The Name of the Game: Predictive Power of Reputations Versus Situational Labels in Determining Prisoner's Dilemma Game Moves

Varda Liberman
The Interdisciplinary Center Herzliya, Israel

Steven M. Samuels
U.S. Air Force Academy

Lee Ross
Stanford University

Two experiments, one conducted with American college students and one with Israeli pilots and their instructors, explored the predictive power of reputation-based assessments versus the stated "name of the game" (Wall Street Game vs. Community Game) in determining players' responses in an N*-move Prisoner's Dilemma. The results of these studies showed that the relevant labeling manipulations exerted far greater impact on the players' choice to cooperate versus defect—both in the first round and overall—than anticipated by the individuals who had predicted their behavior. Reputation-based prediction, by contrast, failed to discriminate cooperators from defectors. A supplementary questionnaire study showed the generality of the relevant shortcoming in naïve psychology. The implications of these findings, and the potential contribution of the present methodology to the classic pedagogical strategy of the demonstration experiment, are discussed.*

Keywords: *Prisoner's Dilemma; lay psychology; name of the game; construal; demonstration experiments*

Psychologists have long recognized that understanding, predicting, and altering human behavior require that we attend to the manner in which the relevant actors interpret or "construe" the stimulus situations that confront them (Asch, 1952; Bruner, 1957; Lewin, 1935; Thomas & Znaniecki, 1918). In this article, we present three studies that deal with the impact of a labeling manipulation designed to influence such construals and with a hypothesized failing in intuitive psychology (Nisbett & Ross, 1980; Ross & Nisbett, 1991). The first two studies pit the power of this labeling manipulation against the predictive power of assessments made about those actors by people who know them well. A third study further explores lay beliefs that individuals hold about the power of situational labels on themselves and their peers.

The particular context employed in our investigation was the familiar Prisoner's Dilemma in which participants must decide whether to "cooperate" or "defect" (Luce & Raiffa, 1957). In selecting this context, we address not only the longstanding dialogue between social psychology and personality psychology but also a newer dialogue and seeming tension between the subjectivist tradition of psychology and the objectivist spirit of game-theory economics.

Construal and the Prisoner's Dilemma

The Prisoner's Dilemma (PD) literature is too extensive to be reviewed here, but we can note that both altruistic and strategic motives have been offered to explain why people might choose to cooperate in the *N*-round version of the game (see Axelrod, 1984; Dawes, 1980; Kerr, 1995; Komorita & Parks, 1996; Messick, 1999). An early finding relevant to our present research is that players are more willing to cooperate, even in a single-round version of the game, when they think that similarly coop-

Authors' Note: Correspondence concerning this article should be addressed to Lee Ross, Stanford University, Dept. of Psychology, Jordan Hall, Bldg. 420, Stanford, CA 94305-2130; e-mail: ross@psych.stanford.edu.

PSPB, Vol. 30 No. 9, September 2004 1175-1185
DOI: 10.1177/0146167204264004

erative choices are likely to be forthcoming from their partner. By the same token, "cooperators" seem to think that cooperation is a more common and normative response than do "defectors" (Dawes, McTavish & Shaklee, 1977; Kelley & Stahelski, 1970).

Even more relevant to our present concerns are previous studies that manipulated the way the payoff matrix, response alternatives, or participants' roles in various games or allocation tasks are "framed" (Allison & Messick, 1985; Dawes, 1980; Samuelson & Allison, 1994; see reviews by Komorita & Ellis, 1995; Schwartz-Shea & Simmons, 1995). Studies contrasting conditions in which the description of the relevant tasks or games evoked norms related to business dealings versus ethical dilemmas, or more generally competitive versus cooperative norms (Allison, Beggan, & Midgley, 1996; Blount & Larrick, 2000; Larrick & Blount, 1997; Tenbrunsel & Messick, 1999; van Dijk & Wilke, 2000), are of particular note.

We expect our present research design to provide further evidence about the malleability of construal processes and to show that influencing such processes can in turn influence behavioral choices of the sort made in the PD game. At the same time, we expect to show that laypeople fail to appreciate the power of such manipulations and instead give undue weight to their perceptions of individual differences in personal attributes (notably, cooperativeness vs. competitiveness) that members of our society infer and rely on in their normal interpersonal dealings. Thus, in our first two studies, we employ an experimental design that allows us to contrast the power of our specific manipulation—that is, the name that we attached to the game in describing the relevant matrix—with the behavioral predictions of people who knew the individuals well and based their predictions on that knowledge.

For half of the participants in the studies we shall describe, the relevant label was the Wall Street Game (or, in the case of the Hebrew-speaking Israelis in Study 2, the Bursa Game). For the other half of our sample, the label in question was the Community Game (or, in Hebrew, the Kommuna Game). The former label, of course, connotes rugged individualism, concern with self-interest, and contexts in which competitive or exploitative norms are likely to operate. The latter label, by contrast, connotes interdependence, collective interest, and contexts wherein cooperative norms are likely to operate. (No explicit mention of social norms was made, of course, and attention was not directed to the significance of the relevant name.) Our primary hypothesis was that the responses of our research participants would defy both the dispositionist expectations of the individuals who predicted those responses and also the dispositionist notions of social science colleagues and of future psychology students who learn about our findings.

STUDY 1: REPUTATIONS VERSUS SITUATIONAL LABELS IN THE PD GAME

Participants

Forty-eight Stanford undergraduate men, selected as described below, were recruited by telephone. They were told that they had been selected to participate in a study on negotiation in which they could "earn money" but nothing about the criterion for that selection.

