University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 82

Attraction and Cooperative Behavior

Donja Darai and Silvia Grätz

Revised version, June 2013

Attraction and Cooperative Behavior*

Donja Darai and Silvia Grätz†

June 26, 2013

Abstract

Being good-looking seems to generate substantial benefits in many social interactions, making the "beauty premium" a not to be underrated economic factor. This paper investigates how physical attractiveness enables people to generate these benefits in the case of cooperation, using field data from a modified one-shot prisoner's dilemma played in a high-stakes television game show. While attractive contestants are not more or less cooperative than less attractive ones, facial attractiveness produces more cooperative behavior by counterparts, but only in mixed-gender interactions. Effects of attractiveness are therefore not exclusively due to "beauty-is-good" stereotyping, but rather operate through a preference-based mechanism.

Keywords: beauty premium, gender, stereotypes, attractiveness, cooperation
JEL classification: C71, D03, D83

*The authors thank Daniel Hamermesh, Ulrich Kaiser, Ragan Petrie, Armin Schmutzler, Michelle Sovinsky, Roberto Weber, the seminar participants at the University of Zurich, Florida State University, and the University of Birmingham, the participants at the ESA International Meeting 2012 in New York (U.S.), the Workshop on Behavioral Economics Mainz (Germany), the Zurich Workshop on Economics 2012 in Luzern (Switzerland), and the ESA European Meeting 2012 in Cologne (Germany) for helpful discussions and suggestions, as well as Pascal Kappeler, Christina Richard, and Holger Scriba for excellent research assistance. Financial support of the Swiss National Science Foundation is gratefully acknowledged. The paper has been circulated under the title "Facing a Dilemma: Cooperative Behavior and Beauty". The data on "Golden Balls" were provided to the authors by the television show producers, courtesy of Endemol UK plc, in May 2009.

†Donja Darai (corresponding author), Department of Economics, University of Zurich, Blümlisalpstrasse 10, CH - 8006 Zurich, Switzerland, donja.darai@econ.uzh.ch and Silvia Grätz, Swiss Competition Commission, Monbijoustrasse 43, CH - 3003 Bern, Switzerland, silvia.graetz@weko.admin.ch.

"All that glitters is not gold;
Often have you heard that told."

William Shakespeare, Merchant of Venice (1596-1598)

1 Introduction

Beginning with the seminal paper by Biddle and Hamermesh (1994), which identifies a wage gap based on physical attractiveness using U.S. and Canadian labor market data, economists have been interested in the effects of physical attractiveness on economic decision making. For instance, using a controlled labor market experiment, Mobius and Rosenblat (2006) find a wage premium for attractive people. Eckel and Petrie (2011) provide evidence that people are willing to pay to see a picture of their counterpart before deciding whether to trust that person, suggesting that physical appearance conveys valuable information for deciding whether to trust. The most prominent explanation for these results is the theory of "beauty-is-good" stereotyping, which postulates that people assign a wide range of positive personality traits and abilities to physically attractive people. That is, people believe that physically attractive people are, for instance, more trustworthy, more likable, or more productive, see, e.g., Dion, Berscheid, and Walster (1972), and these beliefs then alter people's behavior toward attractive people. But, as with the cautionary phrase, *not all that glitters is gold*, there is only very limited evidence that the "beauty-is-good" stereotype is justified. Even though face-to-face interactions are very common in the real world, the mechanism underlying the effects of physical appearance on economic decision making is still not completely uncovered. How are attractive people generating the beauty premium they receive? Does attractiveness establish "unjustified" positive expectations about those people's behavior, kindness, and productivity? Or do people (un)consciously favor attractive people in social interactions due to reasons unrelated to their beliefs about attractive people?

To address the above questions, this paper studies the relationship between attractiveness and cooperative behavior, using field data from the British television game show "Golden Balls". This dataset is particularly interesting, because it allows to study people's behavior in a high-stakes environment involving communication and face-to-face interactions. Cooperation is analyzed in a social dilemma game. Further, the fact that people are on television, i.e., in a public environment, is especially useful for studying effects of appearance, since people are aware that they are observed by others and may therefore want to appear as "beautiful" as possible. The situation is, in this sense, comparable to job interviews, where appearance is considered to be important. We show that

people cooperate more with attractive others. But, importantly, physical attractiveness affects cooperative behavior only in mixed-gender interactions. People's behavior is biased toward cooperation only if facing someone attractive of the opposite sex, and not if facing someone attractive of the same gender. Also, attractive individuals are no more cooperative than those who are less attractive. Therefore, while providing evidence that beauty-is-good stereotyping is, indeed, a bias rather than an accurate expectation, we further provide evidence that the effects of attractiveness on economic decision making likely operate through a preference toward cooperation with others one finds personally attractive.

Two independent datasets are combined. First, to study cooperation, we focus on the end of the game show, where two finalists play for an accumulated jackpot. The rules of this game follow a modified version of a simultaneous one-shot prisoner's dilemma, in which defection is a weakly dominant strategy: if both contestants choose to cooperate, the jackpot is split equally; if one chooses to defect while the other cooperates, the former receives the entire jackpot and the latter gets nothing; if both defect, they both go home empty-handed. The accumulation of the jackpot takes place in two rounds of play prior to the prisoner's dilemma and requires neither the contestants' effort nor cognitive ability. On average, the stake size is £12 912 indicating that decisions in the prisoner's dilemma have substantial distributional consequences for the contestants. Second, to study attractiveness, we collect data on perceptions of the contestants' facial appearance. A sample of independent raters evaluates, among other attributes, the physical attractiveness of the contestants by showing them portrait photographs.

Our analysis shows that contestants are significantly more cooperative toward a facially attractive opponent than toward a facially unattractive opponent, independent of e.g., gender, age, stake size, or communication. Hence, facially attractive contestants are awarded with significantly higher earnings in the prisoner's dilemma. Quantifying this beauty premium, a one-standard deviation increase in facial attractiveness, at the margin, causes the contestant's expected earnings to rise by, on average, £2 153, with a probability of obtaining positive earnings of, on average, 5.9 percentage points. This effect is not driven by non-cooperative behavior of those who are attractive. With minor qualifications for younger and female contestants, we find no correlation between a contestant's own facial attractiveness and the likelihood to cooperate. But effects of attractiveness are limited to mixed-gender interactions. Contestants are only biased toward cooperation if their opponent is attractive and of the opposite sex. This finding applies similarly for men and women and offers a new perspective on the underlying mechanism motivating differential behavior toward attractive and unattractive people. The beauty premium is

not driven by people believing that attractive others are more likely to cooperate, but rather by people having a preference to cooperate more with someone toward whom they are personally attracted. Furthermore, there are no effects of attractiveness during play of the earlier stages of the game show, when there are more than two participants. This is in contrast to Belot, Bhaskar, and van de Ven (2012) and suggests that personal attraction in pairwise interactions constitutes a major source of the effect of attractiveness in economic interactions.

To increase the reliability of our result on attractiveness and gender composition, we deliver supportive evidence using the experimental labor market data of Mobius and Rosenblat (2006). In a gift exchange game with a real effort task they show that effects of attractiveness operate through higher confidence and better oral skills of attractive workers and employers' false beliefs about the attractive's ability. However, when parsing their data by the composition of gender between interacting parties, their identified effects of attractiveness disappear for same-sex interactions. The beauty premium of attractive workers is only significant in interactions between employers and workers of the opposite sex. Thus, in addition to the identified channels by Mobius and Rosenblat (2006), beauty influences people's behavior primarily when they are interacting with someone toward whom they feel personally attracted. The result that effects of attractiveness might be entangled with gender matches with studies by Bertrand, Karlan, Mullainathan, Shafir, and Zinman (2010), Landry, Lange, List, Price, and Rupp (2006), and Dreber, Gerdes, and Gränsmark (2013). They show that men change their behavior toward attractive women. Specifically, they find that men's demand for consumer credits increases if a picture of an attractive female is part of the advertisement, that men's contribution to a charity increases when they face an attractive female solicitor, or that men choose riskier strategies when their opponent is an attractive women in expert chess games. However, they provide no evidence that women may also change their behavior toward attractive males. Therefore, the present study not only provides evidence in support of the beauty premium as an important economic factor, but rather that effects of attractiveness on economic decision making should be viewed in light of the "gender-composition", in which people interact.

The remainder of this paper is structured as follows. In Section 2 related literature is discussed. Section 3 describes the data sets and Section 4 presents the results. In Section 5 potential transmission channels are evaluated, followed by Section 6 that provides evidence in support of our identified mechanism underlying effects of attractiveness. Finally, Section 7 concludes.

2 Literature

A long history of research on physical attractiveness in psychology, sociology, and evolutionary biology and a more recent one in economics and political science shows that physically attractive people are or behave not substantially differently from physically unattractive people, but that physically attractive people receive a preferential treatment by others in many regards. This "beauty premium" rewards attractive people with substantially higher monetary gains or greater economic success. Surprisingly, the mechanism underlying the effect of physical attractiveness is still not completely understood.

The most prominent theory, established in sociology and psychology, is that people link beauty and positive personality traits. In particular, people attribute a variety of positive characteristics and higher abilities to physically attractive people, and negative ones to physically unattractive people, see reviews by Eagly, Ashmore, Makhijani, and Longo (1991), Feingold (1992), and Langlois, Klakanis, Rubenstein, Larson, Hallam, and Smoot (2000). This is referred to as *"beauty-is-good" stereotyping* in the literature.[1] The theory argues that stereotype beliefs cause people to treat attractive ones more favorably, and this in turn results in higher economic gains for the attractive. Hence, beauty-is-good stereotyping can explain why attractive people generally fare better in the labor market, i.e., why they are more likely to be hired, promoted, and earn higher salaries. In their seminal work Biddle and Hamermesh (1998) use a broad household survey of the U.S. and Canada and show that physically attractive employees earn about 10-15 percent more than less attractive employees, independent of their occupation. In an experiment on the effects of beauty in labor markets, Mobius and Rosenblat (2006) find that attractive employees are awarded higher wages, although they are not more productive than unattractive employees, measured by a real-effort task. This suggests that employers hold (inaccurate) stereotypical expectations about the performance of physically attractive employees. Ruffle and Shtudiner (2011) explore the value of beauty in the hiring process, examining response rates to CVs sent to companies in Israel. They uncover a gender-specific effect of attractiveness, namely a beauty premium for males, and a beauty penalty for females.

