Grätz, Silvia; Darai, Donja

Conference Paper

Determinants of Successful Cooperation in a Face-to-Face Social Dilemma

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2011: Die Ordnung der Weltwirtschaft: Lektionen aus der Krise - Session: Coordination and Cooperation, No. E7-V1

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Suggested Citation: Grätz, Silvia; Darai, Donja (2011) : Determinants of Successful Cooperation in a Face-to-Face Social Dilemma, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2011: Die Ordnung der Weltwirtschaft: Lektionen aus der Krise - Session: Coordination and Cooperation, No. E7-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:
https://hdl.handle.net/10419/48702

Determinants of Successful Cooperation in a Face-to-Face Social Dilemma*

Donja Darai and Silvia Grätz†

This Version: November 2010

Abstract

What makes you a successful cooperator? Using data from the British television game show "Golden Balls" we analyze a prisoner's dilemma game and its pre-play. We find that players strategically select their partner for the PD, e.g., they bear in mind whether contestants lied. Players' expectations about the stake size strongly influence the outcome of the PD: The lower the stakes, the more likely players successfully cooperate. Most interestingly, unilateral cooperation is encouraged by mutually promising not to defect and shaking hands on it, but a mere handshake serves as manipulating device and increases successful defection.

*This is a revised version of an earlier paper (July 2010) that circulated under the title "Golden Balls: A Prisoner's Dilemma Experiment". The authors thank Armin Schmutzler, Nick Netzer, Michelle Goeree, Daniel Schunk, Kevin Staub, the seminar participants in Zurich, and the participants of the ESA Meeting 2010 in Copenhagen for helpful discussions and suggestions as well as Max Pfister for excellent research assistance. Financial support of the Swiss National Foundation is gratefully acknowledged. The data were provided to the authors by the television show producers, courtesy of Endemol UK plc, in May 2009.

†Both authors: Department of Economics, University of Zurich, Blümlisalpstrasse 10, CH-8006 Zurich.

1 Introduction

The well known prisoner's dilemma game has become the classic economic example to demonstrate non-cooperative behavior: Two contestants face a "dilemma" in which, independent of the other's action, each player is better off by defection than by cooperation. But, the outcome obtained when both defect is worse for each player than the outcome they would have obtained if both had cooperated. Thus, self-interested behavior does not unequivocally lead to a globally optimal solution. Two players who both pursue rational self-interest may end up worse off than if both act contrary to rational self-interest.

This paper uses data from the television show "Golden Balls" which gives us the opportunity to analyze cooperative behavior in an environment of high stakes and face-to-face communication between players as well as players' behavior in the pre-play. Players are not only allowed to talk to each other, but they have the possibility to play with each other, and thereby build a reputation of being trustworthy or being a liar.

The show consists of three rounds, the first two are pre-play and in the third two contestants play a prisoner's dilemma with defection being a weakly dominant strategy. Starting with four contestants, each round every player is randomly assigned a certain cash value. These values are partly common knowledge and partly private information for the respective player. Then players make truthful or untruthful statements about their values. At the end of each round, each player has to cast a vote against one of the other players. The one who receives the majority of votes has to leave the show empty-handed and her values are taken out of the game. Thus, the selection procedure determines the two finalists and the stake size at the same time and does not involve any effort provision by the contestants. The two final players decide about the division of the stakes via playing a prisoner's dilemma. Immediately before the dilemma is played, they can discuss their intentions with respect to their final decision.

Our contribution consists not only of (i) the analysis of cooperative behavior in the presence of high stakes and face-to-face communication, but also of (ii) the analysis of the players' behavior in the pre-play, especially their voting decisions with respect to its influence on the outcome of the prisoner's dilemma. Concerning cooperative behavior, we

observe a unilateral cooperation rate of 55% and a mutual cooperation rate of 33%. Our analysis shows a negative correlation between the stake size and the cooperation rate with a substantial decline around the level of £500. In games with stakes below that value, the unilateral cooperation rate even increases to 74%, and the mutual cooperation rate to 56%. Further, we can show that player's expectation about the stake size matters. If the jackpot is lower than expected, the players are more likely to cooperate. With respect to communication, certain words and gestures are more important than others. Mutually promising each other to cooperate and shaking hands on it increases the cooperation rate, whereas shaking hands without a promise leads to a decline. Apart from those effects, demographic characteristics, such as age and place of residence, matter.
The analysis of the contestants' behavior in the pre-play shows that players make their voting decision dependent on objective criteria such as their opponent's monetary contribution to the stake size, but also on subjective criteria such as the player's trustworthiness and race. In addition, we observe that the weight given to objective and subjective criteria changes between the first and the second voting decision. In the second round, apart from stake size, player's trustworthiness seems to be more important than player's demographics. Further, we show that there is a strong link between the two rounds of pre-prisoner's dilemma play and the player's decision on cooperation. For instance, we find that whether a player lied about her stakes in the pre-play has a significant effect on the cooperation rate.

While writing the first draft of this paper, published in July 2010, it came to our attention that van den Assem, van Dolder and Thaler (2010) are independently analyzing data from "Golden Balls". They came out first with a working paper in April 2010. The only overlap of both studies is the analysis of unilateral cooperation, focussing on the final round. As we mentioned, we also extensively analyze the pre-play and link it to the mutual and unilateral decision outcome of the prisoner's dilemma game. We additionally include variables describing communication, e.g., whether players shake hands or promise each other to cooperate, or lied before. Because of the independent construction of the data set, both studies differ with regard to the definition and modeling of variables. Therefore our analysis and results differ in various aspects.

The structure of the paper is organized as follows. We start with a brief review of the related literature, Section 2, followed by a description of the game show and data set, Section 3. In Section 4 we explain the strategic considerations of the players and motivate our analysis. Sections 5 and 6 represent the main part of the paper, including the empirical analysis and obtained results on cooperative behavior in the prisoner's dilemma as well as contestants' voting behavior in the pre-play. Section 7 concludes.

2 Literature review

In this section, we first review studies that are closely related to our paper in terms of using television show data and/or in terms of considering a prisoner's dilemma with defection being the weakly dominant strategy. Secondly, we present studies analyzing the effect of stake-size, anonymity, communication, and gestures on cooperative behavior in social dilemma games. Finally, we give a brief overview of results about lying and discrimination in voting decisions.

Television game shows and weak prisoner's dilemma

Related studies that use television game show data are List (2006), Oberholzer-Gee, Waldfogel and White (2010), and Belot, Bhaskar and van de Ven (2010). The first two analyze "Friend or Foe", a US game show, and the latter analyzes "Will (s)he share or not?", a show from the Netherlands. In both shows pairs of players build a jackpot by answering trivia questions together. The teams have to decide about the division of their accumulated jackpot by playing a prisoner's dilemma with weakly dominant strategies. In contrast, in "Golden Balls" stakes are built by a random process and the two final players are selected on the basis of a two-round voting procedure. In addition, "Golden Balls" provides the opportunity to analyze a game with very high stakes, i.e., the average jackpot is more than three times as high as the average jackpot in "Friend or Foe" and its median value is more than three times the median in "Will (s)he share or not". All three studies find a very high cooperation rate (around 50%), but only Belot et al. (2010) find an effect of the stake size.

Little empirical work has been done on cooperation in prisoner's dilemma games in which

defection is a weakly dominant strategy.[1] Ortmann and Tichy (1999) analyze the game with respect to gender differences. They find an overall cooperation rate of 46%, and that females cooperate more frequently than males. Studies related to the idea of Rapoport (1988) show that the cooperation rate is higher in a prisoner's dilemma with weakly dominant strategies than in one with strictly dominant strategies.[2]

Stake size, anonymity, and communication

The effect of the stake size on cooperation rates in dilemma games is widely debated and no clear answer has been found so far. Some experiments show that there is no significant effect, whereas others suggest that the cooperation or contribution rate decreases with the stake size (e.g., Camerer and Hogarth, 1999).

Compared to experiments in the laboratory, contestants in "Golden Balls" do not anonymously play the prisoner's dilemma and are allowed to communicate with each other before choosing their action. The relevance of anonymity in dictator games is shown by Hoffman, McCabe, Shachat and Smith (1994). If people feel observed by the experimenters they are more altruistic than in a double-blind setting. In addition, Rege and Telle (2004) showed that the framing of the instructions of the game may raise the cooperation or contribution rate. Since the players in the game under consideration are filmed and play in front of a large television audience, we can expect a similar effect, i.e., a positive effect on cooperation. The cooperation rate we observe is, however, not much different to the rate reported by Ortmann and Tichy (1999), which might suggest that the effect of the audience is not as strong as expected.

In addition, experimental studies have shown that communication increases the cooperation rate significantly (for surveys see Sally, 1995; Ledyard, 1995), although from a theoretical point of view in a prisoner's dilemma communication is cheap talk (see e.g.,

[1] There is a vast experimental literature on prisoner's dilemma games in which defection is a strictly dominant strategy. There the observed cooperation rate varies between 30-40% (see e.g., Shafir and Tversky, 1992).

[2] Rapoport (1988) finds that the cooperation rate in a prisoner's dilemma without fear ($\widehat{=}$ payoff difference between the mutual defector's and unilateral cooperator's payoff) is higher than in one with fear and predicts a cooperation rate of 50% for a no-fear dilemma that corresponds to the game analyzed in this paper. The prediction is independent of the stake size. Rapoport's findings are supported in experiments conducted by Ahn, Ostrom, Schmidt, Shupp and Walker (2001) as well as Ahn, Ostrom, Schmidt and Walker (2003).

Crawford, 1998; Farrell and Rabin, 1996).[3] Bohnet and Frey (1999) analyze the effects of face-to-face communication on cooperative behavior and show that it is very effective, i.e., they observe an increase in the unilateral cooperation rate up to 78%.

Apart from the effects of face-to-face communication, gestures such as a smile might have an impact on cooperation. Scharlemann, Eckel, Kacelnik and Wilson (2001) investigate the impact of a smiling face on people's behavior in a one-shot trust game. They find that subjects are significantly more likely to trust smiling counterparts. Manzini, Sadrieh and Vriend (2009) address this issue in the minimum effort game and test whether people's propensity to choose high effort is increased if subjects can send a "smile" to the other player instead of pressing an ordinary "ready to play" button. They find that this simple device helps players to coordinate on a higher effort even though players are not able to see or to talk to each other.

Furthermore, studies have shown that the effectiveness of communication differs by the words that are used, for instance, when making a promise. Vanberg (2008) finds that people have a preference for keeping a promise and are not driven by concerns about their expected payoff. Ellingsen and Johannesson (2004) propose that people have a preference for keeping their word per se. In contrast, Charness and Dufwenberg (2006) develop the idea that people keep promises because of guilt aversion.[4] Belot et al. (2010) investigate the effect of voluntary vs. elicited promises and find that players are roughly 50% more likely to cooperate if they made a voluntary promise.

Lying and voting behavior

As already mentioned above, guilt is experienced by subjects if they do not keep a promise or in other words lied about their intention which strategy they plan to play. Gneezy (2005) uses a cheap talk sender-receiver game and shows that people's evaluation of whether to lie or not in a situation depends on the consequences of the lie in terms of payoffs. Thereby not only gains achievable through lying are considered but also pos-

[3] In laboratory experiments free-form written communication is often used instead of face-to-face verbal communication to be able to disentangle the effect of facial expressions from the bare content of communication. Roth (1995) provides a survey of bargaining experiments in which the effect of face-to-face communication is tested. The results suggest that face-to-face communication increases the chance of reaching an agreement even further than free-form messaging.

[4] In related work, Miettinen and Suetens (2008) show that players feel most guilty if they communicated their intention to cooperate, but then defect while the opponent cooperates. Charness and Dufwenberg (2010), however, show that providing subjects merely the possibility of communication by sending the word promise or not has almost no positive effect on the cooperation rate in a trust game.

sible losses that might occur to the other players. The fraction of liars is largest if the resulting gains are high and the costs, i.e., losses for the other players, are low. If players have the opportunity to costly punish the other subjects for playing selfishly, they punish much more often if the selfish action followed a deceptive message (Brandts and Charness, 2003). Another approach to analyzing lying is made by Fischbacher and Heusi (2008) who try to figure out under which circumstances people lie. They find that the distribution of truthful, partially truthful and untruthful people is more or less the same independent of the stake size, the consequences of lying, learning, and the degree of anonymity.

Finally, the partner selection process taking place during the two pre-play rounds in "Golden Balls" draws our attention to the literature on discrimination. There exists a vast economic as well as psychological literature on racial and gender discrimination usually with the focus on the labor market (for a comprehensive survey see Altonji and Blank, 1999). Using the data of the US television game show "The Weakest Link", Levitt (2004) and Anonovics, Arcidiacono and Walsh (2005) test taste-based and information-based theories of discrimination, determining whether contestants discriminate on the basis of gender, age, race, and skill level. While Levitt (2004) finds some patterns consistent with information-based discrimination and taste-based discrimination against older players, Anonovics et al. (2005) reveal taste-based discrimination by women against men.

To summarize, there are various reasons to observe a different cooperation rate than the one predicted by game theory and the one observed in laboratory experiments without communication, high stakes, and endogenous partner selection.

