Split or Steal?
Cooperative Behavior When the Stakes Are Large

Martijn J. van den Assem
Erasmus School of Economics, Erasmus University Rotterdam

Dennie van Dolder
Erasmus School of Economics/Tinbergen Institute, Erasmus University Rotterdam

Richard H. Thaler
Booth School of Business, University of Chicago

Abstract: We examine cooperative behavior when large sums of money are at stake, using data from the TV game show "Golden Balls". At the end of each episode, contestants play a variant on the classic Prisoner's Dilemma for large and widely ranging stakes averaging over $20,000. Cooperation is surprisingly high for amounts that would normally be considered consequential but look tiny in their current context, what we call a "big peanuts" phenomenon. Utilizing the prior interaction among contestants, we find evidence that people have reciprocal preferences. Surprisingly, there is little support for conditional cooperation in our sample. That is, players do not seem to be more likely to cooperate if their opponent might be expected to cooperate. Further, we replicate earlier findings that males are less cooperative than females, but this gender effect reverses for older contestants because men become increasingly cooperative as their age increases.

Cooperation is vital for the functioning of society, and the organizations and communities that form its fabric. Not surprisingly, cooperative behavior is the focus of many studies across a wide range of scientific disciplines, including psychology (Dawes, 1980; Dawes and Messick, 2000), sociology (Marwell and Ames, 1979, 1980; Raub and Snijders, 1997), economics (Ledyard, 1995; Fehr and Gächter, 2000a; Fischbacher and Gächter, 2010), political science (Ostrom, Walker and Gardner, 1992) and biology (Gardner and West, 2004; West, Griffin and Gardner, 2007). The key question in this literature is why humans cooperate even in situations in which doing so is not in line with their material self-interest.

While cooperation is ubiquitous in social life and an important topic for all kinds of economic interaction, field data rarely allow for a clean discrimination among competing theories. Because carefully designed laboratory experiments do allow for such rigorous comparisons, laboratory experiments have provided numerous important insights into cooperative behavior, and the resulting rich literature forms the basis of most of our knowledge on human cooperation. Still, laboratory settings inevitably have limitations that some argue may hinder the generalization of findings to situations beyond the context of the lab (Levitt and List, 2007, 2008). Subjects are often volunteering students who thus constitute a non-random sample of the population at large. Also, they generally have less familiarity with decision tasks in the laboratory than with those in everyday life, no opportunity to seek advice from friends or experts, and they know that their behavior is examined in detail.

From an economic perspective, another obvious drawback to lab studies is that the financial stakes employed tend to be relatively small. Even those experiments that utilize relatively large payoffs do not involve amounts in excess of a few hundred dollars (*e.g.,* Hoffman, McCabe and Smith, 1996a; List and Cherry, 2000; Carpenter, Verhoogen and Burks, 2005), giving rise to the question to what extent findings will generalize to situations of significant economic importance. One solution is to perform experiments in low-income countries, where small nominal amounts carry a larger value. In the domain of social interaction, such experiments are, for example, employed by Slonim and Roth (1998), Cameron (1999), Fehr, Fischbacher and Tougareva (2002), Munier and Zaharia (2002), Johansson-Stenman, Mahmud and Martinsson (2005), and Kocher, Martinsson and Visser (2008). While this might appear an ideal approach, it has its own drawbacks. Culture, for example, has been shown to play an important role in social interaction (Henrich *et al.*, 2001, 2004; Herrmann, Thöni and Gächter, 2008), making it difficult to generalize findings from low-income countries.[1] And while

---

[1] Interestingly, though not generally acknowledged, this argument at the same time questions the universal applicability of the many findings from higher-income countries, including ours. We refer to Henrich, Heine and Norenzayan (2010) for a discussion on this issue.

the stakes in these experiments are larger than commonly employed, they still rarely exceed a few months' wages.

In the current paper, we study cooperative behavior using another source of data, namely the behavior of contestants on the British TV game show "Golden Balls". Although the game show setting is an unusual environment, it has the benefit of employing large and varying stakes. Furthermore, game shows are markedly different from laboratory experiments in terms of participant selection, scrutiny, and familiarity of participants with the decision task. Combined with the strict and well-defined rules, game shows can therefore provide unique opportunities to investigate the robustness of existing laboratory findings.

Because game shows are often competitive in nature and ask contestants to make risky or strategic choices, is it not surprising that they have mostly been used to study decision making under risk (*e.g.*, Gertner, 1993; Metrick, 1995; Beetsma and Schotman, 2001; Post *et al.*, 2008) or strategic reasoning (*e.g.*, Bennett and Hickman, 1993; Berk, Hughson and Vandezande, 1996; Tenorio and Cason, 2002). More recently, however, game shows have also been used to study social interaction, in particular discrimination (Levitt, 2004; Antonovics, Arcidiacono and Walsh, 2005) and cooperative behavior (List, 2004, 2006; Belot, Bhaskar and van de Ven, 2010a; Oberholzer-Gee, Waldfogel and White, 2010). The current paper is in the latter category.

In the final stage of "Golden Balls", contestants make a choice on whether or not to cooperate in a variant of the famous Prisoner's Dilemma. In particular, the two final contestants independently have to decide whether they want to "split" or "steal" the jackpot. If both contestants choose "split", they share the jackpot equally. If one chooses "split" and the other chooses "steal", the one who steals takes the jackpot and the other gets nothing. If they both "steal", both go home empty-handed. On average, the jackpot is over $20,000. The variation is large: from a few dollars to about $175,000.

If we assume that each player only cares about maximizing her immediate financial payoff, the choice problem in "Golden Balls" can be labeled as a "weak" form of the Prisoner's Dilemma.[2] Where in the classic form of the Prisoner's Dilemma defecting strictly dominates cooperating, here defecting only weakly dominates cooperating: choosing "steal" always does at least as well, and sometimes better than choosing "split". Of course, contestants may consider other factors aside from their own monetary payoff when deciding which choice to make. Much experimental research suggests that people have social preferences in the sense that the payoffs to others enter their utility functions. For discussions, see, for example, Fehr and Gächter (1998, 2000b), Fehr and Schmidt (1999), Bolton

---

[2] Rapoport (1988) introduced this terminology. For the sake of brevity, we will simply use the term Prisoner's Dilemma to refer to the game studied here.

and Ockenfels (2000), Charness and Rabin (2002), Camerer (2003), Fehr and Gintis (2007), and Cooper and Kagel (2009).

The fact that the show is aired on TV of course creates another set of rather special circumstances that could affect our results, although there is little existing theory to suggest what the effect of a large TV audience would be. One might argue that players would not want to be seen as a "jerk" on national television and so would be more likely to cooperate, but one can also argue that a player would not want to been seen as a "sucker" (or someone who cannot detect the weakly dominant solution to a simple game) in public.[3] The public nature of the choice could also magnify subtle features created by the fact that the game is a "weak" form of the Prisoner's Dilemma. Specifically, if a player thinks that the other player will steal, she might decide to split on the grounds that it costs her nothing to appear "nice" on TV. These complications do not render our results uninteresting, but do need to be incorporated in any attempt to evaluate how our results should be interpreted in the context of existing theories and experimental findings on cooperation.

Although "Golden Balls" is unique in its format, the show shares the Prisoner's Dilemma element with a few game shows from other countries including "Friend or Foe" (US, aired in 2002-2003) and "*Deelt ie 't of deelt ie 't niet*?" (in English: "Will he share or not?"; Netherlands, 2002). These two shows have been studied in four different papers. List (2004, 2006) and Oberholzer-Gee, Waldfogel and White (2010) analyze data from "Friend or Foe", and Belot, Bhaskar and van de Ven (2010a) use the Dutch show. List focuses on the effects of demographic variables such as gender, race and age. Studying the same game show, Oberholzer-Gee, Waldfogel and White compare the behavior in the first season of the show with later seasons in which the contestants have had a chance to observe prior episodes. Finally, Belot, Bhaskar and van de Ven find that making a promise to cooperate prior to the decision is positively related to cooperation if the promise was voluntary, but not if it has been elicited by the host.

In this paper we replicate many of the earlier investigations, but also undertake several novel analyses that are possible due to some unique features in the format of "Golden Balls". The way the stakes are determined and the very wide range they cover provide the basis for new insights into the effect of stakes and context. The dynamic setting of the show enables us to look at reciprocity in cooperative behavior, and also at the effect of earlier deceitful behavior.

---

[3] Most studies related to the issue of observability indicate that people display more other-regarding behavior when they are or feel more subject to public scrutiny (see, for example, Hoffman, McCabe and Smith, 1996b; Rege and Telle, 2004; Haley and Fessler, 2005), but there is also contradictory evidence (Dufwenberg and Muren, 2006). Kerr (1999) suggests that the effect depends on conditions related to social expectations and sanctions.

In our sample, individual players on average cooperate 53 percent of the time. Although this rate is similar to earlier findings from the experimental literature (Dawes and Thaler, 1988; Sally, 1995), direct comparisons are hampered by systematic differences in the stakes, the visibility of decisions, characteristics of the subjects, and preceding opportunities for communication or other social interaction.

We find only limited support for the notion that cooperation will decrease if the stakes get significant. The cooperation rate is unusually high when the stakes lie in the low range of our sample, perhaps because contestants think that for so little money (relatively speaking) they might as well cooperate in public. Cooperation does decline with the stakes for stakes below the median, but plateaus at around 45 percent for medium to large amounts.

The high cooperation rate for relatively small stakes suggests that context can convert money amounts that would normally be considered consequential or "big" into amounts that are perceived to be small, just "peanuts". This idea is supported by our finding that cooperation is not only based on the actual stakes but also on what the jackpot potentially could have been. This effect is especially pronounced for those who appeared in the earlier episodes of the show and had no or little opportunity to watch the show on TV and learn what sizes are large or small in the context of this game.

A special property of "Golden Balls" is the interaction that occurs among contestants prior to the final. Utilizing the dynamic setting, we find evidence that contestants show some tendency toward reciprocity. Among contestants whose final opponent has attempted to vote them off the show, the propensity to cooperate is significantly lower. Contestants do not appear to reciprocate against opponents who have lied earlier in the game. Lying seems to be accepted here, similar to bluffing in poker. A possible reason for this is that, in contrast to a vote cast against someone, lying is a defensive act that is not aimed at anyone in particular.

Surprisingly, we find little evidence that contestants' propensity to cooperate depends positively on the likelihood that their opponent will cooperate. While an opponent's promise to cooperate is a strong predictor of her actual choice, contestants appear not to be more likely to cooperate if their opponent might be expected to cooperate. Our final main result is that young males cooperate less than young females. This difference decreases and even reverses as age increases and men become increasingly cooperative.

The paper proceeds as follows. Section I describes the game show in more detail, discusses our data and presents descriptive statistics. Sections II – VI cover the various possible factors behind cooperative behavior included in our analysis. Each of these sections provides related literature and

other background, explains the variables that we use, and discusses the results of our Probit regression analyses. Section VII concludes.

I.   Game Show and Data

Description of "Golden Balls"

The TV game show "Golden Balls" was developed by the Dutch production company Endemol. It debuted on the ITV network in the United Kingdom in June 2007 and ran until December 2009. Each episode consists of four rounds and starts with four contestants, usually two men and two women.

