Beauty and the Sources of Discrimination

Michèle Belot
V. Bhaskar
Jeroen van de Ven

ABSTRACT

We analyze discrimination against less attractive people on a TV game show with high stakes. The game has a rich structure that allows us to disentangle the relationship between attractiveness and the determinants of a player's earnings. Unattractive players perform no worse than attractive ones, and are equally cooperative in the prisoner's dilemma stage of the game. Nevertheless, they are substantially more likely to be eliminated by their peers, even though this is costly. We investigate third party perceptions of discrimination by asking subjects to predict elimination decisions. Subjects implicitly assign a role for attractiveness but underestimate its magnitude.

I. Introduction

In a surprising and influential paper, Hamermesh and Biddle (1994) found a correlation between attractiveness and labor market earnings across a variety of occupations. Attractive individuals earn 5 percent more than those with average looks, and less attractive individuals earn 5 to 10 percent less. The source of the

Michèle Belot is a research fellow at the Centre for Experimental Social Sciences of Nuffield College, University of Oxford. V. Bhaskar is full professor of economics at the University College London. Jeroen van de Ven is assistant professor at the Amsterdam Center for Law and Economics and the Department of Economics of the University of Amsterdam. *The authors are very grateful to Dan Hamermesh, Gordon Kemp and two anonymous referees for very useful comments and suggestions, to Debbie Hall for research assistance, and they wish to thank seminar and conference participants at University College Dublin, University of St Andrews, University of Amsterdam, University of Essex (ISER), the GATE conference in Lyon (2007) and the ESPE conference in Chicago (2007). V. Bhaskar thanks the Economic and Social Research Council for support. The data used in this article can be obtained beginning six months after publication through three years hence from Jeroen van de Ven, University of Amsterdam, ACLE, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands, j.vandeven@uva.nl.*
[Submitted March 2009; accepted August 2011]
ISSN 022-166X E-ISSN 1548-8004

beauty premium remains an open question.[1] Attractiveness may be correlated with unobservable productive attributes such as health, education or other types of human capital. Less attractive people may also be less confident, reducing their social skills in the workplace (Mobius and Rosenblat 2006). There may also be an element of reverse causality—individuals who fare well in the labor market may have both the ability and incentive (via greater self esteem) to invest in looking good.[2] Perhaps the simplest explanation is that beauty has "consumption-value," either to the customers of the firm, fellow employees, or the boss. This case is a form of taste-based discrimination (Becker 1957). Establishing discrimination and distinguishing between statistical and taste-based discrimination is difficult, as the literature on racial and gender discrimination shows (see Altonji and Blank 1999; Heckman 1998).

Discrimination based on attractiveness also raises somewhat different issues as compared to race or gender discrimination. There is less public awareness regarding its possible prevalence, attractiveness is less objective and discriminators are less subject to social disapproval. For these reasons, discrimination based on attractiveness may prevail even in public environments, where individuals may be reluctant to indulge in race or gender discrimination.

We investigate discriminatory practices in a very public setting—a television game show—where the performance of contestants is clear-cut and the stakes are high. The game takes place over three rounds, in which players accumulate "earnings" by answering quiz questions. Their earnings depend on the accuracy of their answers, on how quickly they press the buzzer and on their "investment decisions." Earnings therefore depend upon ability as well as a player's confidence. At the end of each round, the lead player—the one with the highest earnings—decides which one of the remaining players to eliminate. When only two players remain, they play a prisoner's dilemma game, allowing us to study the relation between beauty and cooperativeness. The median stake in this prisoner's dilemma game is €1,683, so that the monetary consequences are substantial.

This setting provides an ideal environment to study discriminatory practices since performance is observed and precisely measured. We can study whether, conditional on performance, attractive players are less likely to be eliminated. The rich structure of the game enables us to study the relationship between attractiveness and the various determinants of earnings, such as confidence, cooperativeness and discrimination.

We find that unattractive players fare significantly worse than attractive ones. Players can only make positive earnings by making it to the final stage of the game show. Only 27 percent of the least attractive players make it to the final round, against 49 percent of the most attractive ones. This difference cannot be attributed to differences in any aspect of performance, nor to differences in cooperation rates in the final stage.

Discriminating against less attractive people is not simply used as a tie-breaking rule between otherwise similar players. Unattractive players are more likely to be

1. While we refer to the beauty *premium*, it appears that it is often the unattractive people that are being penalized rather than attractive people being rewarded.

2. Biddle and Hamermesh (1998) address the reverse causality problem in a study on lawyers by using a measure of beauty based on photographs taken at law school.

eliminated even when they have a higher score than others. We estimate that the resulting costs to lead players from discrimination in the final round alone amounts to about €440 on average, that is, about 25 percent of the median stake. This is an underestimate of the overall costs of discrimination over all the rounds. Since attractive people are not more cooperative in the prisoner's dilemma, there is no offsetting financial benefit.

It is noteworthy that we find discrimination on a TV show, where each player's performance is clear-cut and where the lead player's decision is subject to public scrutiny. Other studies using TV shows (Levitt 2004; Antonovics, Arcidiacono, and Walsh 2005; and List 2006) find no evidence of discrimination on the basis of race, gender, or ethnicity, but some evidence of discrimination against older players (these papers do not examine the role of beauty). As we have noted, discrimination based on looks may be less imprinted in social consciousness, so that people are not so aware of the possibility of discrimination—that is, such discrimination may be insidious. To investigate third party perceptions of discrimination, we ran an experiment where subjects watched the game show and had to predict elimination decisions. Our subjects predict that attractive players are less likely to be eliminated, although they substantially underestimate the magnitude of this effect. We also ask our subjects to list qualitative factors that influence elimination decisions, and find that very few subjects mention attractiveness. Thus third parties appear to be aware of the possibility of discrimination subconsciously rather than consciously.

The remainder of this paper is organized as follows. Section II reviews the related literature. Section III describes the game show and our measure of attractiveness. Section IV analyzes behavior on the game show, in terms of performance and cooperation. Section V studies the selection decisions by lead players and establishes that unattractive players suffer. This section also presents our experimental findings on third party perceptions of discrimination. The final section concludes.