Procedure

The study was run in two phases. In Phase 1, dormitory resident assistants (RAs) heard a detailed account of the instructions (including two mentions of the name of the game) and payoff matrix to be used in our study. They then were asked (a) to nominate the students they deemed "most likely to play C" (cooperate) and "most likely to play D" (defect) on the first round of the game and (b) to indicate the relevant likelihoods. They also were asked to reassess the relevant likelihood if the stipulated name of the game changed from the Wall Street Game to the Community Game, or vice versa. In Phase 2, the nominees participated, face to face, in a seven-trial PD game, knowing that they and their counterparts would be paid in accord with the following payoff matrix, which was presented both on a large poster board at the front of the room and on a sheet of paper given to each participant:

		Player 2's Choice	
		C	*D*
Player 1's Choice	C	Player 1 + 40 cents Player 2 + 40 cents	Player 1 – 20 cents Player 2 + 80 cents
	D	Player 1 + 80 cents Player 2 – 20 cents	Player 1 zero cents Player 2 zero cents

All participants chosen for this second phase were individuals whose nominators had assessed them as at least 85% likely to cooperate or 85% likely to defect—not only given the originally stipulated name for the game but also if the alternative name had been stipulated. Furthermore, the two participants paired together to comprise a given dyad always had a similar nomination status—either most likely to cooperate or most likely to defect.

The instructions participants received were the same ones previously presented to the nominators, including a pair of verbal references to either the Wall Street Game

or the Community Game. Aside from this difference in the name of the game, all nominees were treated identically. Before play started, the experimenter made certain that each player understood the relevant matrix by having them specify the payoff each player would receive given each combination of moves. Participants were told at the outset that the game would continue for seven rounds and that after specifying their own move for a given round they would learn their counterpart's move, and hence what their respective payoffs would be.

The game then commenced and continued until the seventh round had been completed. Afterward, participants completed some additional measures (not discussed in this report), were thanked for their efforts, and were paid the sum they had accumulated over the seven rounds.

Results and Discussion

Results both for the first round and the seven rounds overall are presented in Figure 1. It is immediately apparent that the participants' nomination status as most likely to cooperate versus most likely to defect had no predictive power at all. It is equally clear that the name of the game exerted a considerable effect on the participants' choices. When playing the Community Game, 67% of the most likely to cooperate nominees and 75% of the most likely to defect nominees cooperated on the first round. When playing the Wall Street Game, 33% of participants with each nomination status cooperated, chi-square (collapsing across nomination status) = 6.76, $p < .01$ (Fisher exact $p = .02$).

Choices over the entire seven rounds presented essentially the same picture (see Figure 1). This picture reflects the fact that participants who both cooperated and received cooperation on the first round generally continued to cooperate over subsequent rounds, whereas participants who had defected and/or faced defection on the first round tended to defect thereafter (again, without regard to their nomination status). In fact, individuals nominated as most likely to cooperate proved to be slightly *less* cooperative overall (a mean of 3.25 cooperative responses per dyad over the seven rounds) than participants nominated as most likely to defect (a mean of 3.58 cooperative responses), $F < 1.0$. Furthermore, the name of the game presented to participants at the outset of the game continued to make its impact felt. Individuals playing the Community Game (collapsing across the nomination status) offered a mean of 4.63 cooperative responses (i.e., a 66.1% rate of cooperation). Individuals playing the Wall Street Game (again collapsing across nomination status) offered a mean of 2.21 cooperative responses (i.e., a 31.5% rate of cooperation), $F(1, 20) = 11.83$, $p < .005$.

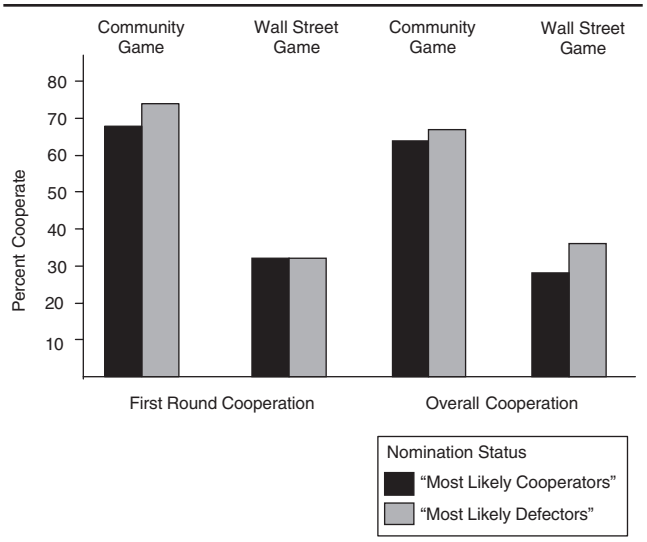

Figure 1 First round and overall cooperation in the Community Game versus Wall Street Game by nominated "most likely cooperators" and "most likely defectors" (Study 1).

More detailed inspection of the responses of individual dyads over the seven rounds revealed that in the Community Game, mutual cooperation occurred on 43 of 84 possible occasions, whereas mutual defection occurred on 16 occasions. By contrast, in the Wall Street Game, mutual cooperation occurred on only 11 of 84 possible occasions, whereas mutual defection occurred on 42 occasions. Significance testing of these differences is problematic because responses across successive rounds obviously are not independent. However, we can note that the mean number of cases of mutual cooperation per dyad over the seven rounds was greater in the Community Game ($M = 3.58$) than in the Wall Street Game ($M = 0.92$), $t(22) = 3.20$, $p < .01$. This pattern, incidentally, was apparent even in the final round, where the strategic demands of the situation might be expected to induce defection among erstwhile cooperators. Whereas 9 of 12 Wall Street Game dyads showed mutual defection on that round, only 3 of 12 Community Game dyads did so, Fisher exact $p = .02$.

Our nominators' confident designations of likely cooperators and defectors, and their conviction that the name of the game would exert little impact on their nominees' moves, were thus belied by the data. However, because we utilized only nominees whose nominators thought they would be relatively unaffected by the name of the game manipulation, our test of nominators' accuracy was obviously biased. Accordingly, in Study 2, we opted to deal with predictions about all of the individuals in a given population of PD players—again, predictions made by people who knew them well. We also attempted to determine the effects of the name of the game on the participants' expectations about the moves

that would be chosen by their counterparts and the association of such expectations with their own choices.