In Round 1, twelve golden balls are randomly drawn from the "Golden Bank", a lottery machine containing one hundred "golden" balls. Each of these balls has a hidden cash amount inside, ranging from a minimum of £10 to a maximum of £75,000.[4] Contestants know that this is the range for the amounts in the balls, but they do not know the precise distribution (though this becomes clearer over time as the show is aired). At a later stage of the game, a subset of the cash balls drawn will contribute to the final jackpot. Also, four balls hiding the word "killer" inside are mixed with the twelve cash balls. Killer balls are undesirable in a way we will explain below. From the sixteen balls, each contestant receives four balls at random. For each contestant, two are placed on the front row with their contents - either a cash amount or the word "killer" - openly displayed; the other two are placed on the back row and their contents are known by the particular contestant alone. (Poker players can think of this as two "up" cards and two "down" cards.) The contestants now have to decide by vote which player will be kicked off the show. Because the balls of voted-off contestants are removed from the game and the remaining contestants' balls matter for the ultimate jackpot, there is a strong incentive to retain players with high value balls and kick off players with low value balls or killer balls.

Before the voting starts, each contestant publicly announces the contents of the balls on her back row (knowing that these values will subsequently be revealed, but only after the vote). Then, the four contestants together have an open discussion in which they can voice their evaluation of other players' statements and their opinion of who should be voted off. Each player then anonymously casts a vote against one specific opponent. After the votes are tallied, the player who received the

---

[4] Values in British pounds can be translated into US dollars using a rate of $1,75 per pound, an approximate average of the exchange rate during the period in which the show ran.

most votes leaves the game.[5] Lastly, all the players reveal the values of their back row balls, and differences between the actual values and the previous claims are noted.

In Round 2, two additional cash balls from the lottery machine and one extra killer ball are added to the twelve remaining balls from Round 1. The fifteen balls are then randomly allocated to the three contestants. Each of them receives two balls on her front row and three on her (hidden) back row. Similar to Round 1, contestants make (cheap talk) statements on the balls on their back row, a round of banter follows, votes are cast anonymously and tallied, the player who receives the most votes leaves the game, and all hidden ball values are revealed.[6] Two players and their ten balls proceed to Round 3.

Round 3 determines the size of the final jackpot. First, one additional killer ball is mixed with the ten balls from Round 2. Then five of the balls are selected sequentially at random. If a ball selected is a cash ball, its face value is added to the jackpot. If a killer ball is drawn, the current cumulative jackpot is divided by ten. For example, if the first two balls were £50,000 and £1,000 and the third is a killer ball, the level of the jackpot is reduced from £51,000 to £5,100. A killer ball does not affect the jackpot contribution of cash balls drawn thereafter. If the fourth and fifth ball in our example were another killer ball and £25,000, respectively, then the actual jackpot would be £25,510. Note that this round is a completely stochastic process, and that contestants have full information on the balls that are in play. Before the five balls are drawn, special attention is always paid to the highest possible jackpot (that is, the sum of the five largest cash values). This value, and the number of killer balls, are explicitly stressed by the game show host.

After Round 3 determined the jackpot, in the fourth and final round the contestants play a variant of the Prisoner's Dilemma. Each contestant receives two golden balls. One of the balls says "split" and the other says "steal" on the inside. The contestants then simultaneously have to decide which ball they want to play. If both choose "split", they share the jackpot equally. If one chooses "split" and the other chooses "steal", the contestant who steals takes the whole jackpot and the other gets nothing. If they both choose "steal", both go home empty-handed. Before each contestant makes her actual decision, a brief time period is reserved for a discussion between the players in which they can make non-binding promises, ask about intentions, or attempt to get assurances of cooperative behavior. This is the final round of cheap talk. Importantly, the contestants have not met before the

---

[5] If two contestants receive two votes each, their opponents openly discuss who they want to keep in the show. If they cannot decide, a decision is made at random. If all four contestants receive one vote each, contestants openly attempt to form a coalition against one specific contestant. Again, if they cannot decide, a decision is made at random.

[6] The procedure in the case of a tie is similar to that in Round 1. Tie-breaking occurs by discussion or by random draw if no agreement is reached.

game starts, and have no opportunity before or during the show to make any kind of collusive agreement.

A relevant question is how the contestants are selected. A spokeswoman of Endemol informed us that anyone can apply to be on "Golden Balls" by submitting a detailed application form. Shortlisted contestants are then invited to an audition in order to determine their skills at playing the game, their character and their suitability to appear on a TV show such as Golden Balls. Producers watch tapings of these auditions and put together shows such that, according to the producers, "a good mix of characters" is represented on each show. Thus while the contestants are not a random sample of society, the selection process does not seem to create any obvious confounds with the analyses we conduct here.

Data and descriptive game characteristics

We examine the "split" and "steal" decisions of 574 final contestants appearing in 287 episodes aired between June 2007 (when the show was introduced) and December 2009. During this period, 288 episodes were aired, and, at the time of writing, no further episodes were aired thereafter. Recordings from the show and additional information such as recording and airing dates were kindly provided by Endemol's local production company Endemol UK. The one missing episode could not be supplied because it was not present in their archives.[7]

For each episode we collected data on the relevant observables in the show, such as the hidden and visible ball values, statements made by contestants, the votes, the jackpot size, and the decision to "split" or "steal" at the end. Some variables were estimated based on contestants' physical appearance and on information provided in the introductory talk and other conversations during the show.

Table 1 displays some descriptive characteristics of the game. Cash balls drawn from the lottery machine during the first two rounds have a mean value of £5,654 and a median of £1,500. Clearly, the distribution is positively skewed. The mean value of the cash balls taken to Round 3 is £6,775, which is statistically significantly greater than the average value of all cash balls in the show, implying that the contestants are successful in using their votes to keep high-value balls in play and eliminate small ones. The average number of killer balls in the game at the start of Round 3 is 3.14, significantly less than the 3.67 we would statistically expect if voting was random. Contestants thus also seem

---

[7] Sixteen episodes in our data set feature returning contestants. In twelve of these, players who previously had lost in the final (opponents "stole" while they themselves chose to "split") get a second chance. In four, unlucky players who had been voted off in the first game round receive a second chance. We do not find that returning contestants behave differently, and, unless stated otherwise hereafter, excluding them from our analyses does not materially affect our results.

successful in eliminating killer balls from the game. Unreported analyses of contestants' voting behavior clearly show that contestants indeed try to vote off the opponents that have the worst set of balls on their front row.

At the start of Round 3, special attention is paid to the highest possible jackpot. Dependent on the cash balls and killer balls taken to this stage, this maximum varies between £5,000 and £168,100, with a mean of £51,493 and a median of £41,150. The actual jackpot for which contestants play the Prisoner's Dilemma game is generally considerably smaller due to the skewed distribution of cash ball values and the effect of killer balls, but still has a mean size of £13,416 and a median of £4,300. These amounts are many times the amounts typically used in laboratory experiments, and also large sums relative to the median gross weekly earnings of £397 in the UK in April 2009 (Source: Statistical Bulletin Office for National Statistics, 12 November 2009). About half of the time, the jackpot in our show exceeds three months of median UK earnings, and 21 percent of the contestants decide over a jackpot that is even larger than a median annual salary (the third quartile in our sample is at £18,350). The stakes are also large compared to the two other game shows employed in earlier analyses of cooperative behavior: in "Friend or Foe", the average is about $3,500 (List, 2004, 2006; Oberholzer-Gee, Waldfogel and White, 2010), and for the Dutch show, Belot, Bhaskar and van de Ven (2010a) report a median of €1,683. The wide range of the jackpot in our sample is caused by its random construction, by the highly skewed distribution of cash ball values, and by the effect of killer balls. The largest jackpot was played for in an exhilarating episode from March 2008; trainee accountant Sarah stole the entire jackpot of £100,150 from collection agent Stephen.[8]

For the jackpot to be awarded, at least one player needs to cooperate. We find that 52.8 percent of the contestants decide to "split". While this might seem high, the rate is actually remarkably similar to earlier experimental evidence (see, for example, Sally, 1995). In our sample, both players split the jackpot 31 percent of the time, one splits while the other one steals occurs in 44 percent of the shows, and in the remaining 25 percent of the shows both players steal. The efficiency rate in terms of the percentage of jackpots that is actually awarded thus amounts to 75 percent. The efficiency rate obtained by dividing the sum of earnings across all episodes by the sum of all jackpots is slightly lower at 72 percent. (The difference in efficiency results from contestants' lower propensity to cooperate when the stakes are larger; we explore this effect in detail later.) These simple statistics are a first indication that contestants do not condition their behavior on that of their opponent. Given the average cooperation rate, we would expect to observe (split, split) in 28 percent of the

---

[8] A video clip of this episode is widely available on the Internet, for example through YouTube.

Table 1: Selected Game Show Characteristics

The table shows selected characteristics for the British TV game show "Golden Balls", extracted from our sample of 287 episodes. *Cash ball (overall)* is the monetary value of a cash ball drawn from the lottery machine in the first or second round of the game. *Cash ball (Round 3)* is the monetary value of a cash ball that is in play at the start of the third round. *No. of killer balls (Round 3)* describes the number of killer balls that are in play at the start of the third round. *Potential jackpot (Round 3)* is the jackpot size that is attained during the third round if the best-case scenario would occur. *Jackpot* describes the actual size of the jackpot. *Decision* is a contestant's decision in the Prisoner's Dilemma at the end of the show, with a value of 1 for "split" and 0 for "steal". *Prize won (if non-zero)* records the take home prize for a contestant who made it to the final (if she did not leave empty-handed). All monetary values are in UK Pounds (£1.00 ≈ $1.75).

	N	Mean	Stdev	Min	Median	Max
Cash ball (overall)	4,018	5,653.92	10,478.49	10.00	1,500.00	75,000.00
Cash ball (Round 3)	2,257	6,775.15	12,204.39	10.00	1,600.00	75,000.00
No. of killer balls (Round 3)	287	3.14	0.90	1.00	3.00	6.00
Potential jackpot (Round 3)	287	51,493.08	31,386.69	5,000.00	41,150.00	168,100.00
Jackpot	287	13,416.09	19,182.98	2.85	4,300.00	100,150.00
Decision (split=1)	574	0.53	0.50	0.00	1.00	1.00
Prize won	574	4,850.55	11,821.06	0.00	38.75	100,150.00
Prize won if non-zero	303	9,188.82	15,004.52	1.83	2,175.00	100,150.00

cases and (steal, steal) 22 percent of the time if the individual decisions in our sample were randomly matched. Although the actual percentages are higher (31 and 25), the differences are relatively small considering that each pair of contestants operates under highly similar conditions (same jackpot, same potential jackpot, and many shared unobserved conditions).

On average, a finalist goes home with £4,851, but the median prize is only £39 because 47 percent of the contestants get nothing. The 303 contestants who end up with a non-zero prize take home £9,189 on average, with a median of £2,175. It is worth noting that would we have run this show as an experiment ourselves, the total costs in subject payoffs alone would have been £2.8 million.