3 Game show and data set

In this section we describe in detail the course of events in the game show (Section 3.1) and the data set (Section 3.2).

3.1 Structure of the game show

The game show "Golden Balls" consists of three rounds of play with the final round being divided into two phases.

Round 1 The game show starts with four players[5], usually two women and two men, who are briefly introduced by the show host, i.e., the players provide some information about themselves including their names, occupation and place of residence. Then the first round starts: 16 golden balls are mixed, twelve of them have written a cash amount (in £) inside and four have written the word "killer" inside. Killer balls are the worst for the players, because these may damage the jackpot in the final round. The balls containing a cash value are drawn from a lottery of 100 golden balls with a minimum ball value of £10 and a maximum ball value of £75,000.[6] Each player arranges the closed golden balls in two rows of two balls in front of herself. The two balls on the front row are opened by each player, and the revealed cash values or number of killers is common knowledge to each player. The content of the remaining two balls is private information to each player, i.e., the players are allowed to secretly look inside but then have to close the balls again. Afterwards the show host asks each player to state what is inside her hidden balls. The order in which players are asked for their statements is exogenously determined by the show host. Some time for discussion follows, in which the players express their distrust about each other's statements. The discussion ends with each player secretly casting a vote against one of the other players. On the basis of the votes, a player is eliminated from the show.[7] After the player who has to leave is determined, all players open their hidden back row balls and thereby reveal whether they stated the truth or not. The four balls of the leaving player are out of the game, while the remaining twelve are carried over to round 2.

Round 2 At the beginning of the second round, two new cash balls are drawn from the lottery and one killer ball is added. These three new balls are mixed with the remaining twelve from round 1, and are equally distributed to the three players at random. Hence, there are at most five killers among the 15 balls. Again the closed balls are arranged in two rows by each player, i.e., two balls are on the front and three balls are on the back

[5]Endemol UK ensured us that the four players do not know each other before the show, and enter and leave the television studio separately (they cannot make any further arrangements after the show).

[6]Players have only limited information about the lottery, i.e., they only know that there may be doubles and they know the margins of the distribution. But they do not know the distribution of the remaining 98 balls.

[7]The player who receives the highest number of votes has to leave the show. In case of a tie the players having received no vote can decide which player has to leave. If all players received one vote each, players discuss openly which player has to leave. If players do not reach a conclusion, ties are broken arbitrarily. In round 2 it is proceeded in the same way.

row. As in round 1 the two balls on the front row are opened and are common knowledge, while the three balls on the back are private information. This time the players determine themselves the order of making statements about the content of their back row balls. Like in the first round, the players then get some time for discussion and afterwards secretly choose a player they want to vote off. After the player to leave has been determined all ball values are revealed, the five balls of the leaving player are out of the game, and the final two players are identified.

Final Round The 10 balls from round 2 are carried over to the final round and one last killer ball is added. The maximal amount the players can gain is the sum of the highest five cash values out of the 11 balls. This amount is called the potential jackpot and its size is announced by the show host.

In the *first phase* of the final round the two players successively select five of the 11 mixed and closed balls, and these five values build the jackpot. The player who brought the highest amount of money from round 2 to the final round starts to select one of the balls to "bin", i.e., to be taken out of the game, and then chooses one ball to "win". The balls are not opened until they have been chosen. Then it is the other player's turn and vice versa until five balls have been selected for the jackpot. If a player chooses a killer ball for the jackpot the accumulated amount up to that point is reduced to one-tenth of the original value.

In the *second phase* of the final round the players play a prisoner's dilemma in which defection is a weakly dominant strategy (see Table I).[8]

Table I: Weak prisoner's dilemma

	Split (C)	Steal (D)
Split (C)	$\frac{1}{2}J$, $\frac{1}{2}J$	0 , J
Steal (D)	J , 0	0 , 0

C $\widehat{=}$ Cooperation, D $\widehat{=}$ Defection

Such a prisoner's dilemma has three pure-strategy Nash equilibria, namely (steal, split),

[8]The show host explains the different outcomes of the game in each episode with the same neutral words (for the exact wording see Supplementary Material A).

(steal, steal), and (split, steal).[9] Thus, each player has an incentive to defect, because she is never monetarily worse off when doing so. Before the players have to decide which strategy to play, they get some additional time, roughly 30 seconds, to discuss with each other what they are going to do.

The dilemma game is played as follows: Each player is assigned two balls, one with the word "steal" and one with the word "split" inside. Then both players choose one of the balls and open it simultaneously. If both players chose the split ball, the jackpot (J) is divided equally between the two players. If one player chooses steal and the other chooses split, the former gets the whole jackpot and the latter receives nothing. If both chose steal, both get nothing.

3.2 Data description

"Golden balls" was first aired on June, 18th 2007 as a late afternoon (5pm) game show and is still running today.[10] In total, we have records of 222 episodes, with 203 regular and 19 special episodes. In the special episodes there are either (i) players that have been on the show before and have "lost" or (ii) only players of the same sex. The regular episodes always consist of two women and two men and all players are on the show for the first time. Importantly, the first series (40 episodes) was filmed prior to the show's television premiere. Hence, all players in these episodes had no chance to observe others playing the show.

For all episodes, we recorded variables describing the players (occupation, hometown, gender, race, and age) and the game (all true and stated ball values in rounds 1 and 2, the order of making statements in both rounds, votes the players received and submitted, the potential jackpot size, values of binned balls by each player, the player's intended strategy[11], the jackpot size, interactions of players before and in the final (handshakes, promises), and the final decision). Table II provides an overview of the data.

[9] Two of the resulting Nash equilibria involve one player to cooperate. Applying the method of iterated elimination of weakly dominant strategies, however, leaves only the (steal, steal) equilibrium, which should be the only one observed.

[10] The show reaches up to 2.2 million people per episode which corresponds to a market share of 21% ("ITV strikes teatime gold", guardian.co.uk, July 3rd, 2007).

[11] Before the show starts, the players are individually and privately asked to explain which strategy they intend to play in the final. The recorded statement is only broadcasted to the television audience, but not to the other players or the audience in the television studio.

Table II: Summary statistics

Variable	Mean	SD	Min	Max	N
Occupation					
Social Job[1] (1 = social job)	0.14	0.34	0	1	887
Student (1 = student)	0.08	0.27	0	1	888
Pensioner (1 = retired)	0.03	0.17	0	1	888
Place of residence					
England (1 = England, 0 = SCO, WAL, NIR, IRL)	0.85	0.36	0	1	886
Large City[2] (1 = population > 268,300)	0.30	0.46	0	1	886
London (1 = London)	0.13	0.34	0	1	888
Gender, race, and age					
Gender (1 = male)	0.50	0.50	0	1	888
Race (1 = white)	0.92	0.27	0	1	888
Age[3] (1 = above 40)	0.43	0.50	0	1	888
Average cash ball in the show	5619.55	10374.12	10	75000	3108
Strategy statement[4] (0 = steal, 1 = split, 2 = other)	1.08	0.86	0	2	612
Round 1					
Value of open balls (balls 1 and 2)[5]	8802.64	13858.91	0	104000	888
Value claimed for balls 3 and 4	14265.86	13908.53	0	83000	888
Value of closed balls (balls 3 and 4)	7852.88	12315.07	0	83000	888
Number of killers in open balls	0.47	0.58	0	2	888
Number of killers claimed	0.23	0.43	0	2	888
Number of killers in closed balls	0.53	0.60	0	2	888
Player lied at least about one ball	0.53	0.50	0	1	888
Player lied at least about one value	0.32	0.47	0	1	888
Player lied at least about one killer	0.28	0.45	0	1	888
Number of killers taken to round 2	2.59	0.76	1	4	888
Round 2					
Value of open balls (balls 5 and 6)	9651.32	14275.73	0	103000	666
Value claimed for balls 7, 8 and 9	18421.19	16683.73	105	95000	666
Value of closed balls (balls 7, 8 and 9)	13352.47	16291.90	0	95000	666
Number of killers in open balls	0.44	0.58	0	2	666
Number of killers claimed	0.44	0.52	0	2	666
Number of killers in closed balls	0.75	0.69	0	3	666
Player lied at least about one ball	0.45	0.50	0	1	666
Player lied at least about one value	0.23	0.42	0	1	666
Player lied at least about one killer	0.28	0.45	0	1	666
Number of killers taken to final round	2.14	0.91	0	5	666
Value of balls taken to final round	23003.79	21134.80	150	143300	666
Final round (1st phase)					
Potential jackpot	51238.36	31261.51	5000	168100	444
Average cash ball	6932.27	12030.86	10	75000	1122
Number of killers	3.21	0.94	1	6	144
Number of killers to bin	1.74	0.92	0	4	144
Number of killers to win	1.47	0.88	0	4	144
Jackpot/Pot. jackpot	0.25	0.28	0.0001	1	444
Final round (2nd phase)					
Jackpot	13343.03	19247.56	3	100150	444
Decision (1 = split)	0.55	0.50	0	1	444
Outcome (0 = steal/steal, 1 = steal/split, 2 = split/split)	1.09	.75	0	2	222
Money taken home	4916.96	12000.86	0	100150	444
Money taken home (steal / split)	15693.11	20087.90	3	100150	94
Money taken home (split / split)	4783.64	8440.02	1.83	43950	148
Money left on the table	14426.34	20255.76	100	92330	108
Discussion (1 = starts discussion)	0.5	0.5	0	1	444
Handshake (1 = shake hands)	0.39	0.49	0	1	444
Mutual promise (1 = say promise)	0.25	0.43	0	1	444

[1] Note that we defined a social job as a job in which people care for other people, e.g., doctors, nurses, child minders, social workers, teachers, police officers, firemen, soldiers.

[2] Large cities are cities with more than 268,300 inhabitants (based on the Mid-2008 Population Estimates published by the Office for National Statistics).

[3] We estimated by personal judgment whether a player is below or above 40.

[4] Players secretly make this statement about the strategy they plan to play in the final before the show starts. It was introduced in episode 19, series 1.

[5] Killer balls are counted as zero for all value variables.

4 Strategic considerations of the players

Following the structure of the game show, we analyze the player's incentives to behave in a particular way. The final goal of each player has to be reaching the final round with a jackpot as high as possible and, most importantly, facing a player who intends to split, independently of whether the player herself prefers to steal or split. Thus, the players have to trade off these goals against each other.

In the pre-play, players base each of their two voting decisions on exogenous as well as endogenous criteria. We define exogenous criteria as characteristics of the players that are determined previously to the show, e.g., the player's age, gender, race, or place of residence. In contrast, endogenous criteria evolve during the course of the game, and are, for instance, the ball values dealt to the players, the order of making statements in round 2, or whether a player lied or not. Besides, the latter two criteria can be strategically used by the players. Players may be able to signal trustworthiness, since they can decide whether to lie or not about the content of their hidden back row balls. Making the statement first in round 2 may influence the other players' statements, e.g., if one player confesses a killer, the others may do the same. Further, during the discussions in round 1 and 2, they can, for instance, state why they distrust a certain player and try to convince the other players to vote this player off.
Once players reach the final round, they have to make sure that their opponent chooses the split ball in the final decision. Players use the discussion in the final round to reassure the opponent that they will cooperate, e.g., players promise each other to share the jackpot and/or shake hands on sharing it.

To summarize, the final decision as well as the voting decisions are functions of exogenous personal and endogenous characteristics of the players. Unfortunately, a game-theoretic analysis of the game is not feasible because the game is too complex and strategies are ill-defined. However, we will use the logic of backward induction to analyze the decisions made in the game. Therefore we start with analyzing the decision in the final round and then successively analyze the voting decisions made in round 2 and 1.

5 Analysis of cooperative behavior in the PD

In this section we identify the influencing factors of cooperative behavior. We observe an average cooperation rate of 54.5%, which is higher than the one found in weak prisoner's dilemma experiments without communication, pre-play, and high stakes (e.g., Ortmann and Tichy, 1999). The rate is also slightly higher than the one observed in "Friend or Foe", another television game show experiment, (e.g., List, 2006). Further, we observe mutual cooperation in 33.3% and successful defection in 42.3% of the cases. Unilateral defectors take home three times as much money as mutual cooperators, £15,693 versus £4,784, and the average amount of money left on the table due to mutual defection is £14,426. Altogether 108 players left £1,558,045 on the table.

We will start the analysis in Section 5.1 with a discussion of the data and the variables that may have an influence on cooperative behavior. In Section 5.2 we will test the derived hypotheses in an empirical analysis. In addition we briefly present an alternative approach in the Supplementary Material B.

5.1 Possible determinants of cooperative behavior

As we already pointed out in Section 2, the prisoner's dilemma under consideration is different in various aspects from the ones usually analyzed in the literature. In this section we will discuss the potential impact of those differences on the observed degree of cooperation, and present first results. The section is divided into four parts, (i) player characteristics, (ii) stake size, (iii) communication, and (iv) pre-play, following the presentation of variables in the regression tables in Section 5.2

(i) Player characteristics

In this category we discuss variables that are exogenously determined. These are demographic player characteristics, which are also used to describe opponent- and team-characteristics, and a variable expressing the player's experience with the show.