STUDY 2: REPLICATION AND EXTENSION WITH ISRAELI PILOTS AND FLIGHT INSTRUCTORS

In this study, we contrasted the PD responses of elite Israeli pilot trainees with predictions made earlier by their flight instructors—individuals who interacted with them extensively and were responsible for assessing their suitability for particular assignments (including assignments requiring successful independent or interdependent functioning). The design of the study also included an additional source of predictions—the participants' self-ratings of potentially relevant traits and similar rating by their instructors.

Overview

Instructors and trainees in the Israeli Air Force Training School participated in this experiment, which was conducted in Hebrew. The trainees played five rounds of a PD game, this time featuring a payoff matrix that offered both players 6 points for mutual cooperation, –6 points for mutual defection, and 8 points and –8 points, respectively, for trainees who defected in the face of cooperation or cooperated in the face of defection.[1] For half of the participants, the relevant game was labeled the Bursa Game (connoting the competitive norms of the stock market), whereas for the other half of our participants, it was labeled the Kommuna Game (connoting more cooperative and interdependent norms). The moves the trainees chose in the face of these two labels were assessed in light of both prior likelihood assessments by the trainees' instructors and of prior trait ratings made by the instructors and the trainees themselves.

Procedure

Several weeks before our participants played the PD game, during a routine paperwork task, they were asked to complete a questionnaire in which they used 7-point scales to evaluate themselves, relative to the other trainees in their squadron, in terms of 11 trait descriptors, including competitiveness. Prior to the experimental sessions, the trainees in each squadron were divided into two groups, one of which would later play the Bursa Game and one of which would later play the Kommuna Game. Half of the instructors made predictions about the trainees they knew best who would play the Bursa Game and half made predictions about the trainees they knew best who would play the Kommuna Game.

Before making their predictions, the instructors received exactly the same information about the game that would later be presented to the players themselves. In addition to making predictions about whether specific trainees would choose C or D on the first round, instructors also were asked to estimate the percentage of all trainees in the school that would choose C on the first round and to rate the traits (including cooperativeness), skills, and habits of the trainees they knew best. Finally, they were asked to estimate the likelihood of cooperation versus defection by the trainees if the alternative name of the game were presented.

During the experimental session itself, trainees were assigned to play the Bursa Game (20 trainees) or Kommuna Game (20 trainees) in accordance with their predetermined experimental condition. The study was run during an in-class meeting in which participants from all squadrons were represented. Trainees were seated in two rows, facing each other, knowing only that their counterpart was in the row facing them. At the front of the room, they could see the matrix of the game clearly visible, with the appropriate name for their session (Bursa or Kommuna) provided. They then were presented with a PowerPoint presentation about the rules of the game, emphasizing that they would not know their counterpart's identity but that they would learn his payoff after each round in accord with the relevant matrix.

They were further informed that they would be playing the game several times with the specific number of rounds determined by the throw of a die (with the game concluding when a six was thrown). "The object of the game," they were told, is to "accumulate, personally, as many points as you can." Participants each got a sheet presenting the payoff matrix, with the relevant name (Bursa or Kommuna) written above it.

The game began with the participants writing down their first move (C or D) on a questionnaire containing an identification number known only to the experimenter. They also were asked what first-round move they had expected from the other player. Then they passed the questionnaires to the experimenter, who shortly thereafter returned the questionnaires to the participants with the other player's move filled in. At that point, trainees recorded both their payoff and their counterpart's payoff for the round and (after the experimenter indicated that he had not thrown a six) the game continued until five rounds of the game had been completed (at which point the experimenter informed the participants that the previous round had been their last). A final questionnaire item then asked participants, "What percentage of all trainees in the school do you think would choose move C in the first round of the game?"

Results and Discussion

Our results (shown in Table 1) reveal that both on Round 1 and on subsequent rounds of the game, the label Bursa Game prompted higher rates of defection than the label Kommuna Game. Whereas only 4 of the

TABLE 1: Trainees' Decision to Choose C (Cooperation) Versus D (Defection) in Playing the Bursa Versus Kommuna Game (Study 2)

First Round Decisions

	Stipulated Name of the Game	
	Bursa	*Kommuna*
Chose C	4 (20%)	11 (55%)
Chose D	16 (80%)	9 (45%)

Decisions Over Five Rounds of the Game

	Total Proportion of Cs Overall	*Total Proportion of Ds Overall*	*Mean # of Cs per Dyad*	*Mean # of CCs per Dyad*	*Mean # of DDs per Dyad*
Bursa game	25/100	75/100	2.50	0.3	2.8
Kommuna game	45/100	55/100	4.50	1.6	2.1

20 Bursa Game participants opted to cooperate on their first move, 11 of the 20 Kommuna Game participants did so (Fisher exact $p = .024$). Also, the mean number of Cs per dyad over the five rounds of the game was 4.50 ($SD = 3.08$) in the Kommuna Game versus 2.50 ($SD = 2.23$) in the Bursa Game, $t(18) = 1.42$, $p = 0.086$ (one-tailed).[2] Moreover, Kommuna Game dyads achieved mutual cooperation on a mean of 1.6 rounds per dyad over the five rounds of the game, whereas Bursa Game dyads achieved mutual cooperation on a mean of only 0.3 such rounds per dyad, $t(18) = 1.89$, $p = .037$.