II. Related Research

Following the work of Hamermesh and Biddle (1994), several papers have replicated its findings, and also attempted to disentangle the components of the premium (or penalty).[3] Biddle and Hamermesh (1998) analyze a sample of lawyers, and find a premium in all areas of expertise, and also among the self-employed. They argue that the most plausible explanation is taste-based discrimination by clients. Mocan and Tekin (2010) find that unattractive people sort into criminal activity due to the existence of a beauty premium on the legal labor market.

There are some difficulties with field studies. First, data on attractiveness is rarely available. Second, it is difficult to disentangle attractiveness from ability, because productivity is rarely observable. This makes it hard to establish whether the beauty premium is due to productivity or discrimination. Such productivity effects are sometimes present. Landry et al. (2006) find that attractive female solicitors are more productive fundraisers. Pfann et al. (2000) study a sample of Dutch advertising firms

3. See also Hamermesh (2011) for a book length discussion.

and find that those with better looking executives have higher revenues. This is particularly pertinent for our paper, since it demonstrates that beauty plays a role in the Netherlands. Third, even in the absence of correlation between ability and performance, attractive people may be "better perceived," by appearing more productive or confident for example. Doran and Hersch (2009) show that the beauty penalty decreases when one controls for measures of other unobservable characteristics such as personality or apparent intelligence.

Laboratory experiments are better suited to disentangling the sources of the beauty premium, since they can be designed for this purpose. Mobius and Rosenblat (2006) take this approach, using university students in Argentina. They find that the beauty premium appears to be partly due to the fact that attractive people are more confident. Since our results differ from theirs—we find that beauty is unrelated to actual confidence, as reflected in behavior, although it is correlated with third party perceptions of confidence—we discuss their paper in more detail in section 4.2, after presenting our findings. Benjamin and Shapiro (2006) find that experimental subjects are able to predict the electoral fortunes of candidates on the basis of ten-second silent video clips, suggesting that attractiveness (or charisma) plays an important political role.

III. Description of the data

A. The Game Show

We use data from all 69 episodes of the game show "Does (S)he Share or Not?"[4] broadcast in the Netherlands in 2002, with 345 contestants in total. The main game starts with five players.[5] Players accumulate earnings over the course of three quiz rounds. Every round has ten regular questions and a bonus question. Players choose at the beginning of each round how much of their capital to "invest" (y_i). Whoever presses the buzzer first gets to answer. A correct answer yields y_i, while an incorrect answer earns $-y_i$. A player whose capital falls below his or her chosen investment may not answer any further questions in that round. Each round ends with a bonus question, where players compete for the right to answer the question by choosing new investments. At the end of the round, the player with the highest score at that point—the "*lead player*" henceforth—must select one of the remaining other players for elimination. An eliminated player has no further role in the game and loses all of his or her earnings. The show then proceeds to the next round, where all players start with a capital that equals the earnings of the lead player in the previous round. At the end of the third round, only two players remain.

The last two remaining players play a prisoner's dilemma game. Let E denote the total prize money, which equals the sum of earnings of the two finalists. After a communication stage, the finalists simultaneously decide whether to share or to grab. If both players choose to share, the prize is split equally among them. If only one

4. The name of the show in Dutch is "Deelt ie 't of deelt ie 't niet?"

5. Prior to this stage, six players choose their initial capital, a number between one and 100, and the player who picks the highest number is eliminated immediately.

Table 1
Monetary Payoffs

	Share	Grab
Share	$\frac{1}{2}E, \frac{1}{2}E$	$0, E$
Grab	$E, 0$	$0, 0$

player shares, the entire prize goes to the player that grabs. The prize is lost if both players grab. The monetary payoffs, as depicted in Table 1, thus correspond to a generalized prisoner's dilemma, where grab is a weakly dominant strategy.

Note that from the perspective of maximizing the amount of prize money, it is in the interest of players to keep the strongest performing players in the game, since the prize money is determined by the performance of players. In addition, it is important for players that the other finalist is a cooperative person.

Table 2 presents summary statistics of the game show and the players. The total prize at stake varies between €380 and €26,600, with a median value of €1,683.[6] Players choose to share 43 percent of the time in the final round. The age of players varies between 18 and 64 with an average of 34.4. About a third of players are women.[7]

B. Measure of Beauty

Hamermesh and Biddle (1994) argue that there are consistent standards of beauty, so that subjective evaluations of attractiveness include a "common component." Our goal is to see how this common component affects performance on the game show. To this end, each of the 345 participants on the game show were rated on a scale from one (very unattractive) to seven (very attractive) by approximately ten raters, balanced by gender.

Raters were recruited among adults at various public spaces, such as parks, bars, trains, and on a ferry between the Netherlands and the United Kingdom. Raters were on average 31.7 years old, which is close to the average age of game show participants, 34.4 years. We recruited 120 raters, and each rated 30 participants. This was based on watching short silent video fragments of the game show in which a player introduced him or herself. We ensured that all five players on any show were rated by the same set of raters, while varying the order in which the shows were presented.

6. These are considerable sums given that the median disposable monthly income of a full-time employed person in the Netherlands was about €1,200 in 2000 (CBS, *Statistics Netherlands*, available at www.cbs.nl).
7. The fact that women are underrepresented is due to self-selection, as no selection criteria were used in the recruitment (see section IIIC). The percentage is comparable to that in the Dutch and German versions of the game show "Deal or no Deal," where participation is also based on self-selection (in the Dutch version 27 percent is female, in the German version 34 percent; see Post et al. 2008). One reason may be that the game show is perceived as competitive, as women are found to shy away from competition (see, for instance Niederle and Vesterlund 2007)

Table 2
Summary Statistics of the Game Show Players

	Mean	Minimum	Maximum
Mean age (years) (N = 345)	34.4	18	64
Percentage women (N = 345)	34.8	—	—
Prize (€) (N = 138)	2,976	380	26,600
Percentage sharing (N = 138)	42.8	—	—
Occupation, percentage	Players	Dutch population	
Agriculture	0.0	1.4	
Manufacturing	1.4	12.0	
Public utilities	2.7	0.4	
Construction	1.4	5.3	
Trade	6.8	16.5	
Hotels/restaurants	2.7	3.7	
Transport	9.6	6.1	
Financial institutions	4.1	3.7	
Business services	20.5	17.0	
Public administration	1.4	7.2	
Education	23.3	6.7	
Health sector	20.5	16.0	
Cultural sector	5.5	4.0	

Source of occupational data for Dutch population is Statistics Netherlands (CBS) for the year 2006.
Note: occupational distribution of players for nonstudents only (N = 78).