Experience We define the players of the first 40 episodes (series 1) as unexperienced players. They had no chance to observe other contestants playing the game. In con-

trast, all later episodes have been filmed after the television premiere of "Golden Balls". Thus one could conjecture that the experienced players are more familiar with the show and therefore better in assessing whether cooperation or defection could be successful or not. But from the raw data, we do not observe a substantial difference, neither in the cooperation rate nor in the distribution of outcomes.[12]

Demographics Player's demographics are defined as exogenous characteristics of a player such as gender, age, race, place of residence, or occupation (descriptive results are reported in Table C.1 and Table C.2 in the Supplementary Material). The relation between demographic characteristics and social behavior seems to be rather ambiguous. Deriving clear-cut hypotheses about the influence of these characteristics on the player's propensity to cooperate is therefore not possible.

Overall, there seems to be no difference between the cooperation rates of men and women. Concerning the rate of successful cooperation we find that it is lowest for female teams (28.6%) and highest for mixed gender teams (35.3%). The null hypothesis of no difference between the overall cooperation rate of men and women cannot be rejected (p=0.435) as well as the one for the mutual cooperation rate (p=0.503). Players above the age of 40 cooperate significantly more than players below 40 (p=0.001).[13] There are only small differences in the success rates of cooperation, teams of players below 40 have the lowest success rate. Whites are more likely to cooperate compared to non-whites, but the difference is not significant (p=0.334).

Based on a player's hometown, we construct variables indicating whether her place of residence is England, London, a small or a big city (see Table II). Players living in England cooperate significantly less than players from other parts of Great Britain (p=0.000). In addition, if neither player lives in England the success rate of cooperation is 50.0% versus 32.5% if both players live in England, and 33.3% if it is a mixed team. The failure

[12]A two-sided binomial probability test can neither reject the null hypothesis of no difference between the cooperation rate of experienced (52.5%) and unexperienced (54,9%) players (p=0.372), nor the null hypothesis of no difference between the probability of mutual cooperation of experienced (34,1%) and unexperienced (30,0%) players (p=0.481), see Table C.1 and Table C.2 in the Supplementary Material. This result is in contrast to the finding of Oberholzer-Gee et al. (2010) who find an effect of learning for players in later episodes of "Friend or Foe". All tests used in this paper are two-sided binomial probability tests, unless stated otherwise.

[13]However, gender conditional on age tends to have an effect on the cooperation rate, i.e., women below 40 cooperate more than men below 40 and vice versa for men and women over 40. Note the results concerning age should not be attached too much weight since the age categories are merely assessed by personal judgment.

rate is highest for mixed teams and significantly higher than the one of English teams (p=0.002).[14]

Further, decisions made in the game might have implications beyond their immediate consequences, because the game is played in front of a large television audience and is therefore possibly being watched by friends, family members and/or colleagues. Depending on the player's occupation, it can be in her interest to appear trustworthy. For instance, police officers act as role models for observing the law and behaving correctly, or teachers are responsible for a moral education of children. These players have an incentive to behave in a fair way, especially when it comes to choosing the strategy in the prisoner's dilemma. We identify roughly 15% of contestants with an occupation for which their reputation is a valuable asset, and construct a variable "social job", including e.g., priests, policemen, firemen, childminders, and teachers. Having a social job might influence cooperative behavior because of two different reasons. On the one hand, players with a social job could be more cooperative because they want to show that they behave socially responsible. On the other hand, the causality could be vice versa: Having a social job could be a sign for being a cooperator itself, because a cooperative person chooses a job in which she can behave according to her preferences. In that case cooperative behavior would not be driven by the opportunity to appear trustworthy. Summarizing, it remains unclear whether we will observe an effect of having a social job.

Social closeness The sociological literature argues that the degree of similarity between players has an impact on their social interactions. While some people might be willing to cooperate without discrimination, others are highly suspicious of people who are not like them and prefer to keep them at arm's length. In many social networks, e.g., friendships or business relations, one observes that individuals associate disproportionately with others who are similar to themselves, i.e., people are more likely to form social ties with others who are alike. This tendency of people to relate to similar types is referred to as "homophily", first defined by Lazarfeld and Merton (1954).[15] Such motivated,

[14]The hypothesis of no difference between the success rates of not-English and English teams (p=0.284), of not-English and mixed teams (p=0.454), as well as of English and mixed teams (p=0.849) cannot be rejected. Neither can the hypothesis of no difference between the failure rates of not-English and English teams (p=0.495) and of not-English and mixed teams (p=0.162) be rejected.

[15]For a survey with respect to sociology see Jackson (2008) and with respect to cooperative game theory see van den Nouweland and Slikker (2001). Homophily is usually based on a variety of characteristics, including gender, race, age, region and education.

we construct an index for the closeness between players by accounting for players' age, gender, race, occupational status, and place of residence (England). The index ranges from 0 to 1, weighting each component by one-fifth. For instance, if both players in the final are male, white, have a social job and live in England, the index takes a value of 0.8. Concerning the distribution, we observe the majority of players to have an index-value of either 0.6 (44%) or 0.8 (32%).

(ii) Stake size

Categorizing the jackpot in five divisions (see Table C.1 and Table C.2 in the Supplementary Material) we find that the cooperation rate decreases with a step-wise increase in the jackpot size. But, surprisingly, the rate declines sharply from 73.6% for jackpots below £500 to roughly 50% for jackpots above £500. The difference, taking the cutoff £500, is highly significant (p=0.000). Concerning the mutual cooperation rate, it is significantly higher if the two players face a jackpot below the level of £500 (p=0.000). This result is even more remarkable if one bears in mind that a stake size around £500 is already much higher than the one used in most laboratory experiments. At the same time, however, the cooperation rate rises with an increase in the potential jackpot, i.e., the highest possible jackpot the players could obtain after the first phase of the final round. Hence, the effects of the actual and the potential jackpot operate in opposite directions. One might presume that the players' perception of the actual jackpot depends on the potential jackpot, i.e., two actual jackpots equal in size will be judged differently depending on their difference to the potential jackpot. In Table III we explore this issue further.
We depict the cooperation rate for the five different jackpot categories and within each we split up the rate by the four categories of the potential jackpot, i.e., the difference between the highest potential and the actual jackpot is increasing within each category. Fixing the jackpot categories, we find that the cooperation rate almost always increases with the size of the potential jackpot. This corroborates the idea of a biased jackpot perception, that we will discuss in the following.

Expectation Players might build an expectation about the size of the actual jackpot depending on the observed size of the potential jackpot. This expectation is used to judge the size of the actual jackpot. But, computing the correct expectation is a rather

Table III: Relation between jackpot and potential jackpot

Jackpot	**Potential Jackpot**	**Split**	
in £	in £	Row %	**N**
[3, 500]	[5000, 30000]	62.5	32
N=72	(30000, 45000]	85.7	14
	(45000, 75000]	75.0	12
	(75000, 168100]	85.7	14
(500, 2500]	[5000, 30000]	41.2	34
N=100	(30000, 45000]	59.5	42
	(45000, 75000]	44.4	18
	(75000, 168100]	100.0	6
(2500, 10000]	[5000, 30000]	50.0	30
N=102	(30000, 45000]	43.3	30
	(45000, 75000]	66.7	18
	(75000, 168100]	58.3	24
(10000, 30000]	[10000, 30000]	38.9	18
N=116	(30000, 45000]	45.7	46
	(45000, 75000]	52.9	34
	(75000, 168100]	55.6	18
(30000, 100150]	[30000, 45000]	50.0	6
N=54	(45000, 75000]	50.0	22
	(75000, 168100]	46.2	26

Note, that the difference between the jackpot and the potential jackpot is increasing with an increasing potential jackpot per jackpot category.

difficult task if no computer is at hand. Therefore, players need some alternative method to calculate their expectation. As mentioned before, the first 40 episodes have been broadcasted before all other episodes were filmed. Henceforth, we assume players to take the observed average ratio between the jackpot and the potential jackpot in series 1 as an estimate to roughly calculate their expectation. The average jackpot in series 1 is £13,066 which corresponds to 27.5% of the average potential jackpot of £47,526. This ratio is multiplied by the observed potential jackpot in each episode, and determines the players' expected jackpot.[16] Depending on whether the jackpot is above or below the player's expectation, the propensity to cooperate changes. The cooperation rate is significantly higher (p=0.002) if the jackpot is below the expectation, and cooperation is much less successful if the expectation threshold is taken, i.e., 18.4% versus 41.1% of mutual cooperation.

[16]The ratio between the jackpot and the potential jackpot observed in the episodes following series 1 is 25.7% which is very similar to the ratio observed in the episodes of series 1.

(iii) Communication

Apart from communication that takes place in round 1 and 2, and in the first phase of the final round, players explicitly get some time to discuss the strategy they intend to play in the prisoner's dilemma. The players use this time to assure each other their willingness to cooperate, i.e., to choose the split ball. As described in Section 2, studies have shown that especially face-to-face communication, involving a mutual agreement to cooperate, increases the cooperation rate significantly. We observe that 24.8% of the players voluntarily promise each other to cooperate. In addition to verbal communication, 39.2% of the players shake hands to corroborate their intention to split, and 41.4% out of those do both, i.e., they shake hands and promise each other to share the jackpot. We define dummy variables to control for mutual promises, handshakes and for whether the contestant starts the final discussion. We expect handshakes and promises to increase successful cooperation.

(iv) Pre-play

The next three potential determinants of cooperative behavior evolve within the pre-play.

Lying Lying is rather common during the pre-play rounds of the game show and might be thought of as an inherent part of the game. Players are concerned about maximizing the stake size, thus having low values or killer balls increases the probability of being voted off. Driven by the fear of being eliminated from the game when having "bad" balls, the contestants bluff their way to the final. But a player is revealed as a liar after each round, and might get a reputation of being not trustworthy, which possibly prevents her from lying. Analyzing our raw data, we find that 43% of the players who reach the final lied about the content of their balls in round 1 and 37% in round 2. A possible implication of lying could be that liars may want to repay their "guilt" and therefore are more tempted to cooperate. But a player who has not lied might perceive a liar as untrustworthy per se and thus is less willing to cooperate. We control for the potential effects of lying by introducing dummy variables for whether the player or her opponent has lied about a cash value or a killer during the two rounds of pre-play, since we believe that the perception of concealing a killer may differ from that about overstating a cash amount.

Kindness In addition, we include variables linked to kindness, experienced kindness, and its repayment in our analysis. Firstly, we account for the impact of the voting decision on the behavior of those contestants who remain in the show. After each round of the pre-play, contestants need to secretly cast a vote against a certain player whom they want to leave the game. It can happen that a player in the final has voted against her opponent during the pre-play. This might influence the player's behavior in the final: A player is less likely to cooperate, since she expressed her dislike against the other player before. In this respect we construct a dummy variable that identifies a player who voted against her opponent.

Secondly, we define a variable termed "should have left the game" in order to investigate whether a player responds to experienced kindness. The variable is constructed as follows: First we rank the three players in round 2 with respect to their weighted sum of cash values and killer balls. The dummy points at the player with the lowest weighted sum. From a purely monetary perspective, this player should be voted off the game. But if such a player nevertheless reaches the final round, she is aware of owing her "survival" to her opponent.[17] In this respect one could surmise that she is more likely to split the jackpot in order to pay back her survival.

Luck At last, we want to focus on the first phase of the final round, in which the jackpot is built by an alternating selection of balls. The player who starts to select the first balls to bin and win is the player who brought along the higher sum of cash values, i.e., contributed most to the potential jackpot. The mean difference between the player's contributions is £1,678 without accounting for the possible damage caused by killer balls. The value of the resulting jackpot is determined purely at random, but one player might be more lucky than the other, i.e., chooses higher values or bins more killer balls. We control for those effects, constructing three dummy variables: One for the player who contributes most to the potential jackpot, one for the player who selects the highest values, and one for the player who bins most killer balls. These players could feel entitled to a larger piece of the pie and are therefore less likely to cooperate. This would be in line with the findings on entitlement and fairness of Rutström and Williams (2000).

[17] In order to rank the players we use the *ex-post* cash-killer-criterion which is described and discussed in detail in Section 6. We assume that a player, who does have the lowest weighted monetary amount is aware of this and does value her survival. Often the players address their pass to the final round during the final discussion and thank their opponent for having taken her so far.

5.2 Regression analysis

In order to explore the individual decision process when playing the prisoner's dilemma game, we estimate bivariate probits of the probability that player i chooses split or steal as a function of own and opponents' characteristics as well as variables determined in the pre-play (see Table IV).[18] Additionally, to analyze team cooperation rates, we estimate ordered probits (see Table V and Table C.3 in the Supplementary Material). We order the outcomes by coding the team outcome as equaling 0 if both players choose steal, equaling 1 if one player chooses steal and the other chooses split, and equaling 2 if both players choose split.