Also, research in political science shows that physical attractiveness has a significant impact on the evaluation of candidates and thereby on electoral outcomes, see the survey by Ottati and Deiger (2002). For instance, attractiveness increases the candidate's

[1] In their pioneering study, Dion, Berscheid, and Walster (1972) claim "what is beautiful is good" by demonstrating that attractive people are believed to have better career prospects, to possess socially desirable traits, to lead happier lives and to be happier overall. This paper spawned a large literature on the physical attractiveness stereotype, demonstrating for both sexes a robust association between physical attractiveness and cognitive ability, competence, sociability, popularity, dominance, sexual experience, mental health, and social skills.

number of votes and thereby the likelihood of being elected, see Berggren, Jordahl, and Poutvaara (2010) and Rosar, Klein, and Beckers (2008). Antonakis and Dalgas (2009) suggest that underlying the effect of physical attractiveness is the voters' belief that attractive candidates are more competent. Recruiting adults and children raters in Switzerland, they find that even children can predict results of the 2002 French parliamentary election retrospectively by rating the competence of the candidates from their photographs.

Furthermore, the theory of beauty-is-good stereotyping in combination with interdependent social preferences can explain findings in various experimental settings with strategic interactions. These findings suggest that people behave more generously or cooperatively toward physically attractive people. Solnick and Schweitzer (1999) show that attractive responders receive significantly higher offers than unattractive responders in the ultimatum game. In dictator games, Rosenblat (2008) shows that allocators treat physically and vocally attractive recipients more generously. Andreoni and Petrie (2008) observe a beauty premium in a repeated public goods game, which is not caused by lower contributions of attractive players, but can be attributed to an increase of the other players' contributions triggered by an attractive group member. In a repeated trust game, Wilson and Eckel (2006) demonstrate that attractive trustees are trusted more and players expect attractive players to be more trustworthy than unattractive ones. The failure to meet these expectations leads to a beauty penalty.[2] Mulford, Orbell, Shatto, and Stockard (1998) examine a repeated prisoner's dilemma, pointing out that players are more cooperative if matched with an attractive partner. In contrast to the other studies, they use attractiveness ratings based on the perception of the player who makes the decision in the experiment, rather than ratings elicited from independent third-party judges.

All above mentioned studies, however, provide only very limited evidence for the accuracy of the beauty-is-good stereotype. Physically attractive people themselves do not behave differently than physically unattractive people.[3]

Besides the theory of stereotyping based upon peoples' visual attractiveness, Zuckerman and Driver (1989) and Zuckerman, Hodgins, and Miyake (1990) proclaim the existence of a *vocal attractiveness stereotype* by showing that physical attractiveness is posi-

[2]In another experimental study on the trust game, Eckel and Petrie (2011) investigate the informational value of a photograph and the differential desire to acquire this information. Subjects are willing to pay to see the photograph of their partner, whom they transact with, indicating that a face has a positive informational value which is used to discriminate between players in their choices.

[3]Jackson, Hunter, and Hodge (1995) show in their meta-analysis that there is a modest correlation between attractiveness and intelligence for children. Mueller and Mazur (1997) use data from a cohort of military officers and find that recruits with a high ranked facial appearance are also high ranked in their physical fitness. Concerning social skillfulness and likability, Goldman and Lewis (1977) and Erwin and Calev (1984) find that physically attractive people indeed possess better social skills and are more likable.

tively correlated with vocal attractiveness. Supporting a vocal-attractiveness mechanism, Mobius and Rosenblat (2006) find that the beauty premium even exists in treatments in which employers are able to talk to the employee without seeing the employee's picture. Rosenblat (2008) provides evidence that physically attractive people only achieve significantly better outcomes in dictator games when the dictator sees the recipient's picture in addition to hear her pre-recorded voice. This suggests voice to be one of the transmission channels of the beauty premium, at least in two-sided communication settings.

Finally, one of the oldest and most well-known theories underlying the effects of attractiveness is the one of *taste-based discrimination*, proposed by Becker (1957). Physically attractive people are favored because people enjoy being or working with them more than with plain looking people. Discrimination is based upon prejudices correlated with people's personal characteristics and is rational in the sense that interactions with such a person generate a (dis)utility for the discriminator in case of positive (negative) discrimination. Belot, Bhaskar, and van de Ven (2012) find that attractive contestants of a Dutch television game show are positively discriminated against unattractive ones in proceeding to the final stage of the show.[4] Castillo, Petrie, and Torero (2011) show in a public good experiment that people discriminate by appearance when forming groups, even if it is costly for them, i.e., they neglect payoff relevant information and thereby receive lower average payoffs. However the theory of taste-based discrimination fails to explain why attractive people are also treated more favorably in one-shot interactions.

3 Data

3.1 The Television Game Show "Golden Balls"

We analyze people's behavior in a simultaneous one-shot prisoner's dilemma of 211 episodes of the British television game show "Golden Balls".[5] The dilemma is played in the final round of the show to allocate the jackpot between the two finalists: each player is assigned two balls, indistinguishable from the outside, but one contains the word "steal" and one contains the word "split". Both contestants choose one of their two balls and then open it

[4] In particular attractive contestants are much more likely to reach the final round of the show, even though they are not performing better or are not more confident than unattractive contestants. Performance in this show means being the first to correctly answer trivia questions. Attractive contestants are believed to be more confident and to be more cooperative. Besides, at the end of the show a prisoner's dilemma – like the one we study – is played. Belot, Bhaskar, and van de Ven (2012) also test for effects of attractiveness on cooperative behavior, however they find no significant effects. We suspect that the reason why they are not able to identify any attractiveness effects is that their sample of finalists becomes too homogeneous with regard to attractiveness, due to the selection bias toward attractive contestants during the pre-play.

[5] The game show was aired in June 2007 and ended in December 2009. We use records of 211 episodes of the first four series of the show; see also the Supplementary Material.

simultaneously. If both choose the split-ball (cooperation), the jackpot is split equally; if one chooses the steal-ball (defection) while the other chooses the split-ball (cooperation), the former receives the entire jackpot and the latter gets nothing; if both choose the steal-ball (defection), they both go home empty-handed (see Figure I). Thus, defection is a weakly-dominant strategy.

Figure I: Prisoner's Dilemma Game

	Split (cooperate)	**Steal** (defect)
Split (cooperate)	$^1/_2$ jackpot , $^1/_2$ jackpot	0 , jackpot
Steal (defect)	jackpot , 0	0 , 0

Before the prisoner's dilemma is played, the contestants have to pass two rounds of pre-play. In these rounds, the finalists are selected and the jackpot is accrued. Accumulating the jackpot does not require contestant's cognitive ability or effort.[6] On average, the jackpot amounts to £12 912 and ranges from a minimum of £3 to a maximum of £93 250. The two finalists are selected through two voting decisions. After each round, the contestants need to cast a vote against one of them to leave the show. The contestant who receives the majority of votes is eliminated. Throughout the show the contestants face each other and are allowed to freely communicate with each other (see the Supplementary Material for a detailed description of the show).

Table A.1 in the Appendix presents summary statistics of the outcomes of the prisoner's dilemma and of the contestants' personal characteristics. The unilateral cooperation rate is 54%, and contestants mutually cooperate (defect) in 32% (25%) of cases. 46% of finalists are male and the mean age is 37 years. For a detailed analysis of the game show with respect to stake size, communication, play prior to the prisoner's dilemma, and commitment see Darai and Grätz (2010).

3.2 Survey on Facial Appearance

We evaluate the contestants' facial appearance using a panel of independent raters. The raters are recruited at the Euro-Airport Basel, at the University of Zurich, and at the University of Zurich for the elderly.[7] All 844 contestants are judged by 728 raters and,

[6]Specifically, in each pre-play round all contestants are randomly assigned a certain cash value and, for instance, do not have to answer trivia questions.

[7]The University of Zurich provides lectures for a senior audience, that are mainly attended by retired people. We only recruited raters who either were native speakers or had a very high proficiency of

from those, 365 raters judged the 422 finalists. Each rater was asked to individually rate the facial appearance of five randomly assigned contestants, of which two or three were male (female). On average, a finalist is judged by 4.3 raters. Table I reports summary

Table I: Summary Statistics of the Finalists' Raters

Rater's variable	**Mean**	**Std. dev.**	**Min.**	**Max.**	**N**
Male	0.5	0.5	0	1	365
Age (in years)	40.88	15.54	17	93	361
Age of male (in years)	41.58	15.60	17	93	183
Age of female (in years)	40.15	15.50	18	86	178
Age of female ($\geq$ 40.15 years)	53.51	10.36	41	86	86
Age of female ($<$ 40.15 years)	27.66	6.38	18	40	92
Age of male ($\geq$ 41.58 years)	55.21	11.42	42	93	85
Age of male ($<$ 41.58 years)	29.76	6.28	17	41	98

statistics for the finalists' raters and Table A.2 in the Appendix for all contestants' raters. The mean age of the 365 raters is 41 years and 50% are male.

