from laboratory experiments that involve small or imaginary stakes.[1] In this study, I provide an initial exploration of strategic behavior in a natural prisoner's dilemma experiment by examining individual decisions from the game show *Friend or Foe?* The game show provides a simple and sharp form of the dilemma and parallels the classic prisoner's dilemma tale: teams are endogenously determined, and players work together to earn money, after which the agents play a one-short prisoner's dilemma game over large stakes—varying from $200 to (potentially) more than $22,000. An added advantage of these data is that they provide a test of whether, and to what extent, social discrimination is practiced.[2]

Several interesting findings emerge. First, in the series of observed shows, thousands of dollars were left on the table: in nearly 25% of the 117 games, both players chose not to cooperate, resulting in a net loss of nearly $100,000. In roughly 25% of the cases both players cooperated. Second, stakes are not found to be a significant determinant of play, lending evidence consistent with the notion that the low stakes typically used in laboratory experiments are not problematic as such. Third, an empirical analysis examining the determinants of individual cooperation rates suggest little social discrimination exists; but because the teams are formed endogenously, the data also permit an exploration of whether players exhibit discrimination when choosing their partners. The data suggest that, in general, players use proper backward induction when selecting partners; but one anomalous finding is that in the partner selection process agents are biased against older participants. The observed pattern of partner choices is consistent with a model of social discrimination wherein certain populations have a general "distaste" for older participants in the Becker (1975) sense.

The remainder of this study proceeds as follows. Section II discusses the *Friend or Foe?* game show and describes important caveats associated with using such data to draw inferences. Section III summarizes the empirical results. Section IV concludes.

II. The Natural Experiment

As previous studies have noted, television game shows provide a natural venue to observe real decisions in an environment with high stakes. For example, Berk, Hughson, and Vandezande (1996) study contestants' behavior on *The Price is Right* to investigate rational decision theory. Gertner (1993) and Beetsma and Schotman (2001) make use of data from *Card Sharks* and *Lingo,* respectively, to examine individual risk preferences. Finally, Metrick (1995) uses data from *Jeopardy!* to analyze behavior under uncertainty and players' ability to choose strategic best responses. Each of these studies has provided a fresh look at important

Received for publication October 7, 2003. Revision accepted for publication August 30, 2005.

* University of Chicago and NBER.

Many thanks to the Editor—Daron Acemoglu—and two anonymous reviewers for very insightful comments that considerably improved this study. I would like to thank Jonathan Alevy, Jennifer List, and Alexandre Vialou for watching numerous *Friend or Foe?* shows and recording the data. Joel Waldfogel kindly lent me his *Friend or Foe?* data to cross-check my data. Comments from Uri Gneezy and Liesl Koch improved the manuscript. Seminar participants at various universities also provided useful comments. Any errors remain my own.

[1] For thoughtful surveys see Roth (1995) and Davis and Holt (1995).

[2] Independently, Oberholzer-Gee, Waldfogel, and White (2005) have used *Friend of Foe?* data to explore social learning and coordination but they do not examine discrimination or other questions addressed herein.

The Review of Economics and Statistics, August 2006, 88(3): 463-471

economic phenomena over large stakes that, short of a large experimental budget or having an experimental laboratory in an underdeveloped country, would be difficult to investigate outside a game-show environment.

Data used in this paper are taken from original *Friend or Foe?* programs. The game show *Friend or Foe?* premiered on June 3, 2002. Importantly, all of these shows were taped in Santa Monica, CA, prior to the show's television premiere. I observed 39 shows, each of which consists of six strangers, who at the beginning of the show are randomly split into two groups of three: group 1 and group 2. Each group 1 agent privately selects one player from group 2 to be his or her partner. This first stage can therefore proceed in one of three ways:

1. All three group 1 agents select different partners. The partnerships are then formed by these choices, and the pairs proceed to the next stage of the show.
2. Two group 1 agents select the same group 2 agent. First, the group 1 agent whose choice did not coincide with another's choice is paired with his or her group 2 selection. Second, the group 2 agent who was chosen by two group 1 agents selects one of those group 1 agents to be his or her partner. The remaining pair is then matched.
3. All three group 1 agents choose the same group 2 agent. In this case, the selected group 2 agent chooses one of the group 1 agents. The remaining group 1 agents each privately select one of the remaining group 2 agents. If they again select the same person, he or she chooses one of the group 1 agents. The remaining pair is then matched.

Teams are therefore determined endogenously: the six strangers become three groups of two via this selection process. Each team is then separated into "isolation chambers" where trivia rounds are played. The newly formed teams work together and agree on answers to trivia questions in order to build a *trust fund,* potentially over three rounds. In the first (second) round, four trivia questions are posed worth $500 ($1,000) each. In the third round, up to ten questions are asked, each worth $500. Given that each team is initially endowed with $200, a team's trust fund can range from $200 to $22,200; in practice, however, the largest sum of trivia earnings in these data is $16,400.