Results for unilateral cooperation

We estimate bivariate probits on the probability that player i chooses split ($y = 1$), or steal ($y = 0$) as a non-linear function of exogenous demographic player characteristics, D, and variables that evolved during the game, G, including a constant and one interaction term, namely the interaction of "handshakes", x_1, and "promise", x_2. The conditional probability that player i chooses split is

$$P(y = 1|x_1, x_2, D, G) = \Phi(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + D\theta + G\nu) = \Phi(m),$$

where $\alpha, \beta, \theta, \nu$ are parameters to be estimated, $\Phi(\cdot)$ is the standard normal cumulative distribution function, and m denotes the index $\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + D\theta + G\nu$. The marginal effect for the k-th independent variable is computed as

$$\frac{\partial P(y = 1|x_1, x_2, D, G)}{\partial x_k} = \phi(m)\beta_k \quad k = 2, ..., K$$

[18] In addition, we estimate the same probits including one additional dummy variable, namely whether player i "voted against (her) opponent" in the pre-play which results in a significantly negative effect on cooperation. We exclude the dummy in our main analysis, since including it reduces the data set by 55 observations which is due to a voting result of 2:1:1:0 in round 1 or 1:1:1 in round 2 (in these cases it is analytically not possible to trace back the players' individual voting decision). Except for the episodes with a tie, the outcome of the voting decision is 2:1:0 in round 2, such that none of the final players received a vote from their opponent. Thus, the control variable only comprises of the voting result in round 1. Additionally, we can only control for the voting decision made by the particular player herself, and not whether this player received a vote by her opponent in the final, since after round 1 the players can only speculate who had cast a vote against them. Therefore we can only identify 18 out of 390 players to have casted a vote against her final opponent. Thus, the variable "voted against (her) opponent" has not much explanatory power.

where $\phi(\cdot)$ is the standard normal density function. The magnitude of the derivative is proportional to $\phi(m)\beta_k$. A change in one of the independent variables results in an effective percentage change in player i's likelihood to cooperate.[19]

Table IV reports the estimation results of four probit models, which differ with respect to the included variables in the group of stake size and pre-play. Model (1) and (3) include the continuous variables jackpot and potential jackpot, while models (2) and (4) include the dummy variable describing whether the expectation of the resulting jackpot is met or not. In addition to the controls for personal players' demographics, stake size, and communication, in models (3) and (4) variables determined in the pre-play are introduced.

Throughout, addressing demographics, we find that age, whether a player lives in England, or whether both players live in a small city have a significant effect on the likelihood to cooperate. If player i is above the age of 40, she is about 16% more likely to cooperate, i.e., split. But, a player who lives in England, compared to any other part of Great Britain, is roughly 27% more likely to be a defector, i.e., steal. Additionally, if both final players live in a small city, then player i's likelihood to cooperate decreases by up to 14%. All three effects are very robust with approximately the same magnitude across the four models. The results on players' gender, race, and occupational status, however, exhibit no significant effects. We also find no effect for player i's experience, i.e., the control for the first series has no significant effect, indicating that players are not able to profit from a learning effect if they could watch the game show on television before.

The results on stake size support our findings gained in Section 5.1: As suggested, the higher the actual jackpot, the less likely player i cooperates; while, the higher the potential jackpot, the more likely player i cooperates. In addition models (2) and (4) highlight our descriptive finding that a player is more willing to cooperate if the actual jackpot is

[19] If x_k is a dummy variable, the marginal effect is computed as the discrete difference

$$\frac{\triangle P(y=1|x_1,x_2,D,G)}{\triangle x_k} = \Phi(m|x_k=1) - \Phi(m|x_k=0), \quad k=2,...,K.$$

Note that the marginal effect of the interacted dummy variable "handshakes" (x_1) and "promise" (x_2) is equal to the discrete double difference

$$\frac{\triangle^2 P(y=1|x_1,x_2,D,G)}{\triangle x_1 \triangle x_2} = \Phi(m) - \Phi(\beta_1+\alpha+D\theta+G\nu) - \Phi(\beta_2+\alpha+D\theta+G\nu) + \Phi(\alpha+D\theta+G\nu).$$

Table IV: Results from binary probit on unilateral cooperation (1)

$y = 1$ (Split)	Marginal Effects							
	Model (1)		**Model (2)**		**Model (3)**		**Model (4)**	
Player Characteristics								
Unexperienced	-0.028	(0.069)	-0.050	(0.069)	-0.019	(0.074)	-0.041	(0.073)
Male	-0.018	(0.059)	-0.017	(0.058)	-0.019	(0.060)	-0.018	(0.060)
Age (>40)	0.164***	(0.064)	0.161**	(0.063)	0.174***	(0.066)	0.172***	(0.065)
White	0.117	(0.148)	0.113	(0.145)	0.052	(0.147)	0.046	(0.143)
England	-0.269***	(0.078)	-0.267***	(0.077)	-0.274***	(0.082)	-0.271***	(0.081)
London	0.001	(0.103)	0.003	(0.100)	-0.059	(0.107)	-0.066	(0.104)
Large City	-0.054	(0.077)	-0.051	(0.075)	-0.024	(0.081)	-0.018	(0.080)
Student	-0.009	(0.096)	0.001	(0.095)	-0.020	(0.095)	-0.014	(0.095)
Pensioner	-0.143	(0.160)	-0.160	(0.163)	-0.126	(0.164)	-0.145	(0.167)
Social Job (Reputation)	-0.017	(0.087)	-0.032	(0.087)	-0.016	(0.090)	-0.033	(0.090)
Team Characteristics								
Index (Social Closeness)	0.458	(0.309)	0.469	(0.306)	0.592*	(0.322)	0.606*	(0.318)
Team Male	-0.147	(0.099)	-0.152	(0.097)	-0.183*	(0.100)	-0.184*	(0.097)
Team Female	-0.054	(0.095)	-0.058	(0.095)	-0.083	(0.097)	-0.084	(0.097)
Team Age> 40	-0.011	(0.103)	-0.009	(0.103)	-0.007	(0.106)	-0.001	(0.107)
Team Age< 40	-0.049	(0.091)	-0.055	(0.090)	-0.073	(0.095)	-0.075	(0.093)
Team White	-0.118	(0.122)	-0.124	(0.118)	-0.108	(0.124)	-0.116	(0.121)
Team England	-0.019	(0.093)	-0.001	(0.093)	-0.021	(0.097)	-0.002	(0.096)
Team Large City	-0.127	(0.101)	-0.128	(0.100)	-0.147	(0.101)	-0.145	(0.101)
Team Small City	-0.141**	(0.067)	-0.127*	(0.067)	-0.143**	(0.069)	-0.126*	(0.068)
Opponent Characteristics								
Opp. Student	0.087	(0.097)	0.095	(0.095)	0.107	(0.098)	0.111	(0.097)
Opp. Pensioner	0.067	(0.162)	0.049	(0.159)	0.053	(0.171)	0.032	(0.168)
Opp. Social Job	0.040	(0.085)	0.027	(0.085)	0.087	(0.086)	0.073	(0.086)
Stake Size								
log(Jackpot)	-0.057***	(0.014)			-0.057***	(0.015)		
log(Pot. Jackpot)	0.097**	(0.048)			0.077	(0.050)		
Jackpot < Expectation			0.162***	(0.056)			0.173***	(0.059)
Communication								
Started Discussion	-0.008	(0.048)	-0.009	(0.047)	-0.005	(0.049)	-0.006	(0.049)
Handshakes	-0.151**	(0.067)	-0.179***	(0.066)	-0.167**	(0.068)	-0.196***	(0.067)
Promise	-0.032	(0.108)	-0.063	(0.107)	-0.060	(0.109)	-0.095	(0.108)
Handshakes*Promise	0.281**	(0.139)	0.309**	(0.170)	0.339**	(0.142)	0.372***	(0.054)
Pre-play								
Acc. Most Money					-0.098*	(0.058)	-0.099*	(0.057)
Selected Higher Values in Bin/Win					-0.148*	(0.079)	-0.157**	(0.077)
Binned Most Killers in Bin/Win					-0.091	(0.090)	-0.092	(0.088)
Lied About Cash Value					-0.105*	(0.059)	-0.106*	(0.059)
Lied About Killer					-0.018	(0.055)	-0.017	(0.055)
Opp. Lied About Cash Value					0.019	(0.058)	0.014	(0.058)
Opp. Lied About Killer					-0.048	(0.056)	-0.041	(0.055)
"Should Have Left The Game"					0.133**	(0.062)	0.144**	(0.061)
Wald $\mathcal{X}^2$	61.23***		55.18***		74.06***		69.14***	
Log-Likelihood	-273.26		-277.81		-264.75		-268.63	
Pseudo R^2	0.10		0.09		0.13		0.11	
N	441		441		440		440	
Number of Clusters	222		222		222		222	

standard errors in parentheses are corrected for episode clusters; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

below her expectation, i.e., her likelihood to cooperate significantly increases by roughly 16%. All effects are independent of player's characteristics or communication, and are highly robust.

Addressing the controls for communication, we find that neither starting the final discussion nor voluntarily promising each other to cooperate are significant determinants. But, we find a negative effect of shaking hands: If both final contestants shake hands during the final discussion, each player actually is more likely to defect. Thus, handshakes seem to serve as an instrument to manipulate the opponent's attitude towards cooperation. Whether shaking hands is perceived differently depending on a promise made at the same time, we interact both dummy variables. As the results show, we find a positive significant interaction effect: Shaking hands in combination with a promise actually increases the player's likelihood to cooperate. We will further explore whether these effects help cooperators to coordinate by looking at team outcomes in the next subsection.

In models (3) and (4) we introduce controls describing pre-play determinants. We find that a lie about a cash value is treated differently with respect to its influence on cooperation. If player i lied about a cash value, she is roughly 10% more likely to defect in the prisoner's dilemma, but concealing a killer has no effect, neither does a lie of her opponent. Regarding the control indicating that player i should have left the game before, player i is roughly 14% more likely to cooperate. As hypothesized a player who should have been voted off the game, but nevertheless made it to the final, is likely to pay back the opponents' confidence. Further, we find that both, having accumulated more money as well as having selected the higher values when building the jackpot, have a significantly negative impact on the players' likelihood to cooperate. As suggested, a player might feel entitled to a larger piece of the pie, since she contributed more to the stake size. Testing for any interaction effect with respect to a player's expectation, we find that a player who accumulated more money to the jackpot and whose expectation is not met, is significantly more likely to steal. This result suggests that a player's perception of the jackpot is correlated with her contribution to the potential jackpot.
Besides, the introduction of the pre-play controls results in two additional effects regarding team characteristics. We find that a male player is significantly less likely to cooperate

with a male opponent, indicating that men are more competitive when facing the same sex. Another significant determinant of cooperative behavior is the index of social closeness between players. The more similar both final players are with respect to their age, gender, race, and place of residence, the more likely player i is to cooperate with her opponent. This finding suggests, that in-group biases may be present.

Results for mutual cooperation

To understand why players arrive at different outcomes in the prisoner's dilemma, we estimate ordered probit models. We observe the discrete variable y that can take on three values, i.e., it equals 0 if both players choose steal, it equals 1 if one player chooses steal and the other chooses split, and it equals 2 if both players choose split. The boundaries between the three cases are determined by the threshold (ξ_i), which needs to be estimated along with the rest of the parameters. The probabilities of the three events $y = 0; 1; 2$ are given by $P(y = 0) = \Phi(\xi_1 - m)$, $P(y = 1) = \Phi(\xi_2 - m) - \Phi(\xi_1 - m)$, $P(y = 2) = \Phi(m - \xi_2)$, where $m = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + D\theta + G\nu$. The marginal effect of d_k (g_k) for the $j-$th response is computed as[20]

$$\frac{\partial P(y = j|x_1, x_2, D, G)}{\partial x_k} = [\phi(\xi_{j-1} - m) - \phi(\xi_j - m)]\beta_k.$$

The results for the three different outcomes are presented in Table V and Table C.3 in the Supplementary Material.

[20] If d_k (g_k) is a dummy variable, then the marginal effect is computed as the discrete difference

$$\triangle P(y = j|x_1, x_2, D, G) = P(y = j|(x_1, x_2, D, G) + \triangle x_k) - P(y = j|(x_1, x_2, D, G)).$$

Note that the marginal effects of the variables that are interacted involve the coefficient of the interaction term. Therefore, the marginal effect of x_1 (analog for x_2) for the j-th response is calculated as

$$\frac{\partial P(y = j|x_1, x_2, D, G)}{\partial x_1} = [\phi(\xi_{j-1} - m) - \phi(\xi_j - m)](\beta_1 + \beta_{12} x_2);$$

and the magnitude of the interaction effect for the j-th response is given by

$$\frac{\partial P(y = j|x_1, x_2, D, G)}{\partial x_1 \partial x_2} = [\phi(\xi_{j-1} - m) - \phi(\xi_j - m)]\beta_{12} - [\phi'(\xi_{j-1} - m) - \phi'(\xi_j - m)](\beta_1 + \beta_{12} x_2)(\beta_2 + \beta_{12} x_1),$$

where $\phi'(\cdot)$ denotes the first derivative of the normal density function w.r.t. its argument. Standard errors are computed by the delta method.