Modeling the decision to "split" or "steal"

In the following sections, we will analyze the decisions to "split" or "steal" the jackpot using a binary Probit model. We assume that when people enter the final round they have a latent propensity to "split" $y*$, where $y* \in (-\infty, \infty)$. Furthermore, we assume that this latent propensity is a linear function of personal demographic characteristics x and context characteristics z, in the form $y* = x'\beta + z'\gamma + u$, where β and γ are parameter vectors and u represents an unobserved stochastic component. We do not observe the latent propensity to "split" directly, but only the actual decision y, where $y = 1$ if a contestant chooses "split" and $y = 0$ if a contestant chooses "steal". We impose the observation criterion $y = \mathbf{1}(y* > 0)$, where $\mathbf{1}(.)$ is the indicator function taking the value of 1 if $y* > 0$ and 0 otherwise. Assuming that the stochastic component has a standard normal distribution, or $u \sim N(0,1)$, leads to the binary Probit model of the form $\Pr(y = 1 | x, z) = \Phi(x'\beta + z'\gamma)$,

where $\Phi(.)$ is the standard normal cumulative distribution function. Using this framework we estimate the parameter vectors β and γ using maximum likelihood estimation. We allow for the possibility that the decisions of contestants within the same episode are correlated by performing a clustering correction on the standard errors (see, for example, Wooldridge, 2003).

Because coefficients in a Probit model do not have an immediate intuitive economic meaning due to the inherent nonlinearities, we will follow the common approach of reporting marginal effects instead. More specifically, the marginal effects we report apply to the medians of the explanatory variables, with two exceptions. To highlight some of the interaction effects that we find in our data, we set Age to be 20 and Transmissions (the number of times the show has aired at the time of recording) to be 0. The resulting "representative agent" is a 20 year old white female without higher education, who lives in a relatively small town and plays the final of our game for a jackpot of £4,300, which potentially could have been £41,150. For dummy variables we consider the effect of a discrete change from 0 to 1. As noted by Ai and Norton (2003), the traditional way of calculating marginal effects and their standard errors is not valid for interaction terms, and we therefore apply the alternative method they propose. For the sake of consistency, we report significance levels that apply to the marginal effects, though these levels do not differ materially from the significance levels for the original regression coefficients. Original coefficients and their significance levels are available from the authors upon request.

II. Demographic Characteristics

First, we investigate how various demographic characteristics are related to the propensity to cooperate. Our later analyses include these demographic variables as control variables.

In previous studies examining the relations between demographic characteristics and cooperative behavior, most attention has been directed to gender. Psychologists have a long history when it comes to investigating the relation between gender and behavior, and, over the past decade, economists have become increasingly interested in gender effects as well. The standard finding is that women act more pro-socially than males, but the reverse is also found.[9] For contextual settings similar to ours, List (2004, 2006), Oberholzer-Gee, Waldfogel and White (2010) and Belot, Bhaskar

---

[9] One possible cause for the varying results is that males and females respond differently to specific contextual settings of the experiments (Croson and Gneezy, 2009). For example, if women are more risk averse than men, this may lead to different social behavior in situations in which risk is involved (Eckel and Grossman, 2008).

and van de Ven (2010a) report that women are more cooperative than men, although some results are only marginally significant.

For other demographic characteristics, the experimental findings are also mixed. Carpenter, Daniere and Takahashi (2004), for example, run public good experiments with symbolic but costly punishment in Bangkok and Ho Chi Minh City. They find that in Bangkok males and higher educated subjects contribute more, while there is no significant age effect. The same experiment in Ho Chi Minh City, however, shows the opposite findings: males and higher educated subjects cooperate less and age increases cooperation. Gächter, Herrmann and Thöni (2004) find no influence of background characteristics in a one-shot public good experiment with Russian subjects.

In order to add to this literature we will explore the effect of various demographic characteristics on cooperative behavior. We employ the following set of variables:

- "Gender" is a dummy variable indicating whether a contestant is male (1) or female (0).

- "Age" is a continuous variable measuring the contestant's age in years. In many instances the contestant's age is not explicitly mentioned during the show. In these cases we estimate age on the basis of physical appearance and other helpful information such as the age of children.

- "Race" is a dummy variable indicating whether a contestant is white (1) or non-white (0). We apply such a broad distinction because the large majority of contestants are white.

- "City" and "London" are two dummy variables that are constructed in order to distinguish contestants that live in major urban areas from those that reside in more rural surroundings. Contestants' city or county of residence is always an integral part of the introductory talk. City indicates whether a contestant lives in a large urban area (1) or not (0). We define a large urban area as a conurbation with a population exceeding 250,000 inhabitants.[10] For some contestants we only know their region and not their exact town or city; we then assume a small domicile. London indicates whether a contestant lives inside (1) or outside (0) the Greater London Urban Area.

- "Education" is a dummy variable for the level of education and differentiates between those with at least a bachelor degree (1) and those without (0). Players generally do not talk about their education during the show. We therefore estimate a contestant's level of education on

---

[10] For England and Wales, the population data and the definitions of conurbations are taken from the UK Office for National Statistics (www.statistics.gov.uk). Similar information for Scotland, Northern Ireland and Ireland is from the General Register Office for Scotland (www.gro-scotland.gov.uk), the Northern Ireland Statistics and Research Agency (www.nisra.gov.uk) and the Central Statistics Office Ireland (www.cso.ie), respectively.

the basis of her occupation, which is always explicitly mentioned when she is introduced, and on the basis of other information given in talks. Contestants who are currently enrolled in higher education and people whose job title suggests work experience equivalent to the bachelor level or higher are included in the higher education category. From the information that we have about each contestant, the proper binary values are generally clear.

- "Student" is a dummy variable indicating whether the contestant currently is a higher education (undergraduate or postgraduate) student (1) or not (0).

Estimates for Age and Education are based on the independent judgments of three research assistants, where each value is based on the assessments of two of them. When the estimates for Education were different, we decided on the most appropriate value ourselves. For Age we took the mean of the two judgments, and included our own assessment as a third input if the values of the coders diverged more than five years.

We have also attempted to collect data on contestants' marital status and the existence of children. These topics were, however, not systematically discussed in the program and values would therefore be unknown for the large majority of our contestants. Table 2 summarizes all the variables that are included in our analyses, including the demographic characteristics.

Table 3, Model 1 shows the regression results for a model that includes demographic characteristics only. To be able to distinguish both general gender and age effects as well as a possible interaction effect, the interaction of gender and age is also included. The results show that, relative to our representative 20 year-old female agent, young males are 22 percentage points less likely to cooperate (p = 0.002). In line with past results by List (2004) and Carpenter, Connolly and Myers (2008), this difference disappears when age increases. The effect of age is significantly different for males and females (p = 0.001). Women do not become significantly more or less cooperative when age increases (p = 0.422). Men, on the other hand, do have a higher propensity to cooperate as they are older: their cooperation rate increases by more than one percentage point per year (p = 0.000; untabulated). [11] Contrary to the two previous studies, we find that the gender difference not only disappears as age increases, but actually reverses; males become significantly more likely to "split" from age 46 onwards. Figure 1 displays observed cooperation rates at different age levels for both

---

[11] The effect of age for males could be related to increasing dependence on others (van Lange *et al.*, 1997), or to hormonal or neurological changes as men grow older, but we are hesitant to draw conclusions in these directions for we cannot exclude that a generational effect (van Lange *et al.*, 1997; List, 2004) or a wealth effect is (partly) driving our finding.

Table 2: Summary Statistics

The table shows descriptive statistics for the explanatory variables in our analyses of cooperative behavior based on the decisions of 574 contestants to either "split" or "steal" the jackpot in the Prisoner's Dilemma at the end of the British TV game show "Golden Balls". *Age* is the contestant's age measured in years. *Gender*, *Race*, *City*, *London*, *Education* and *Student* are dummy variables taking the value of 1 if the contestant is male (*Gender*), is white (*Race*), lives in a conurbation with a population exceeding 250,000 inhabitants (*City*), is a resident of the Greater London Urban Area (*London*), has completed or is enrolled in higher education (bachelor degree or higher) or has equivalent working experience (*Education*), or is a student (*Student*), respectively. *Actual stakes* is the natural logarithm of the size of the jackpot in the Prisoner's Dilemma game. *Potential stakes* is the natural logarithm of the highest possible jackpot at the start of the third round. *Transmissions* expresses the number of episodes that was already aired when the current episode was recorded in the studio. *Vote received from opp.* is a dummy variable taking the value of 1 if the contestant's final opponent has tried to vote her off the program at an earlier stage of the game. *Promise* is a dummy variable taking the value of 1 if the contestant explicitly promised her opponent to "split" (or not to "steal") the jackpot. *Lie Round 1 (Round 2)* is a dummy variable taking the value of 1 if the contestant has misrepresented her back row balls – either by overstating a cash ball or by hiding a killer ball – in the first (second) round. *Lie cash (killer) ball Round 1 (Round 2)* is a dummy variable taking the value of 1 if the contestant has overstated a cash ball (hidden a killer ball) in the first (second) round. Standard deviations for the two stakes variables and the transmissions variable are calculated across episodes (*N* = 287) to avoid the effect of clusters at the episode level. All monetary values are in UK Pounds (£1.00 ≈ $1.75).

	Mean	Stdev	Min	Median	Max
Demographic Characteristics					
Age	36.78	11.76	18.00	34.40	73.00
Gender (male=1)	0.47	0.50	0.00	0.00	1.00
Race (white=1)	0.92	0.27	0.00	1.00	1.00
City (large=1)	0.47	0.50	0.00	0.00	1.00
London (London=1)	0.14	0.35	0.00	0.00	1.00
Education (high=1)	0.34	0.47	0.00	0.00	1.00
Student (student=1)	0.09	0.28	0.00	0.00	1.00
Stakes and Context					
Actual stakes (log)	8.19	2.08	1.05	8.37	11.51
Potential stakes (log)	10.68	0.60	8.52	10.62	12.03
Transmissions	111.68	74.18	0.00	109.00	214.00
Reciprocal Preferences					
Vote received from opp. (yes=1)	0.05	0.22	0.00	0.00	1.00
Expectational Conditional Cooperation					
Promise (promise=1)	0.53	0.50	0.00	1.00	1.00
Past Deceitful Behavior					
Lie Round 1 (lie=1)	0.41	0.49	0.00	0.00	1.00
Lie Round 2 (lie=1)	0.36	0.48	0.00	0.00	1.00
Lie cash ball Round 1 (lie=1)	0.24	0.42	0.00	0.00	1.00
Lie cash ball Round 2 (lie=1)	0.15	0.36	0.00	0.00	1.00
Lie killer ball Round 1 (lie=1)	0.21	0.41	0.00	0.00	1.00
Lie killer ball Round 2 (lie=1)	0.24	0.43	0.00	0.00	1.00

Table 3: Binary Probit Regression Results [1/2]

The table displays results from the Probit regression analyses of contestants' decisions to "split" (1) or "steal" (0) the jackpot in the Prisoner's Dilemma at the end of the British TV game show "Golden Balls". *First (Second) half* is a dummy variable taking the value of 1 if less (more) than 50 percent of the episodes in our sample were already aired when the current episode was recorded in the studio. Definitions of other variables are as in Table 2. For each explanatory variable, the marginal effect is shown for a representative agent who takes the median value on all variables, except for *Age* and *Transmissions*, which are set to 20 and 0, respectively. Standard errors are corrected for clustering at the episode level, *p*-values are in parentheses.