At the end of each round (there are three rounds in total), the lowest-scoring team is eliminated. Before the team is dismissed from the game show, the players must decide how their winnings are divided. In making a decision, each player has a button, which no one else can see. The division of winnings depends on whether the player depresses his or her button. There are three possible outcomes: (1) *Friend-Friend* (that is, cooperate-cooperate)—if both players choose "friend" (do not depress the button), then the total winnings are divided equally between the two; (2) *Friend-Foe* (that is, cooperate–not-cooperate)—if only one player depresses the button, he or she receives the entire amount, leaving the other player with nothing; (3) *Foe-Foe* (that is, not-cooperate–not-cooperate)—if both players press the button, then both players walk away with nothing.

Thus, this final stage of the game naturally sets up a prisoner's dilemma with a weakly dominant strategy: each player has an incentive to play Foe because she is never worse off monetarily for doing so. To my knowledge, little has been done empirically on noncooperative games in such a high-stakes setting.

A. *Caveats*

Before proceeding to the discussion of results, it is important to summarize potentially important caveats associated with using game-show data for such an exercise. First, the contestant pool may not be a representative sample of the underlying population of interest. Second, play takes place in front of a large television audience, potentially attenuating certain behaviors. For example, it is possible that contestants recognize that they are playing a repeated game with members of the audience and therefore future financial results might well be an important consideration when determining whether to cooperate or defect on the game show. Indeed, one would expect that the more closely an individual's actions are being scrutinized (for example, by being televised), the more emphasis fairness concerns are likely to receive. Thus, defecting on a television show might well have very negative consequences, even if the agent has no social preferences.[3]

Besides plausibility, such intuition has some empirical support. For example, decreasing anonymity or confidentiality in classic linear public-goods experiments leads to increased cooperative behavior (see, for example, Masclet et al., 2003; Rege & Telle, 2004). And List et al. (2004) report evidence that social isolation has important effects on stated preferences for environmental goods. Perhaps making this overarching point most clearly are the data in List (2006). That investigation uses a series of laboratory and field experiments to explore the nature of social preferences among real market players in naturally occurring environments. He reports that agents drawn from a well-functioning marketplace reveal strong social preferences in tightly controlled laboratory experiments, but when they are not aware that they are being observed, their behavior in their naturally occurring environment approaches what is predicted by self-interest theory.

Third, the title of the game show, in and of itself, is suggestive, as is the oral introduction that the game show

[3] Thanks to the Editor—Daron Acemoglu—and both reviewers for urging me to explore this line of thought. Levitt and List (2005) present a model that uses such insights to explore the degree of generalizability of results from laboratory experiments to field environments.

host, Kennedy, uses in every show: you will "work together to build a trust fund . . . and ultimately decide to share [it] as friends or fight over [it] as foes." One aspect of the show that does resemble certain characteristics of many naturally occurring environments is the fact that agents endogenously select into the market (the show) and select with whom to interact. In many parallel laboratory experiments, such variables are determined exogenously.

III. Results

The top panel of table 1 provides summary statistics at the individual level. Of the 234 players, 49% were men and 78% were white. The game show, which is taped in California, attracted 38% of its contestant pool from California and the contestants were from a broad age group: from 18 to 61, with an average age of 31.2. This mixture of subjects provides important variation to examine whether participant-specific characteristics influence behavior in prisoner's dilemma games.

The bottom panel of table 1 presents an overview of monies earned in the trivia portion of the game, cooperation rates, and take-home earnings across a few broad classes of agents. These figures show that, on average, teams earned $3,705 during the trivia portion of the game, so that the average prisoner's dilemma game was played over $3,705. On average, men earned the most in the trivia portion of the game ($4,247), and nonwhites earned the least ($2,825). There is no discernible difference across earnings profiles of young ($3,603) and older ($3,839) agents when I split the players along the first moment of the age variable.

In terms of overall cooperation rates, 50% of subjects chose to cooperate. This figure is larger than cooperation rates observed in laboratory one-shot prisoner's dilemma experiments, which are typically around 33% (Shafir & Tversky, 1992). As previously noted, there are several potential explanations for such a disparity, including subject pool differences, television audience effects, that the monies are jointly earned, that the stakes are large, that partnerships are formed endogenously, and that the participants are not anonymous. The combination of trivia earnings with cooperation rates maps into take-home earnings, which were on average much larger for men (for younger agents) than for women (for more mature agents). Indeed, men took home almost 70% more than women. This, of course, is due to two effects: higher cooperation rates among women and higher trivia earnings among men.