About one-third of the raters were non-Dutch and could not be familiar with the show, and only a small minority of the Dutch raters indicated that they had seen any episode of the show. Given this, and the fact that the game show was broadcast five years prior to the ratings, our results are not distorted due to familiarity with the faces.

Our measure of attractiveness is the average of the independent ratings (across raters) for each player.[8] Table 3 reports summary statistics of the ratings. Raters were told to use the benchmark average attractiveness in the population at four. Beauty is negatively correlated with age and women are, on average, rated as being more attractive than men. Average ratings are more variable across women than men, consistent with other studies (Hamermesh and Biddle 1994). We also coded objective participant characteristics that are related to their appearance, such as eye color, hair length, wearing of glasses, etc. These characteristics are correlated with our beauty

8. There is a high degree of concurrence on attractiveness across raters. Across subsamples of raters who rated the same sample of players, the Cronbach's alpha ranged from 0.70 to 0.85, showing high agreement.

Table 3
Summary Statistics of Attractiveness

	Mean	Standard deviation	Minimum	Maximum
All (N = 345)	3.51	0.69	1.7	5.75
Men (N = 225)	3.45	0.63	2.0	5.20
Women (N = 120)	3.62	0.79	1.7	5.75
Age $\geq$ 34(N = 177)	3.30	0.60	1.7	4.80
Age $<$ 34 (N = 176)	3.71	0.71	2.2	5.75

Note: Attractiveness of players is averaged across raters.

measure—for women, beauty is positively correlated with hair-length, while for men, beauty is negatively correlated with having a beard or moustache and wearing glasses. This suggests that our measure of beauty captures aspects of physical attractiveness.

C. Characteristics of the Players

One possible concern with our study is external validity, especially if the participants on the show are not representative of the population.[9] We shall present evidence here to show that the game show participants are not exceptional in terms of attractiveness. They were recruited via announcements on television and radio shows, and applicants could register on the company's website. The producer of the show told us that no explicit criteria were used in recruiting participants. In particular, applicants were not required to submit a photograph, so that the producers seem unconcerned with having good looking people. Nor does it seem that attractive people overly self-select into the show. The average rating of our participants in terms of attractiveness is 3.51, whereas raters were told to use 4 as the benchmark for average attractiveness of the Dutch population. We also obtained ratings for a sample of academics and another of lawyers, for comparison. The average rating of game show participants is only slightly higher than that of the academics (which was 3.41) and somewhat below that of the lawyers (which was 3.97).[10]

Participants also come from a variety of occupations, an advantage that compares favorably with laboratory experiments. Of the third of the players who report their occupation, a third are students, while the others are drawn from all the major sectors (at the two-digit classification level). Table 2 reports the occupational distribution

9. Field data on attractiveness, individual productivity, and pay would be ideal, but it is hard to use this to disentangle the beauty premium. Thus laboratory experiments and our study have a role, see Harrison and List (2004) for a discussion on the external validity of game show data.

10. The two samples consisted of 50 pictures from the Internet: of members of the economics department at a Dutch university, and of employees at a large Dutch law firm. Every picture was rated nine or ten times. The samples resemble the game show population in terms of the age and gender distribution. The average rating across the two samples is 3.69. As in the game show participants, ratings are above average for women (3.89) and younger people (3.82).

of players and the Dutch population. The occupational distributions are fairly similar, although the educational sector seems overrepresented, and manufacturing and trade underrepresented.

IV. Beauty and Behavior

The main question of this study is whether there is any evidence that attractiveness matters in the selection decision, conditional on "productivity" in the game. At the time of selection, the lead player observes perfectly how well the other players have performed in the previous round. Thus, we can study the selection decision conditional on performance in the previous round. Performance in this game depends on a number factors: the choice of initial capital and investment, the decision to press the buzzer or not, and, finally, the answer itself. These decisions depend upon the player's ability, confidence, and risk-aversion. We study the relationship between beauty and these different dimensions of performance, as well as their perceptions. Because the final earnings are conditional on the player's cooperativeness, we also study the relationship between beauty and cooperativeness.

A. Beauty and Performance

We first investigate the relationship between beauty and performance in answering questions. Since players compete to answer each question, their performance in the game is a relative measure. If there is a relation between beauty and performance then this will depend on the composition of players within an episode. Thus, we estimate a model including fixed effects by episode:

$$(1) \quad Y_{ij} - \bar{Y}_j = \beta'(X_{ij} - \bar{X}_j) + u_{ij},$$

where Y_{ij} is the performance measure for player i in show j, X_{ij} is a vector of attributes (age, gender, attractiveness, etc.) and u_{ij} is a white noise error term. $\bar{Y}_j$ and $\bar{X}_j$ are the variable means at the episode level. We use two different measures of performance: the first is the earnings at the end of the first round, the second is the player's earnings rank within the episode. Note that the identification of the effects of variables relies on the variation in characteristics across players within an episode. If there is a show-specific element affecting all the players within the same episode—in this context, affecting their performance—then this does not affect the estimates because it is differenced out. Similarly, this specification eliminates the importance of average differences in rater perceptions across the shows, since all five players have been rated by the same set of independent raters.[11]

We estimate two sets of regressions, one where attractiveness is a continuous measure and the other where we only control for a dummy for the least attractive player among the four players candidate for elimination. If beauty matters, one could expect the least attractive player to be more at risk of being eliminated than the

11. We should point out that other specifications, such as conditional logit model, provide very similar results.

Table 4
Attractiveness and Performance in the First Round

	Score		Score ranking		Perceived Knowledge	
	(1)	(2)	(3)	(4)	(5)	(6)
Attractiveness	−7.09		−0.06		0.02	
	(7.16)		(0.16)		(0.11)	
Least attractive		−1.93		−0.01		−0.13
		(5.86)		(0.19)		(0.13)
Age	−0.82	−0.15	−0.01	−0.01	−.004	−0.001
	(0.54)	(0.34)	(0.01)	(0.01)	(.008)	(0.007)
Female	−5.84	1.69	−0.13	−0.01	−0.48***	−0.44***
	(8.69)	(5.53)	(0.19)	(0.18)	(0.13)	(0.12)
Number of observations	345	276	345	276	345	276
$R-$squared	0.02	0.01	0.01	0.02	0.06	0.08

Notes: FE regressions. In Columns 3 and 4, 1 is the lowest score and 5 the highest. Standard errors in parentheses. Additional controls include dummies for registry of introductory speech (married, children, hobby and profession). Estimations in Columns 2, 4, and 6 exclude the lead player. *** $p < 0.01$.

other players, especially in light of the finding in the literature that unattractive people are penalized. We study whether these players differ systematically from the others. The lead player is excluded from this second set of regressions.