Expectations Regarding Counterpart

In the Bursa Game, 10 trainees thought their counterpart had chosen move D, whereas 9 thought he had chosen C (1 trainee failed to answer the question.) By contrast, only 5 of the 20 Kommuna Game trainees thought their counterparts had chosen move D, whereas 15 thought he had chosen C, Fisher exact $p = .07$. Similarly, Bursa Game trainees estimated that only 24.1% of their peers would opt for first-round cooperation, whereas Kommuna Game trainees estimated that 45.8% would do so, $t(38) = 3.36$, $p = .004$. Furthermore, there was a clear association between these percentage estimates and participants' own moves ($r = .56$, $p < .02$, for trainees in the Kommuna Game; $r = .63$, $p < .005$, for trainees in the Bursa Game). In this regard, however, there was also a notable between-condition difference. Not surprisingly, all participants (in both conditions) who anticipated defection from their counterpart opted to defect themselves. But whereas 6 of the 9 Bursa Game trainees who anticipated cooperation from their counterpart chose D, only 4 of the 15 Kommuna Game trainees who anticipated cooperation chose D, Fisher exact $p = .067$.

Instructors' Predictions

Our findings regarding the impact of the name of the game on trainees' first-round PD choices take on particular significance in light of the relevant likelihood estimates made by the trainees' instructors. As Table 2 makes apparent, the mean likelihood estimates offered regarding first-round cooperation by trainees were just as high when the instructors thought they would be playing the Bursa Game (48.7%) as when they thought they would be playing the Kommuna Game (48.2%).

What is equally clear from Table 2 is that the instructors' likelihood estimates for individual trainees (similar to nomination status of participants in Study 1) had no predictive utility; that is, overall, the estimated likelihood of cooperation for the trainees in Study 2 who did in fact choose C (43.3%) was no higher—in fact, it was somewhat but not significantly lower—than the estimated likelihood of cooperation for trainees who chose D (51.6%). Furthermore, this lack of association between actual and predicted choices was equally evident with respect to predictions about Bursa Game trainees and those about Kommuna Game trainees.

Another indication that instructors gave insufficient weight to the name of the game was provided by their estimates about the responses of "all trainees in the school." The mean estimate made by instructors who heard the label Kommuna Game was 47.4%, whereas the mean estimate by instructors who heard the label Bursa Game was 46.4%. These estimates, it should be noted, stand in marked contrast to the corresponding estimates made by the trainees who had actually participated in the two experimental conditions. Trainees in the Kommuna Game condition, it will be recalled, predicted a 45.8% first-round cooperation rate for other trainees in the school, whereas those in the Bursa Game condition predicted a 24.1% rate. One interpretation of this result is that trainees used their own responses as a proxy in estimating what their peers would do under the same circumstances, and their predictions accordingly reflected the impact of the name-of-the-game manipulation with reasonable accuracy. Instructors, having no

TABLE 2: Instructors' Likelihood Estimates That Trainees' First Round Would Be C Rather Than D (Study 2)

	Bursa Game		*Kommuna Game*		*Combined*	
Estimates Regarding	N	*Mean Likelihood Estimate Re C*	N	*Mean Likelihood Estimate Re C*	N	*Mean Likelihood Estimate Re C*
Trainees who chose C	4	40.0%	11	44.5%	15	43.3%
Trainees who chose D	16	50.9%	9	52.8%	25	51.6%
All trainees	20	48.7%	20	48.2%		

NOTE: C = cooperation, D = defection.

such proxy because they had not actually played the game (and probably did not consider what move they personally would make), disregarded the stipulated name and based their estimate on their impressions about overall trainee cooperativeness.[3]

An alternative interpretation is that the name of the game altered the participants' perception of what constituted normative play in the game and, hence, both how they played and how they expected the majority of their peers to play (but did not do so in the case of instructors who focused on the traits of their trainees rather than on the relevant situational norms).

The instructors' response to a stipulated *change* in the name of the game was further revealing. Although the mention of an alternative name may have focused their attention on behavioral norms, it did not lead them to make adequate allowance for the relevant difference. They estimated a mean change in likelihood of cooperation of 12.9 points—again, much less than the actual 35 percentage point effect (55% for Kommuna Game and 20% for the Bursa Game) that was found.

Predictive Utility of Instructors' and Trainees' Trait Ratings

Neither the trait ratings offered by the instructors nor the trainees' self-ratings proved useful in predicting either first round or overall PD play. Notably, the mean instructor ratings of trainee cooperativeness were actually marginally lower for trainees who cooperated in the first round of the game ($M = 3.91$) than for trainees who defected on that first round ($M = 4.75$), $t(38) = 1.68$, $p = .09$. Similarly, the mean self-ratings of competitiveness for trainees who defected on the first round of the game ($M = 3.85$) were actually a little lower (although not significantly so) than the mean competitiveness self-ratings for trainees who cooperated on that first round ($M = 4.23$). Furthermore, the self-characterizations by the trainees proved to be virtually uncorrelated both with the instructors' ratings of the trainees' cooperativeness and with instructors' specific predictions about their first round choice of moves.

The findings of our first two studies suggest that laypeople give little weight to the name of the game (or, by implication, other external factors that influence the way a particular situation is likely to be construed by the actor responding to that situation). Our third and final study explores this failure in lay psychology in more detail. In particular, it asks individuals to make predictions about the impact of the name-of-the-game manipulation not only on peers in general and on likely cooperators and defectors but also on themselves and their best friend. The design of Study 3 also allows us to contrast the weight given to the name of the game when that name is embedded within a general description of the procedure to be employed and the weight given to it when the two alternative names are explicitly contrasted.

STUDY 3: LAY PREDICTIONS ABOUT THE IMPACT OF THE NAME OF THE GAME ON SELF AND OTHERS

Method

In Study 3, a group of students again made likelihood estimates about the first-round PD game move for a peer they nominated as most likely to cooperate and for a peer they nominated as most likely to defect. They also made similar estimates for self, for best friend, and for Stanford students in general. In half of the cases, the name of the game stipulated again was the Wall Street Game, and in half of the cases, the name of the game stipulated was the Community Game. (In this study, however, because our concern lay solely with the nature of the relevant predictions, we did not proceed to have the relevant nominees actually play the game in question.)