The survey is questionnaire based. Each questionnaire contains two portrait photographs of the same contestant and is divided into three parts. To receive non-biased evaluations of the contestants and to reduce measurement error, all photographs are selected from the same two sequences of the game show such that one photograph shows a neutral facial expression with a view to the camera and the other a neutral facial expression with a view to the side of the camera.[8]

In the first part of the questionnaire, the raters are asked to judge the contestant with respect to her age by fitting the contestant into one of seven age categories, "<20", "20-30", "30-40", "40-50", "50-60", "60-70", or ">70" years. The second part includes assessments of the contestant's appearance using four opposite word pairs, i.e., "attractive - unattractive", "likable - unlikable", "trustworthy - untrustworthy", and "honest - dishonest". These items are rated on a 1-to-7 point Likert scale, where 1 equals very unattractive, 4 comprises a neutral position and 7 equals very attractive.[9] In the last part of the questionnaire we asked the raters to give a binary response (yes/no) to two statements: "this person's appearance helps him/her in life" and "this person strikes me as calculating". These statements are included to capture a rater's overall impression of the respective contestant's appearance, i.e., whether a contestant is either attributed a beauty premium or a strategic

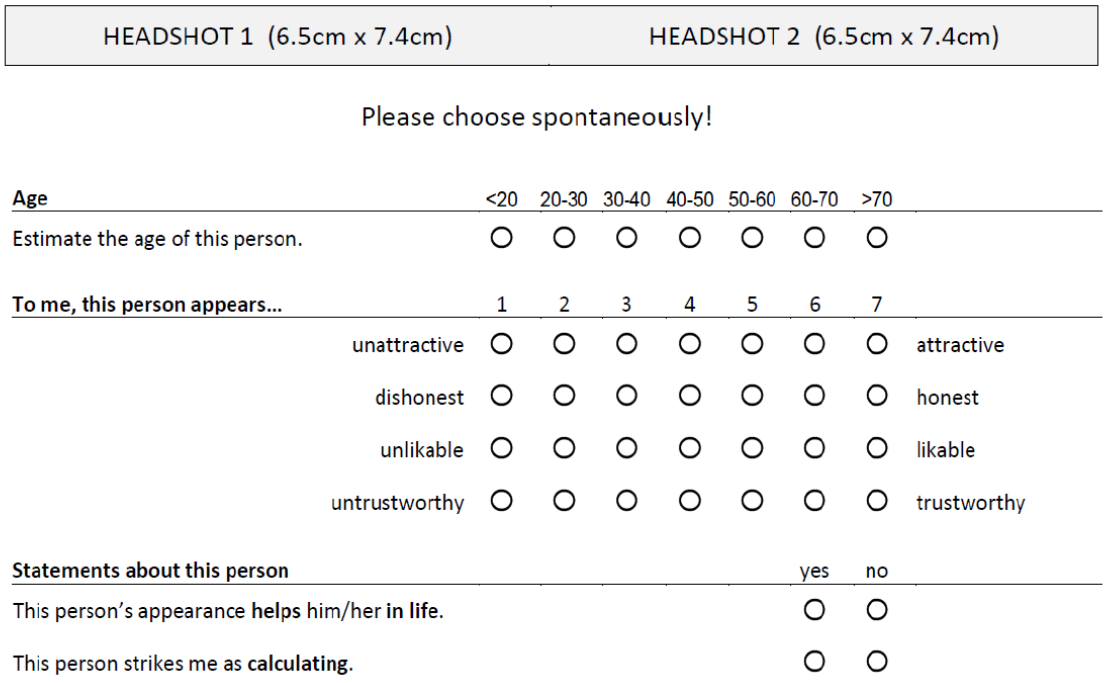
German, French, or English and provided the questionnaires in the respective language.

[8]If possible we chose a neutral facial expression of the contestant, otherwise a positive one was chosen, but never a negative or disadvantageous one.

[9]The use of a 7-point Likert scale, which includes a neutral element, allows for sufficient diversification and is standard in the relevant literature, see, e.g., Alreck and Settle (1995), pp.113-114.

intention per se. For an illustration of a sample questionnaire see Figure II.

Figure II: Sample Questionnaire



3.3 Evaluation of Facial Appearance Ratings

The raters' evaluations are used to construct a facial appearance measure for each of the four items: attractiveness, honesty, likability, and trustworthiness. We account for rater specific variation in the perception of the respective appearance item in the construction of each measure.[10] For each rater j we calculate the average across all contestants rated by j, $\bar{x}_j$. Then we mean-center the ratings by subtracting the respective rater's mean rating, $\bar{x}_j$, from her individual rating of contestant i, x_{ij}. These demeaned ratings are now anchored at 0 for each rater and are therefore corrected for rater-specific fixed effects

[10]Raters may differ in the interpretation of the 7-point scale, and may vary in anchoring their average rating above, below or close to 4. Student t-tests of differences between the means of the facial appearance measure by the raters' gender are: $t = 2.799$, $p = 0.005$ for attractiveness, $t = 4.746$, $p = 0.000$ for likability, $t = 4.913$, $p = 0.000$ for trustworthiness, and $t = 4.591$, $p = 0.000$ for honesty. This shows that female raters tend to give higher ratings than male raters. This phenomenon is also persistent with respect to the raters' age. Further, it is known that images of females tend to be rated higher than images of males, see, e.g., Jackson, Hunter, and Hodge (1995) and Wilson and Eckel (2006). Since we control directly for gender-related effects in the regressions, the attractiveness measure can be used uncorrected for the gender of the contestants. As a robustness check we construct a measure by making use of the rater-specific fixed effects, rather than correcting for them. In particular, we attach higher weight to ratings made by raters that are similar to the contestant with respect to age and gender, and we attach lower weight to ratings made by raters that are dissimilar to the contestant. All results remain unchanged.

(see Figure A.1 in the Appendix). Finally, we take the mean of all mean-centered ratings of contestant i, resulting in the particular facial-item measure for each contestant. To account for biased ratings with respect to gender and age, we include contestants' gender and age in the regression analysis.

$$[\text{Facial-item}]_i = E_j[x_{ij} - \bar{x}_j] \quad \text{with} \quad \begin{cases} i = & \text{i-th contestant} \\ j = & \text{j-th rater} \end{cases}$$

Additionally, we construct two variables from the binary statements "appearance helps in life" and "appearance strikes as calculating" by taking the mean rating for each contestant.[11] In the following we refer to these variables as "statement variables". Table II provides summary statistics of all appearance measures.

Table II: Summary Statistics Facial Appearance

Variable	**Mean**	**Std. dev.**	**Min.**	**Max.**	**N**
Attractiveness (mean-centered, cont.)	0	0.79	-1.8	2.65	422
Honesty (mean-centered, cont.)	0	0.6	-1.8	1.45	422
Likability (mean-centered, cont.)	0	0.66	-1.95	1.65	422
Trustworthiness (mean-centered, cont.)	0	0.6	-1.75	1.65	422
Appearance helps in life (cont.)	0.58	0.30	0	1	422
Appearance strikes as calculating (cont.)	0.35	0.25	0	1	422

All four facial appearance variables are highly positively correlated (see Table III). This is also reflected in a sufficiently high Cronbach coefficient alpha ($\alpha = 0.82$).[12] The statement variable "appearance helps in life" is also positively correlated with all four facial appearance variables. However, the statement variable "appearance strikes as calculating" is not correlated with attractiveness, and even negatively correlated with the remaining facial appearance variables. In the following, we focus on facial attractiveness, since it is a crucial part of the first impression of a person and a stable characteristic which is almost impossible to mimic, see, e.g., Grammer, Fink, Møller, and Thornhill (2003). The two statement variables are used as additional controls.

There are many ways to define an attractive person. We use the following classifi-

[11] Again, there may be differences between the means of the statement variables' ratings by the raters' gender. Student t-tests w.r.t. gender are $t = 0.851$, $p = 0.395$ for "appearance helps in life" and $t = -5.437$, $p = 0.000$ for "appearance strikes as calculating", i.e., male raters are more likely to rate a contestant as calculating than female raters.

[12] We use Cronbach's alpha for standardized variables to measure the inter-item reliability for facial appearance. The measure adjusts for item specific mean and variance. Also a nonparametric test for testing whether samples originate from the same distribution cannot be rejected (Kruskal-Wallis $K = 0.9868$).

Table III: Correlation Matrix for Facial Appearance Variables, N=422

	Attractiveness	Honesty	Likability	Trustworthiness	Helps in life
Attractiveness[a]	1				
Honesty	0.286***	1			
Likability	0.461***	0.651***	1		
Trustworthiness	0.340***	0.760***	0.680***	1	
Helps in life[b]	0.643***	0.229***	0.389***	0.299***	1
Strikes as calculating	0.007	−0.313***	−0.288***	−0.315***	0.034

[a] Mean-centered, continuous variables (for attractiveness, honesty, likability, trustworthiness).
[b] Averaged, continuous variables (for "appearance helps in life", "appearance strikes as calculating").

cations: first, we classify a contestant as *facially attractive* if her facial attractiveness rating lies above or is equal to the mean over all facial attractiveness ratings, and as *facially unattractive* if her facial attractiveness rating lies below this mean. The average mean-centered rating of facially attractive contestants is 0.599, and the one of facially unattractive contestants is −0.650. We also used the median over all facial attractiveness ratings as a classification device. The average mean-centered rating of a facially attractive contestant, who is rated above or equal to (below) the median is 0.613 (−0.635). Second, we define extreme measures of facial attractiveness. A contestant is classified as *most attractive* if her facial attractiveness rating lies within the top 10% (25%) percentile of the distribution of facial attractiveness and as *least attractive* if her facial attractiveness rating lies within the bottom 10% (25%) percentile. Contestants who are rated as most attractive receive, on average, a mean-centered rating of 1.389 (0.993), and those who are rated as least attractive receive a mean-centered rating of −1.386 (−1.003). For an illustration of the distribution of facial attractiveness see Figure A.1 in the Appendix. Third, we define a contestant's *appearance to be helpful in life* (*to strike as calculating*) if her "helps in life"-rating ("strikes as calculating"-rating) lies above or is equal to the mean over all "helps in life"-ratings ("strikes as calculating"-ratings), and as *not to be helpful in life* (*not to strike as calculating*) if her "helps in life"-rating ("strikes as calculating"-rating) lies below this mean. Table IV summarizes the binary attractiveness and statement variables. A detailed description of the distribution of attractive and unattractive finalists with respect to gender and age is provided by Table A.3 in the Appendix.

Table IV: Summary statistics facial attractiveness

Variable	Mean	Std. dev.	Min.	Max.	N
Attractiveness (mean, d)	0.52	0.50	0	1	422
Attractiveness (median, d)	0.51	0.50	0	1	422
Most attractive (90% percentile, d)	0.11	0.31	0	1	422
Most attractive (75% percentile, d)	0.25	0.44	0	1	422
Least attractive (10% percentile, d)	0.10	0.30	0	1	422
Least attractive (25% percentile, d)	0.25	0.43	0	1	422
Helps in life (d)	0.50	0.50	0	1	422
Strikes as calculating (d)	0.47	0.50	0	1	422

(d) for dummy variable.

4 Results

4.1 Facial Attractiveness

In order to investigate the relationship between cooperative behavior and facial attractiveness we use several binary probit models with the decision to cooperate as the dependent variable (with $y_i = 1$ equal cooperate; $y_i = 0$ equal defect). Throughout the analysis, we control for effects and interactions related to the contestant's gender and age, other demographic characteristics, as well as variables of stake size, communication, and variables describing the course of events of the game show previous to the prisoner's dilemma (pre-play), see Table A.1 in the Appendix.[13] Further, we use a range of different classifications of attractiveness, for instance, the binary and continuous measures of attractiveness, a standardized attractiveness measure in line with Mobius and Rosenblat (2006), in which our constructed attractiveness measure is normalized across all contestants, and a standardized attractiveness measure in line with Biddle and Hamermesh (1998), in which the normalization is across all raters. All measures produce qualitatively the same results.[14]

[13]Besides effects of attractiveness on cooperation, we extensively analyzed the impact of the control variables, in particular, demographics, stake size, communication, and pre-play on cooperative behavior, see Darai and Grätz (2010), and also van den Assem, van Dolder, and Thaler (2012).