	Model 1	Model 2	Model 3	Model 4	Model 5
Demographic Characteristics					
Age	0.002 (0.422)	0.002 (0.345)	0.003 (0.311)	0.002 (0.352)	0.002 (0.403)
Gender (male=1)	-0.221 (0.002)	-0.241 (0.000)	-0.233 (0.001)	-0.236 (0.001)	-0.233 (0.001)
Race (white=1)	0.134 (0.101)	0.143 (0.082)	0.139 (0.091)	0.142 (0.089)	0.149 (0.065)
City (large=1)	-0.039 (0.396)	-0.043 (0.359)	-0.045 (0.335)	-0.039 (0.405)	-0.036 (0.439)
London (London=1)	0.059 (0.402)	0.066 (0.348)	0.059 (0.400)	0.054 (0.444)	0.058 (0.406)
Education (high=1)	0.088 (0.062)	0.093 (0.053)	0.094 (0.050)	0.094 (0.049)	0.091 (0.058)
Student (student=1)	0.012 (0.888)	-0.002 (0.983)	-0.008 (0.923)	-0.013 (0.884)	-0.007 (0.933)
Age x Gender	0.011 (0.001)	0.010 (0.001)	0.010 (0.001)	0.011 (0.001)	0.010 (0.001)
Stakes and Context					
Actual stakes (log)		-0.043 (0.000)	-0.048 (0.000)	-0.049 (0.000)	-0.048 (0.000)
Potential stakes (log)			0.057 (0.139)	0.174 (0.006)	
Transmissions				-0.000 (0.722)	
Potential stakes x Transmissions				-0.001 (0.037)	
Second half (second=1)					0.030 (0.508)
Potential stakes x First half					0.139 (0.005)
Potential stakes x Second half					-0.035 (0.543)
Wald chi^2 (df)	34.87(8)	51.61(9)	52.57(10)	57.87(12)	62.35(12)
Log pseudo-likelihood	-379.78	-371.55	-370.42	-368.29	-367.51
Pseudo R^2	0.043	0.064	0.067	0.072	0.074
N	574	574	574	574	574
Number of clusters	287	287	287	287	287

males, females and aggregates, clearly depicting an age effect for men. Further analyses show that there is no evidence of a quadratic age effect, neither for men nor for women. We have also experimented with specifications where the (semi-) continuous age variable is replaced by a set of dummy variables that represent various age groups. The results are economically and statistically similar.

When it comes to race we find weak evidence that whites are more likely (about 13 percentage points) to cooperate than non-whites ($p = 0.101$; $p < 0.10$ in the models discussed hereafter). List (2004, 2006) and Oberholzer-Gee, Waldfogel and White (2010) report a similar pattern, yet in more conventional experiments the reverse is often found (see, for example, Cox, Lobel and McLeod, 1991). Since possible but unobservable wealth effects could contribute to this result, it should be interpreted with caution.
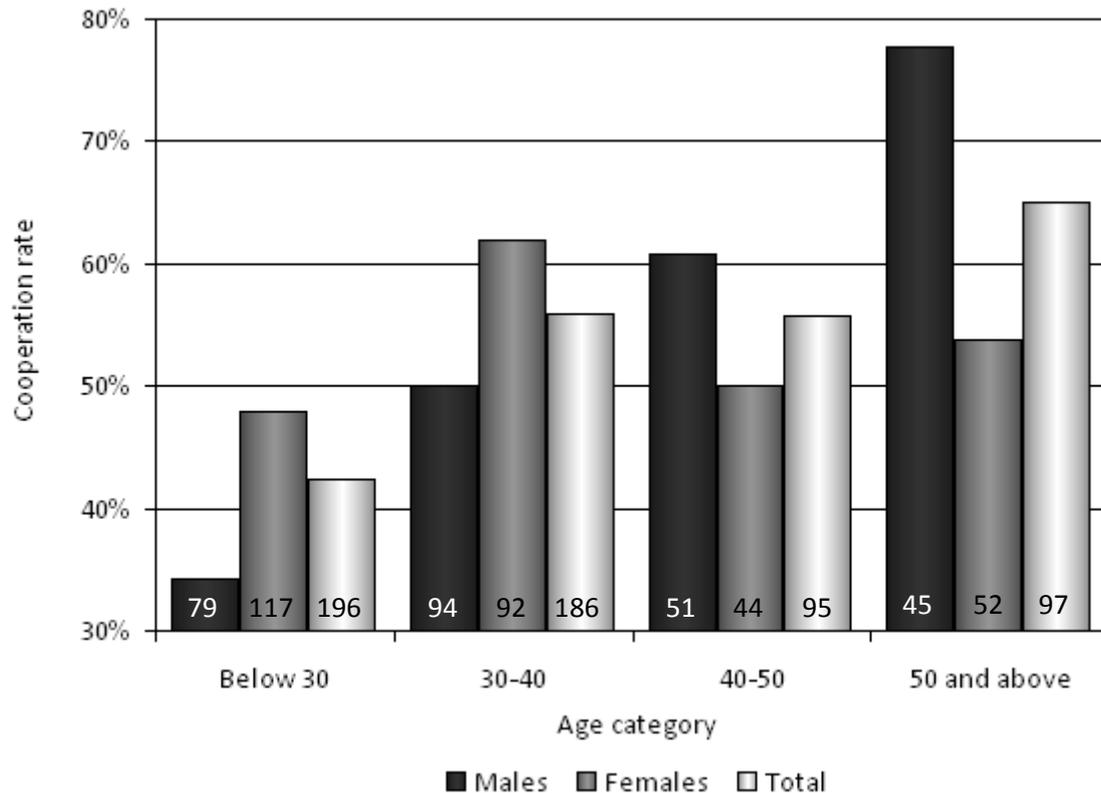
Figure 1: Age and the Propensity to Cooperate for Males and Females. The figure displays the relative frequency of contestants who decide to "split" across various age intervals. Bars depict the percentage of cooperators within specific age brackets for males, females and the aggregate, respectively. For each category, the number of contestants is displayed at the bottom of the bar.

Higher educated contestants are about 9 percentage points more cooperative ($p = 0.062$), although this effect is only marginally significant in the current model and not consistently significant across the various regression models discussed hereafter ($0.041 < p < 0.070$). Similar to the effect of race, the effect of education could be spurious due to an unobservable wealth effect.

Students are frequently used as subjects in experiments, and the reliance on such a specific subject pool is often criticized. Sears (1986), for example, extensively describes how the use of student subjects might produce misleading or mistaken conclusions about social behavior. It is therefore interesting to investigate whether there is evidence that students behave differently from others, holding other observable characteristics constant. This turns out not to be the case. Controlling for demographics such as age and education, our regression results yield no indications of a different attitude toward cooperation among students ($p = 0.888$).[12]

---

[12] In a similar vein, van Lange *et al*. (1997) and Bellemare and Kröger (2007) do not detect a difference, whereas Carpenter, Connolly and Myers (2008) and Egas and Riedl (2008) do report a negative bias.

None of the residence dummy variables have a significant effect. Possibly, relatively small social differences between urban and more rural areas in the UK explain this null result.[13]

III.   Stakes and Context

Economists typically argue that behavior will converge toward the prediction of rational self-interest if the stakes increase (*e.g.*, Rabin, 1993; Telser, 1995; Levitt and List, 2007). The evidence from lab and field experiments is, however, not generally supportive of this view. Except for the finding that people seem to become more willing to accept relatively low offers in ultimatum bargaining games when the stakes are high, empirical research generally finds no evidence that stake size affects behavior, even when the stakes are increased up to several months' wages.[14]

Given that the stakes in "Golden Balls" are widely ranging, and, on average, considerably larger than in previous studies, the show provides an excellent opportunity to re-examine the relation between cooperation and stakes. In addition, compared to earlier game-show studies on cooperation, an advantage of "Golden Balls" is that the stakes are mainly built up by a random process and not by contestants' answers to trivia questions. The latter may lead to a spurious correlation because the ability to answer trivia questions may be related to unobserved background characteristics such as income, which in turn may well be related to the propensity to cooperate.

The variable that we use in our regressions is labeled "Actual stakes" and defined as the natural logarithm of the size of the jackpot.

Model 2 in Table 3 displays the regression results when the stakes are included. Clearly, cooperative behavior in our show is sensitive to the amount that is at stake. To illustrate this effect, Figure 2 depicts the actual and estimated cooperation rates for different stake levels. The fitted line based on our full regression model (Model 6 presented later on) appears to capture the pattern rather well. Cooperation is high when the stakes are relatively small: for amounts up to £500, people on average

---

[13] In experiments conducted in a region of Russia where there is a large gap, Gächter and Herrmann (2011) do find that rural residents are more cooperative than urban residents.

[14] See, for example, Hoffman, McCabe and Smith (1996a), Slonim and Roth (1998), Cameron (1999), List and Cherry (2000, 2008), Fehr, Fischbacher and Tougareva (2002), Munier and Zaharia (2002), Carpenter, Verhoogen and Burks (2005), Johansson-Stenman, Mahmud and Martinsson (2005), and Kocher, Martinsson and Visser (2008). For TV game show data, List (2004, 2006) and Oberholzer-Gee, Waldfogel and White (2010) also find that cooperative behavior is practically invariant to the stakes. Belot, Bhaskar and van de Ven (2010a) report some counter-intuitive evidence that cooperation actually increases with stakes.
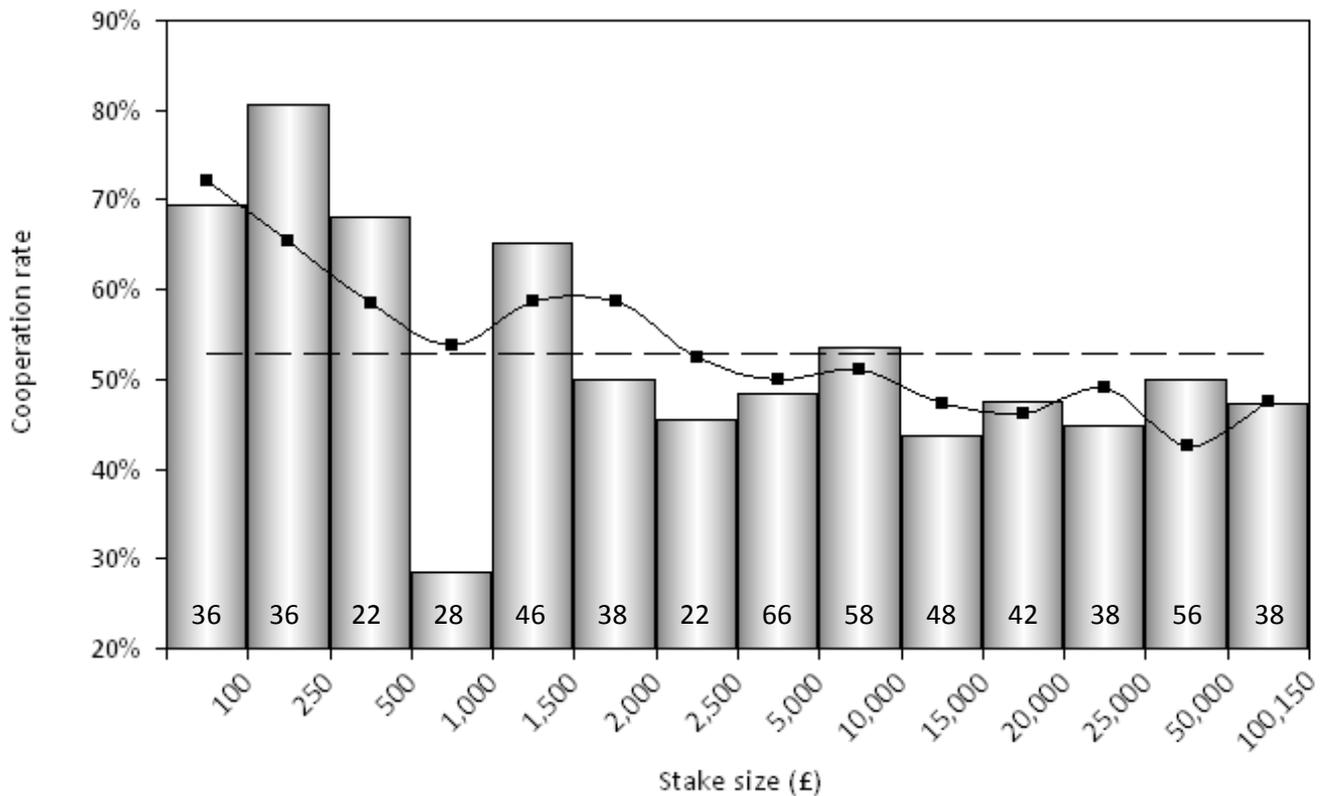
Figure 2: Stakes and the Propensity to Cooperate. The figure displays the relative frequency of contestants who decide to "split" across various stake intervals. Tick mark values represent the endpoints of the intervals. Each bar depicts the percentage of cooperators within a specific stake bracket. The dashed line reflects the average cooperation rate across our full sample, while the solid line connects the average estimate of the propensity to cooperate for each stake bracket. The estimates are computed using our "full model", which is Model 6 in Table 4. For each interval, the number of contestants is displayed at the bottom of the bar.