Constructing a simple matrix using gender, race, and age as cross-tab variables yields table 2. Table 2 provides team-level information pertaining to the level of trivia earnings (called *stakes*), cooperation rates, average take-home earnings, and efficiency rates (take-home earnings divided by trivia earnings). For example, insights contained in panel A complement table 1, and suggest that women who are on all-female teams have a 56% cooperation rate (row 2, column 2), whereas findings contained in row 1, column 1 suggest that men who are on all-male teams cooperate in only 48% of the observations. This leads to a higher efficiency rate for all-female teams. Interestingly, in mixed-gender teams, cooperation rates are much different across men and women: whereas men cooperate at a rate of 43%, women choose to cooperate at a rate of 55%. In this case, men are showing a slight discrimination against women, cooperating slightly less (48% versus 43%) when paired with a female.

Likewise, panels B and C in table 2 provide insights across race and age cohorts. The overall picture is that white agents cooperate more than nonwhites, regardless of whether they are on all white teams or interracial teams. And older agents are found to cooperate to a much greater degree than younger agents, regardless of the age of the partner. The distribution of earnings in the mixed-age pairs, as well as the efficiency rates across the three cells, highlights the considerable effects of such behavior.

TABLE 1.—SELECTED CHARACTERISTICS OF PARTICIPANTS

	Mean (Std. Dev.)	Minimum	Maximum
Gender (% male)	0.49 (0.50)	0	1
Race (% white)	0.78 (0.42)	0	1
Age	31.2 (8.63)	18	61
% Californians	0.38 (0.49)	0	1
n	234	—	—
Outcomes	Trivia Earnings	Cooperation Rate	Take-Home Earnings
Overall (*n* = 234)	$3,705 (2,977)	0.50 (0.50)	$1,455 (2,308)
Men (*n* = 115)	$4,247 (3,294)	0.45 (0.50)	$1,834 (2,734)
Women (*n* = 119)	$3,183 (2,541)	0.56 (0.50)	$1,088 (1,738)
White (*n* = 182)	$3,957 (3,105)	0.53 (0.50)	$1,417 (2,322)
Nonwhite (*n* = 52)	$2,825 (2,294)	0.42 (0.50)	$1,587 (2,274)
Young (age <31) (*n* = 132)	$3,603 (2,805)	0.41 (0.49)	$1,592 (2,475)
Mature (age ≥ 31) (*n* = 102)	$3,839 (3,195)	0.63 (0.49)	$1,276 (2,070)

Notes:
1. *Gender* (*Race*) denotes a categorical variable: 1 if male (white), 0 otherwise; *Age* denotes actual age in years.
2. "Trivia earnings" means the amount earned in the trivia portion of the game show; "cooperation rate" means the percentage of subjects that chose to cooperate; "take-home earnings" means the amount of money actually taken home by the subject.

TABLE 2.—SUMMARY GROUP OUTCOMES

	A. Gender	
	Male	Female
Male	Stakes: $5,291 Cooperation rate: 0.48 Avg. take-home: $2,201 Efficiency: 0.83 $n = 48$	Stakes: $3,558 Cooperation rates: male, 0.43; female, 0.55 Avg. take-home: male, $1,572; female, $1,046 Efficiency: 0.74 $n = 134$
Female		Stakes: $2,623 Cooperation rate: 0.56 Avg. take-home: $1,142 Efficiency: 0.87 $n = 52$
	B. Race	
	White	Nonwhite
White	Stakes: $4,382 Cooperation rate: 0.51 Avg. take-home: $1,626 Efficiency: 0.74 $n = 134$	Stakes: $2,772 Cooperation rates: white, 0.58; nonwhite, 0.44 Avg. take-home: white, $832; nonwhite, $1,601 Efficiency: 0.88 $n = 96$
Nonwhite		Stakes: $3,450 Cooperation rate: 0.25 Avg. take-home: $1,425 Efficiency: 0.83 $n = 4$
	C. Age	
	Young	Mature
Young (less than 31 years old)	Stakes: $3,714 Cooperation rate: 0.40 Avg. take-home: $1,307 Efficiency: 0.70 $n = 68$	Stakes: $3,484 Cooperation rates: young, 0.42; mature, 0.63 Avg. take-home: young, $1,896; mature, $907 Efficiency: 0.80 $n = 128$
Mature (31 years old or older)		Stakes: $4,436 Cooperation rate: 0.63 Avg. take-home: $1,900 Efficiency: 0.86 $n = 38$

Rather than belabor the raw data further, to provide insights into team cooperation rates, I make use of the inherent ordering of the outcomes by coding team outcomes as equaling 0 if both players choose "not cooperate," equaling 1 if one player chooses "not cooperate" and the other chooses "cooperate," and equaling 2 if both players choose "cooperate," and I build a model around a latent regression of the form

$$Y_i^* = X_i'\beta + \varepsilon_i, \qquad (1)$$

where Y_i^* is unobserved, X_i is a vector of team-specific variables, β is the estimated response coefficient vector, and ε_i is the well-behaved random error component. Although I do not directly observe Y_i^*, I do observe an approximation to it:

$$Y_i = \begin{cases} 0 & if \quad Y_i^* \leq 0, \\ 1 & if \quad 0 < Y_i^* \leq \phi_1, \\ 2 & if \quad \phi_1 < Y_i^* \leq \phi_2. \end{cases} \qquad (2)$$

The ϕ_i are unknown parameters that are estimated jointly with β.