[14]Table V reports the regression results including a dummy variable for the attractive contestant. For robustness of all our results, we also estimate the regressions including (i) the mean over the four appearance variables, (ii) the predicted factors obtained in a confirmatory factor analysis of the four appearance variables, and (iii) the mean and median attractiveness ratings of the raw data, where the mean (median) is 0.443 (0.728) with an average rating of facially attractive contestants above the mean (median) of 5.552 (4.914), and below the mean (median) of 3.196 (2.357). Again, all measures produce qualitatively the same results.

Own Attractiveness The results depicted in Table V, model (1) to (5), show that facially attractive contestants do not behave differently with respect to cooperativeness than facially unattractive contestants, independent of the specification of the attractiveness measure. This finding is in line with our conjecture from the literature, see, e.g., Eagly, Ashmore, Makhijani, and Longo (1991): physically attractive people are not more pro-social and therefore not more cooperative than physically unattractive people.

Whereas, overall, we find no difference between the cooperative behavior of attractive and unattractive people, there are some qualifications with respect to gender and age.[15] Both gender and age seem to mediate the effect of a contestant's own attractiveness on cooperation, see Table A.5 in the Appendix. Attractive females (model (1)) and attractive younger contestants (model (3)) show more cooperative behavior, whereas attractive males (model (1)) and attractive older contestants (model (3)) cooperate less.

Opponent's Attractiveness We now turn to the impact of the opponent's facial attractiveness on a contestant's willingness to cooperate. The regression results in Table V show that contestants are $10 - 16$ percentage points more likely to cooperate when facing an attractive opponent than when facing an unattractive opponent. Hence, attractive contestants are rewarded with greater cooperativeness, and this provides attractive contestants a beauty premium. The premium is independent of the opponent's gender and age, see Table A.5, model (2) and (4) in the Appendix. Furthermore, our results show that least attractive contestants suffer a beauty penalty due to lower cooperativeness toward them. As model (5) reports, a contestant is less likely to cooperate if the opponent is rated to be least attractive than if the opponent is neither rated to be most nor least attractive. There is no significant effect on cooperative behavior if the opponent is rated to be most attractive, independently of the definition of the extreme measure of attractiveness. These findings suggest that contestants rather focus on the opponent's "negative" than "positive" appearance in the decision to cooperate. The results remain unchanged when adding the two binary statement variables (see model (4) of Table V). There is additional evidence that a contestants is less likely to cooperate if the opponent's appearance is rated as to "help her in life" than if it is not. But we find no interaction effect of the binary statement variables and our attractiveness measures, which indicates

[15] Irrespective of a contestant's own attractiveness, we find a very strong and significant correlation between age and cooperative behavior (see Table V): older contestants (≥ 37 years) are much more likely to cooperate than younger contestants (< 37 years), regardless of the age of the opponent. This result is in line with List (2006) who finds that contestants ≥ 31 years are significantly more likely to cooperate than younger contestants. Concerning gender, we find no direct effect, which is contrary to the studies of e.g., Kahn, Hottes, and Davis (1971) and Ortmann and Tichy (1999). But we can show that younger males are significantly more likely to defect; and, as age increases, males are more likely to cooperate than females (see Table A.4 in the Appendix).

Table V: Results from Binary Probit Regressions of the Decision to "Cooperate" ($y_i = 1$) or "Defect" ($y_i = 0$) in the Prisoner's Dilemma

	Marginal effects									
	Model (1)		Model (2)		Model (3)		Model (4)		Model (5)	
Attractiveness (mean-centered, d) (d)	0.067	(0.049)	0.067	(0.052)	0.048	(0.053)	0.026	(0.060)		
Opponent attractiveness (mean-centered, d) (d)	0.112**	(0.046)	0.125**	(0.049)	0.101**	(0.050)	0.162***	(0.057)		
Appearance helps in life (d)							0.062	(0.066)	0.075	(0.061)
Opponent appearance helps in life (d)							-0.110*	(0.064)	-0.050	(0.060)
Appearance strikes as calculating (d)							0.017	(0.056)	0.012	(0.055)
Opponent appearance strikes as calculating (d)							-0.001	(0.056)	0.000	(0.056)
Most attractive (90% percentile) (d)									-0.134	(0.086)
Opp. most attractive (90% percentile) (d)									-0.096	(0.085)
Least attractive (10% percentile) (d)									-0.077	(0.090)
Opp. least attractive (10% percentile) (d)									-0.230***	(0.081)
Male (d)	-0.036	(0.048)	-0.046	(0.050)	-0.058	(0.051)	-0.052	(0.056)	-0.060	(0.051)
Age (cont.)	0.081***	(0.023)	0.082***	(0.025)	0.088***	(0.027)	0.093***	(0.030)	0.085***	(0.028)
Opp. male (d)							-0.008	(0.056)		
Opp. age (cont.)							0.016	(0.028)		
Promise or vow (d)					0.240***	(0.051)	0.254***	(0.052)	0.258***	(0.052)
Opp. promise or vow (d)					0.006	(0.054)	-0.002	(0.054)	-0.013	(0.054)
Handshake (d)					-0.102*	(0.059)	-0.098	(0.060)	-0.117**	(0.059)
Demographics	no		yes		yes		yes		yes	
Stake size	no		yes		yes		yes		yes	
Pre-play	no		no		yes		yes		yes	
Wald χ^2	18.59***		43.01***		63.96***		69.21***		68.21***	
Log-Likelihood	-282.75		-264.39		-250.74		-248.53		-246.79	
Adjusted R^2	0.013		0.039		0.055		0.042		0.048	
N	422		419		419		419		419	
Number of clusters	211		211		211		211		211	

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

(d) for discrete change of dummy variable from 0 to 1

that the statement variables have not much additional explanatory power.[16]

Similarity Since contestants behave more cooperatively toward the attractive counterpart, the question arises whether pairs of attractive contestants behave differently in the prisoner's dilemma than pairs of unattractive contestants or pairs who are mixed in terms of facial attractiveness. We denote pairs of contestants as *similar team* if either both contestants are facially attractive or both are facially unattractive, and we identify pairs of contestants as an *attractive team* if both are facially attractive, and as an *unattractive team* if both are facially unattractive. We find no evidence that contestants who are similar with respect to attractiveness are more or less likely to cooperate or defect, see model (1) of Table A.6 in the Appendix. However, we find that an unattractive contestant is less likely to cooperate with her unattractive counterpart compared to teams of contestants who are both attractive or mixed, see model (2) and (3) of Table A.6 in the Appendix. This improves our attractiveness-results: contestants not only behave more cooperatively toward an attractive partner, but also more deceitfully toward an unattractive partner, and this unattractive-penalty is likely to dominate.[17]

Thus, our results provide evidence for a causal relationship between the opponent's attractiveness and cooperative behavior. Facially attractive contestants are able to provoke cooperation from their counterpart, independent of their gender or age. But we do not find a significant difference in behavior between facially attractive and facially unattractive contestants.

4.2 Beauty Premium

The results of the previous section should also translate into a monetary beauty premium, i.e., into higher earnings for the attractive than for the unattractive. In order to quantify the marginal beauty premium we use a standard censored tobit model. The outcome "taking no money home" from the prisoner's dilemma is interpreted as a corner solution

[16]Note, the effect of the variable "the contestant's appearance helps her in life" without the inclusion of the facially attractiveness variables is positive, but not significant. We also tested for interaction effects between the two statement variables and gender or age (table unreported). We find that a female opponent, whose appearance is rated as helping her in life, lowers the contestant's propensity to cooperate; but this effect is not robust with respect to different specifications of the model.

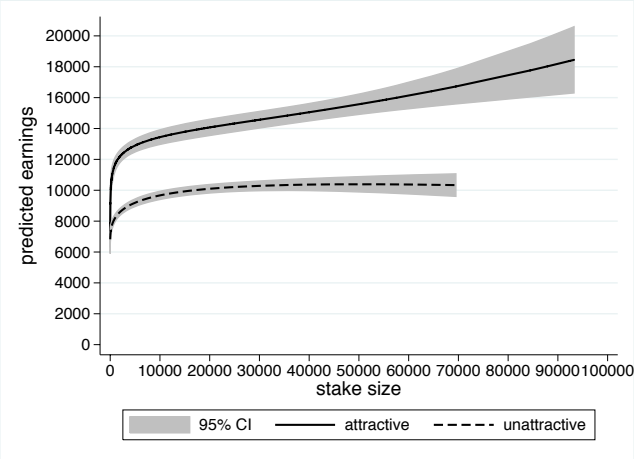
[17]For robustness of our results, we repeat the analysis using ordered probits on the contestants' likelihood to mutually cooperate. Again, similar teams, as well as teams of attractive contestants are no more likely to reach a certain outcome. Only pairs of unattractive contestants are significantly less likely to reach mutual cooperation (tables unreported). Additionally, we estimate regressions including the relative difference between both final contestants' attractiveness, i.e., the distance in attractiveness between both contestants, and including an index of the contestants' similarity with respect to facial attractiveness, age, and gender, weighting each component by one-third. All measures do not matter for the contestant's decision to cooperate.

outcome, where the response variable y_i describes the observable outcome of a contestant, which takes on the value zero with positive probability (if the opponent defects), and which is a continuous variable over strictly positive values (if the opponent cooperates).

We estimate marginal effects for contestant i's expectation of receiving positive earnings $y_i > 0$, conditional on i's own and her opponent's facial attractiveness (standardized measures), demographic characteristics, the log value of the stake size, as well as opponent characteristics, which are summarized in $\mathbf{X_i}$, $E(y_i|\mathbf{X_i}, y_i > 0)$.