The participants in Study 3 later made a second set of predictions about their "nominees," this time with a stipulated change in the name of the game either from the Wall Street Game to the Community Game, or vice versa. In half of the cases, predictions were made about all five target individuals with a given name of the game stipulated and then about the same individuals with the other name of the game stipulated; in half of the cases, the change in the name of the game was stipulated with respect to each target in turn. The order in which participants were asked about the relevant targets ("most likely cooperators" and "most likely defectors" before self, best friend, and Stanford students in general vs. "most likely

cooperators" and "most likely defectors" after self, best friend, and Stanford students in general) was also varied. The result was a $2 \times 2 \times 2$ design that allowed us to examine both the impact of the initially stipulated name of the game on the relevant likelihood estimates and the degree to which those estimates changed when the rater was asked about the effect of a change in the name of the game. The former thus involved a between-subject comparison, whereas the latter involved a within-subject comparison.

Results and Discussion

Statistics summarizing our participants' predictions regarding first-round choices of moves are shown in Tables 3 and 4 below. Table 3 pertains to differences in likelihood estimates given the initially stipulated names of the game. Table 4 pertains to changes in estimates participants made when asked about the effect of a change in the name of the game from the Wall Street Game to the Community Game, or vice versa.

The data presented in Table 3 suggest that participants gave relatively little weight to the initially stipulated name of the game. The relevant between-condition difference in estimates was about 11 percentage points in the case of estimates for best friends and for Stanford undergrads in general and about 7 percentage points for estimates about "most likely defectors" ($p < .05$ in all three cases). The corresponding difference was 4 percentage points in the case of estimates for "most likely cooperators" and slightly less than 7 percentage points in estimates for self ($p > .10$). The actual choices made by the nominated "most likely cooperators" and "most likely defectors" in Study 1, it will be recalled, had produced a much greater between-condition difference (34 percentage points and 43 percentage points, respectively).

Making the difference in the stipulated name of the game salient and explicit (i.e., by asking participants first for predictions given one name of the game and then for predictions given the alternative name of the game) increased the weight given to the name of the game (see Table 4). Overall (combining across the order in which the two names were stipulated), the differences in estimates for the Wall Street versus Community Game were 17.0 percentage points for estimates regarding self, 17.4 percentage points for estimates regarding best friends, and 18.7 percentage points for estimates regarding students in general. The differences in estimates were smaller but still highly significant for estimates regarding "most likely cooperators" (13.6 percentage points) and "most likely defectors" (11.8 points).

Given our present concerns, however, the most important thing to note about the means summarized in Table 4 involves the relationship between the estimated and actual impact of the name of the game. Even with the difference in name of the game highlighted by explicit mention of the alternatives, the allowance for that difference made by our Study 3 respondents (similar to that made by our Study 2 nominators) continued to be too small. In particular, it continued to be considerably smaller than the relevant 34 and 43 percentage-point differences in actual percentages of cooperation versus defection suggested by our Study 1 results.

TABLE 3: Likelihood Estimates Regarding First Round Cooperation of Stipulated Players in Wall Street Game Versus Community Game When Names of Game Embedded (Study 3)

	Community Game (n = *48) Mean*	*Wall Street Game* (n = *48) Mean*	*Estimated Difference*
ML cooperators	81.75	77.50	4.25
ML defectors	23.02	15.98	7.04*
Diff (MLC – MLD)	58.73	61.52**	
Self	63.19	56.33	6.86
Best friend	66.62	55.94	10.68*
Stanford undergrads	52.44	41.04	11.40*

NOTE: ML = most likely, MLC = most likely cooperators, MLD = most likely defectors.
*$p < .05$.

TABLE 4: Likelihood Estimates Regarding First Round Cooperation of Various Stipulated Players When Alternative Names of Game Explicitly Contrasted[a] (Study 3)

	Community Game Mean	*Wall Street Game Mean*	*Estimated Difference*
ML cooperator	84.52	70.95	13.57***
ML defector	28.23	16.40	11.83***
Diff (MLC – MLD)	56.29	54.55	
Self	67.80	50.80	17.00***
Best friend	70.06	52.61	17.45***
Stanford students	56.51	37.82	18.69***

NOTE: ML = most likely, MLC = most likely cooperators, MLD = most likely defectors.
a. Both orders for originally stipulated and alternative name of game combined ($n = 96$).
***$p < .001$.

GENERAL DISCUSSION

The most important result of our two experiments is the discrepancy between the actual effects of our "name of the game" manipulation and the effects anticipated by those predicting the responses on the basis of extensive contact with the players. Indeed, predictions about individual players proved valueless, as did trait ratings—including trait ratings made by the players themselves. These studies also provide evidence about the mediating role played by the participants' expectations about the

play of their counterparts. When told they were playing the Bursa Game, participants expected defection to be the most likely response; when told they were playing the Kommuna Game, they expected cooperation to be the most likely response. Furthermore, regardless of which game they were playing, participants who expected defection from their counterpart overwhelmingly opted to reciprocate that defection. Expectations of cooperation, however, had a less straightforward consequence. Whereas Kommuna Game participants expecting cooperation generally opted to cooperate in return, Bursa Game participants expecting cooperation generally opted to exploit that cooperation by defecting. In other words, the effect of expectations regarding other's choices on own choices depended on the name of the game, and thus on the way the participants construed the game.

The result of these tendencies over successive rounds, in which defection begat defection and cooperation was sustained only when it was mutual, was inevitable (Axelrod, 1984); that is, first-round responses tended to dictate later-round responses, and as a consequence, overall rates of cooperation—especially mutual cooperation—were significantly higher in the Community/Kommuna Game than in the Wall Street/Bursa Game.