A few aspects of the estimation procedure merit further consideration. First, because the ϕ_i's are free parameters, there is no significance to the unit distance between the set of observed values of *Y*. In fact, the outcomes represent a *cooperation ranking,* and therefore 1-unit changes are not directly comparable. Second, estimates of the marginal effects in the ordered probability model are quite involved, because there is no meaningful conditional mean function. I therefore compute the effects of changes in the covariates on the *j*th probability:

$$\partial X_i \partial \text{Prob}[\text{cell } j] = [f(\phi_{\text{j}-1} - X_i'\beta) - f(\phi_j - X_i'\beta) \times \beta, \qquad (3)$$

where $f(\bullet)$ is the standard normal density, and the other variables are as defined above. By definition, these effects

TABLE 3.—ORDERED PROBIT ESTIMATION RESULTS[a,b,c]

Variable	Parameter Estimate	Marginal Effects		
		$P(0)$	$P(1)$	$P(2)$
Team Male	0.05 (0.23)	−0.01	0.00	0.01
Team Female	0.10 (0.22)	−0.03	−0.00	0.03
Team White	−0.29 (0.18)	0.08	0.00	−0.08
Team Nonwhite	−1.89 (1.12)	0.55	0.02	−0.57
Team Young	−0.50 (0.19)	0.15	0.00	−0.15
Team Mature	0.51 (0.22)	−0.15	−0.00	0.15
Team California	0.01 (0.001)	−0.00	0.00	−0.00
Team Non-California	−0.001 (0.001)	0.00	0.00	0.00
1st Selected	0.39 (0.18)	−0.11	−0.00	0.12
1st Selected (Tie)	0.77 (0.21)	−0.22	−0.01	0.23
Stakes	0.002 (0.001)	0.00	0.00	0.00

[a] "Team" variables are dichotomous and equal 1 if the team is so composed and 0 otherwise. For example, *Team Male* equals 1 if both players on the team are male, 0 otherwise.

[b] Marginal effects are calculated as changes in the covariates on the *jth* probability: ∂X_i ∂Prob[cell j]=$[f(\phi_{j-1} - X_i'\beta) - f(\phi_j - X_i'\beta)] \times \beta$.

[c] Figures in parentheses are standard errors.

must sum to 0, because the probabilities sum to 1. Accordingly, I obtain threshold levels of cooperation rates by measuring how the exogenous variable vector X_i affects the ranked responses Y_i^*. Here X_i includes dichotomous team regressors to explore the effects of team gender, race, age, and geographic residence on outcomes. For example, *Team Male* (*Team White*) equals 1 if both team members are male (white), and 0 otherwise; *Team Mature* (*Team California*) equals 1 if both team members are 31 years old or older (reside in California), and 0 otherwise.[4]

The vector X_i also includes two regressors to explore the potential importance of endogenous partner selection. As aforementioned, in markets many times partners are chosen endogenously and this fact might be important. The first variable indicates whether the individual became a partner by choice in the first selection round: *1st Selected* = 1 for both team members if the group 1 agent was the only person who selected the group 2 agent, and 0 otherwise. The second variable, *1st Selected(Tie),* equals 1 for both group members if a group 2 agent was chosen by at least two group 1 agents and subsequently was forced to choose his or her group 1 partner, and 0 otherwise. Intuition would suggest the reinforcement involved in the tie-breaking procedure would yield an even stronger bond between partners. Finally, I include a variable that controls for the stakes (the amount of money at risk) in X_i.

Estimation results are contained in table 3. Column 1 contains the coefficient estimates from the model, which is statistically significant at the $p < 0.05$ level. The results corroborate insights gained from the raw data: older teams are more likely to cooperate, and all-white teams cooperate less than mixed-race teams. The results also highlight the significance of allowing partners to be determined endogenously: those who were selected by one another to be teammates cooperate to a much greater extent than those who did not select to be teammates on the first attempt. Although these coefficient estimates provide interesting insights, not much information beyond their statistical significance can be used, because they are not marginal effects.

To amend this situation, I present marginal effects in columns 2–4 of table 3. The estimates can be read as follows: An all-female team is 3% more likely to be in cell 2 (both players cooperate) than a mixed-gender team. Interestingly, mature (young) teams are 15% more (15% less) likely to cooperate fully than mixed-age teams. The importance of the selection process is notable as well: those teams that were formed by both agents choosing one another are 23% more likely to fully cooperate. Interestingly, though the effect of stakes was marginally significant in column 1, it is negligible.