Table V: Results from ordered probit on outcomes in the PD (1)

$y = 0; 1; 2$	Marginal Effects					
	Steal/Steal (0)		**Split/Steal (1)**		**Split/Split (2)**	
Team Characteristics						
Team Unexperienced	0.000	(0.063)	0.000	(0.013)	0.000	(0.077)
Team Male	0.091	(0.089)	0.006	(0.011)	-0.097	(0.083)
Team Female	0.010	(0.064)	0.002	(0.012)	-0.012	(0.075)
Team > 40	-0.068	(0.065)	-0.023	(0.032)	0.091	(0.096)
Team < 40	0.103	(0.066)	0.012	(0.011)	-0.115	(0.070)
Team England	0.121**	(0.050)	0.041	(0.026)	-0.162**	(0.071)
Team Small City	0.079	(0.048)	0.016	(0.013)	-0.095	(0.058)
Index (Social Closeness)	-0.302*	(0.162)	-0.063	(0.635)	0.364*	(0.195)
Pre-Play						
Team Never Lied	-0.046	(0.059)	-0.013	(0.022)	0.059	(0.081)
Team Lying	0.028	(0.053)	0.005	(0.010)	-0.033	(0.063)
Communication						
Handshakes	0.078	(0.054)	0.016*	(0.009)	-0.094	(0.058)
Promise	-0.081	(0.053)	0.127***	(0.016)	0.098	(0.064)
Handshakes*Promise	-0.295**	(0.129)	0.029	(0.046)	0.324**	(0.137)
Stake Size						
log(Jackpot)	0.051***	(0.013)	0.011	(0.107)	-0.062***	(0.015)
log(Pot. Jackpot)	-0.081*	(0.043)	-0.017	(0.178)	0.098*	(0.051)
Wald $\mathcal{X}^2$	36.90***					
Log-Likelihood	-219.55					
Pseudo R^2	0.08					
N	54		94		74	
Number of Clusters	222					

Standard errors in parentheses are corrected for episode clusters; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Concerning team characteristics, Table V shows that teams of English players are more likely to defect. But the more similar both team players are with respect to their age, gender, race, and place of residence, i.e., the higher the index value, the more likely they manage to successfully cooperate. Addressing the division of team by gender, age or having lied in the pre-play, there are no considerable differences. The results also highlight the significance of handshakes and promises. Teams that shake hands or promise each other to cooperate, are more likely to miscoordinate. In addition, handshakes in combination with a promise increase the likelihood of mutual cooperation and decrease the one of mutual defection. As the analysis of the raw data suggests, we find a highly significant and inverse effect of the actual and potential jackpot. Both players are 5% more likely to defect when the stakes are large, and are 6% less likely to cooperate. In contrast, teams are almost 10% more likely to cooperate when the potential jackpot increases. The results in Table C.3 (Supplementary Material) support the discussed findings, highlighting that the players' expectation about the stake size is also a significant influencing factor of mutual cooperation. If both players' expectation about the jackpot is above the actual one, they are 17% more likely to cooperate, but 16% more likely to defect.

Summary We identify various player characteristics (e.g., age, living in England, or social closeness), stakes size, and communication, as well as pre-play to be significant influencing factors of unilateral and mutual cooperation. Most noticeable we find a robust and substantial negative effect of the actual stake size on cooperation, i.e., the higher the jackpot the more likely players are to defect. Controlling for the effect of handshakes and voluntarily stated promises we find that players who shake hands are more likely to defect, while handshakes in combination with a promise are likely to result in cooperation and successful coordination. Both effects are robust and independent of player characteristics and pre-play determinants. We are not aware of any other study having shown similar effects.

6 Pre-play decision making: Voting behavior

As the results from the probit regressions suggest, the opponents' characteristics play a decisive role for cooperative behavior. Therefore, we now draw our attention to the pre-prisoner's dilemma play, i.e., the selection of the two finalists. Starting with four contestants, each player faces the decision for whom to vote to leave the game in round 1 and 2. The players' voting behavior will be a function of observable characteristics and subjective criteria that maybe inconsistent with one another: Firstly, players have powerful monetary incentives and would like to vote off the player who has the lowest cash values or most killers in her golden balls. But, at the same time, players need to evaluate the opponent's character in view of the final round, i.e., assess her trustworthiness, sympathy, or susceptibility to manipulation. Finally, the players need to vote in a way that increases their own survival, thus their optimal action depends critically on the belief about how other players vote.

In the next three sections we will show how the players balance their voting decision, bearing in mind the stake size, the opponent's character and their own survival. In section 6.1, we discuss the expected voting behavior with the help of three objective evaluation criteria. In Section 6.2 we briefly describe the impact of personal player characteristics as well as round-specific determinants. Finally, in Section 6.3 we empirically test the validity of the objective criteria.

6.1 Objective criteria

Each episode, before the players have to decide against whom to cast a vote, the show host reminds the players to "keep in the cash, and kick out the killers". Following this prompt, we describe the player's voting decision by means of three objective criteria assuming that it is a player's aim to maximize the potential stake size. Each episode and round these criteria predict one particular player who *should* be voted off: The **cash-(CC)** and **cash-killer-criterion (CKC)** are constructed on the basis of the (weighted) monetary values of balls and declare the player with the lowest amount of money to be voted to leave the game. While we count a killer ball as a ball with zero value in the CC, we attach the killer balls a weight of 0.1 in the CKC, i.e., one killer ball reduces the monetary value to one-tenth of the original value, a second killer ball reduces it to one-hundredth of the original value and so forth. The **killer-criterion (KC)** focuses on the number of killer balls per player, and accordingly declares the player with the highest number of killers to be voted off.

Further, within each criterion we distinguish three different time-dimensions, i.e., we determine the prediction of each criterion separately taking into account (i) the two opened balls on the front row (**ex-ante**), (ii) the two open balls and the statements of the hidden back row balls (**stated**), and (iii) all revealed balls (**ex-post**[21]).

By means of these three, respectively nine criteria, we analyze to what extent each criterion explains the player's voting decision within and between round 1 and 2. Descriptive results are reported in Table C.4, Table C.5, and Table C.6 in the Supplementary Material.[22]

First we want to look at the proportions of players who are effectively voted off in line with the three criteria (see Table C.4 in the Supplementary Material). Focusing on the time-division of each criterion, in round 1 we find that most players vote in line with the prediction of the *ex-ante* CKC and *ex-ante* CC, as well as of the *stated* KC. In round 2 instead, the *ex-post* CKC and the *ex-post* CC dominate, but the *stated* KC again yields the best prediction. Overall the KC, especially when looking at the *stated* values, fits

[21]The *ex-post* criteria serve to test whether the players use the *ex-ante* criteria as a best estimation of the true state.

[22]Note that we exclude those episodes which have a voting result of 2:1:1:0 in round 1 or a tie in round 2, since it is analytical impossible to reconstruct the players' individual decision. Additionally, in Table C.5 and Table C.6 we restrict the sample to only those players who take part in both rounds (with an almost equal share of males (48.5%) and females (51.5%)); thereby we can compare the voting results for both rounds taking the same player's decisions into account.

best: In round 1, 81.1% of players who are voted off have the highest number of killer balls both on their front row as well as stated on their hidden back row balls; in round 2 this proportion slightly reduces to 70.3%, but still exceeds the CC and CKC.
The findings are confirmed when considering for each criterion the proportions of players who received a vote when predicted, additionally distinguished by gender (see Table C.5 in the Supplementary Material).
As above, we focus on the time-dimension of each criterion and find that the players most frequently vote in line with the *ex-ante* CKC and CC in round 1, but in line with the *ex-post* CKC and CC in round 2. Concerning the KC, players vote in line with the *stated* KC in round 1 and the *ex-ante* KC in round 2. Separating these findings by gender, we observe that significantly more males than females vote in line with the *stated* KC and *ex-ante* CKC in round 1 ($p = 0.024$ and $p = 0.034$) whereas in round 2 more females vote in line with the *ex-post* CC ($p = 0.021$).
Most noticeable, in both rounds the players take the statements about killer balls seriously, although one might argue that statements are only cheap talk and should therefore be ignored. But a statement about a killer ball has to be treated different from stating a particular value. In our setting, players have a strong incentive to lie about a killer ball, since it threatens them progressing in the game. If players nevertheless state to have one, this message is "self-signaling" and not cheap talk: If a player states to have a killer ball it is the truth.

A first result is that players seem to base their voting decision on objective criteria, but switch within the time-dimension of the criteria from round 1 to 2. This switch maybe explained by the different amount of information a player has at a certain time.

Information based separability of players

In the two pre-play rounds the players have different information about the distribution of values. In round 1, the players face a situation in which they base their voting decision on an *ambiguous* distribution of outcomes, i.e., only the values of the revealed front row balls are common knowledge. In round 2 the players have additional information. After the voting decision in round 1, all contestants need to reveal their true values on the back row. The twelve balls carried over from round 1 to round 2 are now common knowledge to

all remaining contestants, and only the two new added cash values are unknown. Hence, a player can be in two different states: First, if both new values are within the revealed balls on the front row, or a particular player has at least one new value on her hidden back row and the other is observable on any other player's front row, she knows the exact distribution of values in play. From an informational point of view a player who knows all ball values in play, makes her voting decision in a situation where only the precise allocation of each value is *uncertain*. Second, if both new values remain unobservable, a player again lacks information, but not as much as in round 1.[23] Thus, using the statements about the hidden ball values, the players are able to infer - up to a certain extent - whether a contestant lies.

We expect that a player, who faces *uncertainty* about the allocation of balls makes her voting decision in consideration of all "true" values, and weights the utility of the outcome by the probability of obtaining it. These players tend to vote by means of the *ex-post* criteria. This effect should be more pronounced compared to the one observe for players in a situation of *ambiguity*. Here, they should vote most frequently by means of the *ex-ante* criteria.

In what follows we refer to Table C.6 in the Supplementary Material that presents proportions of the players' voting decision by means of the three criteria and its time-divisions as well as player's informational background. We restrict the data set to compare the same 573 players in round 1 and 2, of which we identify 50 players to be in an *uncertain* state, and 241 players to be in an *ambiguous* one in round 2. In round 1 all 573 players are in the same *ambiguous* situation.

As suggested, we find that the proportion of *uncertain* players who vote in line with the *ex-post* CKC is significantly higher (p=0.004 and p=0.000) than of *ambiguous* players. Additionally, the spread between proportions of players who consider the *ex-post* or the *stated* prediction within the CKC is much larger for players facing *uncertainty* than

[23]We also analyzed the case when only one of the new ball values is known to a player, termed *partial ambiguity* (see Table C.6 in the Supplementary Material). But for the sake of clarity we limit the discussion to the cases when both new values are either known or not. Naturally the extent of information about a state's probability influences the players' willingness to bet on the state. Camerer and Weber (1992) review the empirical and theoretical literature on ambiguity in decision making.

ambiguity.[24] This indicates that those players are able to infer which player overstates her hidden values on the back row, and in return cast a vote against her. The KC is again special, firstly, due to the self-signaling message when stating a killer ball, and secondly because the number of killer balls seem to be the strongest determinant for player's voting behavior.

Summary With the help of the time-dimension within the three criteria we find a possibility to explain the decision making of more than two thirds of the players. The players' decision may origin in an objective evaluation of their opponents, giving most weight to killer balls. Given the predictions of all criteria, we find contrary voting patterns due to informational differences between both rounds of pre-play, i.e., players switch from *ex-ante* criteria in round 1 to *ex-post* criteria in round 2.[25]

6.2 Subjective criteria and pre-play determinants

Regarding players' choice of their counterpart for the final round, besides the objective criteria, subjective personal valuations as well as observed behavior in the first two rounds are likely to play a decisive role.

First, it is likely that people have different "tastes" for others. We find that women are more likely to vote against men and vice versa. In round 1, males cast a vote against females in 65% and females against males in even 75%. The difference to vote against the opposite sex is highly significant ($p = 0.000$). In round 2 we find that only females are more likely to cast a vote against males (52%, $p = 0,096$), but that males are significantly more likely to vote against their same sex (54%, $p = 0,064$). This finding suggests that in-group biases may be prevalent, especially in round 1.[26] Apart from gender differences, we find that non-whites reach the final round significantly less frequently: 63% of non-whites do not reach the final ($p = 0.000$).

[24]For instance, in round 2 70% of all player facing *uncertainty* vote by means of the *stated* CKC, but only 55% of all players in round 1, compared to 64% of players facing *ambiguity* in round 2. The same holds for the *ex-post* CKC, as well as for males and females in both criteria. Besides we find a different voting pattern for males and females facing *uncertainty*: A much larger proportion of males votes in line with the *ex-ante* CC and CKC than females in round 2, while these are almost equal when players are *ambiguous*, or in round 1.

[25]Over the whole sample, we find only 22 contestants (2.8%) who do not vote in line with neither criterion, and only 65 contestants (8.3%) who never vote in line with an *ex-ante* criterion.

[26]In addition, we find that males receive a vote more frequently than females. In round 1, 55% of males receive a vote ($p = 0.000$), in round 2 53% ($p = 0.012$).