cooperate 73.4 percent of the time. The rate drops to approximately 45 percent as the stakes increase and remains relatively stable for the largest amounts. An unreported test shows that we cannot reject that the relation becomes essentially flat for stakes larger than £1,500.

While the absolute level of the stakes thus appears to have some influence on the propensity to cooperate, behavioral research suggests that people do not always evaluate prospects just in absolute terms, but rather they sometimes use relative comparisons to determine subjective values. This way, what comprises the context can strongly influence choices (Kahneman, Ritov and Schkade, 1999). In choice tasks, for example, one can increase the likelihood that a given option is chosen by adding an alternative to the choice set that is dominated by the given option but not by the other alternatives available (Huber, Payne and Puto, 1982). Also, as a consequence of the use of relative judgments, seemingly irrelevant anchors can influence how people value goods of various kinds (Green *et al.*, 1998; Ariely, Loewenstein and Prelec, 2003; Simonson and Drolet, 2004) and even (risky) monetary prospects (Johnson and Schkade, 1989).

[18]

For our purposes, the question of interest is whether the game show influences the contestants' perceptions of what constitutes "serious money". Suppose that some contestants decide that for "serious" money they are willing to bear the reputational costs, if any, of defecting on national TV, but if the stakes are small, so-called "peanuts", then they will just cooperate to look good. In this scenario, we would observe the pattern of cooperation in our data: high cooperation rates for low stakes and lower cooperation for high stakes. The interesting point, however, is that the "small" stakes on this show, several hundred Pounds, are quite large relative to most experiments. So even when the contestants are playing for what seems to be peanuts, these are big peanuts!

In "Deal or No Deal", another game show that has even larger stakes than Golden Balls, Post *et al.* (2008) also find strong evidence of such a "big peanuts" phenomenon. Namely, when unlucky contestants faced decisions near the end of the show that were "merely" for thousands of euros, they displayed little or no risk aversion. In fact, some of their contestants made risk-seeking choices in such situations. The authors provide further evidence for this behavior in classroom experiments designed to mimic the show at two levels of stakes, call them "low" and "medium". In the low stakes treatment the average prize was €40 with a maximum of €500, while in their medium stakes treatment the average prize was €400 with a maximum of €5,000. While risk aversion increased with stakes within each treatment, such an effect was not found across treatments: despite the very different money amounts, risky choices were similar for the low and medium stakes session. Choices in both conditions were even remarkably similar to those made in the actual TV show, despite the huge stakes used there (average €400,000, maximum €5,000,000).

These results suggest that a context can convert a sum of money that would normally be considered consequential into perceived peanuts. In the "Golden Balls" scenario, earlier expectations about the jackpot size or a specific value from the game might operate as an anchor or reference value by which the actual size of the jackpot is evaluated.[15] The most obvious benchmark contestants may use seems to be the maximum possible jackpot at the beginning of Round 3. Though the expected jackpot size might be an alternative candidate, it is neither salient nor easily calculated. In fact, even a rough assessment is rather complicated, particularly because of the influence of killer balls. The maximum potential jackpot, however, is always visually displayed and explicitly stressed by the game show host.

In order to test for such an effect, we include the variable "Potential stakes" in our analyses, defined as the highest possible jackpot at the start of Round 3. As with the actual stakes, we take the natural

---

[15] We intentionally do not use the term "reference point" in order to avoid associations with prospect theory here, since we are hesitant to translate the elements of prospect theory to preferences in this game and to derive testable predictions.

logarithm. Not surprisingly, the actual and the potential stakes variable are significantly correlated, but due to the effect of killer balls and the skewed distribution of cash balls the degree is rather limited; the Pearson correlation coefficient is $\rho = 0.30$.

Model 3 shows the new results. The positive sign of the potential stakes coefficient is in line with what we would expect, but the effect is statistically insignificant for the entire sample ($p = 0.139$). Interestingly, the effect becomes marginally significant if we exclude the (returning) contestants who already appeared in a previous episode of the show ($p = 0.066$; untabulated). This gives rise to the idea that the anchoring effect of the potential jackpot may decrease over time as contestants become more familiar with the show by watching it on TV. Prior shows will give contestants an impression of expected payoffs, which may help them to evaluate whether the stakes they face themselves are high or low in the context of the game and reduce the role of an episode-specific reference value such as the maximum possible jackpot size.

We explore the effect of experience by testing whether the effect of the potential jackpot changes as contestants have watched more episodes on TV. As a (noisy) proxy for how many shows a contestant has watched, we define the variable "Transmissions" as the number of different episodes broadcast on TV prior to the studio recording of the current episode. Model 4 in Table 3 displays the results. There is no significant main effect of this variable ($p = 0.722$), indicating that there is no evidence of a trend in the cooperation rate over time. However, the interaction effect of the number of transmissions and the potential stakes is significantly negative ($p = 0.037$), and implies that the anchoring effect of the maximum possible jackpot decreases by 0.10 percentage point for each previously aired episode. Controlling for this interaction effect, the effect of the maximum potential jackpot is highly significant in the early episodes ($p = 0.006$), where doubling the maximum potential jackpot increases cooperation by more than 12 percentage points.[16]

The pattern of a pronounced effect of the potential jackpot size in the earlier but not in the later shows also becomes apparent if we include dummy variables that subdivide our sample. Model 5, for example, uses a natural subdivision and employs a dummy variable for the first 149 episodes (0-112 transmissions prior to the recordings) and for the remaining 138 (149-214 prior transmissions). Clearly, the effect is significant across the first half our data ($p = 0.005$) and insignificant thereafter ($p = 0.543$).

---

[16] Because the potential stakes and the actual stakes are correlated, the interaction of Potential stakes and Transmissions might pick up an effect of the interaction of Actual stakes and Transmissions. As a robustness check, we have therefore also added the latter to Model 4. The effect of this additional control variable is insignificant ($p = 0.573$; untabulated), confirming our interpretation.

IV.   Reciprocal Preferences

Reciprocity refers to a tendency to repay kindness with kindness and unkindness with unkindness. Reciprocal behavior in the field is generally embedded in long term social interaction, and reputation concerns therefore form a plausible explanation for virtually all instances where people reciprocate (Sobel, 2005). Reciprocal actions can sometimes also be explained by preferences over outcome distributions, most notably by a desire for equity or equality (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). Interestingly, laboratory experiments in which such motivations are controlled for have convincingly demonstrated that people also have a real intrinsic preference for reciprocity.[17]

"Golden Balls" provides a neat opportunity to investigate the presence of reciprocal preferences outside of the laboratory and for substantial stakes. In particular, in the first two rounds contestants cast votes to determine who has to leave the show. Each vote carries a significant weight due to the small number of contestants, and voting against somebody can be viewed as unkind, particularly when the other votes were cast against other players. If people indeed have reciprocal preferences, we would expect that a contestant who makes it to the final in spite of her opponent's vote against her, has a lower propensity to cooperate.[18]

Although the voting is anonymous, it is often straightforward to deduce who has voted against whom. If a contestant in Round 1 (Round 2) receives three (two) votes it is obvious that all others have voted against her and that she herself has voted against the contestant who received one vote. For the other possible distributions of votes, we can usually deduce the individual votes from the banter preceding the vote, or, in the case of a tie, from the discussion following the vote. In the banter leading up to a vote, contestants generally make abundantly clear whom they intend to vote against (possibly out of an attempt to coordinate voting with other contestants). In the case of a tie, contestants openly discuss who they want to leave the program; if it was not already clear from the banter whom they had originally voted against, this post-vote discussion generally makes it apparent. This procedure allows us to determine a contestant's vote 95 percent of the time. For various

---

[17] See, for example, Kahneman, Knetsch and Thaler (1986), Blount (1995), Brandts and Solà (2001), Kagel and Wolfe (2001), McCabe, Rigdon and Smith (2003), Charness and Rabin (2002), Cox (2002, 2004), Falk, Fehr and Fischbacher (2003, 2008), Offerman (2002), Charness (2004), de Quervain *et al.* (2004). For theoretical accounts, see, for example, Rabin (1993), Levine (1998), Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006), Cox, Friedman and Gjerstad (2007), and Gul and Pesendorfer (2010).

[18] Studying voting patterns in the internationally successful TV game show "The Weakest Link", both Levitt (2004) and Antonovics, Arcidiacono and Walsh (2005) find evidence that people reciprocate against people who voted against them in past rounds. However, because players in this show have an incentive to vote off players that are more likely to vote against them, they cannot rule out that this strategic concern drives the effect.

reasons it is much more difficult to determine clear instances of someone going out of their way to be nice to another player, so we limit our analysis to negative rather than positive reciprocity.

Based on the voting information, we create a dummy variable entitled "Vote received from opponent", taking the value of one if a contestant received a vote from her final opponent and zero otherwise. If we could not establish whether a contestant received a vote from her final opponent, she is assigned the value of zero as well (exclusion of these cases does not change our results). Since contestants who receive votes often do not make it to the next round, relatively few contestants qualify: as displayed in Table 2, 5 percent (28 subjects) of the final-round contestants received a vote from their opponent.

Model 6 in Table 4 includes the new dummy variable. In line with the idea that people have reciprocal preferences, the likelihood of a contestant to cooperate with her final opponent plummets by approximately 21 percentage points if this opponent has voted against her earlier in the game ($p = 0.019$). There are, however, three alternative explanations for such a behavioral pattern that are unrelated to a genuine preference for reciprocity. Although we cannot rule out that these alternative explanations explain part of the effect, they do not appear particularly strong.