One can dig one level deeper into the individual decision process by exploring how individuals play the game. In this spirit, I estimate the following regression model at the individual level: $cooperate_i = g(\alpha + \beta_z Z)$, where $cooperate_i$ equals 1 if agent i chooses to cooperate, and 0 otherwise; $g(\bullet)$ is the standard logistic function (results are similar if I assume normally distributed errors); and Z is identical to X but at the individual, rather than the team, level.[5] Because these results merely corroborate evidence from tables 1–3, I suppress them, but note that individual-level attributes are correlated with cooperation rates: males cooperate less than females, whites cooperate less than nonwhites, older participants cooperate less than younger participants, and participants from California cooperate less than non-Californians. Again, stakes are found not to matter, but the partner selection process does. Finally, when I augment Z with observable partner characteristics, they are found to be statistically insignificant at conventional levels in the pooled data.

Overall, from results in tables 1–3 and the insights gained from the individual-level model, I draw the following conclusions:

Result 1. Stakes do not have an important effect on play.

[4] One rationale for including a control for whether the contestant resides in California is that this subset of players might be experienced game-show participants.

[5] It would have been nice to also include a regressor that indicated whether, and to what extent, each player contributed to the pool of earnings in the trivia stage. Given that each team of two must agree on an answer before final submission (that is, both team members must submit the same answer) and that the actual television broadcast only included select parts of the discussion between teammates, I found it too subjective to extrapolate the information aired into the construction of a "production variable."

Result 2. Individual-level attributes are correlated with cooperation rates: women, whites, and older participants cooperate more.

Result 3. Partner attributes do not influence play significantly.

Result 4. How partners are determined influences play: players cooperate more if they are able to select each other as partners.

Result 1 provides insights into the validity of a number of laboratory experiments that use low stakes. As Clark and Sefton (2001, p. 54) note, "In the experimental literature there is no consensus on the relationship between cooperation and stakes in social dilemmas." In this sense, the data herein show that over much larger stakes variations than previously examined, play is not considerably affected by changes in stakes. These results are consonant with recent results from ultimatum laboratory games that indicate that the level of stakes has negligible effect on proposers' behavior (see, for example, Slonim & Roth, 1998; Cameron, 1999; Carpenter et al., 2005). In terms of trust and gift exchange games, the laboratory evidence is mixed: Fehr, Fischbacher, and Tougereva (2005) report that fairness concerns play an important role for both low- and high-stakes games, whereas Parco et al. (2002) find that raising financial incentives causes a breakdown in mutual trust in centipede games.[6]

Besides its empirical significance, result 2 might also serve as an external validity check of recent results from laboratory experiments. For example, two recent studies that examine behavior in much different settings find similar results on gender effects. Using laboratory experimental data, Andreoni and Vesterlund (2001) find that in dictator games women are much more likely than men to be "equalitarians" (choose a division that gives equal payoffs). Likewise, Eckel and Grossman (1998) eliminate all factors other than gender-related differences in selfishness and find that women are less selfish than men in dictator games.

Result 3 indicates that partner attributes do not influence play. This result is inconsistent with previous laboratory results that suggest that trusting behavior and trustworthiness rise with social connection (see, for example, Glaeser et al., 2001; Fershtman & Gneezy, 2001; Andreoni & Petrie, 2004). Indeed, Andreoni and Petrie (2004, p. 6) note that "working with familiar others can reduce transactions costs, as familiarity can enhance trust." Yet, this finding is consistent with players not discriminating against other players due to the "television" effect, which makes sense in that discriminatory behavior should be attenuated in the game-show environment because the agent's actions are observed literally by millions of viewers.

Result 4 is interesting in that it highlights the importance of partner determination, a factor that is not often discussed and manipulated in laboratory experiments. Thus, it is clear that partner selection is crucial in determining play. And, though discrimination is not observed at the level of the dilemma, the data are sufficiently rich to allow an exploration of whether discrimination is evident in the partner selection process. In particular, I can analyze whether players systematically bias their choices in a manner that is consistent with models of discrimination.

To begin the examination, I estimate an earnings function of the following form:

$$Earnings_i = \delta' \mathrm{X_j} + \varepsilon_{ij}, \tag{4}$$

where $Earnings_i$ equals the amount person i would have earned had he played the strategy that maximizes his own payoff function for this single game: *Trivia earnings* × *Cooperate,* where *Cooperate* equals 1 if the individual's partner cooperated, and 0 otherwise.[7] In X_j, I follow the empirical analysis above and include observable characteristics for person *j:* gender, race, age, and whether the person was a Californian. Summary statistics in table 1 and empirical estimates discussed above suggest that each of these characteristics influences trivia earnings and the propensity to cooperate.

Thus, estimation of equation (4) provides insights into the effects of a partner's characteristics on the earnings of an agent who does not cooperate. I present marginal effects estimates from the tobit earnings model in column 1 of table 4. The results suggest a first finding:

Result 5. Partner attributes influence players' profits: white, older, and non-Californian partners increase earnings.