We find that a one standard deviation increase in facial attractiveness significantly increases a contestant's expected positive earnings by, on average, £617 to £741 depending on the controls included, see Table A.7 in the Appendix. Figure III provides an illustration of the beauty premium based on model (2) of Table A.7 in the Appendix, and depicts the predicted positive earnings for attractive and unattractive contestants as a function of stake size. The figure shows that the predicted earnings of facially attractive contestants are always clearly above the earnings of unattractive contestants, given the size of the jackpot.[18]

Figure III: Predicted Positive Earnings by Attractive and Unattractive Contestants, Given Stake Size



Decomposing the beauty premium with respect to gender and age, we can show that there are no differences in receiving positive earnings between attractive male and female, as well as attractive older and younger contestants. However, a contestant's expected earnings increase the older the opponent, and a contestants expected earnings decrease when the opponent is male (see model (2) of Table A.7 in the Appendix).

[18]We obtain a similar result for the overall conditional expectation that a contestant receives the outcome y_i, given $\mathbf{X_i}$, $E(y_i|\mathbf{X_i})$. A one standard deviation increase in facial attractiveness significantly increases a contestant's expected earnings by, on average, £1 784 to £2 153. Further, we estimate that the probability of obtaining positive earnings, given X_i, $Pr(y_i > 0)|\mathbf{X_i})$, increases by, on average, 5.9 percentage points (tables unreported).

5 Transmission Channels

Our results show that attractive people are able to produce more cooperative behavior from their counterpart and, since they are not more or less cooperative than unattractive people, they obtain a beauty premium in the prisoner's dilemma. In this section we will address the potential transmission channels suggested by the literature (see Section 2) and evaluate their explanatory power for our observed effects of attractiveness.

5.1 Beauty-is-Good Stereotyping and Taste-Based Discrimination

According to the most prominent theory called "beauty-is-good" stereotyping, effects of attractiveness on behavior are caused by stereotype believes about attractive people. People are said to believe that attractive ones behave more pro-socially than less attractive ones. In the presence of interdependent social preferences, people holding this belief may then behave more cooperatively toward attractive people with the intention to reciprocate cooperation. Hence, these stereotype beliefs can mediate people's behavior. Recall, that we observe a higher likelihood of cooperation toward attractive contestants in the prisoner's dilemma. If the theory of "beauty-is-good" stereotyping is driving our results, this effect should be independent of the selected sample. However, theories from evolutionary psychology argue that effects of physical attractiveness on behavior originate in primeval partner selection and therefore predict the effects to be more prevalent in mixed-gender interactions, see, e.g., Cosmides and Tooby (1987). To scrutinize this argument, we parse the data in mixed-gender and same-gender interactions (63% and 37% of all interactions). The two subsamples are not different regarding the observed cooperation rate (54% in mixed- vs. 53% in same-gender interactions). Within the two subsamples, we run several probit regressions to evaluate the influence of facial attractiveness on a contestant's propensity to cooperate, controlling for demographic characteristics and stake size.

As the regression results of Table VI show, attractiveness only matters in mixed-gender interactions, but not in same-gender ones. Males and females are about 15 percentage points more likely to cooperate if facing a facially attractive opponent of the other sex than if they face an unattractive opponent.[19] The finding applies across sexes, both females and males are biased toward attractiveness if facing someone of the other sex.[20] In same-

[19] This result is also displayed in the fact that facially attractive contestants enjoy a beauty premium in mixed-gender interactions, but not in same-gender ones (tobit regression tables unreported). A one standard deviation increase in attractiveness results in a significantly higher gains for the attractive in mixed-gender (p=0.003), but not in same-gender interactions (p=0.840).

[20] This extends the finding of Ashmore and Longo (1995), who note that only attractive females tend to have the ability to make males more likely to do them a favor. We show that males have the same ability.

Table VI: Results from Binary Probit Regressions of the Decision in the Prisoner's Dilemma in Mixed- and Same-Gender Interactions

	Marginal effects			
	(1) Mixed-gender		(2) Same-gender	
Attractiveness (mean-centered, d)	0.047	(0.067)	0.101	(0.094)
Opp. attractiveness (mean-centered, d)	0.151**	(0.062)	0.013	(0.092)
Male (d)	-0.049	(0.059)	-0.104	(0.107)
Age (cont.)	0.076**	(0.033)	0.111***	(0.042)
Demographics	yes		yes	
Stake size	yes		yes	
Wald χ^2	47.27***		20.01*	
Log-Likelihood	-161.07		-96.46	
Adjusted R^2	0.043		-0.036	
N	265		154	
Number of clusters	133		78	

Note: Binary probit regressions of the decision to "cooperate" ($y_i = 1$) or "defect" ($y_i = 0$) in the prisoner's dilemma, restricting the sample to (1) mixed-gender and (2) same-gender interactions. The marginal effect of the respective explanatory variable determines the effective change of this variable on player i's predicted probability to "split". Standard errors are reported in parentheses and are corrected for episode clusters. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

gender interactions we find no effects of attractiveness on cooperation, independent of whether the pair consists of two women (60.3% of same-gender pairs) or two men.

Further, we find no effect of attractiveness in the two group-decisions taken by the contestants prior to the prisoner's dilemma.[21] The absence of a beauty premium or plainness penalty in the pre-play as well as in same-gender interactions in the finale suggests that neither beauty-is-good stereotyping nor taste-based discrimination can explain our results consistently. It rather seems that the ability of attractiveness to mediate behavior is entangled to the opponent's sex and that people have a preference to cooperate with

[21] Taste-based discrimination (Becker (1957)) could be the underlying mechanism if we identify a beauty premium in the pre-play, but not in all interactions in the prisoner's dilemma, because contestants might favor to be in the final round with a facially attractive person. Using binary and ordered probit models, we estimate the effect of attractiveness on the likelihood to be voted off the game and on the likelihood to receive a certain number of votes after the first and second round. We find no significant effect of attractiveness on the voting outcome (tables unreported). This is also reflected in the almost equal share of attractive contestants in the final and initial round of the game show (52.4% vs. 51.8%). In all regressions we control for demographic characteristics such as gender, age, race and place of residence as well as for objective voting criteria such as the stake size a contestant accumulated, and whether a contestant lied in a previous round of the pre-play. We also follow Belot, Bhaskar, and van de Ven's approach and rank the contestants by their facial attractiveness to explain the likelihood to be voted off during the game, which yields no significant effects either (table unreported). Our finding is in contrast to Belot, Bhaskar, and van de Ven (2012), who find that attractive people are positively discriminated against unattractive people.

someone to whom they are personally attracted.

5.2 Vocal Attractiveness and Social Skills

In addition to the theories discussed above, the literature reports that physical and vocal attractiveness are highly correlated and thus suggests the existence of a vocal rather than a visual attractiveness stereotype, see, e.g., Zuckerman and Driver (1989). Furthermore, physically attractive people are also assigned stronger verbal and social skills, see, e.g., Goldman and Lewis (1977) and Erwin and Calev (1984). In the presence of other regarding preferences, vocal attractiveness and potentially strong verbal and social skills might enable attractive people to trigger cooperative behavior off their opponents. Although our data does not allow us to directly test for the impact of (perceived) vocal attractiveness on cooperative behavior, we can indirectly test for a correlation between facial attractiveness and communication which comprises verbal and social skills.

Shortly before the prisoner's dilemma is played both contestants are given some extra time to talk to each other. In these short conversations which are, on average, 38 seconds long, each contestant tries to convince her opponent to cooperate. Since it has been shown that promises effect people's behavior in experiments, see, e.g., Charness and Dufwenberg (2006) and Vanberg (2008), and in the field, see, e.g., Belot, Bhaskar, and van de Ven (2010), we code whether a contestant explicitly promises her opponent to cooperate.[22] Furthermore, we observe that contestants use handshakes to corroborate their mutual intention to cooperate and therefore we also code whether two contestants shake hands. As Table V shows, promises have a significantly positive impact on cooperative behavior, whereas handshakes have a negative one.[23] Comparing models (1) and (3) of Table V shows that the effect of attractiveness on cooperation remains almost unchanged when we add the variables of communication as controls. We find also no evidence that the effects of attractiveness and communication are interacted. However, we find limited evidence that contestants are significantly more likely to state a promise if the opponent is attractive (table unreported).[24] Even though we cannot entirely exclude that promises

[22]We count all statements as a promise when they contain either the word 'promise" or 'swear" or they are a statement of intent. Examples are 'I promise to split", 'I promise I will not steal", 'I swear I will split", 'I swear I will not steal", 'I will split", or 'I will not steal".

[23]The effect of the communication variables is robust to various specifications of the regression model. Including only demographics, the likelihood to cooperate of a contestant who made a promise to cooperate is 25 percentage points (p=0.000) higher than the one of a contestant who did not promise to cooperate. When contestants use a handshake they are actually 13 percentage points (p=0.031) less likely to cooperate than when they do not use a handshake (table unreported).

[24]If attractive contestants are better in terms of verbal and social skills, a promise or handshake of a facially attractive contestant might be more convincing than a promise or handshake of a facially unattractive contestant. Further, we could expect that facially attractive contestants are more likely to elicit a promise from their opponent and less likely to engage in a handshake and thereby provoke more

are the underlying transmission channel, the theory of better social and verbal skills or vocal attractiveness can also not explain the absence of the beauty premium in the play prior to the prisoner's dilemma.

6 Supportive Evidence

To justify our hypothesis of a preference based mechanism underlying the effect of attractiveness on cooperative behavior, we use the data from Mobius and Rosenblat (2006)'s experiments on the beauty premium in the labor market. They show that attractive people earn significantly higher wages even though they are not more (or less) productive than unattractive people. The experimental setup they use is a gift exchange between employers and workers. Subjects are randomly assigned roles and are then asked as a worker to exert real effort in a maze solving task and as an employer to set wages. One employer and one worker interact only once. Therefore, their data resembles our setting of a one-shot interaction in pairs without complete information about the other's type. Mobius and Rosenblat argue that stereotype beliefs about attractive people's productivity, higher confidence of attractive people in their own productivity, and better oral or communication skills of attractive people are driving the effects of attractiveness. We use the total of 820 interactions, of which 47.56% are mixed- gender interactions and 52.44% are same-gender interactions, and re-estimate their full model (Table 6, p.234 in Mobius and Rosenblat (2006)), parsing the data into mixed- and same-gender interactions. The results are presented in Table A.8 in the Appendix. We show that the effects of attractiveness remain only prevalent in interactions between either a female worker and a male employer or a male worker and a female employer (Model (2)); in the sample of same-gender interactions (model (3)), all effects of attractiveness disappear. This result clearly corroborates our hypothesis that effects of attractiveness depend on the gender composition, and are no artifact of our data.