First, the causality may not run from receiving a vote to cooperativeness, but the other way around, players voting against contestants with a less cooperative disposition. This would imply that cooperation is also related to the number of votes received from other players, which appears not to be the case ($p = 0.231$; untabulated).

Second, a contestant may like to match her opponent's choice for reasons other than reciprocal concerns, and interpret the earlier vote against her as a signal that her opponent dislikes her and will not cooperate. However, her interpretation would generally not be legitimate: players do not cooperate less with someone they voted against ($p = 0.403$; untabulated). Moreover, the next section finds little support for such expectational conditional cooperation.

Last, a contestant's lower propensity to cooperate with someone who voted against her may be out of reputation concerns ("I am not to be messed with") instead of an intrinsic preference for reciprocity. However, when asked to explain their choice after the final decisions, contestants never use this costless opportunity to strengthen their message and point to their reciprocal nature.

V.   Expectational Conditional Cooperation

There is considerable evidence that many people have a preference for conditional cooperation, defined as the desire to match the cooperation of others. In laboratory and field experiments, about half of the subjects are more willing to cooperate if others do so as well (*e.g.*, Fischbacher, Gächter and Fehr, 2001; Frey and Meier, 2004). Conditional cooperation can arise from reciprocal preferences, but also for other reasons. Social norms or a desire for conformity might account for it, and, especially in the laboratory, egalitarian motives can often explain conditional cooperative behavior because equality in payoffs generally only arises if players coordinate on their level of cooperation, as is also the case in our show.

Experimental studies typically investigate conditional cooperation in settings where subjects have the possibility to condition their behavior directly on the behavior of others. In everyday life, such clear-cut conditioning is usually not possible, especially in one-shot situations. Conditional cooperation then has to be based on expectations about the behavior of others, and the degree of coordination would depend on the predictive power of available information and on whether and how this information is interpreted. A natural question is whether conditional cooperation can be observed when the conditioning is only on an expectation of cooperation rather than on actual cooperation.

In "Golden Balls" it is not possible for a contestant to condition directly on her opponent's behavior since the two are playing a simultaneous move game. However, we can investigate whether contestants condition their behavior on factors that form reliable predictors of their opponent's behavior. That is, we can investigate the joint hypothesis that players make rational forecasts of their opponent's behavior *and* then condition their behavior on those expectations.

The first step in such an analysis is determining the factors that a contestant could use to form an expectation about their opponent's likelihood of cooperation. One such factor is whether an opponent made a promise to "split". While the literature on conditional cooperation is rather recent, literature investigating the role of communication and, especially, promises in social dilemma situations already pointed towards tendencies of conditional cooperation. In a meta-analysis of Prisoner-Dilemma experiments, Sally (1995), for example, finds that cooperation occurs more often when the other player makes an explicit though non-binding promise that she will cooperate. The combination of a preference for conditional cooperation and a reluctance to lie (*e.g.*, Charness and Dufwenberg, 2005, 2006; Gneezy, 2005) can explain why promises have such an effect: people like to cooperate if others do, and a promise is a reliable signal of others' behavior if they have a reluctance to lie.

In the conversation prior to the decision to either "split" or "steal", many contestants explicitly promise to "split" or otherwise make a definitive statement of their intention to do so. Based on the statements made in this small talk, we create a dummy variable labeled "Promise", indicating whether the contestant made an explicit, unambiguous promise or announcement that she will choose "split" (1) or not (0).[19] As shown in Table 2, about half (53 percent) of the contestants make such a promise. We investigate both whether observing a promise is predictive of the cooperative behavior of the contestant making the promise, and whether a contestant conditions her behavior on whether or not her opponent made a promise.

As shown in Table 4, Model 7, a player's promise is a highly significant predictor of her propensity to cooperate ($p = 0.000$). Those who make a promise are about 31 percentage points more likely to cooperate. In fact, an explicit promise is the single most reliable predictor of whether someone will cooperate.[20]

While a promise is a strong signal of cooperation, contestants whose opponent made a promise do not have a higher propensity to choose split. In fact, as Model 7 also shows, if an opponent promises to be cooperative, the other player even displays a marginally significant *decrease* in the likelihood of choosing "split". Belot, Bhaskar and van de Ven (2010a) obtain a similar result.

An explicit promise is the strongest predictor of cooperation, but, as we have previously shown, there are also demographic factors that a contestant could use to forecast cooperation. For example, we have seen that young males cooperate less than young females. However, inferences from this sort of analysis have to be tentative since there could be an additional confound if opponents have a taste for cooperating with someone with particular demographic characteristics.

As Table 4, Model 8 shows, we find little evidence that contestants condition their behavior on their opponents' background characteristics. The only weak evidence for conditional cooperation is that people cooperate significantly more frequently with higher educated opponents ($p = 0.017$), but education is not a very strong predictor of behavior. If we assess the joint significance of the various

---

[19] If a contestant responds affirmative to a question whether she will choose "split" or if she announces that she will not choose "steal", Promise takes the value of one as well. The value is zero in all other cases, including when people give the impression that they plan to split but do not explicitly express themselves as such, when they just refer to earlier intentions (for example, "I came here to split"), when they confine themselves to statements like "you can trust me" and "I will not let you down", and when they only express their preference for a coordinated outcome ("I want us to split"; "I do not want both of us to go home empty-handed").

[20] Of course, we do not interpret the promise as causing the cooperation. The direction of the causation could go the other way.

Table 4: Binary Probit Regression Results [2/2]

The table displays results from the Probit regression analyses of contestants' decisions to "split" (1) or "steal" (0) the jackpot in the Prisoner's Dilemma at the end of the British TV game show "Golden Balls". The opponent variables measure the demographic characteristics of the contestant's opponent and are defined similar to the contestant's own demographic variables. Other definitions are as in previous tables.

	Model 6	Model 7	Model 8	Model 9	Model 10
Age	0.002 (0.387)	0.003 (0.283)	0.002 (0.457)	0.002 (0.372)	0.002 (0.355)
Gender (male=1)	-0.249 (0.001)	-0.292 (0.000)	-0.258 (0.001)	-0.252 (0.001)	-0.247 (0.001)
Race (white=1)	0.149 (0.079)	0.162 (0.077)	0.146 (0.078)	0.147 (0.089)	0.148 (0.083)
City (large=1)	-0.034 (0.467)	-0.027 (0.552)	-0.035 (0.463)	-0.034 (0.463)	-0.037 (0.433)
London (London=1)	0.041 (0.565)	0.039 (0.540)	0.050 (0.501)	0.042 (0.551)	0.037 (0.597)
Education (high=1)	0.088 (0.068)	0.089 (0.041)	0.089 (0.067)	0.089 (0.065)	0.089 (0.066)
Student (student=1)	0.001 (0.988)	-0.025 (0.768)	0.008 (0.924)	-0.002 (0.983)	0.001 (0.988)
Age x Gender	0.011 (0.000)	0.008 (0.010)	0.010 (0.009)	0.011 (0.001)	0.011 (0.001)
Actual stakes (log)	-0.050 (0.000)	-0.054 (0.000)	-0.052 (0.000)	-0.051 (0.000)	-0.052 (0.000)
Potential stakes (log)	0.183 (0.004)	0.174 (0.004)	0.170 (0.006)	0.180 (0.004)	0.186 (0.004)
Transmissions	-0.000 (0.660)	-0.000 (0.106)	-0.000 (0.608)	-0.000 (0.584)	-0.000 (0.497)
Potential stakes x Transmissions	-0.001 (0.026)	-0.001 (0.022)	-0.001 (0.030)	-0.001 (0.025)	-0.001 (0.029)
Reciprocal Preferences					
Vote received from opp. (yes=1)	-0.215 (0.019)	-0.237 (0.015)	-0.202 (0.020)	-0.214 (0.026)	-0.216 (0.025)
Exp. Conditional Cooperation					
Promise (promise=1)		0.311 (0.000)			
Promise opp. (promise=1)		-0.080 (0.053)			
Age opp.			0.001 (0.732)		
Gender opp. (male=1)			-0.118 (0.140)		
Race opp. (white=1)			-0.026 (0.775)		
City opp. (large=1)			0.051 (0.288)		
London opp. (London=1)			0.027 (0.713)		
Education opp. (high=1)			0.116 (0.017)		
Student opp. (student=1)			0.069 (0.442)		
Age opp. x Gender opp.			0.004 (0.279)		
Past Deceitful Behavior					
Lie Round 1 opp. (lie=1)				-0.013 (0.782)	
Lie Round 2 opp. (lie=1)				-0.027 (0.578)	
Lie cash ball Round 1 opp. (lie=1)					0.037 (0.491)
Lie cash ball Round 2 opp. (lie=1)					-0.064 (0.295)
Lie killer ball Round 1 opp. (lie=1)					-0.055 (0.330)
Lie killer ball Round 2 opp. (lie=1)					0.007 (0.894)
Wald chi^2 (df)	59.65(13)	98.03(15)	62.40(21)	59.19(15)	62.72(17)
Log pseudo-likelihood	-365.86	-343.31	-359.26	-365.64	-364.52
Pseudo R^2	0.078	0.135	0.095	0.079	0.082
N	574	574	574	574	574
Number of clusters	287	287	287	287	287

opponent background characteristics, we also find that they collectively do not have a significant effect on cooperation.[21] Analyzing the show "Friend or Foe?", Oberholzer-Gee, Waldfogel and White (2010) find no conditioning on opponent background characteristics in the first season, but they do find it in later seasons. They interpret this as conditional cooperation on the basis of learned expectations. We too have investigated whether conditioning arises as more episodes were transmitted, but we find no indication for such an effect.

As an alternative to the two models discussed above, we also examined a two-step approach. We first estimated each opponent's propensity to cooperate given her background characteristics and promise behavior, and then added the estimated propensity of opponents as an explanatory variable to our regression model. Again, we found no indication of conditionally cooperative behavior.

In summary, we find no evidence of expectational conditional cooperation. Apparently, either players cannot or do not forecast the behavior of their opponents, or they do not have conditionally cooperative preferences. Our evidence for reciprocal preferences in the previous section hints that it is the former rather than the latter interpretation that underlies this result.

Belot, Bhaskar and van de Ven (2010b) also provide evidence that predicting one's opponent's behavior is difficult. They had subjects watch clips from the Dutch counterpart to "Golden Balls" and asked them to assess the likelihood of each contestant's cooperation. While the estimated likelihood for cooperators was significantly higher than for defectors, the difference was only seven percentage points.

VI. Past Deceitful Behavior

In this section, we investigate whether lies influence opponents' willingness to cooperate. In the early rounds of the show, contestants have numerous opportunities to lie about the values on their hidden balls, lies that are quickly revealed to everyone. These lies can be consequential. If someone hides low value and killer balls and in so doing manages to remain in the game, she will have reduced the potential payoff to the remaining contestants.

In the final, contestants might be less likely to cooperate with opponents who have lied, either out of reciprocal concerns (*e.g.*, Brandts and Charness, 2003) or because they interpret lying as evidence of a self-interested nature and a sign of an imminent "steal" decision (*e.g.*, van Lange and Kuhlman,

---

[21] We also looked at more complex mechanisms related to the similarity of the contestant's own background characteristics and those of her opponent, such as whether people cooperate more with those who are more similar to them ("social-distance" effects). In our data, there is no evidence of such behavior.