Second, we ask whether the order in which players announce the content of their hidden balls impacts on the players' voting decision. The determination of the order differs between rounds: In round 1, the show host calls on a particular player to start telling what is on her hidden back row balls. Usually this is the player with the weakest front row. Thus, in round 1 the order of statements is exogenously determined by the show host. In round 2 instead it is endogenous. The show host asks the players who wants to open up the round. On the one hand starting to report the content of the hidden back row balls helps to state high values, especially if it is the truth, and thereby turning the focus on the opponents' balls. On the other hand, the player who begins to state her hidden values cannot make her statement dependent on her opponents' statements, e.g., if all other players confess to have a killer ball a player who states her values afterward might be more likely to confess a killer as well. After round 2, we observe that only 26.1% of the players who announce first are effectively voted to leave the game, but the second or third player is voted off in 38.7% and 35.1%, respectively.[27] Therefore, we expect that the order in round 2 has explanatory content in the sense that a player has a lower propensity to receive a vote or is less likely voted off if she makes her statement first. In round 1 we do not expect the order to have additional explanatory power, since the player is either selected randomly or due to her weak front row. The effect of a weak front row on the likelihood of receiving a vote is already captured by the objective criteria.

6.3 Regression analysis

In this section we determine which of the objective as well as subjective determinants, explain the observed voting pattern best and thus affect a player's survival within the two rounds of pre-play. Addressing the objective criteria, which we discussed in Section 6.1, we construct variables for each criterion and their corresponding time-dimension. For each criterion the variables are composed of (i) the sum of the two open balls, (ii) the sum of the stated values on the back row, and (iii) a dummy indicating whether the player made a truthful statement about her hidden back row balls or not.

We expect that the values or number of killers in the open balls on a player's front row affect the decision about whom to cast a vote against in both rounds. But, the statements

[27] The test for the difference between the order of statements is highly significant between the first and second player ($p = 0.000$), and the first and third player ($p = 0.001$). There is no difference between being the second or third announcer ($p = 0.241$).

of the hidden back row balls are only considered in the voting decision in round 2, since in round 2 the players have some information about their possible content. More precisely, the higher the amount of cash in the two open balls, the more likely a player survives the pre-play, but the more killer balls a player has, the more likely she is eliminated from the game.

In addition, we control for whether it is a good strategy of players to tell the truth about the content of their hidden back row balls. Although in round 1 the players have no information about the distribution of balls in play, an honest player might be able to convince the others that her statement is true, because she is e.g., not nervous. Further, we control for the effect of lying by including a dummy "lied in round 1" in the regression for round 2, i.e., the dummy indicates whether a player lied in round 1. A player who lied in round 1 takes the risk of getting a reputation of being untrustworthy, which she may suffer from in round 2. We observe that a substantial amount of players lies in round 1 (53%) as well as in round 2 (45%). The average overstatement is £6,413 in round 1, and £5,069 in round 2.

Concerning the subjective criteria and pre-play determinants (Section 6.2), we control for the effects of players' age, gender, race, and place of residence as well as the effects of lying and the order of statements in the regression analysis. We expect that voting incentives switch between rounds such that players attach more weight to the objective criteria in round 1, and in round 2 shift weight to personal judgment about the opponent's sympathy or trustworthiness with regard to the final.

We estimate a probit on the likelihood that a player is voted off the game (see Table VI) as well as an ordered probit where we look at the determinants impacting on the likelihood a player receives 0,1,2, or 3 votes in round 1 (see Table C.7 in the Supplementary Material), and 0,1, or 2 votes in round 2 (see Table C.8 in the Supplementary Material). In all regressions we simultaneously control for the CC and KC (model (1)), and separately for the CKC (model (2)), which is a mix between the other two. Again, we present marginal effects instead of coefficient estimates.

Table VI: Binary Probits on Voting Behavior in Round 1 and 2

$y_i = 1$ (player i has to leave the show)

Round 1	Marginal Effects			
	Model (1)		**Model (2)**	
Player Characteristics				
Male	0.040	(0.035)	0.040	(0.035)
Age > 40	-0.002	(0.035)	-0.004	(0.034)
White	-0.117*	(0.067)	-0.109*	(0.066)
England	0.023	(0.044)	0.023	(0.043)
Order of Statements	0.024	(0.042)	0.028	(0.041)
Cash-Criterion (CC)				
log(Value Open Balls)	-0.031***	(0.011)		
log(Value Stated Balls)	-0.002	(0.017)		
Truthful Statements	-0.146***	(0.049)		
Killer-Criterion (KC)				
No. Killers Open Balls	0.142***	(0.036)		
No. Killers Stated Balls	0.040	(0.046)		
Truthful Statements	-0.091**	(0.042)		
Cash-Killer-Criterion (CKC)				
log(Whgt. Value Open Balls)			-0.074***	(0.006)
log(Whgt. Value Stated Balls)			-0.003	(0.010)
Truthful Statements			-0.068**	(0.033)
Wald $\mathcal{X}^2$	161.47***		232.20***	
Log-Likelihood	-360.51		-364.04	
Pseudo R^2	0.24		0.23	
N	842		842	
Number of clusters	211		211	

Round 2	Marginal Effects			
	Model (1)		**Model (2)**	
Player Characteristics				
Male	0.027	(0.045)	0.026	(0.044)
Age > 40	0.051	(0.045)	0.055	(0.044)
White	-0.123	(0.084)	-0.110	(0.082)
England	0.050	(0.049)	0.051	(0.049)
Order of Statements	-0.101**	(0.046)	-0.097**	(0.046)
Lied in Round 1	0.057	(0.038)	0.055	(0.038)
Cash-Criterion (CC)				
log(Value Open Balls)	-0.031***	(0.011)		
log(Value Stated Balls)	-0.002	(0.017)		
Truthful Statements	-0.146***	(0.049)		
Killer-Criterion (KC)				
No. Killers Open Balls	0.052	(0.049)		
No. Killers Stated Balls	0.071*	(0.040)		
Truthful Statements	-0.147***	(0.049)		
Cash-Killer-Criterion (CKC)				
log(Wght. Value Open Balls)			-0.034***	(0.008)
log(Wght. Value Stated Balls)			-0.020*	(0.012)
Truthful Statements			-0.191***	(0.041)
Wald $\mathcal{X}^2$	61.55***		69.99***	
Log-Likelihood	-368.25		-367.39	
Pseudo R^2	0.08		0.09	
N	631		631	
Number of Clusters	211		211	

standard errors in parentheses are corrected for episode clusters; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Note: 11 special episodes in which all players have the same sex are excluded.

Results

Looking at model (1) and (2) in Table VI, we find a significant race effect in round 1. Non-whites are voted off more frequently than whites. In round 2, the race effect disappears. But other personal player characteristics, as gender, age or place of residence, have no impact on the likelihood that a player is voted off the game in both rounds. Thus, the gender differences described in Section 6.2 seem to cancel out each other.

Addressing the order in which players make their statements each round, making a statement first has a significant impact on the likelihood to be voted to leave the game in round 2: If a player decides to state the content of her hidden back row balls first, the likelihood to stay in the game increases by roughly 10%. Thus our hypothesis of making the statement first allows players to credibly state "good" balls, is supported. As suggested, there is no significant effect of the order of statements in round 1.

Apart from that, we find no significant effect on the probability of being voted off in round 2 if a player lied in round 1, independent of the lie's content (overstatement or concealing a killer).

The variables representing the objective criteria have a strong explanatory power, with the highest Pseudo R^2 in model (2). As expected, concerning the open ball values on a players' front row, the higher the cash amount the lower a player's likelihood to be voted off. The reverse it true when addressing killer balls. The higher the number of killer balls in the open balls, the more likely a player is voted to leave the game. The statements about the hidden back row balls have no effect in round 1, and in round 2 only a statement about a killer ball is significant. This supports our suggestion that statements about cash values are meaningless, but that statements about killer balls are taken into account in the players' voting decision: Confessing a killer ball increases the likelihood of being voted off the game. Further, the variable indicating whether a player stated the truth, either about her hidden back row cash values or killer balls has a significant effect. In both rounds it is worthwhile to be honest, stating the truth reduces the likelihood to be voted off by roughly 10%. In round 2, the effect is even more pronounced, which might be due to the players' informational advantage about the ball values in play. Although the players have no information about the distribution of the hidden ball values in round

1, a player who states the truth seems to be able to signal credibility.[28]

For a more detailed analysis of the voting process, we draw our attention to the results from the ordered probit estimation, see Table C.7 and Table C.8 in the Supplementary Material. There we can identify the effects on the number of votes a player receives. Overall, the results provided by the probit estimation above are confirmed.[29] In addition, we find that whether a player lied in round 1 matters for the probability of receiving zero or two votes in round 2. A player who lied before receives two votes 6% more likely than a player who made an honest statement.

Finally, we find that all effects are more pronounced in round 2 than in round 1. But, the explanatory power of both models decreases sharply between both rounds. In round 1, the cash- and killer-criterion in model (1) explain a 16% higher mass of the variance than in round 2 (Pseudo R^2=0.24 to Pseudo R^2=0.008), and the cash-killer-criterion in model (2) explains a 14% higher mass of the variance than model (2) in round 2 (R^2=0.23 to R^2=0.09). This decline serves as a further indicator for the switch between players' voting behavior. It is likely that players decide on whom to vote off the game by means of sympathy or trustworthiness. Unfortunately, we are not able to directly control for those effects.

Summary Our conjectures regarding player's strategy are largely confirmed in the data. As we expected, in round 1 neither player takes into account the statements about the hidden back row balls, and only decides on whom to vote by means of the *ex-ante* criteria. However, players discriminate against non-whites. On the contrary, in round 2 taste characteristics become meaningless, but players punish liars, i.e., liars have to bear the danger of being voted off more likely. Additionally, a player who decides to state the content of her hidden back row balls first is more likely to survive round 2. Surprisingly, the trustworthiness of players is highly valued in both rounds. Hence, strategic considerations, such as accumulating a high jackpot and selecting a cooperator as the final

[28]From a psychological perspective, people may recognize a liar with the help of certain body signals, for instance, avoiding eye contact, sweating, or blushing.

[29]The results on the probability of receiving two or three votes in round 1 or 2 are very similar to the ones from the probit model on the player's likelihood to be effectively voted off. This confirms the results, since receiving two or three votes (likely) results in the player's elimination from the game.

opponent, rather than player's characteristics appear to be the primary determinants of voting behavior in round 2.

7 Conclusion

In this paper we analyze cooperative behavior in a prisoner's dilemma game in the presence of high stakes, communication, and two rounds of pre-play, involving two voting decisions. Using data from 222 episodes of the British television game show "Golden Balls", we observe a unilateral cooperation rate of 55% and a mutual cooperation rate of 33%.

Summarizing our main results, we find that stake size, communication as well as pre-play have a significant impact on cooperation. Stake size is inversely related to player's likelihood to cooperate, i.e., the higher the jackpot the more likely players are to defect. Further we can show that player's expectation about the stake size matters: If the jackpot is above (below) the player's expectation, the propensity to cooperate significantly decreases (increases), and mutual cooperation is less (more) successful. With respect to communication, certain words and gestures are more important than others. We test for the effect of handshakes and voluntarily stated mutual promises, and find that players who shake hands are more likely to defect, while handshakes in combination with a promise are likely to result in cooperation and successful coordination. The effects of stake size and communication are robust and independent of player characteristics and pre-play determinants.

The analysis of contestants' behavior in the pre-play shows that players make their voting decision dependent on objective criteria, i.e., their monetary contribution to the stake size, as well as on subjective personal characteristics of their opponents. We show that there is a strong link between the two rounds of pre-play and the players' decision in the prisoner's dilemma. Whether a player lied in the pre-play or contributed more to the stake size has a negative influence on cooperative behavior, whereas whether a player enjoys her opponent's goodwill has a positive one.

We are aware that there are potential drawbacks associated with the use of television game show data. The first addresses anonymity: Players on the show interact face-to-

face and in front of large audience, including their family, friends, and colleagues. This might amplify cooperative behavior, e.g., a selfish-person might choose to cooperate only to avoid embarrassment or punishment by her peer group. The second addresses the selection of players for the show: Contestants are not randomly selected, but have to apply to the show. But, with respect to players exogenous personal characteristics, the sample can be considered, at least to some extent, as representative of the underlying (British) population.

This paper has shown that decisions in the prisoner's dilemma are influenced by the stake size, the player's expectation about stakes, and communication. We are not aware of any other study using handshakes and promises, as well as a player's expectation about stake size to explain cooperative behavior.

References

Ahn, T. K., Elinor Ostrom, David Schmidt, and James Walker, "Trust in Two-Person Games: Game Structure and Linkages," in Elinor Ostrom and James Walker, eds., *Trust and Reciprocity*, Vol. VI, Russell Sage Foundation Series on Trust, 2003, chapter 12, pp. 323–351.

—— , —— , —— , **Robert Shupp, and James Walker**, "Cooperation in PD Games: Fear, Greed, and History of Play," *Public Choice*, 2001, *106* (1-2), 137–155.

Altonji, Joseph and Rebecca Blank, "Race and Gender in the Labor Market," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, 1 ed., Vol. 3, New York: Elsevier, 1999, pp. 3143–3159.