These estimates also suggest an economically significant effect of the attributes. For example, being paired with a white participant increases expected earnings by nearly $900; and, measured at the overall sample means, an additional year of age for your partner maps into nearly $100 larger expected earnings.

These estimates suggest that individual-specific observables matter, but the necessary next step is to model the partnership selection process. In doing so, I take a structural

[6] The interested reader should see the excellent surveys on stakes effects in laboratory games in Camerer and Hogarth (1999) and Hertwig and Ortmann (2001). Note that this result does not speak to the comparison between hypothetical and real stakes. For a discussion of this comparison see List (2001).

[7] In the discussion below I take *proper* backward induction as representing choices that maximize one's own payoff function for this single game. It should be noted that backward induction for a conditional cooperator may lead to a different result than for such a player (a conditional cooperator is an agent who prefers to cooperate with those who cooperate and prefers to punish those who do not cooperate). And, given the caveats in section II, I ignore the requirement that proper backward induction should take account of the unobserved "larger game" that is taking place between the game-show participants and members of the audience. In such a case, proper backward induction would include consideration of future financial effects.

TABLE 4.—EARNINGS AND PARTNER SELECTION ESTIMATES[a,b]

	Earnings Equation	Selection Equation[1]				
		Pooled	Men	Women	Young	Mature
Male	−91.07 (343)	−0.50 (0.21)	−0.48 (0.31)	−0.60 (0.31)	−0.35 (0.27)	−0.93 (0.39)
White	890.15 (426)	0.60 (0.29)	1.01 (0.44)	0.26 (0.39)	0.65 (0.36)	0.70 (0.51)
Age	236.70 (130)	−0.12 (0.08)	−0.05 (0.13)	−0.19 (0.11)	−0.18 (0.10)	0.09 (0.18)
Age^2	−2.39 (1.81)	0.002 (0.001)	0.001 (0.002)	0.003 (0.002)	0.003 (0.001)	−0.002 (0.003)
Californian	−762.31 (364)	−0.76 (0.25)	−0.62 (0.38)	−0.86 (0.33)	−0.68 (0.32)	−1.03 (0.42)
n	234	117	51	66	74	43

[a] Column 1 estimates are obtained via a tobit model and reveal how a partner's characteristics influence the take-home earnings of a Nash player. Thus, the dependent variable equals *trivia earnings* × *cooperate,* where *cooperate* equals 1 if the partner cooperated, and 0 otherwise. Columns 2–6 present empirical estimates from a discrete choice logit model. The parameter estimates reveal how characteristics of group 2 agents influence group 1 agents' selections. "Young" ("Mature") are choosers who are less than (greater than or equal to) 31 years old.

[b] Figures in parentheses are standard errors.

modeling approach and assume that person i selects person j if the expected payoffs π_{ij} exceed the expected payoffs π_{ik} for all K alternative persons. This structural approach models the payoff for person i choosing person j as

$$\pi_{ij} = \beta' \mathrm{X_j} + \mu_{ij}, \tag{5}$$

where $\mathrm{X_j}$ is a vector of observable person-specific characteristics that potentially affect payoffs, which include variables that measure expected trivia earnings and willingness to cooperate in the prisoner's dilemma game.

A well-known property of equation (5) is that if the μ_{ij} follow a Weibull distribution and are independently and identically distributed, the probability that person i will select person j is given by[8]

$$P_{ij} = \sum_{k=1}^{K} \exp(\beta' \mathrm{X_k}) \exp(\beta' \mathrm{X_j}), \tag{6}$$

where K is the number of alternatives (the three group 2 agents), and the parameters β are estimated using maximum likelihood techniques.

Empirical results from estimating this discrete choice model are contained in columns 2–6 of table 4. I present estimates from a pooled model, and then delineate by gender and age.[9] Empirical estimates suggest the following insight:

Result 6. There is evidence that agents discriminate statistically.

Estimates in column 2 of table 4 suggest that women, whites, and non-Californians tend to be chosen as partners more often than their counterparts. The empirical result that women are preferred to men is ubiquitous: across every specification but one, this result is significant at better than the $p < 0.10$ level. A similar finding occurs for white agents: every cohort prefer their partner to be white. A similar result follows for agents not from California.

These empirical results are consistent with the view that agents have correct beliefs about the empirical results from the earnings equation in column 1 of table 4 and choose their partners correctly. Such evidence is consistent with the notion that agents use observable characteristics to make statistical inference about their prospective partners in the sense of Arrow (1972) and Phelps (1972) (statistical discrimination). In this spirit, these results are consistent with participants using individual observables in an effort to maximize their own earnings. In light of the literature that suggests people are not well calibrated in their probability judgments (see, for example, Camerer, 1995), these results are quite surprising. Perhaps even more surprising is that participants perform this well considering that all of the shows were taped prior to the television premiere.

As well as agents statistically discriminate along certain observables, one result at odds with this insight is:

Result 7. Agents do not properly anticipate the effects of their partner's age on potential earnings.