7 Conclusion

We can show that attractiveness biases economic decision making of men and women when facing someone attractive of the other sex. This paper analyzes the relationship between attractiveness and cooperative behavior in a high-stakes field setting with two- sided face-

cooperative behavior from their counterpart. Testing for interaction effects between facial attractiveness and the communication variables reveals no additional effects (table unreported). Using binary probit regressions on the contestant's propensity to promise or to shake hands, we find that facially attractive contestants are not more likely to state a promise or to shake hands than facially unattractive contestants (table unreported).

to-face communication. Two independent data sets are combined. One is on cooperation, collected from decisions made in a slightly modified prisoner's dilemma played in the final round of a television game show. The other one is on the game show's contestants' facial attractiveness using a sample of independent third-party raters. In the prisoner's dilemma two finalists play for an accumulated jackpot by deciding either to "split" the jackpot or to "steal" it, where stealing implies keeping the entire amount to oneself . Our results show a strong and robust effect of attractiveness on cooperative behavior: facially attractive contestants are able to secure cooperation by their counterparts. Attractive contestants, however, are not more or less likely to be cooperative compared to less attractive contestants. This rewards attractive contestants with a beauty premium, which, at the margin, amounts, on average, to £2 153 for an increase in attractiveness of one standard deviation. The attractiveness effect is robust to, for instance, gender, age, stake size, or communication, but might be amplified by the attractive contestant's ability to talk their opponent into promising to cooperate. Decomposing the data into same- and mixed-gender interactions reveals a new insight on effects of attractiveness. The ability of attractive contestants to elicit cooperative behavior from their opponent vanishes in interactions between two contestants of the same sex. Contestants are only biased by the facial attractiveness of their opponent when the opponent is of the other sex. This suggests that stereotype beliefs about attractive people, such as them being more pro-social, cannot explain our results, since then the effect should prevail in all interactions. In line with our finding rather are theories from evolutionary psychology arguing that effects of physical attractiveness originate in primeval partner selection and should therefore be only or at least more present in mixed-gender interactions, see, e.g., Cosmides and Tooby (1987). In addition, the absence of a beauty premium or plainness penalty in earlier stages of the game show suggests that physical appearance is particularly important as soon as people are lacking objective information. Hence, we propose a preference-based mechanism as the underlying transmission channel of attractiveness on cooperative behavior: people are more likely to cooperate with someone toward whom they are personally attracted. We justify this hypothesis by challenging the experimental labor market data from Mobius and Rosenblat (2006). We can show that their identified effects of beauty remain only prevalent in mixed-gender interactions, i.e., in interactions between a female employer and a male worker or a male employer and a female worker. Our results are relevant and applicable to all one-shot face-to-face interactions and are particularly important when objective information is scarce. Such situations are, for instance, job interviews or face-to-face negotiations. And, Biddle and Hamermesh's finding of attractive people having advantages in the labor market might be reinforced by the fact that attractive people benefit from greater cooperativeness toward them.

References

Alreck, P. L., and R. B. Settle (1995): *The Survey Research Handbook*. Chicago: Irwin.

Andreoni, J., and R. Petrie (2008): "Beauty, Gender and Stereotypes: Evidence from Laboratory Experiments," *Journal of Economic Psychology*, 29, 73–93.

Antonakis, J., and O. Dalgas (2009): "Predicting Elections: Child's Play!," *Science*, 323, 1183.

Ashmore, R. D., and L. C. Longo (1995): "Accuracy of Stereotypes: What Research on Physical Attractiveness can Teach us," in *Stereotype Accuracy: Toward Appreciating Group Differences*, ed. by Y. Lee, L. Jussim, and C. McCauley, chap. 3, pp. 63–86. American Psycological Association.

Becker, G. (1957): *The Economics of Discrimination*. University of Chicago Press.

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

——— (2012): "Beauty and the Sources of Discrimination," *Journal of Human Resources*, 47(3), 851–872.

Berggren, N., H. Jordahl, and P. Poutvaara (2010): "The Looks Of A Winner: Beauty And Electoral Success," *Journal of Public Economics*, 94, 8–15.

Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir, and J. Zinman (2010): "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment," *Quarterly Journal of Economics*, 125(1), 263–305.

Biddle, J. E., and D. S. Hamermesh (1994): "Beauty and the Labor Market," *American Economic Review*, 84(5), 1174–1194.

——— (1998): "Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre," *Journal of Labor Economics*, 16(1), 172–201.

Castillo, M., R. Petrie, and M. Torero (2011): "Beautiful or White? Discrimination in Group Formation," *Working Paper*.

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

Cosmides, L., and J. Tooby (1987): "From Evolution to Behavior: Evolutionary Psychology as the Missing Link," in *The Lates on the Best: Essays on Evolution and Optimality*, ed. by J. Dupré, chap. 13, pp. 277–306. Cambridge: MIT Press.

Darai, D., and S. Grätz (2010): "Determinants of Successful Cooperation in a Face-to-Face Social Dilemma," *Working Paper*.

Dion, K., E. Berscheid, and E. Walster (1972): "What Is Beautiful Is Good," *Journal of Personality and Social Psychology*, 24, 285–290.

Dreber, A., C. Gerdes, and P. Gränsmark (2013): "Beauty Queens and Battling Knights: Risk Taking and Attractiveness in Chess," *Journal of Economic Behavior and Organization*, 90, 1–18.

Eagly, A. H., R. D. Ashmore, M. G. Makhijani, and L. C. Longo (1991): "What is Beautiful is Good, But...: A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype," *Psychological Bulletin*, 110(1), 109–128.

Eckel, C. C., and R. Petrie (2011): "Face Value," *American Economic Review*, 101, 1497–1513.

Erwin, P. G., and A. Calev (1984): "Beauty: More than Skin Deep?," *Journal of Social and Personal Relationships*, 1, 359–361.

Feingold, A. (1992): "Good-Looking People Are Not What We Think," *Psychological Bulletin*, 111(2), 304–341.

Goldman, W., and P. Lewis (1977): "Beautiful is Good: Evidence that the Physically Attractive are More Socially Skillful," *Journal of Experimental Social Psychology*, 13, 125–130.

Grammer, K., B. Fink, A. P. Møller, and R. Thornhill (2003): "Darwinian Aesthetics: Sexual Selection and the Biology of Beauty," *Biological Review*, 78, 385–407.

Jackson, L. A., J. H. Hunter, and C. N. Hodge (1995): "Physical Attractiveness and Intellectual Competence: A Meta-Analysis Review," *Social Psychology Quarterly*, 58(2), 108–122.

Kahn, A., J. Hottes, and W. L. Davis (1971): "Cooperation and Optimal Responding in the Prisoner's Dilemma Game: Effects of Sex and Physical Attractiveness," *Journal of Personality and Social Psychology*, 17(3), 267–279.

Landry, C. E., A. Lange, J. A. List, M. K. Price, and N. G. Rupp (2006): "Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 121(2), 747–782.

Langlois, J. H., L. Klakanis, A. J. Rubenstein, A. Larson, M. Hallam, and M. Smoot (2000): "Maxims or Myths of Beauty? A Meta-Analysis and Theoretical Review," *Psychological Bulletin*, 126(3), 390–423.

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

Mobius, M. M., and T. S. Rosenblat (2006): "Why Beauty Matters," *The American Economic Review*, 96(1), 222–235.

Mueller, U., and A. Mazur (1997): "Facial Diominance in Homo Sapiens as Hinest Signaling of Male Quality," *Behavioral Ecology*, 8, 569–579.

Mulford, M., J. Orbell, C. Shatto, and J. Stockard (1998): "Physical Attractiveness, Opportunity and Success in Everyday Exchange," *The American Journal of Sociology*, 103(6), 1565–1593.

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

Ottati, V. C., and M. Deiger (2002): "Visual Cues and the Candidate Evaluation Process," in *Social Psychology of Politics*, ed. by V. C. Ottati, and R. S. Tindale, pp. 75–87. New York: Kluwer Academic/Plenum.

Rosar, U., M. Klein, and T. Beckers (2008): "The Frog Pond Beauty Contest: Physical Attractiveness and Electoral Success of the Constituency Candidates at the North Rhine-Westphalia State Election of 2005," *European Journal of Political Research*, 47, 64–79.

Rosenblat, T. S. (2008): "The Beauty Premium: Physical Attractiveness and Gender in Dictator Games," *Negotiation Journal*, 24(4), 465–481.

Ruffle, B. J., and Z. Shtudiner (2011): "Are Good-Looking People More Employable?," *Working Paper*.

Solnick, S. J., and M. E. Schweitzer (1999): "The Influence of Physical Attractiveness and Gender on Ultimatum Game Decisions," *Organizational Behavior and Human Decision Process*, 79(3), 199–215.

van den Assem, M., D. van Dolder, and R. Thaler (2012): "Split or Steal? Cooperative Behavior When the Stakes are Large," *Management Science*, 58, 2–20.

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

Wilson, R. K., and C. C. Eckel (2006): "Judging a Book by its Cover: Beauty and Expectations in the Trust Game," *Political Research Quarterly*, 59(189), 188–202.

Zuckerman, M., and R. E. Driver (1989): "What sounds beautiful is good: The vocal attractiveness stereotype," *Journal of Nonverbal Behavior*, 13, 67–82.

Zuckerman, M., H. Hodgins, and K. Miyake (1990): "The vocal attractiveness stereotype: Replication and elaboration," *Journal of Nonverbal Behavior*, 14, 97–112.