Person Versus Situation and the Didactic Strategy of the Demonstration Experiment

Although our studies obviously speak to the traditional issue of the "person versus the situation" (see Ross & Nisbett, 1991), a note of clarification is in order. The results of our studies illustrated the power of the situation, or at least the power of situational labels and construals, but they did not prove (nor were they designed to prove) that individual differences have no predictive power in the Prisoner's Dilemma or other games presenting players with similar choices and tradeoffs. Indeed, a search of the literature reveals a number of studies showing significant cross-situational consistency in responses to non-zero-sum games that oblige individuals to choose among the goals of maximizing joint outcome, maximizing own outcome, or maximizing advantage of self over others (e.g., Bem & Lord, 1979[4]; Bennett & Carbonari, 1976; Kuhlman & Marshello, 1975).

There are even some studies showing significant correlations between PD play and paper-and-pencil measures of specific cognitive abilities (Pincus & Bixenstine, 1979) and/or measures of specific traits, including locus of control, self-monitoring, Type-A behavior, and sensation-seeking (Boone, De Brabander, & van Witteloostuijn, 1999); adherence to Protestant Ethic values (Furnham & Quilley, 1989); both Factor G (which involves moral values and concerns) and Factor E (which deals with dominance-submissiveness) on the 16 PF inventory (Gillis & Woods, 1971); as well as various measures of cooperative, accommodative, or prosocial (vs. competitive, egoistic, or exploitative) personal motivations and orientations (e.g., Houston, Kinnie, Lupo, Terry, & Ho, 2000; Parks & Rumble, 2001; Vinacke, 1974). However, it is worth noting that to produce such statistical significance, some aggregation of the relevant choice measures typically has been required (whereas in both of the present experiments the name-of-the-game manipulation proved powerful enough to produce significant between-condition differences in response on a single occasion, that is, the first round of the game).

What the present studies dealt with was not the (lack of) predictive value of individual differences relative to that of a situational manipulation but rather the (lack of) predictive value of the type of lay assessments that people make about their peers—even peers they think they know well and about whom they are willing to make relatively confident predictions vis-à-vis play in the Prisoner's Dilemma. In this context, it is worth noting the relevance of our present study to the traditional didactic or pedagogic strategy of the "demonstration experiment" (see Devine & Brodish, in press; Ellsworth & Gonzalez, in press). A list of classic experiments in this tradition would include Milgram's (1963, 1974) obedience studies, Asch's (1951, 1956) conformity studies, Rosenthal and Jacobson's (1968) studies of self-fulfilling expectations in the classroom, Freedman and Fraser's (1966) foot-in-the-door study, Darley and Batson's (1973) bystander intervention studies, and other crown jewels of experimental social psychology. These studies did not test falsifiable general propositions about human behavior, and they did not link the relevant outcomes to specific cognitive or motivational mediators. Instead, they illustrated the power of particular situational influences that proved more powerful than our students or other laypeople (and many of our colleagues in psychology, as well as those in other disciplines) had heretofore appreciated. In so doing, they taught us to make less dispositionist attributions about actors who are responding to the demands and constraints of their social contexts (see Nisbett & Ross, 1980; Ross, Greene, & House, 1977; Ross & Nisbett, 1991).

The studies noted above did not offer this lesson explicitly. None included measures of lay expectations as part of their original design (although Milgram, famously, did collect predictions from colleagues—who consistently underestimated the relevant level of obedience). Instead, the investigators invited their readers to weigh the relevant findings against their own expectations and underlying assumptions about human psychology and then to accommodate those views to the relevant empirical lessons.[5] The present studies, by contrast, were

designed to offer the same lessons (both to future students of social psychology and to our colleagues in economics and other social science disciplines) more explicitly—that is, by contrasting actual effects of the manipulations in question with the effects anticipated by individuals who know the actors well and base their predictions and likelihood assessments on that knowledge. In a sense, what we have provided is an empirical parable. This parable dealt with the importance of attending to subjective interpretations or construals, the folly of overemphasizing the role of objective economic outcomes in human decision making, and the dubious value of impressions about personal dispositions—especially when one is called on to make predictions about the behavior of actors whom one has never observed in the relevant situations.

Implications for Everyday Attributions and Searches for Common Ground

In the case of trained social scientists, suggestions about the importance of subjective construal and the possible impact of manipulations of such construal are hardly necessary—at least not when we are perusing the pages of a scholarly journal. Indeed, when reminded of earlier classic studies and the situationist and subjectivist traditions of our field and asked to venture a prediction about the outcome of the present experiments, many of our professional peers would predict a significant effect of the name of the game, and some would even anticipate the limited success of the lay personologists who offered assessments in our studies. Rather, such suggestion becomes most relevant when people, laypeople and social scientists alike, encounter each other in our everyday lives, discuss issues of social policy, or read the daily news.

Our present findings, moreover, pertain as much to attributions as they do to social predictions. They imply that upon observing behavior that challenges our expectations about a given actor or group we should avoid making broad dispositional inferences. Instead, we should entertain the possibility that we have failed to appreciate the structural demands and constraints of the situation at hand (Blount & Larrick, 2000; Morris, Larrick, & Su, 1999) and/or failed to take into account the way those demands and constraints were interpreted and linked to preexisting cognitive structures and situational norms by the relevant actors. These findings in no way negate the fact that different individuals may interpret particular types of situations differently—that the same discussion may be an invigorating exercise for one party and an identity-threatening debate for another, or that one person's jolly coed softball game may be another's test of athletic prowess. Rather, the findings of our present research suggest that most of us are equipped with a range of different cognitive schemas that could be applied to a given situation and that our behavior may depend on whatever factors determine which schema happens to be recruited by chance or by subtle features of the context at hand.