This finding can be gleaned from empirical estimates in table 4. Column 1 of table 4 shows that older partners provide higher expected profits than younger ones do, an effect that is statistically significant at the $p < 0.05$ level. As tables 1 and 2 highlight, this result is entirely due to older agents cooperating more than younger agents, rather than to superior performance in the trivia questions.

Columns 2–6 in table 4 illustrate that these higher earnings are generally not reflected in partner choices. For example, in every specification, except for the case when mature participants are choosing, younger agents are preferred. For women and younger agent selectors, the coefficients are statistically significant at conventional levels. The pattern of choices causes agents to lose thousands of dollars in earnings.

Overall, table 4 shows that agents typically perform quite well in using observables to select partners to maximize

[8] The strong assumption that the error terms (μ_{ij}) are independently and identically distributed imposes the *independence of irrelevant alternatives* (IIA) restriction on the predicted values. This assumption poses problems in that it stretches the bounds of credulity to assume that, for example, a person's decision not to choose one player is independent of her decision to reject another player. I tested for IIA via Hausman and McFadden's (1984) specification test and found that the maintained assumption of independence of the stochastic terms in the utility function was valid in those cases where the Hessian did not become singular.

[9] I suppress discussion of empirical estimates from other groupings of the data (whites, nonwhites, Californians, non-Californians, and so on), as they are similar to the broader group of empirical estimates presented in table 4.

their earnings. Yet, when it comes to drawing inference from prospective partners' ages, all agents, except for the older ones, miscalculate the attractiveness of partnering with more mature players. Indeed, selectors, particular women and younger agents, even favor younger agents in the partner selection process. This behavior is consistent with several underlying models, one of which is a *taste* for discrimination in the Becker (1975) sense: economic actors are willing to pay a financial price to avoid interacting with older agents. Yet, as with many empirical studies in the literature on discrimination, though these findings *appear* to be more consistent with some agent types having a general "distaste" for others, it cannot be ruled out (for example) that selectors incorrectly apply statistical inference about partners.[10] However, as Levitt (2004; p. 431) points out, a simple model of incorrect stereotyping, although perhaps descriptively valid, is "not a very satisfying economic model because it implies that individuals are making systematic errors."

IV. Concluding Comments

Adam Smith's notion of the invisible hand—if each individual acts in a manner that maximizes his or her own profit, the total profit for the community will be maximized—remains an important influence within the behavioral sciences. Yet game theorists have proposed a simple situation in which promotion of self-interest does not unequivocally lead to globally optimal solutions: the prisoner's dilemma game. Every year thousands of new undergraduate and graduate students are introduced to noncooperative game theory via simple illustrations and examples of the classic prisoner's dilemma game. To my best knowledge, however, there does not exist an empirical examination of behavior in a high-stakes game that mirrors the classic prisoner's dilemma tale.

This study fills this gap by examining data drawn from a game show that is stunningly similar to the classic example. If one were to conduct such an experiment in the laboratory, the cost to gather the data would be well over $350,000. The results suggest a few major themes: in a one-shot prisoner's dilemma game over high stakes, play is not unduly influenced by the level of stakes. Indeed, at every level of monetary stakes thousands of dollars are left on the table by agents playing the weakly dominant strategy of not cooperating. A further insight concerns discriminatory behavior, which is indicated in the data. For example, there is evidence that agents use observable characteristics to make statistical inference about their prospective partners. In this spirit, whereas participants make proper use of variables such as race, gender, and geographic residence in an effort to maximize their own earnings, they tend to underselect older participants. Such behavior is consistent with agents having a general "distaste" for such players. Although these findings have important implications, I view it apropos to highlight that any interpretation of results should keep in mind the various caveats discussed in section II.

[10] Using game-show data from *The Weakest Link,* Levitt (2004) also finds evidence of taste-based discrimination against older players. One way to parse these theories is to use a series of field experiments wherein beliefs are elicited (List, 2004).

REFERENCES

Andreoni, James, and Lise Vesterlund, "Which Is the Fair Sex? Gender Differences in Altruism," *Quarterly Journal of Economics* 116:1 (2001), 293–312.

Andreoni, James, and Ragan Petrie, "Beauty, Gender, and Stereotypes: Evidence from Laboratory Experiments," University of Wisconsin working paper (2004).

Arrow, Kenneth, "The Theory of Discrimination," in Orley Ashenfelter and A. Rees (Eds.), *Discrimination in Labor Markets* (Princeton, NJ: Princeton University Press, 1972).

Becker, Gary, *The Economics of Discrimination,* 2nd ed. (Chicago: University of Chicago Press, 1975).

Beetsma, Roel M. W. J., and Peter C. Schotman, "Measuring Risk Attitudes in a Natural Experiment: Data from the Television Game Show Lingo," *Economic Journal* 111:474 (2001), 821–848.