A Appendix Figures and Tables

Figure A.1: Raw [Left] and Demeaned [Right] Variables of Facial Appearance, N=422

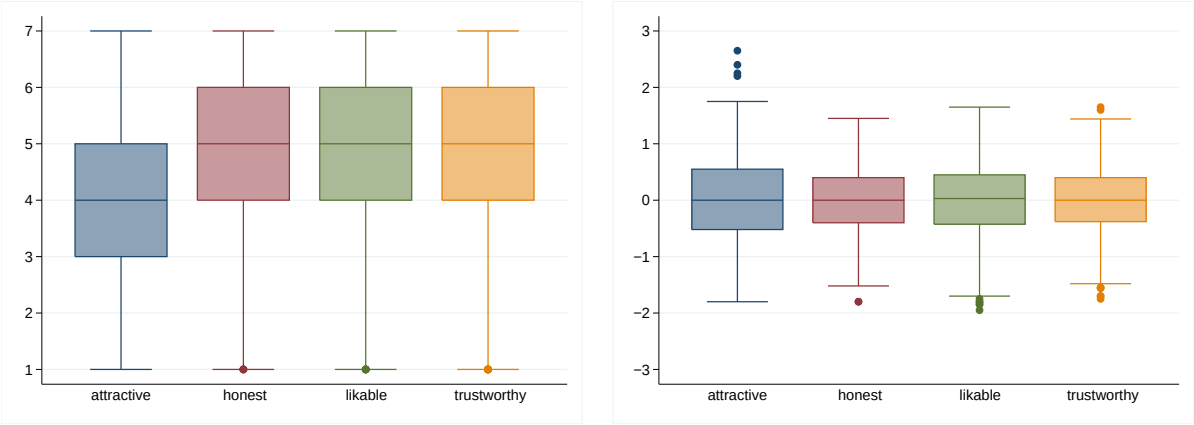


Table A.1: Summary Statistics Prisoner's Dilemma

Variable	Mean	Std. dev.	Min.	Max.	N
Decision variables					
Cooperate	0.54	0.5	0	1	422
Mutual decision	1.07	0.76	0	2	422
(0="steal-steal", 1="steal-split", 2="split-split")					
Amount money taken home	4 614.8	10 799.03	0	93 250	422
Demographics					
Male	0.46	0.5	0	1	422
Age[a](cont.)	3.21	1.1	1	6.5	422
Age of male (cont.)	3.26	1.07	1	6.4	422
Age of female (cont.)	3.19	1.09	1.25	6.6	422
White	0.94	0.25	0	1	422
London	0.11	0.32	0	1	422
England (1 = ENG, 0 = SCO, WAL, NIR, IRL)	0.85	0.36	0	1	420
Social job (reputation)[b]	0.16	0.36	0	1	421
Inexperienced (contestants of series 1)	0.19	0.39	0	1	422
Experienced (contestants of series 4)	0.19	0.39	0	1	422
Stake size					
Jackpot	12912.33	18213.95	3	93250	422
Potential jackpot[c]	50329.69	29946.46	5000	168100	422
Communication					
Promise or vow	0.42	0.494	0	1	422
Handshake	0.33	0.47	0	1	422
Pre-play					
Accumulated most money	0.5	0.5	0	1	422
Selected higher values in bin/win	0.5	0.5	0	1	422
Selected most killers in bin/win	0.33	0.47	0	1	422
"Should have left the game"[d]	0.26	0.44	0	1	422
Lied during pre-play	0.62	0.49	0	1	422

[a] Age is judged on a 7-item scale (see questionnaire and section 3.2), where 3="30-40", and 4="40-50" implying that the scale average of 3.21 equals a mean age of 37 years.

[b] A social job is defined as a job in which people care for other people, e.g., doctors, nurses, child minders, social workers, teachers, police officers, firemen, soldiers.

[c] The potential jackpot is announced by the show host before the actual jackpot is determined. It displays the maximal amount the actual jackpot could potentially comprise of.

[d] The variable "should have left the game" points at the player who is the "weakest" in material terms in round 2.

Table A.2: Summary Statistics all Raters

Variable	**Mean**	**Std. dev.**	**Min.**	**Max.**	**N**
Age (in years)	41.76	18.46	17	93	720
Age of male (in years)	39.57	17.70	17	93	371
Age of female (in years)	44.10	19.00	18	86	349
Male	0.51	0.5	0	1	728
Female ($\geq$ 44.1 years)	62.07	10.10	45	86	162
Female ($<$ 44.1 years)	28.52	7.84	18	44	187
Male ($\geq$ 39.6 years)	58.02	11.88	40	93	153
Male ($<$ 39.6 years)	26.62	5.17	17	39	218

Table A.3: Distribution of Attractive and Unattractive Finalists

	Facially attractive	Facially unattractive
Male ($<$ 37 years)	21%	27%
Male ($\geq$ 37 years)	15%	30%
Female ($<$ 37 years)	41%	18%
Female ($\geq$ 37 years)	23%	25%
	100% (N=221)	100% (N=201)

Table A.4: Results from Binary Probit Regressions of the Decision to "Cooperate" ($y_i = 1$) or "Defect" ($y_i = 0$) in the Prisoner's Dilemma, Including the Interaction Term "Male*age"

	Marginal effects									
	Model (1)		Model (2)		Model (3)		Model (4)		Model (5)	
Attractiveness (mean-centered, d)			0.062	(0.052)	0.048	(0.058)	0.025	(0.059)		
Opp. attractiveness (mean-centered, d)			0.124**	(0.049)	0.168***	(0.056)	0.161***	(0.057)		
Appearance helps in life (d)					0.036	(0.062)	0.056	(0.065)	0.067	(0.061)
Opp. appearance helps in life (d)					-0.090	(0.061)	-0.108*	(0.064)	-0.049	(0.060)
Appearance strikes as calculating (d)					0.011	(0.054)	0.018	(0.057)	0.056	(0.056)
Opp. appearance strikes as calculating (d)					0.018	(0.055)	0.001	(0.056)	0.002	(0.056)
Most attractive (90% percentile, d)									-0.141	(0.087)
Opp. most attractive (90% percentile, d)									-0.090	(0.087)
Least attractive (10% percentile, d)									-0.088	(0.089)
Opp. least attractive (10% percentile, d)									-0.235*	(0.081)
Male*age (cont.)	0.071**	(0.042)	0.067*	(0.042)	0.065*	(0.042)	0.047	(0.043)	0.059*	(0.043)
Male (d)	-0.304**	(0.147)	-0.286*	(0.149)	-0.273*	(0.149)	-0.232	(0.165)	-0.288*	(0.160)
Age (cont.)	0.037	(0.033)	0.048	(0.034)	0.050	(0.034)	0.066*	(0.039)	0.073	(0.050)
Opp. male (d)							-0.004	(0.056)		
Opp. age (cont.)							0.015	(0.028)		
Promise or vow (d)							0.252***	(0.051)	0.255***	(0.052)
Opp. promise or vow (d)							-0.008	(0.055)	-0.020	(0.055)
Handshake (d)							-0.092	(0.060)	-0.109*	(0.059)
Demographics	yes		yes		yes		yes		yes	
Stake size	yes		yes		yes		yes		yes	
Pre-play	no		no		no		yes		yes	
Wald χ^2	40.35***		47.21***		48.56***		69.17***		70.26***	
Log-Likelihood	-266.43		-263.10		-261.74		-247.87		-245.75	
Adjusted R^2	0.035		0.040		0.030		0.040		0.048	
N	419		419		419		419		419	
Number of clusters	211		211		211		211		211	

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

(d) for discrete change of dummy variable from 0 to 1

Table A.5: Results from Binary Probit Regressions of the Decision to "Cooperate" ($y_i = 1$) or "Defect" ($y_i = 0$) in the Prisoner's Dilemma, Including the Interaction Terms of Attractiveness w.r.t. Gender and Age

	Marginal effects							
	Model (1)		Model (2)		Model (3)		Model (4)	
Attractiveness*male (d)	-0.153*	(0.104)						
Opp. attractiveness*male (d)			0.099	(0.103)				
Attractiveness*age (cont.)					-0.079**	(0.044)		
Opp. attractiveness*age (cont.)							-0.011	(0.044)
Attractiveness (mean-centered, d)	0.133*	(0.078)			0.334**	(0.150)		
Opponent attractiveness (mean-centered, d)			0.068	(0.073)			0.154	(0.161)
Male (d)	0.044	(0.072)			-0.036	(0.050)	-0.054	(0.050)
Opp. male (d)			-0.055	(0.075)				
Age (cont.)	0.082***	(0.027)	0.075***	(0.024)	0.123***	(0.033)		
Opp. age (cont.)							0.006	(0.033)
Demographics	yes		yes		yes		yes	
Stake size	yes		yes		yes		yes	
Wald χ^2	36.01***		44.63***		43.93***		35.32***	
Log-Likelihood	-266.08		-265.29		-265.65		-270.02	
Adjusted R^2	0.033		0.035		0.034		0.019	
N	419		419		419		419	
Number of clusters	211		211		211		211	

Standard errors in parentheses are corrected for episode clusters; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; (d) for discrete change of dummy variable from 0 to 1

Table A.6: Results from Binary Probit Regressions of the Decision to "Cooperate" ($y_i = 1$) or "Defect" ($y_i = 0$) in the Prisoner's Dilemma when Teams of Contestants are Considered

	Marginal effects					
	Model (1)		Model (2)		Model (3)	
Similar team (d)	-0.043	(0.054)				
Attractive team (d)			0.045	(0.067)	0.027	(0.070)
Unattractive team (d)			-0.154**	(0.064)	-0.113*	(0.067)
Male (d)	-0.056	(0.050)	-0.041	(0.050)	-0.064	(0.056)
Opp. male (d)					-0.014	(0.056)
Age (d)	0.166***	(0.050)	0.187***	(0.051)	0.184***	(0.057)
Opp. age (d)					-0.034	(0.056)
Demographics	yes		yes		yes	
Stake size	yes		yes		yes	
Communication	no		no		yes	
Pre-play	no		no		yes	
Wald χ^2	40.40***		46.68***		68.32***	
Log-Likelihood	-266.77		-263.49		-250.48	
Adjusted R^2	0.034		0.042		0.049	
N	419		419		419	
Number of clusters	211		211		211	

Note: The "team variables" are indicators and equal 1 if the team is so composed and 0 otherwise, e.g., "Similar Team" equals 1 if both contestants are either attractive or unattractive, and 0 otherwise. Standard errors are reported in parentheses and are corrected for episode clusters. * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$)

Table A.7: Censored Tobit Regressions Results for Receiving Positive Earnings in the Prisoner's Dilemma

	Marginal effects			
	Model (1)		Model (2)	
Attractiveness (standardized)	616.927*	(334.835)	741.419**	(325.057)
Opp. attractiveness (standardized)			457.207	(285.004)
Male (d)	-209.774	(525.072)	-328.794	(540.835)
Age (cont.)	-218.725	(283.775)	-8.920	(268.037)
Opp. male (d)			-1126.281*	(573.093)
Opp. age (cont.)			1087.797***	(363.794)
Log(jackpot)	689.636***	(157.020)	615.929***	(146.146)
Demographics	yes		yes	
Opp. demographics	no		yes	
F-Statistic	3.186***		2.843***	
Log-Likelihood	-2593.35		-2578.46	
Adjusted R^2	0.003		0.006	
$\hat{\sigma}$	15059.81		14483.33	
N	419		419	
Number of clusters	211		211	

