Insidious Discrimination? Disentangling the Beauty Premium on a Game Show*

Michèle Belot[†], V. Bhaskar[‡] & Jeroen van de Ven[§]

April 23, 2007

Abstract

This paper analyzes behavior on a TV game show where players' monetary payoffs depend upon an array of factors, including ability in answering questions, perceived cooperativeness and the willingness of other players to choose them. We find a substantial beauty premium and are able to disentangle contributing factors. Attractive players perform no differently from less attractive ones, on every dimension. They also exhibit and engender the same degree of cooperativeness. Nevertheless, attractive players are substantially less likely to be eliminated by their peers. Our results suggest a consumption value basis for the beauty premium, and it appears that this is a form of insidious discrimination.

Keywords: beauty premium, discrimination.

JEL Classification Numbers: C93, D63, J15, J16.

*We are grateful to Dan Hamermesh for very useful comments and suggestions, and to Debbie Hall for research assistance.

[†]Department of Economics and Institute for Social and Economic Research (ISER), University of Essex, mbelot@essex.ac.uk

[‡]Department of Economics, University College London, v.bhaskar@ucl.ac.uk

[§]Department of Economics and Amsterdam Centre for Law and Economics (ACLE), University of Amsterdam. j.vandeven@uva.nl

1 Introduction

In a surprising and influential paper, Hamermesh and Biddle (1994) found a substantial beauty premium in the labor market, of the order of 15%. While there are several competing explanations for this premium, its source remains an open question. Attractiveness may be correlated with other unobservable productive attributes such as health, education or other types of human capital. Attractive people may be more confident, thus enhancing their social skills in the workplace.[1] 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) in investing in looking good.[2] Perhaps the simplest (and least palatable) explanation is that attractiveness has a "consumption-value", either to the customers of the firm, fellow employees, or the boss. The beauty premium in this case is a form of taste-based discrimination, as discussed in Becker (1957). While anecdotal evidence on the importance of consumption value considerations in the hiring of air hostesses or waitresses certainly exists, the question remains whether this is a more general phenomenon. More generally, as the literature on racial/gender discrimination shows (see Altonji and Black, 1999; Heckman, 1998), establishing discrimination and distinguishing between statistical and taste-based discrimination is difficult.

The main contribution of this paper is to disentangle the reasons for the beauty premium. We do this in the context of a TV game show where participants are engaged in a variety of "tasks". The premium we find is substantial – attractive players, irrespective of gender, earn on average twice as much as unattractive players, i.e. about €300 more.[3] While the finding

[1] See Mobius and Rosenblat (2006) for interesting experimental evidence for this hypothesis. A fascinating study by Persico et al. (2004) finds that adolescent height has an important effect on wages, suggesting an important role for social and psychological factors in teenagers.

[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.

[3] Attractive players are those with a beauty rating at least one standard deviation above

of a beauty premium is not novel, the main advantage of our setting is that we are able to disentangle its components. Specifically, we can ask: are attractive people more productive, and do they exhibit greater confidence? Are they more cooperative or do they engender cooperation? Are they more likely to be chosen by their peers when a selection decision has to be made? We are able to answer these questions since the game show has a rich structure, with participants being involved in a number of different tasks and decisions. It takes place over three rounds, in which players accumulate "earnings" by answering quiz questions, and their earnings depend on the accuracy of their answers, on how quickly they press the buzzer and also on their "investment decisions". Earnings therefore depend upon ability as well as a player's confidence. This allows us to study the effect of attractiveness as well as other player characteristics upon performance. At the end of each round, the lead player – the one with the highest earnings – decides which one of the remaining players to eliminate. This allows us to study the role of attractiveness and gender (in addition to performance) upon the selection decision. After the final round, when only two players remain, they play a prisoner's dilemma game, allowing us to study the role of beauty on cooperation. The median stake in this prisoner's dilemma game is €1,683, so that the monetary consequences of players' decisions are substantial.

We find that the beauty premium cannot be attributed to any aspect of performance – attractive people fare no differently from unattractive ones in answering questions or in investment behavior. They are no more likely to cooperate in the final stage, and opponents also behave no differently vis-a-vis them. Nevertheless, when one person has to be eliminated by the lead player in any round, the least attractive player of a show is significantly more likely to be chosen. This is illustrated by Figure 1, which shows the average attractiveness of all players in each round, and that of the players

the mean, while unattractive ones have a rating at least one standard deviation below the mean.

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.