1994). Thus, past deceitful behavior is not a separate possible determinant of cooperation, but rather a special case of either reciprocity or conditional cooperation, or both.

We collected data on the statements made by contestants and the actual values of the balls that they possessed, allowing us to specify various measures for deceitful behavior. The analyses reported here are restricted to the use of dummy variables. We have also tried more complex, continuous variables for lying, but these approaches yielded similar results.

We apply separate variables for each game round. The general variables take the value of one if the contestant lied, irrespective of whether she overstated the monetary value of a cash ball or failed to disclose a killer ball. To distinguish between these two types of lies, we also use specific variables for each type separately. It is not obvious which type of lie would be considered more objectionable. On the one hand, lying about killers is much more harmful to others than exaggerating the value of a cash ball, and, assuming a preference to reciprocate, doing so could then be expected to have a greater negative effect on an opponent's propensity to cooperate. On the other hand, lying about killer balls might also be more understandable, since killer balls have a much greater impact on a contestant's chances to be voted off the show. Players may realize that nearly everyone will fail to disclose a hidden killer ball, and thus not be inclined to punish such behavior. Lying about cash ball values might be more like gratuitous lying and be viewed more harshly, and, consequently, have a greater effect on cooperation.

As shown in Table 2, lying is rather common on the show: 41 percent of the contestants who made it to the final lied about their back row balls in Round 1, while 36 percent lied in Round 2 (some did both). Furthermore, in the first round, 24 percent overstated the value of a cash ball, while 21 percent hid a killer ball (some did both). For the second round, these figures are 15 and 24 percent, respectively.[22]

Table 4 displays the regression results when we add the general dummy variables (Model 9) and the dummy variables that distinguish between lying about killer balls and lying about cash balls (Model 10). We find that past lies of an opponent do not affect a contestant's propensity to cooperate: each of the six variables is insignificant ($0.295 < p < 0.894$). In addition to these simple tests, we also investigated whether lying is considered less fair and has more impact the more it is unexpected or "abnormal" given the circumstances, but again we found no significant effect.[23] Lying neither

---

[22] Conditional on having at least one killer ball on their back row, contestants hid a killer ball 50 (43) percent of the time in Round 1 (Round 2).

[23] We used a two-stage procedure to express the abnormality of a lie. For each round, we estimated a regression model that explains a contestant's propensity to lie, given the ball values on her back row, the ball

predicts a contestant's own cooperative behavior.[24] One plausible interpretation of these results is that lying is seen as an inherent part of "Golden Balls" and therefore unobjectionable behavior, much as bluffing is considered in the game of poker (Charness and Dufwenberg, 2005).

VII. Conclusions and Discussion

"Golden Balls" provides us with the possibility to examine cooperative behavior outside the conventional context of the laboratory with large sums of money at stake.

Our results provide support for the view that attitudes are strongly influenced by context. We find unusually high rates of cooperation when the luck of the game reduces the stakes to "merely" a few hundred Pounds. Such amounts are tiny in the light of the thousands and even tens of thousands the game is often played for, but would be considered very large in any laboratory setting. In the early days of the show, when the contestants have not had an opportunity to watch the show on TV and are still learning what kind of stakes are to be expected, cooperation rates appear to be influenced by the salient but normatively irrelevant value representing the maximum they could have been playing for with a lucky selection of balls. Over time, this effect vanishes, suggesting that expectations about stakes become well-informed.

We label the tendency to be unusually cooperative for what would normally be considered high stakes a "big peanuts" result. Players seem to feel that when making a choice about a few hundred dollars when they might otherwise have been dividing tens of thousands, they are playing for "peanuts", and cooperate, perhaps thinking that it is not worth stealing for what they perceive to be so little money. This finding reinforces a similar result for risk taking behavior in another game show, "Deal or No Deal". In that context, where the stakes were even higher, amounts of money in the tens of thousands of dollars became perceived as peanuts, since hundreds of thousands of dollars had been on the line. These are very big peanuts indeed.

Using the interaction that occurs among contestants prior to the final, we also examined the effects that past opposition and lying have on cooperation. Using the votes we find evidence to support the

---

values on her front row, and the rank of her front row balls relative to those of the other players. For each final contestant, the "abnormality" of a lie we then measured as the difference between unity and this estimated lie propensity.

[24] Such a relation might be expected if the propensity to be honest and the propensity to cooperate are influenced by a similar preference for "pro-social", "kind" or "fair" behavior. It has, for example, been argued that the reluctance to lie is driven by guilt aversion (Charness and Dufwenberg, 2005, 2006; Gneezy, 2005), and empirical analysis suggests that guilt aversion is also a strong driving force behind cooperative behavior (Dufwenberg, Gächter and Hennig-Schmidt, 2010).

view that people have reciprocal preferences. Contestants are less likely to cooperate if their opponent has tried to vote them off the show at an earlier stage of the game. Lying on the other hand has no significant effect. We investigated several measures, but none was significantly related to cooperation. Lying is evidently not frowned upon in "Golden Balls", perhaps because it is expected. The different impact of opposition and lying might be related to their different nature in this game. Voting is a directed and aggressive act towards one specific contestant. Lying, on the other hand, is an undirected and defensive act.

With "Golden Balls" we are also able to investigate an interesting, expectational form of conditional cooperation. Specifically, since explicit promises to cooperate are strong predictors of actual cooperation, we can see whether players are more likely to cooperate with someone who has made such a promise. We find no evidence of such behavior. More generally, we find that players do not appear to condition their choice of whether to cooperate on factors that predict the cooperation likelihood of their opponent. Players may lack the ability or ignore the possibility to reliably interpret information about the expected behavior of others, or they may not have a preference for matching the other's choice. Given our finding that people reciprocate votes against them, the former explanation seems more likely. For situations beyond the context of our game these results suggest that conditional cooperation is not a very important phenomenon, at the least when direct conditioning is not possible and people would need to form expectations about the behavior of others.

We conclude with some comments on the generalizability of our results. There are three primary concerns. First, selection procedures may have affected the average cooperation rate in our study. Subjects self-select into auditions, are then selected by the producer, and during the game they themselves have the opportunity to vote off opponents they would rather not play the final with. For some demographic variables, selection may perhaps also have affected the correlation with cooperation. Unfortunately, we cannot substantiate our intuition that such effects are negligible, nor could we have prevented them if they would exist. Note, however, that selection procedures are inevitable in any lab experiment or field setting. Moreover, the subjects in our sample vary widely in terms of their demographic characteristics and as a group they seem to resemble a (middle-class)[25] cross-section of the general population more closely than subjects in most conventional experiments.

Second, subjects' behavior in a game show might be influenced by what could be called "a drive to win the contest". However, an important but hard to answer question would then be what "winning"

---

[25] For whatever reason, whether it is the interest in applying or the preferences of the producers, contestants are rarely very rich or very poor.

actually means in this context. After all, it seems like a matter of personal social preferences whether winning is equivalent to a successful stealing attempt or to a successful coordination attempt.

Last, our contestants are not strictly playing a one-shot game. In the setting we study decisions are made on national TV, under the scrutiny of a studio audience and millions of viewers. This undoubtedly influences the behavior we observe. However, we do not feel that these special circumstances render our findings less interesting or less predictive of behavior in other settings. The truth is that every setting is, in some way, special. Subjects in a laboratory experiment know that their behavior is being scrutinized to some extent as well. Field settings are also "special"; bargaining over the price of a car or a house is different from negotiating compensation with a new employer or the division of household chores with a spouse. Although it would be fascinating for researchers to be able to surreptitiously study the outcomes of these sorts of interactions from the "real world", the researchers would still only be able to speculate on how their results would generalize to different real world settings. TV game shows offer a unique opportunity to study theoretically interesting behavior at stakes that are impossible to replicate in the lab. How the results compare with other contexts will be determined by future research.

In the absence of an ability to conduct such surreptitious field experiments in many domains, researchers are left with two alternatives: run experiments in the lab or the field, or study naturally occurring behavior in an interesting setting. This paper is an example of the latter strategy. Although a game show may seem like a strange environment, we think it may be closer to the situations that occur in the workplace than many other settings in which cooperation has been studied. Co-workers often must choose whether to cooperate, and their actions are often at least semi-public.

Finally, the big peanuts phenomenon, perhaps the most interesting finding in this paper, is one that does not appear to depend in any important way on the specific game show environment. As a US Senator once famously said, "a billion here, a billion there, pretty soon you're talking real money."

References

Ai, Chunrong, and Edward C. Norton. 2003. "Interaction Terms in Logit and Probit Models." *Economics Letters*, 80(1): 123-129.

Antonovics, Kate, Peter Arcidiacono, and Randall Walsh. 2005. "Games and Discrimination: Lessons from 'The Weakest Link'." *Journal of Human Resources*, 40(4): 918-947.

Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. "Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences." *Quarterly Journal of Economics*, 118(1): 73-105.

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

Bellemare, Charles, and Sabine Kröger. 2007. "On Representative Social Capital." *European Economic Review*, 51(1): 183-202.

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

Belot, Michèle, V. Bhaskar, and Jeroen van de Ven. 2010b. "Can Observers Predict Trustworthiness?" *Review of Economics and Statistics*, forthcoming.

Bennett, Randall W., and Kent A. Hickman. 1993. "Rationality and the 'Price is Right'." *Journal of Economic Behavior & Organization*, 21(1): 99-105.

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

Blount, Sally. 1995. "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences." *Organizational Behavior and Human Decision Processes*, 63(2): 131-144.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1): 166-193.

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

Brandts, Jordi, and Carles Solà. 2001. "Reference Points and Negative Reciprocity in Simple Sequential Games." *Games and Economic Behavior*, 36(2): 138-157.

Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press.

Cameron, Lisa A. 1999. "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia." *Economic Inquiry*, 37(1): 47-59.

Carpenter, Jeffrey P., Cristina Connolly, and Caitlin Knowles Myers. 2008. "Altruistic Behavior in a Representative Dictator Experiment." *Experimental Economics*, 11(3): 282-298.

Carpenter, Jeffrey P., Amrita G. Daniere, and Lois M. Takahashi. 2004. "Cooperation, Trust, and Social Capital in Southeast Asian Urban Slums." *Journal of Economic Behavior & Organization*, 55(4): 533-551.

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

Charness, Gary. 2004. "Attribution and Reciprocity in an Experimental Labor Market." *Journal of Labor Economics*, 22(3): 665-688.

Charness, Gary, and Martin Dufwenberg. 2005. "Deception: The Role of Guilt." Working paper. UC Santa Barbara and University of Arizona.

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

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences With Simple Tests." *Quarterly Journal of Economics*, 117(3): 817-869.

Cooper, David J., and John H. Kagel. 2009. "Other Regarding Preferences: A Selective Survey of Experimental Results." Working paper.

Cox, James C. 2002. "Trust, Reciprocity, and Other-Regarding Preferences: Groups vs. Individuals and Males vs. Females." In *Experimental Business Research*, ed. Rami Zwick and Amnon Rapoport, 331-350. Kluwer Academic Publishers.

Cox, James C. 2004. "How to Identify Trust and Reciprocity." *Games and Economic Behavior*, 46(2): 260-281.

Cox, James C., Daniel Friedman, and Steven Gjerstad. 2007. "A Tractable Model of Reciprocity and Fairness." *Games and Economic Behavior*, 59(1): 17-45.