Anonovics, Kate, Peter Arcidiacono, and Randall Walsh, "Games and Discrimination: Lessons From The Weakest Link," *Journal of Human Resources*, 2005, *XL (4)*, 918–947.

Belot, Michèle, V. Bhaskar, and Jeroen van de Ven, "Promises and Cooperation: Evidence from a TV Game Show," *Journal of Economic Behavior and Organization*, 2010, *73* (3), 396–405.

Bohnet, Iris and Bruno S. Frey, "The Sound of Silence in Prisoner's Dilemma and Dictator Games," *Journal of Economic Behavior and Organization*, 1999, *38* (1), 43–57.

Brandts, Jordi and Gary Charness, "Truth or Consequences: An Experiment," *Management Science*, 2003, *49* (1), 116–130.

Camerer, Colin and Martin Weber, "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity," *Journal of Risk and Uncertainty*, 1992, *5* (4), 325–370.

Camerer, Colin F. and Robin M. Hogarth, "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework," *Journal of Risk and Uncertainty*, 1999, *19 (1-3)*, 7–42.

Charness, Gary and Martin Dufwenberg, "Promises and Partnership," *Econometrica*, 2006, *74* (6), 1579–1601.

___ **and** ___ , "Bare Promises: An Experiment," *Economic Letters*, 2010, *107* (2), 281–283.

Crawford, Vincent, "A Survey on Experiments on Communication via Cheap Talk," *Journal of Economic Theory*, 1998, *78*, 286–298.

Ellingsen, Tore and Magnus Johannesson, "Promises, Threats and Fairness," *The Economic Journal*, 2004, *114* (495), 397–420.

Farrell, Joseph and Matthew Rabin, "Cheap Talk," *Journal of Economic Perspectives*, 1996, *10 (3)*, 103–118.

Fischbacher, Urs and Franziska Heusi, "Lies in Disguise. An Experimental Study on Cheating," *Thurgau Institute of Economics TWI Working Paper 40*, 2008.

Gneezy, Uri, "Deception: The Role of Consequences," *American Economic Review*, 2005, *95 (1)*, 384–394.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith, "Preferences, Property Rights, and Anonymity in Bargaining Games," *Games and Economic Behavior*, 1994, *7* (3), 346–380.

Jackson, Matthew, *Social and Economic Networks*, Princeton, Princeton University Press, 2008.

Lazarfeld, Paul and Robert K. Merton, "Friendship as a Social Process: A Substantive and Methodological Analysis," in Morroe Berger, Theodore T. Abel, and Charles C. Page, eds., *Freedom and Control in Modern Society*, New York: Van Nostrand, 1954.

Ledyard, John O., "Public Goods: A Survey of Experimental Research," in John H. Kagel and Alvin E. Roth, eds., *The Handbook of Experimental Economics*, The Handbook of Experimental Economics, Princeton University Press, 1995, pp. 111–194.

Levitt, Steven, "Testing Theories of Discrimination: Evidence From Weakest Link," *Journal of Law and Economics*, 2004, *XLVII*, 431–452.

List, John A., "Friend or Foe? A Natural Experiment of the Prisoner's Dilemma," *Review of Economics and Statistics*, 2006, *88* (3), 463–471.

Manzini, Paola, Abdolkarim Sadrieh, and Nicolaas J. Vriend, "On Smiles, Winks and Handshakes as Coordination Devices," *The Economic Journal*, 2009, *119* (537), 826–854.

Miettinen, Topi and Sigrid Suetens, "Communication and Guilt in a Prisoner's Dilemma," *Journal of Conflict Resolution*, 2008, *52* (6), 945–960.

Oberholzer-Gee, Felix, Joel Waldfogel, and Matthew W. White, "Friend or Foe? Cooperation and Learning in High-Stakes Games," *Review of Economics and Statistics*, February 2010, *92* (1), 179–187.

Ortmann, Andreas and Lisa K. Tichy, "Gender differences in the laboratory: evidence from prisoner's dilemma games," *Journal of Economic Behavior and Organization*, 1999, *39*, 327–339.

Rapoport, Anatol, "Experiments with N-Person Social Traps I," *Journal of Conflict Resolution*, 1988, *32* (3), 457–472.

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

Roth, Alvin E., "Bargaining Experiments," in John H. Kagel and Alvin E. Roth, eds., *The Handbook of Experimental Economics*, Princeton University Press, 1995, chapter 4, pp. 253–348.

Rutström, E. Elisabet and Melonie B. Williams, "Entitlements and Fairness: An Experimental Study of Distributive Preferences," *Journal of Economic Behavior and Organization*, 2000, *43*, 75–89.

Sally, David, "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992," *Rationality and Society*, 1995, *7* (1), 58–92.

Scharlemann, Jörn P. W., Catherine C. Eckel, Alex Kacelnik, and Rick K. Wilson, "The Value of a Smile: Game Theory with a Human Face," *Journal of Economic Psychology*, 2001, *22*, 617–640.

Shafir, Eldar and Amos Tversky, "Thinking Through Uncertainty: Nonconsequential Reasoning and Choice," *Cognitive Psychology*, 1992, *24* (4), 449–474.

van den Assem, Martijn, Dennie van Dolder, and Richard Thaler, "Split or Steal? Cooperative Behavior When the Stakes are Large," *SSRN Working Paper*, 2010.

van den Nouweland, Anne and Marco Slikker, *Social and Economic Networks in Cooperative Game Theory*, Berlin: Springer, 2001.

Vanberg, Christoph, "Why Do People Keep Their Promises? An Experimental Test of Two Explanations," *Econometrica*, 2008, *76* (6), 1467–1480.

Not for Publication: Supplementary Material

A Instructions of the prisoner's dilemma

The show host Jasper Carrott explains the "weak" prisoner's dilemma in every episode with almost the same words:

> "It is time to split or steal. You have got two final golden balls left, you have each got a golden ball with the word *split* written inside, you have got each a golden ball with the word *steal* written inside. I will ask you to make a conscious choice and you will choose either the split or the steal ball, neither of you will know what the other has chosen. If you both choose the split balls, you split today's jackpot of $£J$ and you both go home with $£J/2$. If one of you splits and one of you steals, whoever steals goes home with all the money $£J$, whoever splits goes home with nothing. If you both decide to steal and you are very greedy, you both go home with nothing. Before I ask you to choose, Player A, B just check the two balls to make sure you know which is to split and which is to steal. Do not show to each other. It is very important that you know which is which. [PLAYERS CHECK THE BALLS] Are you happy to know which is split and which is steal? Okay, before I ask you to choose, I will give you some time to talk to each other about what has happened today and how you feel. [PLAYERS DISCUSS] Okay, player A, B choose the split or steal ball now. [PLAYERS CHOOSE BALLS] Hold it up, make sure that when you open it, the other player can see it. Player A, B split or steal? [PLAYERS OPEN BALLS]"

B An alternative empirical strategy

In Section 5, we have shown which factors significantly influence cooperative behavior. In this section we present an alternative approach. Before the show starts, contestants are individually asked to make a private statement about the strategy they plan to play in case they reach the final round (see Section 3.2). These filmed statements are broadcasted

to the television audience, but cannot be observed by the contestants.

We observe an unambiguous strategy-statement by 59% of the final players. Given these individual statements, we can infer whether contestants stick to their announced strategy, i.e., behave consistently or not. If players are either defectors or cooperators, independent of the situation and their opponent, we should neither observe switching strategies nor significant effects of any explanatory variables.

In Table B.1 we depict the average cooperation rate depending on the players' strategy-statement.

Table B.1: Relation between commitment and the cooperation rate

			Decision	
Statement	N	%	Steal (%)	Split (%)
Steal	136	33,3	64.7	35.3
Split	105	25,7	22.9	77.1
Ambiguous	167	43,1	43.1	56.9
Total	408[a]	100	45.1	54.9

[a] Note, that the strategy-statement is not filmed in the first 18 episodes. This reduces the data set to 204 episodes (408 players).

Interestingly, the raw data show that contestants more often state to steal than to split (33,3% versus 25,7%). But we find that players switch their strategy significantly more often (p=0.008) if they initially planned to steal (35.3% split) than if they planned to split (22.9% steal). Addressing the players who do not explicitly state their strategy, we find that 56.9% actually split. The difference between those and the observed average cooperation rate (54,5%) is not significant (p=0.587).

Thus, one could surmise that a substantial fraction of players are of a certain type, either cooperators or defectors. The ones that change their strategy make their strategy dependent on the events in the game as well as on opponent characteristics. We therefore can conclude that at least 17.6% of the players have situation dependent social preferences.

C Tables

Table C.1: Cooperation rates by gender, series, demographics, and stake size

	Split			
	Men (N=207)	Women (N=237)	All (N=444)	
	Row %	Row %	Row %	**N**
Experience				
Unexperienced (Series 1)	44.4	59.1	52.5	80
Experienced (Series 2-4)	55.0	54.9	54.9	364
Age				
≤ 40	44.3	53.1	49.2	260
> 40	64.1	59.8	62.0	184
Race				
Non-White	46.2	42.9	44.4	27
White	53.6	56.5	55.2	417
England				
Not from England	73.3	70.7	71.8	71
From England	49.4	52.3	50.9	371
Jackpot in £				
$[3, 500]$	71.4	75.7	73.6	72
$(500, 2500]$	50.9	55.6	53.0	100
$(2500, 10000]$	56.0	50.0	52.9	102
$(10000, 30000]$	43.2	51.4	48.3	116
$(30000, 100150]$	43.5	51.6	48.1	54
Potential Jackpot in £				
$[5000, 30000]$	46.3	51.7	49.1	114
$(30000, 45000]$	49.3	58.5	53.6	138
$(45000, 75000]$	52.3	58.3	55.8	104
$(75000, 168100]$	72.2	53.8	61.4	88
Expectation				
Jackpot $<$ Expectation	58.3	59.5	58.9	292
Jackpot $\geq$ Expectation	41.3	49.4	46.1	152
Total	53.1	55.7	54.5	444

Table C.2: Mutual decision outcomes by series, demographics and stake size

	(steal, steal)	**(steal, split)**	**(split, split)**	
	Row %	Row %	Row %	**N**
Experience				
Unexperienced (Series 1)	25.0	45.0	30.0	80
Experienced (Series 2-4)	24.2	41.8	34.1	364
Gender				
Male Team	35.3	32.4	32.4	68
Female Team	16.3	55.1	28.6	98
Mixed Team	24.5	40.3	35.3	278
Age				
Old Team (> 40)	11.8	50.0	38.2	68
Young Team (≤ 40)	23.6	48.6	27.8	144
Mixed Team	28.4	36.2	35.3	232
England				
Not English Team	25.0	25.0	50.0	8
English Team	28.6	39.0	32.5	308
Mixed Team	14.3	52.4	33.3	126
Jackpot				
$\leq £500$	8.3	36.1	55.6	72
$> £500$	27.4	43.5	29.0	372
Expectation				
Jackpot $<$ Expectation	23.3	35.6	41.1	292
Jackpot $\geq$ Expectation	26.3	55.3	18.4	152
Total	24.3	42.3	33.3	444
Average Winnings	(0, 0)	(15693, 0)	(4784, 4784)	444

Table C.3: Results from ordered probit on outcomes in the PD (2)

$y = 0; 1; 2$	Marginal Effects[a]					
	Steal/Steal (0)		**Split/Steal (1)**		**Split/Split (2)**	
Player Characteristics						
Team Unexperienced	0.015	(0.066)	0.003	(0.010)	-0.018	(0.075)
Team Male	0.094	(0.088)	0.005	(0.011)	-0.099	(0.082)
Team Female	0.020	(0.064)	0.003	(0.010)	-0.023	(0.074)
Team > 40	-0.065	(0.067)	-0.021	(0.030)	0.086	(0.097)
Team < 40	0.104	(0.065)	0.012	(0.011)	-0.116	(0.069)
Team England	0.109**	(0.051)	0.034	(0.023)	-0.143**	(0.070)
Team Small City	0.069	(0.049)	0.013	(0.012)	-0.082	(0.058)
Index	-0.332**	(0.165)	-0.065	(0.047)	0.397**	(0.197)
Pre-Play						
Team Never Lied	-0.055	(0.059)	-0.016	(0.023)	0.070	(0.081)
Team Lying	0.029	(0.052)	0.005	(0.009)	-0.035	(0.061)
Communication						
Handshakes	0.097	(0.097)	0.019	(0.015)	-0.116	(0.110)
Promise	-0.062	(0.101)	0.125***	(0.021)	0.075	(0.119)
Handshakes*Promise	-0.327***	(0.109)	0.035	(0.085)	0.361***	(0.134)
Stake Size						
Jackpot $<$ Expectation	-0.161***	(0.055)	-0.013	(0.015)	0.174***	(0.056)
Wald $\mathcal{X}^2$	28.98**					
Log-Likelihood	-223.65					
Pseudo R^2	0.06					
N	54		94		74	
Number of clusters	222					

Standard errors in parentheses are corrected for episode clusters; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: Voting decision by means of objective criteria

Criteria to predict a player who should be voted to leave[a]	After round 1[b] in Row %	After round 1[b] out Row %	After round 2 in Row %	After round 2 out Row %
Cash-Criterion (CC)				
ex-ante				
stay	87.2	12.8	72.3	27.7
vote to leave	38.3	61.7	55.4	44.6
stated				
stay	76.7	23.3	70.0	30.0
vote to leave	69.8	30.2	59.9	40.1
ex-post				
stay	83.0	17.0	74.8	25.2
vote to leave	50.9	49.1	50.5	49.5
Killer-Criterion (KC)				
ex-ante				
stay	92.0	8.0	83.6	16.4
vote to leave	23.9	76.1	32.9	67.1
stated				
stay	93.7	6.3	85.1	14.9
vote to leave	18.9	81.1	29.7	70.3
ex-post				
stay	86.6	13.4	80.9	19.1
vote to leave	40.1	59.9	38.3	61.7
Cash-Killer-Criterion (CKC)				
ex-ante				
stay	87.7	12.3	73.2	26.8
vote to leave	36.9	63.1	53.6	46.4
stated				
stay	83.3	16.7	74.1	25.9
vote to leave	50.0	50.0	51.8	48.2
ex-post				
stay	83.2	16.8	74.5	25.5
vote to leave	50.5	49.5	50.9	49.1
N	666	222	444	222

[a] We take into account that the prediction might not be unique per episode, i.e., more than one player might have a prediction to be eliminated.