A study coauthored by the third author of this article (Kay & Ross, 2003)—one directly inspired by the first of the three studies we reported here—illustrates this point. It showed that a simple priming manipulation (having participants unscramble words to form sentences) could influence the way individuals construed the norms of the PD game, and hence, the responses they subsequently made, in the same way as our labeling manipulation. (Unfortunately, however, no attempt was made in that study to compare these actual effects with lay presumptions about the power of the relevant manipulation.)

An obvious question raised by the present studies is their relevance to contexts in which participants respond not to a simple payoff matrix but to a more complex and realistic task involving social interaction and negotiation. Can one change the types of offers negotiators make, and the responses they make, to the offers of their counterparts, by changing the manner in which the task at hand is presented (e.g., as a search for justice in the light of conflicting claims and entitlement vs. a search for mutual advance over the status quo)? Although it is difficult to do rigorous random assignment experiments to answer such questions in consequential negotiations that occur outside the laboratory, the memoirs of skilled mediators (e.g., Saunders, 1991) suggest the importance of changing perceived norms and thus the meaning of compromise. There is also encouraging evidence from within the laboratory tradition that different framings of Ultimatum Game and other resource allocation tasks—both via explicit labels and descriptions of the task or the player's roles (e.g., Blount & Larrick, 2000; Larrick & Blount, 1997) or by implicit priming (Kay, Bargh, Wheeler, & Ross, in press) cues—can make individuals more generous in their treatment of other players. Let us hope that collaboration between practitioners and researchers in this domain will lead both to greater sophistication and even some new insights about how to produce "communities of cooperators" (Axelrod, 1984) who thrive and encourage others to follow their example.

NOTES

1. This payoff matrix was decided on after pretesting indicated that, in this context, for this particular population, the payoff matrix that had been used in Study 1 (with no label attached to it) yielded much more defection than cooperation.

2. Results reflecting the name of the game on choices of cooperation versus defection constitute a direct replication of Study 1 (necessarily with smaller *N*s because only members of the relevant pilot train-

ing program could be tested). Accordingly, here and wherever else we present Study 2 findings relevant to the name-of-the-game effect, the data were subjected to one-tailed instead of two-tailed tests of significance. It further should be noted that in the case of predictions regarding play over the entire game, as opposed to those pertaining only to Round 1, the number of degrees of freedom is halved (because dyads rather than individuals necessarily become the unit of analysis), which of course reduces the significance level of any between-condition differences.

3. One of our anonymous reviewers observed that it would be interesting to investigate the corresponding attributions made by the players, that is, whether, the experience of having played the Prisoner's Dilemma (PD) game with a given label protects the players from making unwarranted inferences about other actors who have opted to cooperate versus defect in the face of the relevant Community Game or Wall Street Game label—unwarranted inferences to which mere observers of the relevant behavior would presumably fall prey (see Morris, Larrick, & Su, 1999).

4. The Bem and Lord study is one investigation that did investigate, albeit somewhat indirectly, the value of peer assessments in predicting PD play. The investigators first had a group of five graduate student raters produce Q-sorts for the traits they thought might characterize players with particular goals (i.e., maximizing own payoff vs. maximizing joint payoff vs. maximizing the difference of payoff for self relative to other) in PD-like, non-zero sum games. They then identified a subset of players who showed high consistency in pursuing those particular goals when presented with relevant sets of outcome matrices for self and other. Finally, they showed significant commonalities between the Q-sorts for those consistent individuals that had been obtained from peers who knew them well and the Q-sorts produced by the graduate student raters. The methodology employed and results of this study offered a thoughtful contribution to then-contemporary debate between Bem and his critics about the use of more "idiographic" strategies for discovering cross-situation consistency in behavior. But the authors neither showed nor claimed to show that laypeople could directly predict the responses of their peers in a given PD game—certainly not well enough to justify the levels of confidence shown by the individuals who made predictions in our present studies. (In addition, there is no reason to imagine that the authors would claim that the predictive value of the Q-sort procedure they devised was as great as the power of our present labeling manipulation.)

5. A few subsequent investigators examined whether laypeople underestimate the power of the particular situational manipulations involved in these classic studies and accordingly make unwarranted dispositional inferences. Safer (1980) and Bierbrauer (1979) provided such evidence with respect to the Milgram situation and Pietromonaco and Nisbett (1982) did so with respect to the Darley and Batson manipulation.

REFERENCES

Allison, S. T., Beggan, J. K., & Midgley, E. H. (1996). The quest for "similar instances" and "simultaneous possibilities": Metaphors in social dilemma research. *Journal of Personality and Social Psychology, 71*, 479-497.

Allison, S. T., & Messick, D. M. (1985). *Behaviors or non-behaviors as the cooperative or defecting responses in two types of social dilemmas.* Unpublished manuscript.

Asch, S. E. (1951). Effects of group pressures upon the modification and distortion of judgment. In H. Guetzkow (Ed.), *Groups, leadership, and men.* Pittsburgh, PA: Carnegie Press.

Asch, S. E. (1952). *Social psychology.* New York: Prentice Hall.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs, 70*(9, Whole No. 416).

Axelrod, R. (1984). *The evolution of cooperation.* New York: Basic Books.

Bem, D. J., & Lord, C. G. (1979). Template matching: A proposal for probing the ecological validity of experimental settings in social psychology. *Journal of Personality and Social Psychology, 37*, 833-846.

Bennett, R. P., & Carbonari, J. P. (1976). Personality patterns related to own-, joint-, and relative-gain maximizing behaviors. *Journal of Personality and Social Psychology, 34*, 1127-1134.

Bierbrauer, G. (1979). Why did he do it? Attribution of obedience and the phenomenon of dispositional bias. *European Journal of Social Psychology, 9*, 67-84.

Blount, S., & Larrick, R. P. (2000). Framing the game: Examining frame choice in bargaining. *Organizational Behavior and Human Decision Processes, 81*, 43-71.