The results are reported in Columns 1–4 of Table 4. We find no clear correlation between any of these characteristics and performance. In particular, there is no evidence that attractive people perform differently from unattractive ones.

To take into account the possibility that the lead player also observe other signals of each player's ability, we asked independent raters to rate each player's "perceived general knowledge"—the relevant measure of ability in our context.[12] The measure is highly correlated with actual performance in the first round (the correlation is 0.70 and is significant at the 1 percent level). Columns 5 and 6 show the results of an OLS regression of mean perceived knowledge on the player's characteristics. We find no correlation between attractiveness and perceptions of ability either.

B. Beauty and Confidence

In an experimental study, Mobius and Rosenblat (2006) argue that confidence explains a substantial part of the beauty premium. They find that subjects who are assigned the role of "employers" have higher estimates of the productivity of more attractive "employees," even though they are given independent evidence on pro-

12. The raters saw videos of the show up to the end of the first round. Each player was rated on a seven point scale by three to six raters, and each rater saw 12 shows.

ductivity. An intriguing finding is that attractive subjects are estimated to have higher productivity even when their interaction with the employer is only oral, not visual. Mobius and Rosenblat attribute this to the higher self-confidence of attractive subjects, and this explains about 20 percent of the beauty premium. Translated to the context of our game show, one might expect confident players to invest more and to answer more questions. Such a player will also make more mistakes than a player of similar ability who has less confidence. Table 5 presents the results of OLS estimates[13] of the determinants of initial capital, share invested, number of answers and percentage of good answers (Columns 1 to 8 respectively).[14] We do not find that attractive players are more confident. If they were, they should invest more, be more likely to answer, and, conditional on answering, perform worse and we find no significant difference in any of these dimensions. We do find a systematic difference according to gender: Women are much less likely to answer a question but conditional on answering, they actually do not perform better than men, what you would expect with risk-aversion or lack of confidence.[15]

Even if there is no correlation between beauty and confidence, it could be the case that attractive players *appear* more confident by other players. To investigate this, we constructed a measure of perceived confidence, by having independent raters assess the confidence of a player on a seven point scale (the raters were different from those used for the measure of perceived ability). The raters saw videos of the players introducing themselves at the beginning of the show. Each player was rated by five to ten raters, and each rater saw 35 videos of players. In both cases, the player-specific measure is the mean across raters.

We find that perceived confidence is positively correlated with attractiveness, and this is statistically significant (Columns 9 and 10). Attractive players *appear* more confident to outside observers, but do not behave more confidently in the game. This may be due to the fact that players receive continuous feedback on their performance in the game show. This contrasts with the setup of Mobius and Rosenblat, where subjects did not receive systematic feedback on their performance throughout the experiment.

C. Beauty and Cooperativeness

Players can earn money in this game only by reaching in the final stage. At this point, the two remaining players play a prisoner's dilemma game and decide simultaneously to share the accumulated money or not. Table 6 reports our results on the relation between beauty and the decision to share.[16] We find no correlation

13. Alternative econometric specifications (conditional logit estimates for the probability of answering and logit estimates for the probability of answering correctly) give similar results to the ones presented here.

14. We have analyzed separately the decisions made for the bonus question. Again, we found that attractiveness is uncorrelated with the share invested or the probability of giving a correct answer. The results are not reported for the sake of brevity.

15. These results suggest that the reason why women are less likely to answer is not due to a wrong perception of their ability or a higher degree of risk-aversion, but rather because they are less able to answer the type of questions on the show.

16. Belot, Bhaskar, and Van de Ven (2010) conduct a comprehensive analysis of the determinants of sharing behavior. The key findings are that own characteristics matter—specifically, women are more likely to share than men. However, the characteristics of the opponent turn out to be irrelevant to the sharing decision.

Table 5
Attractiveness and Confidence in the First Round

	Initial capital		Share invested (percentage)		Number of answers		Percentage correct answers		Perceived confidence	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Attractiveness	0.26		−0.031		−0.03		0.01		0.20***	
	(0.73)		(0.022)		(0.16)		(0.04)		(0.05)	
Least attractive		−0.32		−0.02		0.030		−0.01		−0.19**
		(1.14)		(0.03)		(0.184)		(0.07)		(0.08)
Age	0.04	0.04	−0.002	−0.001	0.001	0.009	−0.003	0.001	0.012***	0.009**
	(0.06)	(0.07)	(0.002)	(0.002)	(0.012)	(0.011)	(0.003)	(0.004)	(0.004)	(0.004)
Female	0.12	−0.25	−0.01	0.003	−0.49**	−0.50***	0.02	0.04	−0.28***	−0.28***
	(0.89)	(1.08)	(0.03)	(0.030)	(0.19)	(0.17)	(0.05)	(0.07)	(0.06)	(0.08)
Number of observations	345	276	345	276	345	276	271	203	345	276
$R-$squared within	0.01	0.01	0.06	0.05	0.04	0.05	0.01	0.01	0.12	0.10

Notes: FE regressions. Standard errors in parentheses. Additional controls include: dummies for registry of introductory speech (married, children, hobby, and profession). Columns 2, 4, 6, 8, and 10 exclude the lead player. *** $p < 0.01$, ** $p < 0.05$.