[b] Each round, 222 contestants are eliminated. In round 1 (2) 55.4% (52.2%) of the eliminated players are men.

Table C.5: Voting decision per player by means of objective criteria

Players voted by means of	**all (%)**	**men (%)**	**women (%)**
Cash-Criterion (CC)			
ex-ante			
Round 1	72.4	74.5	70.5
Round 2	66.5	67.3	65.8
stated			
Round 1	36.0	38.5	33.6
Round 2	64.4	62.6	66.1
ex-post			
Round 1	55.3	55.4	55.3
Round 2	70.2	66.9	73.2
Killer-Criterion (KC)			
ex-ante			
Round 1	82.1	83.0	81.3
Round 2	81.9	84.9	79.1
stated			
Round 1	83.6	85.5	81.8
Round 2	78.9	80.2	77.6
ex-post			
Round 1	70.9	72.7	69.2
Round 2	76.3	77.0	75.6
Cash-Killer-Criterion (CKC)			
ex-ante			
Round 1	71.0	74.1	68.1
Round 2	65.8	67.3	64.4
stated			
Round 1	55.0	58.3	51.9
Round 2	63.5	61.9	65.1
ex-post			
Round 1	60.4	61.5	59.3
Round 2	66.8	65.5	68.1
N[a]	573	278 (48.5%)	295 (51.5%)

[a] Note: For purpose of comparability, we restrain the sample to 573 observations including only those players, who are not being eliminated in round 1. Further we consider only those decisions for which we can trace back for whom a player voted, i.e., we exclude episodes with a voting of 2:1:1:0 in round 1 and 1:1:1 in round 2.

Table C.6: Voting decision per player under risk or ambiguity (by means of objective criteria)

	Round 2			Round 1
Players vote by means of	**Uncertainty** (%)	**Partial Ambiguity** (%)	**Ambiguity** (%)	**Ambiguity** (%)
Cash-Criterion (CC)				
ex-ante				
all	64.0	66.3	67.2	72.4
men	82.4	64.9	67.7	74.5
women	54.5	67.6	66.7	70.5
stated				
all	58.0	67.7	61.8	36.0
men	52.9	65.7	60.6	38.5
women	60.6	69.6	63.2	33.6
ex-post				
all	74.0	70.2	69.3	55.3
men	70.6	66.4	66.9	55.4
women	75.8	73.6	71.9	55.3
Killer-Criterion (KC)				
ex-ante				
all	84.3	79.8	84.0	82.1
men	82.4	81.3	89.1	83.0
women	85.3	78.4	78.3	81.3
stated				
all	84.0	75.9	81.3	83.6
men	94.1	76.9	81.9	85.5
women	78.8	75.0	80.7	81.8
ex-post				
all	78.0	75.2	77.2	70.9
men	88.2	76.9	75.6	72.7
women	72.7	73.6	78.9	69.2
Cash-Killer-Criterion (CKC)				
ex-ante				
all	60.0	64.9	68.0	71.0
men	76.5	62.7	70.9	74.1
women	51.5	66.9	64.9	68.1
stated				
all	70.0	61.7	64.3	55.0
men	70.6	59.0	63.8	58.3
women	69.7	64.2	64.9	51.9
ex-post				
all	76.0	66.3	65.6	60.4
men	88.2	64.9	63.0	61.5
women	69.7	67.6	68.4	59.3
N[a]	50	282	241	573

[a] Note: The sample is restricted to the same 573 players in round 1 and 2.

Table C.7: Ordered probit regression results on the number of votes (round 1)

$y_{ie} = 0; 1; 2; 3$ (Number of votes a player receives in round 1 per episode)

		Marginal Effects			
No. of Votes	**Variables**	Model (1)		Model (2)	
0	**Player Characteristics**				
	Male	-0.051	(0.035)	-0.051	(0.035)
	Age > 40	0.012	(0.034)	0.015	(0.034)
	White	0.135**	(0.058)	0.126**	(0.059)
	England	-0.009	(0.045)	-0.002	(0.046)
	Order of Statements	0.021	(0.042)	0.016	(0.042)
	Cash-Criterion (CC)				
	log(Value Open Balls)	0.097***	(0.010)		
	log(Value Stated Balls)	-0.025	(0.019)		
	Truthful Statements	0.111***	(0.036)		
	Killer-Criterion (KC)				
	No. Killers Open Balls	-0.159***	(0.040)		
	No. Killers Stated Balls	-0.007	(0.050)		
	Truthful Statements	0.148***	(0.037)		
	Cash-Killer-Criterion (CKC)				
	log(Wght. Value Open Balls)			0.110***	(0.007)
	log(Wght. Value Stated Balls)			-0.011	(0.012)
	Truthful Statements			0.138***	(0.036)
1	**Player Characteristics**				
	Male	0.007	(0.005)	0.008	(0.006)
	Age > 40	-0.002	(0.005)	-0.002	(0.005)
	White	-0.005	(0.005)	-0.008**	(0.004)
	England	0.001	(0.006)	0.000	(0.007)
	Order of Statements	-0.003	(0.006)	-0.003	(0.007)
	Cash-Criterion (CC)				
	log(Value Open Balls)	-0.013***	(0.002)		
	logsumvalueclaims1	0.003	(0.003)		
	Truthful Statements	-0.011***	(0.004)		
	Killer-Criterion (KC)				
	No. Killers Open Balls	0.021***	(0.007)		
	No. Killers Stated Balls	0.001	(0.007)		
	Truthful Statements	-0.012***	(0.003)		
	Cash-Killer-Criterion (CKC)				
	log(Wght. Value Open Balls)			-0.017***	(0.003)
	log(Wght. Value Stated Balls)			0.002	(0.002)
	Truthful Statements			-0.022***	(0.007)
2	**Player Characteristics**				
	Male	0.025	(0.017)	0.024	(0.017)
	Age > 40	-0.006	(0.016)	-0.007	(0.016)
	White	-0.064**	(0.027)	-0.060**	(0.028)
	England	0.004	(0.022)	0.001	(0.022)
	Order of Statements	-0.010	(0.020)	-0.008	(0.020)
	Cash-Criterion (CC)				
	log(Value Open Balls)	-0.047***	(0.006)		
	logsumvalueclaims1	0.012	(0.009)		
	Truthful Statements	-0.053***	(0.018)		
	Killer-Criterion (KC)				
	No. Killers Open Balls	0.076***	(0.020)		
	No. Killers Stated Balls	0.003	(0.024)		
	Truthful Statements	-0.071***	(0.019)		
	Cash-Killer-Criterion (CKC)				
	log(Wght. Value Open Balls)			-0.052***	(0.005)
	log(Wght. Value Stated Balls)			0.005	(0.006)
	Truthful Statements			-0.066***	(0.018)
3	**Player Characteristics**				
	Male	0.020	(0.014)	0.019	(0.013)
	Age > 40	-0.005	(0.013)	-0.005	(0.013)
	White	-0.066*	(0.036)	-0.058*	(0.034)
	England	0.003	(0.017)	0.001	(0.017)
	Order of Statements	-0.008	(0.016)	-0.006	(0.015)
	Cash-Criterion (CC)				
	log(Value Open Balls)	-0.037***	(0.005)		
	logsumvalueclaims1	0.010	(0.007)		
	Truthful Statements	-0.046***	(0.016)		
	Killer-Criterion (KC)				
	No. Killers Open Balls	0.061***	(0.016)		
	No. Killers Stated Balls	0.003	(0.019)		
	Truthful Statements	-0.066***	(0.019)		
	Cash-Killer-Criterion (CKC)				
	log(Wght. Value Open Balls)			-0.041***	(0.004)
	log(Wght. Value Stated Balls)			0.004	(0.004)
	Truthful Statements			-0.051***	(0.013)
	Wald $\mathcal{X}^2$	284.59***		390.05***	
	Log-Likelihood	-875.00		-870.64	
	Pseudo R^2	0.18		0.19	
	N	842		842	
	Number of clusters	211		211	

standard errors in parentheses are corrected for episode clusters; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: 11 special episodes are excluded (all players have the same sex)

Table C.8: Ordered probit regression results on the number of votes (round 2)

$y_{ie} = 0; 1; 2$ (Number of votes a player receives in round 1 per episode)

		Marginal Effects			
No. of Votes	**Variables**	Model (1)		Model (2)	
0	**Player Characteristics**				
	Male	-0.002	(0.037)	-0.002	(0.037)
	Age > 40	-0.030	(0.035)	-0.033	(0.035)
	White	0.039	(0.074)	0.031	(0.076)
	England	-0.045	(0.041)	-0.047	(0.042)
	Order of Statements	0.175***	(0.046)	0.170***	(0.045)
	Lied in Round 1	-0.065**	(0.031)	-0.062**	(0.031)
	Cash-Criterion (CC)				
	log(Value Open Balls)	0.036***	(0.009)		
	log(Value Stated Balls)	0.005	(0.015)		
	Truthful Statements	0.177***	(0.047)		
	Killer-Criterion (KC)				
	No. Killers Open Balls	-0.047	(0.040)		
	No. Killers Stated Balls	-0.044	(0.033)		
	Truthful Statements	0.014	(0.054)		
	Cash-Killer-Criterion (CKC)				
	log(Wght. Value Open Balls)			0.040***	(0.007)
	log(Wght. Value Stated Balls)			0.017*	(0.010)
	Truthful Statements			0.182***	(0.034)
1	**Player Characteristics**				
	Male	0.000	(0.000)	0.000	(0.000)
	Age > 40	-0.000	(0.001)	-0.000	(0.001)
	White	0.002	(0.008)	0.001	(0.006)
	England	0.002	(0.004)	0.002	(0.004)
	Order of Statements	-0.014*	(0.007)	-0.013*	(0.007)
	Lied in Round 1	-0.000	(0.001)	-0.000	(0.001)
	Cash-Criterion (CC)				
	log(Value Open Balls)	-0.000	(0.001)		
	log(Value Stated Balls)	-0.000	(0.000)		
	Truthful Statements	0.004	(0.003)		
	Killer-Criterion (KC)				
	No. Killers Open Balls	0.000	(0.001)		
	No. Killers Stated Balls	0.000	(0.001)		
	Truthful Statements	0.000	(0.001)		
	Cash-Killer-Criterion (CKC)				
	log(Wght. Value Open Balls)			-0.000	(0.001)
	log(Wght. Value Stated Balls)			-0.000	(0.000)
	Truthful Statements			0.004	(0.003)
2	**Player Characteristics**				
	Male	0.002	(0.037)	0.002	(0.037)
	Age > 40	0.030	(0.036)	0.034	(0.036)
	White	-0.041	(0.082)	-0.032	(0.082)
	England	0.043	(0.037)	0.044	(0.038)
	Order of Statements	-0.161***	(0.039)	-0.157***	(0.039)
	Lied in Round 1	0.066**	(0.031)	0.062**	(0.031)
	Cash-Criterion (CC)				
	log(Value Open Balls)	-0.036***	(0.010)		
	log(Value Stated Balls)	-0.005	(0.015)		
	Truthful Statements	-0.181***	(0.049)		
	Killer-Criterion (KC)				
	No. Killers Open Balls	0.047	(0.040)		
	No. Killers Stated Balls	0.044	(0.033)		
	Truthful Statements	-0.014	(0.055)		
	Cash-Killer-Criterion (CKC)				
	log(Wght. Value Open Balls)			-0.040***	(0.007)
	log(Wght. Value Stated Balls)			-0.017*	(0.010)
	Truthful Statements			-0.186***	(0.036)
	Wald $\mathcal{X}^2$	102.47***		110.81***	
	Log-Likelihood	-634.79		-633.74	
	Pseudo R^2	0.08		0.09	
	N	631		631	
	Number of clusters	211		211	

standard errors in parentheses are corrected for episode clusters; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: 11 special episodes are excluded (all players have the same sex)