Cox, Taylor H., Sharon A. Lobel, and Poppy Lauretta McLeod. 1991. "Effects of Ethnic Group Cultural Differences on Cooperative and Competitive Behavior on a Group Task." *Academy of Management Journal*, 34(4): 827-847.

Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature*, 47(2): 448-474.

Dawes, Robyn M. 1980. "Social Dilemmas." *Annual Review of Psychology*, 31: 169-193.

Dawes, Robyn M., and David M. Messick. 2000. "Social Dilemmas." *International Journal of Psychology*, 35(2): 111-116.

Dawes, Robyn M., and Richard H. Thaler. 1988. "Anomalies: Cooperation." *Journal of Economic Perspectives*, 2(3): 187-197.

de Quervain, Dominique J.-F., Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Ulrich Schnyder, Alfred Buck, and Ernst Fehr. 2004. "The Neural Basis of Altruistic Punishment." *Science*, 305(5688): 1254-1258.

Dufwenberg, Martin, Simon Gächter, and Heike Hennig-Schmidt. 2010. "The Framing of Games and the Psychology of Play." *Games and Economic Behavior*, forthcoming.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, 47(2): 268-298.

Dufwenberg, Martin, and Astri Muren. 2006. "Generosity, Anonymity, Gender." *Journal of Economic Behavior & Organization*, 61(1): 42-49.

Eckel, Catherine C., and Philip J. Grossman. 2008. "Differences in the Economic Decisions of Men and Women: Experimental Evidence." In *Handbook of Experimental Economics Results; Volume 1*, ed. Charles R. Plott and Vernon L. Smith, 509-519. Elsevier.

Egas, Martijn, and Arno Riedl. 2008. "The Economics of Altruistic Punishment and the Maintenance of Cooperation." *Proceedings of the Royal Society B: Biological Sciences*, 275(1637): 871-878.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2003. "On the Nature of Fair Behavior." *Economic Inquiry*, 41(1): 20-26.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2008. "Testing Theories of Fairness-Intentions Matter." *Games and Economic Behavior*, 62(1): 287-303.

Falk, Armin, and Urs Fischbacher. 2006. "A Theory of Reciprocity." *Games and Economic Behavior*, 54(2): 293-315.

Fehr, Ernst, Urs Fischbacher, and Elena Tougareva. 2002. "Do High Stakes and Competition Undermine Fairness? Evidence from Russia." Working paper. Institute for Empirical Research in Economics, University of Zurich.

Fehr, Ernst, and Simon Gächter. 1998. "Reciprocity and Economics: The Economic Implications of Homo Reciprocans." *European Economic Review*, 42(3-5): 845-859.

Fehr, Ernst, and Simon Gächter. 2000a. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90(4): 980-994.

Fehr, Ernst, and Simon Gächter. 2000b. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, 14(3): 159-181.

Fehr, Ernst, and Herbert Gintis. 2007. "Human Motivation and Social Cooperation: Experimental and Analytical Foundations." *Annual Review of Sociology*, 33: 43-64.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817-868.

Fischbacher, Urs, and Simon Gächter. 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments." *American Economic Review*, 100(1): 541-556.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, 71(3): 397-404.

Frey, Bruno S., and Stephan Meier. 2004. "Social Comparisons and Pro-Social Behavior: Testing 'Conditional Cooperation' in a Field Experiment." *American Economic Review*, 94(5): 1717-1722.

Gächter, Simon, Benedikt Herrmann. 2011. "The Limits of Self-Governance When Cooperators Get Punished: Experimental Evidence from Urban and Rural Russia." *European Economic Review*, 55(2): 193-210.

Gächter, Simon, Benedikt Herrmann, and Christian Thöni. 2004. "Trust, Voluntary Cooperation, and Socio-Economic Background: Survey and Experimental Evidence." *Journal of Economic Behavior & Organization*, 55(4): 505-531.

Gardner, Andy, and Stuart A. West. 2004. "Cooperation and Punishment, Especially in Humans." *American Naturalist*, 164(6): 753-764.

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

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

Green, Donald, Karen E. Jacowitz, Daniel Kahneman, and Daniel McFadden. 1998. "Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods." *Resource and Energy Economics*, 20(2): 85-116.

Gul, Faruk, and Wolfgang Pesendorfer. 2010. "Interdependent Preference Models as a Theory of Intentions." Working paper.

Haley, Kevin J., and Daniel M. T. Fessler. 2005. "Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game." *Evolution and Human Behavior*, 26(3): 245-256.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin F. Camerer, Ernst Fehr, and Herbert Gintis (ed.). 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies.* Oxford University Press.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin F. Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *American Economic Review*, 91(2): 73-78.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences*, 33(2-3): 61-83.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial Punishment Across Societies." *Science*, 319(5868): 1362-1367.

Hoffman, Elizabeth, Kevin A. McCabe, and Vernon L. Smith. 1996a. "On Expectations and the Monetary Stakes in Ultimatum Games." *International Journal of Game Theory*, 25(3): 289-301.

Hoffman, Elizabeth, Kevin A. McCabe, and Vernon L. Smith. 1996b. "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review*, 86(3): 653-660.

Huber, Joel, John W. Payne, and Christopher Puto. 1982. "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *Journal of Consumer Research*, 9(1): 90-98.

Johansson-Stenman, Olof, Minhaj Mahmud, and Peter Martinsson. 2005. "Does Stake Size Matter in Trust Games?" *Economics Letters*, 88(3): 365-369.

Johnson, Eric J., and David A. Schkade. 1989. "Bias in Utility Assessments: Further Evidence and Explanations." *Management Science*, 35(4): 406-424.

Kagel, John H., and Katherine Willey Wolfe. 2001. "Tests of Fairness Models Based on Equity Considerations in a Three-Person Ultimatum Game." *Experimental Economics*, 4(3): 203-219.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. "Fairness and the Assumptions of Economics." *Journal of Business*, 59(4): 285-300.

Kahneman, Daniel, Ilana Ritov, and David A. Schkade. 1999. "Economic Preferences or Attitude Expressions?: An Analysis of Dollar Responses to Public Issues." *Journal of Risk and Uncertainty*, 19(1-3): 203-235.

Kerr, Norbert L. 1999. "Anonymity and Social Control in Social Dilemmas." In *Resolving Social Dilemmas*, ed. Margaret Foddy, Michael Smithson, Sherry Schneider and Michael Hogg, 103-119, Psychology Press.

Kocher, Martin G., Peter Martinsson, and Martine Visser. 2008. "Does Stake Size Matter for Cooperation and Punishment?" *Economics Letters*, 99(3): 508-511.

Ledyard, John O. 1995. "Public Goods: A Survey of Experimental Research." In *Handbook of Experimental Economics*, ed. John H. Kagel and Alvin E. Roth, 111-194. Princeton University Press.

Levine, David K. 1998. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics*, 1(3): 593-622.

Levitt, Steven D. 2004. "Testing Theories of Discrimination: Evidence from Weakest Link." *Journal of Law and Economics*, 47(2): 431-453.

Levitt, Steven D., and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives*, 21(2): 153-174.

Levitt, Steven D., and John A. List. 2008. "Homo Economicus Evolves." *Science*, 319(5865): 909-910.

List, John A. 2004. "Young, Selfish and Male: Field Evidence of Social Preferences." *Economic Journal*, 114(492): 121-149.

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

List, John A., and Todd L. Cherry. 2000. "Learning to Accept in Ultimatum Games: Evidence from an Experimental Design that Generates Low Offers." *Experimental Economics*, 3(1): 11-29.

List, John A., and Todd L. Cherry. 2008. "Examining the Role of Fairness in High Stakes Allocation Decisions." *Journal of Economic Behavior & Organization*, 65(1): 1-8.

Marwell, Gerald, and Ruth E. Ames. 1979. "Experiments on the Provision of Public Goods. I. Resources, Interest, Group Size, and the Free-Rider Problem." *American Journal of Sociology*, 84(6): 1335-1360.

Marwell, Gerald, and Ruth E. Ames. 1980. "Experiments on the Provision of Public Goods. II. Provision Points, Stakes, Experience, and the Free-Rider Problem." *American Journal of Sociology*, 85(4): 926-937.

McCabe, Kevin A., Mary L. Rigdon, and Vernon L. Smith. 2003. "Positive Reciprocity and Intentions in Trust Games." *Journal of Economic Behavior & Organization*, 52(2): 267-275.

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

Munier, Bertrand, and Costin Zaharia. 2002. "High Stakes and Acceptance Behavior in Ultimatum Bargaining: A Contribution from an International Experiment." *Theory and Decision*, 53(3): 187-207.

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

Offerman, Theo. 2002. "Hurting Hurts More Than Helping Helps." *European Economic Review*, 46(8): 1423-1437.

Ostrom, Elinor, James Walker, and Roy Gardner. 1992. "Covenants With and Without a Sword: Self-Governance is Possible." *American Political Science Review*, 86(2): 404-417.

Post, Thierry, Martijn J. van den Assem, Guido Baltussen, and Richard H. Thaler. 2008. "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show." *American Economic Review*, 98(1): 38-71.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5): 1281-1302.

Rapoport, Anatol. 1988. "Experiments With N-Person Social Traps I: Prisoner's Dilemma, Weak Prisoner's Dilemma, Volunteer's Dilemma, and Largest Number." *Journal of Conflict Resolution*, 32(3): 457-472.

Raub, Werner, and Chris Snijders. 1997. "Gains, Losses, and Cooperation in Social Dilemmas and Collective Action: The Effects of Risk Preferences." *Journal of Mathematical Sociology*, 22(3): 263-302.

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

Sears, David O. 1986. "College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature." *Journal of Personality and Social Psychology*, 51(3): 515-530.

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

Simonson, Itamar, and Aimee Drolet. 2004. "Anchoring Effects on Consumers' Willingness-to-Pay and Willingness-to-Accept." *Journal of Consumer Research*, 31(3): 681-690.

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

Sobel, Joel. 2005. "Interdependent Preferences and Reciprocity." *Journal of Economic Literature*, 43(2): 392-436.

Telser, Lester G. 1995. "The Ultimatum Game and the Law of Demand." *Economic Journal*, 105(433): 1519-1523.

Tenorio, Rafael, and Timothy N. Cason. 2002. "To Spin or Not To Spin? Natural and Laboratory Experiments from 'The Price is Right'." *Economic Journal*, 112(476): 170-195.

van Lange, Paul A. M., and D. Michael Kuhlman. 1994. "Social Value Orientations and Impressions of Partner's Honesty and Intelligence: A Test of the Might Versus Morality Effect." *Journal of Personality and Social Psychology*, 67(1): 126-141.

van Lange, Paul A. M., Wilma Otten, Ellen M. N. de Bruin, and Jeffrey A. Joireman. 1997. "Development of Prosocial, Individualistic, and Competitive Orientations: Theory and Preliminary Evidence." *Journal of Personality and Social Psychology*, 73(4): 733-746.

West, Stuart A., Ashleigh S. Griffin, and Andy Gardner. 2007. "Evolutionary Explanations for Cooperation." *Current Biology*, 17(16): R661-R672.

Wooldridge, Jeffrey M. 2003. "Cluster-Sample Methods in Applied Econometrics." *American Economic Review*, 93(2), 133-138.