Table 6
Attractiveness, Sharing, and Perceptions of Sharing

	Probability of sharing		Perceived probability of sharing
	(1)	(2)	(3)
Own attractiveness	−0.03	−0.04	−0.031***
	(0.07)	(0.06)	(0.012)
Opponent's attractiveness		−0.01	
		(0.11)	
Age	0.00	0.00	0.001
	(0.01)	(0.01)	(0.001)
Female	0.19**	0.21**	0.091***
	(0.09)	(0.10)	(0.015)
Contribution to prize money (percentage)	−0.72**	−0.82**	−0.066**
	(0.31)	(0.35)	(0.032)
Total prize (×1,000)	0.03**	0.03***	0.008***
	(0.01)	(0.01)	(0.002)
Constant			0.481***
			(0.100)
Number of observations	138	138	1,672
Method	Bivariate probit	Bivariate probit	OLS

Notes: Specifications in Columns 1 and 2 are bivariate probit estimates. Standard errors (in parentheses) are clustered by episode. The marginal effects in Columns 1 and 2 are computed at the means of the lead player's characteristics. Column 3 includes subject fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

between beauty and cooperative behavior. Attractive players are no more (or less) likely to share—indeed, the coefficient is very close to zero. Attractive opponents are also no more (or less) likely to induce sharing behavior from their opponents. This is interesting, although our overall results suggest that players obtain consumption value from having attractive coplayers, they are no more likely to share with them.[17]

Perceptions of cooperativeness are likely to be very important in determining the selection decision, and also possibly in the decision by a player to share.[18] Belot,

17. Other experimental evidence on the relation between attractiveness and cooperation is mixed, see Mulford et al. (1998), Solnick and Schweitzer (1999), Eckel and Wilson (2004), and Andreoni and Petrie (2008).

18. Indeed, given the importance of the prisoner's dilemma stage of the TV show, they are of a higher order of importance than perceptions of other variables such as confidence or ability.

Bhaskar, and Van de Ven (2012) experimentally investigate third party perceptions of trustworthiness. Experimental subjects watched a random sample of shows and were asked to report a probability that a player would share. Column 3 of Table 6 reports our findings on the relation between beauty and perceived cooperativeness. Our subjects predict that attractive people will be *less* cooperative—this is statistically significant, although the effect is small. Thus there appears to be little reason to discriminate in favour of attractive subjects from the point of view of selecting a more cooperative partner.

V. Beauty and Selection

We now study the elimination decision, having established that there is no objective reason to discriminate against unattractive players either on the grounds of performance, perceptions about ability or confidence, or because they are or are thought to be more cooperative. Thus any bias against unattractive players in lead player selection decisions can plausibly be attributed to the lead players obtaining consumption value from having attractive co-players.

An important advantage of the rules of our game show is that in making the elimination decision, the lead player in any round is faced with a relatively simple *decision* problem, rather than a game. If the lead player chooses to eliminate player i then the lead player is decisive and i will play no further part in the game. In contrast, elimination decisions in other game shows (such as *The Weakest Link*, analyzed by Levitt 2004; and Antonovics, Arcidiacono, and Walsh 2005) are often made by majority voting, involving all the participants remaining at that stage. If a player j votes to eliminate i, then i may not be eliminated, and may in turn vote against j at a later stage. This implies that players have a strong incentive to vote to eliminate whoever they think others are going to vote against. Thus, there are multiple voting equilibria, and this is coupled with a strategic dynamic motive to vote with the majority. This may induce a significant role for irrelevant characteristics as possible focal points, even when players do not have any preference for discriminating on the basis of such a characteristic. In our game, these strategic considerations do not apply, since only the lead player votes and his vote is decisive. Thus evidence of discrimination can be attributed to lead player preferences.

Figure 1 shows the average attractiveness of all players in each round, and that of the players who are eliminated in that round. In each round, eliminated players are less attractive than average, and in consequence, average attractiveness increases steadily over the rounds. Other summary statistics confirm this picture. If a player is average-looking (that is, within one standard deviation of the mean), he or she has 0.4 probability of reaching the final round. An attractive player has a substantially higher probability of 0.51, while an unattractive player's probability is only 0.31.

We investigate in more detail the role of physical attractiveness in the selection decision by the lead player. We estimate a conditional logit model for the probability of being eliminated. The conditional logit is a natural framework for modelling choices from a set of alternatives. In our context, the alternatives are the players in the round, and each player i in show$j$ has a vector of attributes X_{ij} (gender, age, attractiveness). The conditional logit model has the form:

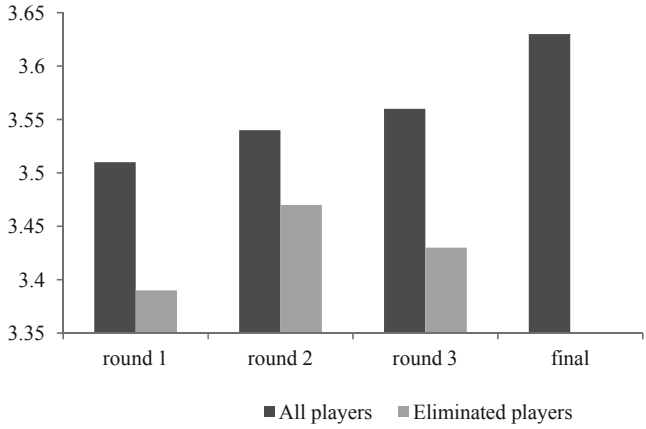

Figure 1
Average Attractiveness of All Players and Eliminated Players by Round.

$$(2) \quad p(y_{ij}=1)=\frac{\exp(\beta' X_{ij})}{\sum_i \exp(\beta' X_{ij})} \quad \text{for } i=1,\ldots 5,$$

where y_{ij} is an indicator variable, which takes value one when the player is eliminated from the show.[19] Note that the identification of the effects of variables relies on the variation in characteristics across players within an episode. If there is a show-specific element affecting all the players within the same episode then this is differenced out (so that there is no need to cluster the standard errors at the episode level). Similarly, this specification implicitly eliminates the importance of average differences in rater perceptions across the shows, since all five players have been rated by the same set of independent raters.

A. Discrimination in the First Round

Table 7 shows our results for the first round (we discuss subsequent rounds later). The odds ratio corresponding to attractiveness is below one and significant at the 10 percent level, implying that attractive players are less likely to be eliminated. In Columns 2 to 4 we use a dummy for the least attractive player within the candidates for elimination of an episode. We find that the least attractive player appears to be twice as likely to be eliminated than the other players.[20]

19. One important assumption for the validity of the conditional logit estimates is independence of irrelevant alternatives. Our tests show that this assumption is not rejected.
20. Our results are robust to alternative specifications. For example, we can consider players with attractiveness one standard deviation above the mean (> 4.2), and one standard deviation below the mean (< 2.8). The latter group is three times more likely to be eliminated than those in the middle group (between 2.8 and 4.2). The most attractive ones are less likely to be eliminated but the effect is not significant. Also, the most attractive player in the show is not significantly less likely to be eliminated. These results are available upon request from the authors.