Berk, Jonathan B., Eric Hughson, and Kirk Vandezande, "The Price Is Right, but Are the Bids? An Investigation of Rational Decision Theory," *American Economic Review* 86:4 (1996), 954–970.

Camerer, Colin, "Individual Decision Making" (pp. 587–683), in J. H. Kagel and E. R. Alvin (Eds.), *The Handbook of Experimental Economics* (Princeton, NJ: Princeton University Press, 1995).

Camerer, Colin, and Robin M. Hogarth, "The Effect of Financial Incentives on Performance in Experiments: A Review and Capital-Labor Theory," *Journal of Risk and Uncertainty* 19:2 (1999), 7–42.

Cameron, Lisa, "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia," *Economic Inquiry* 31:1 (1999) 47–59.

Carpenter, Jeffrey, Eric Verhoogen, and Stephen Burks, "The Effect of Stakes in Distribution Experiments," *Economics Letters* 86:3 (2005), 393–398.

Clark, Kenneth, and Martin Sefton, "The Sequential Prisoner's Dilemma: Evidence on Reciprocation," *Economic Journal* 111:468 (2001): 51–68.

Davis, Douglas, and Charles Holt, *Experimental Economics* (Princeton: Princeton University Press, 1995).

Eckel, Catherine C., and Philip J. Grossman, "Are Women Less Selfish than Men? Evidence from Dictator Experiments," *Economic Journal* 108:448 (1998), 726–735.

Fehr, Ernst, Urs Fischbacher, and Elena Tougereva, "Do High Stakes and Competition Undermine Fairness? Evidence from Russia," University of Zurich working paper (2005).

Fershtman, Chaim, and Uri Gneezy, "Discrimination in a Segmented Society: An Experimental Approach," *Quarterly Journal of Economics* 116:1 (2001), 351–377.

Gertner, Robert, "Game Shows and Economic Behavior: Risk-Taking on Card Sharks," *Quarterly Journal of Economics* 108:2 (1993), 507–521.

Glaeser, Edward L., David I. Laibson, Jose A. Scheinkman, and C. L. Soutter, "Measuring Trust," *Quarterly Journal of Economics* 105:2 (2001), 811–846.

Hausman, Jerry, and Daniel McFadden, "Specification Tests for the Multinomial Logit Model," *Econometrica* 52:5 (1984), 1219–1240.

Hertwig, Ralph, and Andreas Ortmann, "Experimental Practices in Economics: A Methodological Challenge for Psychologists?" *Behavioral and Brain Sciences* 24:1 (2001), 433–451.

Levitt, Steven D., "Testing Theories of Discrimination: Evidence from Weakest Link," *Journal of Law and Economics* 47:2 (2004), 431–452.

Levitt, Steven D., and John A. List, "What Do Laboratory Experiments Tell Us about the Real World?" University of Chicago working paper (2005).

List, John A., "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards," *American Economic Review* 91:5 (2001), 1498–1507.

——— "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field," *Quarterly Journal of Economics* 108:1 (2004), 45–90.

——— "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions," *Journal of Political Economy* 114:1 (2006), 1–37.

List, John A., Robert Berrens, Alok Bohara, and Joseph Kerkvliet, "Examining the Role of Social Isolation on Stated Preferences," *American Economic Review* 94:3 (2004), 741–752.

Masclet, David, Charles Noussair, Steven Tucker, and Marie-Claire Villeval, "Monetary and Nonmonetary Punishment in the Voluntary-Contributions Mechanism," *American Economic Review* 93:1 (2003), 366–380.

Metrick, Andrew, "A Natural Experiment in Jeopardy!" *American Economic Review* 85:1 (1995), 240–253.

Oberholzer-Gee, Felix, Joel Waldfogel, and Matthew W. White, "Friend or Foe? Coordination, Cooperation, and Learning in High Stakes Games," Wharton School working paper (2005).

Parco, J. E., A. Rapoport, and W. E. Stein, "Effects of Financial Incentives on the Breakdown of Mutual Trust," *Psychological Science* 13 (2002), 292–297.

Phelps, Edmund, "The Statistical Theory of Racism and Sexism," *American Economic Review* 62:4 (1972), 659–661.

Rege, Mari, and Kjetil Telle, "The Impact of Social Approval and Framing on Cooperation in Public Good Situations," *Journal of Public Economics* 88:7–8 (2004), 1625–1644.

Roth, Alvin E., "Introduction to Experimental Economics," in John H. Kagel and Alvin E. Roth (Eds.), *The Handbook of Experimental Economics* (Princeton, NJ: Princeton University Press, 1995).

Shafir, Eldar, and Amos Tversky, "Thinking Through Uncertainty: Nonconsequential Reasoning and Choice," *Cognitive Psychology* 24:1 (1992), 449–474.

Slonim, Robert, and Alvin E. Roth, "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic," *Econometrica* 66:3 (1998), 569–596.