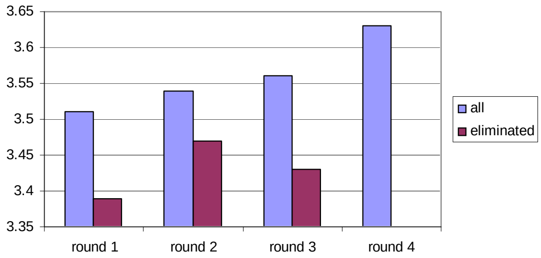

Figure 1: Average beauty per round of all players and eliminated players.

Our results support a "consumption value" basis for the beauty premium in our context. Discriminating in favor of attractive players is costly to the lead player, implying an adverse selection in terms of earning potential. These costs are significant; a back-of-the-envelope calculation suggests that the losses amount to €350 on average, roughly 20 percent of the median stake. However, the costs inflicted upon the discriminated are even larger.

These findings suggest that discrimination on the basis of looks without any performance rationale may be a significant social phenomenon, warranting further investigation. Furthermore, such discrimination may be *insidious*[4] — both discriminators and the discriminated against may be unaware of the phenomenon. Indeed, those who are perceived as being unattractive may well consider themselves to be "above average" – the social psychology literature (e.g. Myers, 1996) has established that individuals have positive self assessment on a number of dimensions that are subjective and socially

[4] Insidious is defined as "working or spreading harmfully in a subtle or stealthy manner" (answers.com).

desirable. [5] It is likely that insidious discrimination has very different effects from discrimination on the basis of clearly recognized categories such as race or gender.

The remainder of this paper is organized as follows. Section 2 describes the game show and our construction of the measure of attractiveness. Section 3 analyzes behavior on the game show, in terms of performance and cooperation. Section 4 studies the selection decisions by lead players and establishes that attractive players benefit in this context. Section 5 briefly considers whether there is discrimination on the basis of a more obvious characteristic – gender – in the context of our show. We find significant differences in performance and behavior between men, and conclude that there does not appear to be any evidence of discrimination against women. The final section reviews the related literature and concludes.

2 Description of the data

2.1 The game show

We use data from all 69 episodes of the game show *Shafted*, broadcast in the Netherlands in 2002. The game starts with five players, who accumulate earnings by answering quiz questions. At the end of each of three rounds, the player with the highest earnings in that round – the *lead player* henceforth – must choose one of the remaining players for elimination. An eliminated player has no further role in the game and loses all of his or her earnings. After three rounds, the 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. The finalists simultaneously decide whether to share (S) or to grab (G). The monetary payoffs, as depicted in Table 1, correspond to a generalized prisoner's dilemma, where G is a weakly dominant strategy.[6]

[5] Positive self-assessment in the context of attractiveness has not been explored much; however, Gabriel et al. (1994) present some evidence for its existence.

[6] Players share 43 percent of of the time, suggesting that non-monetary motivations are important. The prisoner's dilemma stage is similar to that in the game show *Friend or*

Table 1: Payoff matrix

	S	G
S	$\frac{1}{2}E, \frac{1}{2}E$	$0, E$
G	$E, 0$	$0, 0$

Players earn money by answering questions in each round as follows. At the beginning of each round a player must decide how much of his capital to "invest" in answering each question. Only the player who first presses the buzzer gets to answer the question. A correct answer yields y_i, his chosen level of investment, while an incorrect answer earns $-y_i$. A player whose capital falls below his or her chosen investment, may not answer any further questions. Each round ends with a bonus question, for which players choose new investments. The player with the highest investment must answer the bonus question, again gaining or losing the investment. The player with the highest score at the end of the round chooses one of the remaining players for elimination. Play then proceeds to the next round, where the initial capital for all players equals the earnings of the lead player in the previous round.[7]

Table 2 presents summary statistics of the game show and the participants. The average age is 34.4 and about a third of participants are women. The total prize at stake varies between €380 and €26,600, with a median value of €1,683. These are considerable sums given that the median monthly income in the Netherlands is roughly €1,200.[8] Players choose to play share about 40% of the time in the final round.

Foe (see List (2004, 2006), Kalist (2005) and Oberholzer-Gee et al. (2004)). However, the overall game is significantly different from the point of view of the concerns in the present paper. In particular, our game allows us to measure *individual* performance, and selection decisions by the lead player in the light of this performance.

[7] A player's capital in the first round is determined as follows. Six prospective participants simultaneously choose a number between 1 and 100. The player choosing the highest number is eliminated, and the remaining players have an initial capital equal to their chosen number. The format of this preliminary elimination stage was slightly different in the first few episodes.

[8] This is the median disposable income of a full-time employed person in 2000 (*Statistics Netherlands*, available at www.cbs.nl).