Table 7
Probability of Being Eliminated at the End of the First Round

	(1)	(2)	(3)	(4)
Attractiveness				
Mean attractiveness	0.66*			
	(0.16)			
Least attractive		2.08***	2.14**	2.16***
		(0.59)	(0.64)	(0.64)
Performance				
Fourth highest	0.50**	0.49**	0.47**	0.49**
	(0.15)	(0.15)	(0.15)	(.16)
Third highest	0.28***	0.30***	0.32***	0.33***
	(0.10)	(0.11)	(0.13)	(0.13)
Second highest	0.19***	0.18***	0.21***	0.20***
	(0.08)	(0.08)	(0.11)	(0.09)
Behavioral				
Capital invested (percentage)			0.73	0.79
			(0.54)	(0.59)
Number of answers			1.05	1.02
			(0.16)	(0.14)
Perceived confidence			1.29	
			(0.38)	
Perceived least confident				0.68
				(0.23)
Perceived knowledge			0.81	
			(.022)	
Perceived least knowledgeable				0.79
				(0.59)
Female	0.76	0.74	0.75	0.71
	(0.23)	(0.22)	(0.25)	(0.23)
Age	1.00	1.00	0.99	1.00
	(0.02)	(0.02)	(0.02)	(0.02)
Number of observations	276	276	276	276

Notes: conditional logit estimates, reporting odds ratios. Standard errors in parentheses corresponding to the odds ratios. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

After attractiveness, the score ranking is the next best predictor of elimination. The player with the lowest score (the reference category) is twice as likely to be eliminated as the one ranked fourth, and more than five times as likely to be eliminated as the one with the highest score among the candidates for elimination. This also suggests that the lead player does not seek to eliminate "competition" for leadership by eliminating the player with next highest score, but instead seeks to max-

Table 8
Probability of Being eliminated at the End of the Second and Third Round

	Round 2		Round 3	
	(1)	(2)	(3)	(4)
Attractiveness	1.05		0.59	
	(0.33)		(0.22)	
Least attractive		1.05		1.84*
		(0.33)		(0.67)
Number of observations	207	207	138	138

Notes: Conditional logit estimates, reporting odds ratios. All regressions control for the same variables as Table 7, Column 3. Standard errors in parentheses corresponding to the odds ratios. * $p < 0.10$.

imize the overall score, by eliminating weak players. It is also intriguing that attractiveness plays such an important role despite the strong strategic motive for eliminating attractive players and retaining unattractive ones—since unattractive players are more likely to be discriminated against in future rounds, they provide an insurance value to the lead player, and since they are no less able than more attractive players, they are also not costly in terms of maximizing the total score over rounds.

As for other variables, age and gender are irrelevant in the selection decision. Controls for behavior during the game and measures of perceived ability and perceived confidence do not change the results and do not matter in the selection decision (Columns 3 and 4). Thus, even if perceived confidence was found to be positively correlated with attractiveness, this does not explain why the least attractive player is more likely to be eliminated. Column 3 controls for perceived confidence as a continuous variable, while Column 4 controls for dummies for the player perceived least confident among the players candidate for elimination. Thus, less attractive players are discriminated against for reasons that are uncorrelated with their performance or behavior during the game or with negative stereotypes.

B. Discrimination over the Rounds

Because less attractive players are eliminated in the first round, the population of players becomes more homogeneous in terms of attractiveness, which should reduce discriminatory practices. This will be a factor making us less likely to find discrimination in subsequent rounds. Table 8 show the results for Rounds 2 and 3. We find no evidence of discrimination based on attractiveness in the second round. In the third round, we find evidence of discrimination against the least attractive player, and the coefficient is roughly comparable to that in the first round.

C. Taste-based Discrimination?

Our results provide evidence that less attractive people suffer a penalty due to discrimination. We have established that this is unrelated to performance on the game show, so that such discrimination is not statistical. Thus discrimination in favor of attractive people appears to reflect consumption value considerations—discriminators enjoy being in the company of attractive people and are prepared to pay a price for this. Note that this consumption value could also arise indirectly, since TV show participants may enjoy being watched when they are surrounded by attractive individuals.[21] We cannot distinguish between these two different sorts of consumption value.

D. Costs of Discrimination and the Benefits of Beauty

We now provide an estimate of the benefits of beauty and the costs of discrimination. As in many other game shows, most contestants (more than 80 percent) go home empty handed. Moreover, even among those with positive earnings, the variance is very large (the standard deviation is €2,758 compared to the mean of €2,570). These factors imply that there is a high degree of variability in earnings.

Column 1 in Table 9 presents OLS estimates of the determinants of log earnings. An increase in attractiveness of one standard deviation increases earnings by 41 percentage points. Despite the large standard error, this is significantly different from zero at the 10 percent level. An alternative estimate of the beauty premium is based on estimating the probability of reaching the final round. Columns 2 and 3 provide estimates of the probability of reaching the final round with attractiveness as a continuous variable (Column 2) or as a discrete variable, based on the ranking of players within each show according to attractiveness (Column 3). The results confirm our previous findings: the two most attractive players are about twice as likely to reach the final round than the least attractive player (50 percent against 27 percent with a standard error of eight percentage points). This means that, on average, their expected earnings are twice as large. Note that if we take a very conservative approach and consider the lower bound of the 95 percent confidence interval for the probability of reaching the final round, we still have a difference of seven percentage points in terms of probability of reaching the final round between the most and least attractive players. In this case, the estimated lower bound on the beauty premium equals 9 percent, which remains substantial.

Turning to the costs of discrimination, stakes are substantial on the game show. By eliminating the least attractive players instead of players who would maximize their monetary payoff, players implicitly pay a price for keeping more attractive players in the game. While we cannot directly calculate the price for keeping more attractive players in the game in the first or second round (since we do not observe the earnings of those who are eliminated), we can do a back-of-the-envelope cal-

21. The evidence in psychology is mixed on whether individuals benefit from being amongst attractive peers. Sigall and Landy (1973) provide evidence that people are more favorably evaluated when they are in the company of an attractive person. However, other evidence shows that comparing oneself unfavorably to another on the basis of appearance may lead to lower self-esteem and dissatisfaction with one's own appearance—see Myers and Crowthers (2009) for an overview study.