Boone, C., De Brabander, B., & van Witteloostuijn, A. (1999). The impact of personality on behavior in five Prisoner's Dilemma games. *Journal of Economic Psychology, 20*, 343-377.

Bruner, J. (1957). *Contemporary approaches to cognition.* Cambridge, MA: Harvard University Press.

Darley, J. M., & Batson, C. D. (1973). From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology, 27*, 100-119.

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31*, 169-193.

Dawes, R. M., McTavish, J., & Shaklee, H. (1977). Behavior, communication, and assumptions about people's behavior in a common dilemma situation. *Journal of Personality and Social Psychology, 35*, 1-11.

Devine, P. G., & Brodish, A. B. (in press). Modern classics in social psychology. *Psychological Inquiry.*

Ellsworth, P. C., & Gonzalez, R. (in press). Questions, comparisons, and preparation: Methods of research in social psychology. In M. Hogg & J. Cooper (Eds.), *Sage handbook of social psychology.*

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology, 4*, 195-202.

Furnham, A., & Quilley, R. (1989). The protestant work ethic and the Prisoner'S Dilemma game. *British Journal of Social Psychology, 28*, 79-87.

Gillis, J. S., & Woods, G. T. (1971). The 16 PF as an indicator of performance in the Prisoner's Dilemma game. *Journal of Conflict Resolution, 15*, 393-402.

Houston, J. M., Kinnie, J., Lupo, B., Terry, C., & Ho, S. S. (2000). Competitiveness and conflict behavior in simulation of a social dilemma. *Psychological Reports, 86*, 1219-1225.

Kay, A. C., Bargh, J. A., Wheeler, J. S., & Ross, L. (in press). Material priming: The influence of mundane physical objects on situational construal and competitive behavioral choice. *Organizational Behavior and Human Decision Processes.*

Kay, A. C., & Ross, L. (2003). Nonconscious influences and conscious reflection: The interplay of implicit cues and explicit situational construals in the Prisoner's Dilemma. *Journal of Experimental Social Psychology, 39*, 634-643.

Kelley, H. H., & Stahelski, A. J. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology, 16*, 66-91.

Kerr, N. L. (1995). Norms in social dilemmas. In D. A. Schroeder (Ed.), *Social dilemmas: Perspectives on individuals and groups.* Westport, CT: Praeger.

Komorita, S. S., & Ellis, A. L. (1995). Reward structure and cooperation in the *N*-person Prisoner's Dilemma. In D. A. Schroeder (Ed.), *Social dilemmas: Perspectives on individuals and groups* (pp. 15-30). Westport, CT: Praeger.

Komorita, S. S., & Parks, C. D. (1996). *Social dilemmas.* Boulder, CO: Westview.

Kuhlman, D. M., & Marshello, A. (1975). Individual differences in the game motives of own, relative, and joint gain. *Journal of Research in Personality, 9*, 240-251.

Larrick, R. P., & Blount, S. (1997). The claiming effect: Why players are more generous in social dilemmas than in ultimatum games. *Journal of Personality and Social Psychology, 72*, 810-825.

Lewin, K. (1935). *Dynamic theory of personality.* New York: McGraw-Hill.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey.* New York: John Wiley.

Messick, D. M. (1999). Alternative logics for decision making in social settings. *Journal of Economic Behavior and Organization, 39*, 11-28.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67*, 371-378.

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York: Harper & Row.

Morris, M., Larrick, R. P., & Su, S. K. (1999). Misperceiving negotiation counterparts: When situationally determined bargaining behaviors are attributed to personality traits. *Journal of Personality and Social Psychology, 77*, 52-67.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Parks, C. D., & Rumble, A. C. (2001). Elements of reciprocity and social value orientation. *Personality and Social Psychology Bulletin, 27*, 1301-1309.

Pietromonaco, P., & Nisbett, R. E. (1982). Swimming upstream against the fundamental attribution error: Subject's weak generalizations from the Darley and Batson study. *Social Behavior, 20*, 1-4.

Pincus, J., & Bixenstine, V. E. (1979). Cognitive factors and cooperation in the Prisoner's Dilemma game. *Psychological Record, 29*, 463-471.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. New York: Holt.

Ross, L., Greene, D., & House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279-301.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill.

Safer, M. A. (1980). Attributing evil to the subject, not the situation: Student reaction to Milgram's film on obedience. *Personality and Social Psychology Bulletin, 6*, 205-209.

Samuelson, C. D., & Allison, S. T. (1994). Cognitive factors affecting the use of social decision heuristics in resource-sharing tasks. *Organizational Behavior and Human Decision Processes, 58*, 1-27.

Saunders, H. H. (1991). *The other walls: The Arab-Israeli peace process in a global perspective*. Princeton, NJ: Princeton University Press.

Schwartz-Shea, P., & Simmons, R. T. (1995). Social dilemmas and perceptions: Experiments on framing and inconsequentiality. In D. A. Schroeder (Ed.), *Social dilemmas: Perspectives on individuals and groups* (pp. 87-103). Westport, CT: Praeger.

Tenbrunsel, A. E., & Messick, D. M. (1999). Sanctioning systems, decision frames, and cooperation. *Administrative Science Quarterly, 44*, 684-707.

Thomas, W. I., & Znaniecki, F. (1918). *The Polish peasant in Europe and America*. Chicago: University of Chicago Press.

van Dijk, E., & Wilke, H. (2000). Decision-induced focusing in social dilemmas: Give-some, keep-some, take-some, and leave-some dilemmas. *Journal of Personality and Social Psychology, 78*, 92-104.

Vinacke, W. E. (1974). Accommodative strategy and communication in a three-person matrix game. *Journal of Personality and Social Psychology, 29*, 509-525.

Received March 19, 2003
Revision accepted December 1, 2003