Note: Censored tobit regressions for the conditional expectation, that player i receives a positive earning $y_i > 0$ from the prisoner's dilemma game: $E[y_i|\mathbf{X_i}, y_i > 0] = \mathbf{X_i}\beta + \sigma\left[\frac{\phi(\mathbf{X_i}\beta/\sigma)}{\Phi(\mathbf{X_i}\beta/\sigma)}\right]$, where the inverse Mills ratio is evaluated at $\frac{\mathbf{X_i}\beta}{\sigma}$. The marginal effect for the jth independent (continuous) variable on $E[y_i|\mathbf{X_i}, y_i > 0]$ is computed as $ME_j = \frac{\partial E[y_i|\mathbf{X_i}, y_i>0]}{\partial x_j} = \beta_j\left[1 - \frac{\phi(\mathbf{X_i}\beta/\sigma)}{\Phi(\mathbf{X_i}\beta/\sigma)}\left(\frac{\mathbf{X_i}\beta}{\sigma} + \frac{\phi(\mathbf{X_i}\beta/\sigma)}{\Phi(\mathbf{X_i}\beta/\sigma)}\right)\right]$, and quantifies the expected increase in earnings, conditional on being positive. (d) for discrete change of dummy variable from 0 to 1. Standard errors are reported in parentheses and are corrected for episode clusters. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Fixed Effects Regression of Mobius and Rosenblat (2006)

	Model (1) All interactions (Table 6, p. 234)		Model (2) Mixed gender Interactions		Model (3) Same gender Interactions	
LNPROJECTED	0.409***	(0.043)	0.423***	(0.066)	0.337***	(0.066)
LNPROJECTED*VISUAL	0.007	(0.059)	0.022	(0.088)	0.047	(0.089)
LNPROJECTED*AUDIO	-0.129**	(0.059)	-0.101	(0.122)	-0.091	(0.081)
LNPROJECTED*VISUAL*AUDIO	0.056	(0.084)	-0.092	(0.155)	0.039	(0.117)
LNPROJECTED*FTF	-0.069	(0.060)	0.003	(0.104)	0.038	(0.079)
LNACTUAL	-0.004	(0.027)	-0.094*	(0.048)	0.032	(0.045)
BEAUTY	-0.010	(0.031)	-0.076*	(0.046)	0.073	(0.055)
BEAUTY*VISUAL	0.094**	(0.043)	0.170**	(0.066)	-0.007	(0.080)
BEAUTY*AUDIO	0.103***	(0.035)	0.162**	(0.067)	0.030	(0.059)
BEAUTY*VISUAL*AUDIO	-0.097*	(0.050)	-0.161*	(0.093)	0.076	(0.085)
BEAUTY*FTF	0.052	(0.035)	0.105	(0.070)	0.019	(0.053)
LNESTIMATED	0.018	(0.065)	0.157	(0.096)	0.048	(0.116)
LNESTIMATED*VISUAL	0.034	(0.083)	-0.126	(0.118)	0.131	(0.149)
LNESTIMATED*AUDIO	0.265***	(0.083)	0.084	(0.142)	0.205	(0.138)
LNESTIMATED*VISUAL*AUDIO	-0.056	(0.117)	0.110	(0.200)	-0.102	(0.194)
LNESTIMATED*FTF	-0.116	(0.083)	0.065	(0.141)	-0.253**	(0.126)
F-Statistic	21.939***		8.018***		11.773***	
R^2	0.627		0.664		0.709	
N	812		384		423	

Note: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is LNWAGE, standard errors are in parentheses. The base university is UNIVERSITY1.The regression includes the following resume controls: demographic variables (sex, age and age squared, Internet at home, participation in team sports, choice of college major, hobby variables, and previous job market experience). The regression also includes SETWAGE, and SETWAGE interacted with BEAUTY and LNESTIMATED.

B Online Appendix – Supplementary Material

B.1 The Structure of the Game Show

The game show "Golden Balls" consists of three rounds of play and starts with four contestants, regularly two women and two men. At the beginning of the show host Jasper Carrott briefly introduces each contestant (first name, occupation, and place of residence).[25]

Round 1 12 golden balls are drawn from a bingo cage with 100 golden balls that all contain a cash value. Contestants have only limited information about the lottery, i.e., they only know that there may be doubles among the 100 balls and they know that the margins of the distribution are included, namely a £10 and £75,000 ball. But they do not know the distribution of the remaining 98 balls. 4 different golden balls are then added to the 12 golden balls. These balls do not contain a cash value, but the word 'killer". Killer balls may decrease the jackpot in the final round and therefore contestants try to avoid them if possible. The total of 16 golden balls, indistinguishable from the outside, are shuffled and then evenly divided between the four contestants. Each contestant arranges her still closed 4 golden balls in two rows of two balls in front of herself. The two balls on the front row are then opened by each contestant such that the content, cash value or killer, is common knowledge to all contestants. The content of the remaining two balls on the back row is private information to each contestant, i.e., the contestants are allowed to privately look inside but then have to close the balls again. Afterwards the show host asks each contestant to state what is inside her closed balls. The show host exogenously determines the order in which contestants are asked for their claims. Some time for discussion between the contestants follows, in which the contestants express their distrust about each other's claims. The discussion ends with each contestant secretly casting a vote against one of the other contestants. The contestant who received most votes has to leave the show. In case of a tie the contestants having received no vote can decide which of the two contestants has to leave. If all contestants received one vote each, they again discuss who has to leave. If contestants do not reach a conclusion, ties are broken arbitrarily. After the contestant who has to leave the show is determined, all contestants open their still closed back row balls. The show hosts then points out whether a contestant's claim has been truthful or not. The four balls of the leaving contestant are out of the game, while the remaining 12 are carried over to round 2.

[25] Endemol UK ensured us that the four contestants do not know each other before the show and enter and leave the television studio separately such that they cannot make any further arrangements after the show.

Round 2 At the beginning of the second round, two new cash balls, drawn from the lottery, and one killer ball are added to the remaining 12 balls from round 1. The 15 balls are shuffled and equally distributed to the remaining three contestants. Again each contestant arranges the closed balls in two rows, i.e., two balls on the front and three balls on the back row. Like in round 1 the two balls on the front row are opened such that their content becomes common knowledge to all contestants, while the three balls on the back row are private information to each contestant. This time the contestants determine the order in which they make their statement about the content of their back row balls. As in round 1, contestants then get some time to discuss the statements that have been made. Afterwards contestants again vote who has to leave the game. After this contestant has been determined all ball values are revealed and the five balls of the leaving contestants are taken out of the game.

Final Round The ten balls from round 2 are carried over to the final round and one killer ball is added. The final round is divided into two parts, (i) determining the jackpot and (ii) playing the modified prisoner's dilemma. Before part (i) starts, the show host announces the maximal amount the contestants can gain. This amount, called 'potential jackpot", is the highest possible amount the contestants can get when they randomly draw five of their collected eleven golden balls. In part (i) the two contestants successively select five of their eleven closed balls. Specifically, the contestant who brought the highest amount of money from round 2 to the final round starts part (i) by selecting one of the eleven balls to "bin", i.e., to be taken out of the game, and then chooses one ball to "win". The balls are not opened until they have been chosen. Then it is the other contestant's turn and vice versa until five balls have been selected for the jackpot. If a contestant chooses a killer ball for the jackpot the accumulated amount up to that point is reduced to one-tenth of the original value. These five randomly drawn balls determine the actual jackpot.

In part (ii) of the final round the players play for this jackpot in a prisoner's dilemma in which defection is a weakly dominant strategy. The dilemma game is played as follows: each contestant is assigned two golden balls, one with the word "steal" and one with the word "split" inside. The show host explains the game with almost the same words in each episode:

> "It is time to split or steal. You have got two final golden balls left, you have each got a golden ball with the word *split* written inside, you have got each a golden ball with the word *steal* written inside. I will ask you to make a conscious choice and you will choose either the split or the steal ball, neither of you will know what the other has chosen. If you both choose the split balls,

you split today's jackpot of $£J$ and you both go home with $£J/2$. If one of you splits and one of you steals, whoever steals goes home with all the money $£J$, whoever splits goes home with nothing. If you both decide to steal and you are very greedy, you both go home with nothing. Before I ask you to choose, NAME OF CONTESTANT A AND B just check the two balls to make sure you know which is to split and which is to steal. Do not show to each other. It is very important that you know which is which. [CONTESTANTS CHECK THE BALLS] Are you happy to know which is split and which is steal? Okay, before I ask you to choose, I will give you some time to talk to each other about what has happened today and how you feel.

Then contestants get, on average, 38 seconds to discuss their intentions with each other. This communication is free-format. The show host continues:

> Okay, NAME OF CONTESTANT A AND B choose the split or steal ball now. [CONTESTANTS CHOOSE BALLS] Hold it up, make sure that when you open it, the other player can see it. NAME OF CONTESTANT A AND B split or steal? [CONTESTANTS OPEN BALLS]"

After both contestants revealed their decision by opening their balls, the show host thanks the contestants for taking part in the show and then the show ends.

B.2 Game Show Data

"Golden Balls" was first aired on June, 18th 2007 as a late afternoon (5pm) game show and ended December, 18th 2009. The show reached up to 2.2 million people per episode, which corresponds to a market share of 21% ("ITV strikes teatime gold", guardian.co.uk, July 3rd, 2007). In total, we have records of 221 episodes, divided into four series, with 202 regular and 19 special episodes. Ten special episodes consist of contestants who are on the show for the second time. We exclude these ten episodes from the analysis in order to avoid any bias from repeated interactions. The other nine special episodes comprise of contestants of the same sex. The regular episodes always consist of two women and two men and all contestants are on the show for the first time. Importantly, the first series (40 episodes) was filmed prior to the show's television premiere, i.e., all contestants in these episodes had no chance to observe others playing the game. In total we use a data set of 211 episodes (844 contestants), divided into 4 series, where series 1, 2, 3, and 4 consist of $160, 232, 288$, and 156 contestants, respectively.