Table 9
The Beauty Premium

	Log of earnings	Probability of reaching final round	
	(1)	(2)	(3)
Attractiveness	0.41*	1.42*	
	(0.24)	(0.26)	
Beauty Rank 5 (least attractive)			
Beauty Rank 4			1.57
			(0.52)
Beauty Rank 3			1.29
			(0.45)
Beauty Rank 2			2.14**
			(0.73)
Beauty Rank 1 (most attractive)			1.89**
			(0.64)
Age	−0.005	0.98	0.98
	(0.018)	(0.01)	(0.01)
Gender	−0.46	1.07	1.12
	(0.31)	(0.23)	(0.24)
Constant	0.15		
	(1.13)		
Number of observations	345	345	345
Pseudo *R*-squared	0.01	0.03	0.04

Notes: Column 1: OLS estimates allowing for a show-specific random effect. Columns 2 and 3: conditional logit estimates, reporting odds ratios. Standard errors in parentheses. ** $p < 0.05$, * $p < 0.1$.

culation of the price they pay by eliminating the least attractive player in the third round, where we do observe perfectly the earnings of the contestants. However, the sample of observations is relatively small and the results should be taken with caution.

We can identify 10 of the 69 episodes where in the final round, the lead player eliminates the least attractive player, and where this player has a higher score than the player who is chosen to stay. The average difference in scores between the eliminated and chosen player is €440 in these episodes. Thus the average cost of eliminating the least attractive player is €440, in terms of the total prize money at stake.[22] This estimate of the costs of discrimination is a lower bound since we are

22. It might be that lead players expect those with a lower score to be more cooperative, for which there is evidence (Belot, Bhaskar, and Van de Ven 2010). But this does not explain why the player with a lower score is rarely chosen to play the final if he is the least attractive player.

only looking at costs associated with the third round, and not earlier rounds, since a precise imputation of financial costs in earlier rounds is more difficult.

E. Insidious Discrimination?

Are third parties aware of the possibility of discriminatory behavior? And if they are aware, is this knowledge conscious or subconscious? These questions are relevant—if discrimination is not perceived by third parties, it may persist even under public scrutiny.[23] To investigate these questions, we adopt a novel experimental procedure, by asking subjects to predict the elimination decision at the end of the first round of quiz questions. We focus on the first round since we do not want subject predictions to be influenced by learning from observing selection decisions in previous rounds. Our subjects (71 students from the University of Amsterdam) were shown a trailer and given a handout, setting out the overall structure of the game show, and then shown a random sample of seven episodes. They were informed in advance of the identity of the lead player in the first round, so that they could focus on predicting the lead player's decision. The subject was asked to assign a probability to each of the other players being eliminated by the lead player. Subjects were rewarded by using a quadratic scoring rule, giving them strict incentives to report their true beliefs. Subjects were not informed of the actual elimination decision in any episode, so as to prevent any learning. At the end, we asked subjects a qualitative question: What, in their opinion, are the most important factors determining the elimination decision? Subjects were permitted to list up to four such factors.[24]

Table 10 reports ordinary least squares estimates, where the dependent variable is the prediction of subject i regarding the elimination probability of player j. We have 71 subjects, making predictions for four players in each of seven episodes, giving us 1,988 observations.

Our main finding is that our subjects do perceive that less attractive players are more likely to be eliminated.[25] That is, our subjects are at least implicitly aware that discrimination on the grounds of attractiveness is likely to occur. They predict that an increase in one standard deviation in attractiveness should decrease the probability of being chosen by 1.7 percentage points. This is substantially less than what we have estimated in the actual data (the corresponding number is nine percentage points). Interestingly, attractiveness is not mentioned by anyone among the two most important criteria influencing the selection decision, and only four subjects list attractiveness at all. Of course, it is possible that students are relatively inexperienced with discrimination based on attractiveness since they have little or no experience

23. This is related but not identical to the notion of implicit discrimination (see Greenwald, McGhee, and Schwartz 1998, and Bertrand and Mullainathan 2005), which uses psychological tests to measure implicit bias, say, against African-Americans. It is worth noting that researchers have found an implicit bias against older people and the overweight.

24. The subjects also filled out a questionnaire on their background characteristics. They earned on average €18 (for 90 minutes), including a €4 show up fee. The full set of instructions is in the appendix.

25. Second-ranked players are predicted to be more likely to be eliminated, an effect which is not present in the actual selection decision. This appears to be due to subjects incorrectly thinking that the first ranked players want to eliminate rivals for leadership.

Table 10
Predictions of Elimination in the First Round

	(1)	(2)	(3)	(4)
Attractiveness	−2.58**	−2.39***		
	(1.03)	(0.84)		
Least attractive			2.42	2.17*
			(1.61)	(1.32)
Age	−0.09	−0.03	−0.04	0.01
	(0.07)	(0.06)	(0.07)	(0.06)
Gender	−3.58***	−3.55***	−3.96***	−3.92***
	(1.39)	(1.11)	(1.38)	(1.10)
Score Rank 4th		−6.64***		−6.63***
		(1.39)		(1.39)
Score Rank 3d		−5.38***		−5.09***
		(1.39)		(1.37)
Score Rank 2d		8.35***		8.53***
		(1.74)		(1.75)
Constant	38.29***	36.73***	27.09***	26.23***
	(5.19)	(4.47)	(2.64)	(2.31)
Number of observations	1,988	1,988	1,988	1,988

Notes: OLS estimates. Standard errors in parentheses, clustered by player. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

on the labour market. Nevertheless, this provides suggestive evidence that discrimination is insidious, since subjects appear to be implicitly aware of its existence, but not consciously so. This is particularly noteworthy since a subject is detailing the motivations of players on the game show (rather than himself), and would therefore have little reason to lie or practice self-deception.[26]

Discrimination on the basis of looks appears to meet with less social disapproval than other forms of discrimination. Following a report in "New Scientist" summarizing the findings of the present paper, it was covered by newspapers in several countries including the United Kingdom and the Netherlands. Press coverage one of the largest Dutch newspapers "De Telegraaf" triggered a large public response—within a day, 156 readers had posted a response on the newspaper's website. Many people simply found the findings somewhat amusing rather than a matter for social concern. We found that 36 of these suggested that beautiful people "deserve" to be treated better than the "ugly"—some comments were quite offensive about unattractive people. Only six people explicitly condemned this kind of discriminatory behavior.

26. Fershtman and Gneezy (2001) find that experimental subjects who discriminate seem to be unaware that they do so.

Our experimental findings (and the public response) raise important issues of social concern regarding the persistence of discrimination. Following Becker (1957), economists have focused on the extent to which the forces of competition and profit maximization eliminate discrimination. It is arguable that the social stigma associated with racial or gender discrimination is no less important a force in its elimination. To the extent that discrimination on the basis of looks or age occurs, but is either insidious or is "acceptable" even when perceived, its persistence is enhanced.

VI. Concluding Comments

To summarize, our main finding is that beauty is only skin-deep, and has no implications for a person's performance or their cooperativeness. Nevertheless, it is an attribute well worth having. Attractive players earn a premium, that arises from the reluctance of other players to eliminate them. This seems to reflect consumption value considerations on the part of the other players in the game. The preference for the beautiful is therefore a form of taste based discrimination. Our finding is also noteworthy since participants on a TV show may be reluctant to discriminate, since their behavior is subject to public scrutiny. While discrimination on the basis of gender, or race are rightly frowned upon, discrimination based on a person's physical appearance is less remarked upon. Indeed, it is likely that discriminators, the discriminated, and third parties are less aware of the phenomenon, so that it is, to some extent, insidious. Our experimental evidence on third party perceptions of discrimination is suggestive in this regard. We believe that discrimination on the basis of less obvious criteria such as attractiveness is likely to have qualitatively different characteristics from discrimination based on recognized categories. Without overstating the external validity of our results, we believe that this raises important questions for society and social policy, and merits further research.

References

Altonji, Joseph G., and Rebecca M. Blank. 1999. "Race and Gender in the Labor Market." In *Handbook of Labor Economics*, ed. Orley Ashenfelter and David Card, 3143–259. Amsterdam: North Holland.

Andreoni, James, and Ragan Petrie. 2008. "Beauty, Gender and Stereotypes: Evidence from Laboratory Experiments." *Journal of Economic Psychology* 29(1):73–93.

Antonovics, Kate, Peter Arcidiacono, and Randall Walsh. 2005. "Games and Discrimination: Lessons from 'The Weakest Link'." *Journal of Human Resources* 40(4):918–47.

Becker, Gary. 1957. *Economics of Discrimination.* Chicago University Press, Chicago.

Belot, Michèle, V. Bhaskar, and Jeroen van de Ven. 2010. "Promises and Cooperation: Evidence from a TV Game Show." *Journal of Economic Behavior & Organization* 73(3):396–05.

———. 2012. "Can Observers Predict Trustworthiness?" *Review of Economics and Statistics.* 94(4):246–59.

Benjamin, Daniel J., and Jesse M. Shapiro. 2006. "The Rational Voter, Thinly Sliced: Personal Appeal as an Election Forecaster." Harvard University. Unpublished.

Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan. 2005. "Implicit Discrimination." *American Economic Review* 95(2):94–8.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94(4):991–1013.
Biddle, Jeffe, and Daniel Hamermesh. 1998. "Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre." *Journal of Labor Economics* 16(1):172–201.
CBS. 2006. Central Bureau of Statistics Netherlands, statline.cbs.nl.
Doran, Kirk, and Joni Hersch. 2009. "The Beauty Premium is Not Robust." Unpublished
Eckel, Catherine, and Rick Wilson. 2004. "Detecting Trustworthiness: Does Beauty Confound Intuition?" Dallas: University of Dallas. Unpublished.
Fershtman, Chaim, and Uri Gneezy. 2001. "Discrimination in a Segmented Society: An Experimental Approach." *Quarterly Journal of Economics* 116(1):351–77.
Greenwald, Anthony, Debbie McGhee, and Jordan Schwartz. 1998. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology* 74(6):1464–80.
Hamermesh, Daniel. 2011. *Beauty Pays: Why Attractive People Are More Successful.* Princeton: Princeton University Press.
Hamermesh, Daniel, and Jeffe Biddle. 1994. "Beauty and the Labor Market." *American Economic Review* 84(5):1174–94.
Harrison, Glenn, and John List. 2004. "Field Experiments." *Journal of Economic Literature* 42(4):1009–55.
Heckman, James. 1998. "Detecting Discrimination." *Journal of Economic Perspectives,* 12(2):101–16.
Landry, Craig, Andreas Lange, John List, Michael Price, and Nicholas Rupp. 2006. "Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment." *Quarterly Journal of Economics* 121(2):747–82.
Levitt, Steven. 2004. "Testing Theories of Discrimination: Evidence from Weakest Link." *Journal of Law and Economics* 47(1):431–52.
List, John. 2006. "Friend or Foe? A Natural Experiment of the Prisoner's Dilemma." *Review of Economics and Statistics* 88(3):463–71.
Mocan, Naci, and Erdal Tekin. 2010. "Ugly Criminals." *Review of Economics and Statistics* 92(1):15–30.
Mobius, Markus, and Tanya Rosenblat. 2006. "Why Beauty Matters." *American Economic Review* 96(1):222–35.
Mulford, Matthew, John Orbell, Catherine Shatto, and Jean Stockard. 1998. "Physical attractiveness, opportunity, and success in everyday exchange." *American Journal of Sociology* 103(6):1565–93.
Myers, Taryn, and Janis Crowther. 2009. "Social comparison as a predictor of body dissatisfaction: A meta-analytic review." *Journal of Abnormal Psychology* 118(4): 683–98.
Niederle, Muriel, and Lisa Vesterlund. 2007. "Do Women Shy away from Competition? Do Men Compete too Much?." *Quarterly Journal of Economics* 122(3):1067–101.
Pfann, Gerard, Jeff Biddle, Daniel Hamermesh, and Ciska Bosman. 2000. "Business Success and Businesses' Beauty Capital." *Economics Letters* 67(2):201–207.
Post, Thierry, Marcel van den Assem, Guido Baltussen, and Richard Thaler. 2008. "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show." *American Economic Review* 98(1):38–71.
Sigall, Harold, and David Landy. 1973. "Radiating Beauty: Effects of Having a Physically Attractive Partner on Person Perception." *Journal of Personality and Social Psychology* 28(2):218–24.
Solnick, Sara, and Maurice Schweitzer. 1999. "The Influence of Physical Attractiveness and Gender on Ultimatum Game Decisions." *Oganizational Behavior and Human Decision Process* 79(3):199–215.