Table 2: Summary statistics

	Mean	Min	Max
Mean age (years) (N = 345)	34.4	18	64
Percentage women (N = 345)	34.8	-	-
Prize (€) (N = 138)	2,976	380	26600
Percentage sharing (N = 138)	42.8	-	-

2.2 Measure of beauty

We construct a measure of beauty based on independent ratings. The 345 participants on the game show were rated on a scale from 1 (very unattractive) to 7 (very attractive) by approximately 10 raters, balanced by gender. Raters were recruited in public spaces to obtain a representative sample of the adult population. Raters were on average 31.7 years old, which is close to the average age of 33 years among participants on the game show. In total 120 raters were recruited, and each rater rated 30 participants. This was based on watching short silent video fragments of the game show in which a candidate introduced herself. We ensured that all five participants on any show were rated by the same set of raters, while varying the order in which the shows were presented.

Our simple measure of attractiveness is the average of ratings (averaged across raters) for each candidate.[9] There is a high degree of concurrence on attractiveness across raters. Across sub-samples of judges who rated the same sample of participants, the Cronbach's alpha ranged from 0.70 to 0.85, showing high agreement. Table 3 reports some summary statistics

[9] Raters may have different perceptions on the average beauty. To correct for such differences, some other studies use standardized measures. Each rating is adjusted for the mean rating of that rater, and is then normalized by dividing by the standard error (see e.g. Mobius and Rosenblat, 2006). We chose to have many different raters rating different subjects. The means are not comparable among raters, because the sample of episodes they rated were partly different. We therefore prefer to use the raw data. However, if we standardize ratings for the subsamples of ratings made for the same players, the results remain very similar. Similarly, since we have many raters for each participant, we can weight the observations in our sample by the degree of agreement among raters. All our results were robust to such a weighting process, where each observation was weighted according to the inverse of the standard deviation in ratings.

of the ratings. Raters were told to use the benchmark 4 for the average attractiveness in the population. Somewhat surprisingly, the average rating across participants is 3.51. Beauty is negatively correlated with age and women are on average rated as being more attractive than men. Average ratings are also more variable across women than men, consistent with other studies (Hamermesh and Biddle, 1994).

Table 3: Attractiveness - Summary statistics

	Mean	(st. dev.)	Min	Max
All ($N = 345$)	3.51	(.69)	1.7	5.75
Men ($N = 225$)	3.45	(.63)	2.0	5.20
Women ($N = 120$)	3.62	(.79)	1.7	5.75
Age ≥ 34 ($N = 177$)	3.30	(.60)	1.7	4.80
Age < 34 ($N = 176$)	3.71	(.71)	2.2	5.75

3 Decomposing the beauty premium

Consistent with earlier research, we find evidence of a sizeable beauty premium. Players with an attractiveness level more than one standard deviation above the mean earn roughly €300 more than those whose attractiveness is more than one standard deviation below the mean, i.e. the earnings of the former are twice as large as those of the latter. This large premium is partly due to the fact that those who are eliminated have zero earnings. Given the many determinants of payoffs in the show, this premium could be attributable to a variety of factors, and we study each of these in turn.

3.1 Beauty and performance

We first investigate the relationship between beauty and performance in answering quiz questions. The total earnings of a player in a round is probably the most important measure of overall performance. The player with the highest earnings becomes the lead player for that round, making the elimination decision. In the first two rounds the lead player's earnings determine the initial capital for all remaining players in the next round. In the third

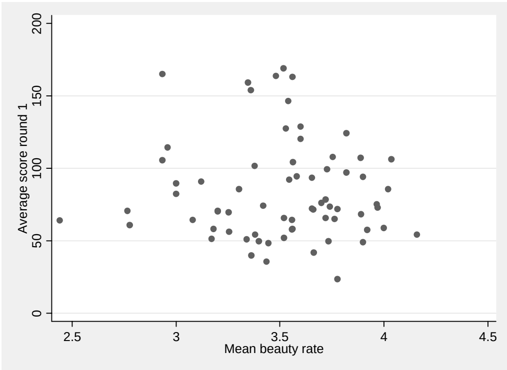

Figure 2: Average final score in round 1 by mean beauty rate of players per episode.

round, a player's earnings are added to the total stake, if he is either the lead player or not eliminated by the lead player.

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. Hence, we cannot draw conclusions from simple cross-correlations over episodes. But if attractive people have a better overall ability, shows with more attractive participants should be associated with a better performance overall. Figure 2 plots the average final score of players at the end of the first round, against the average attractiveness of participants in the show. We find no significant correlation between the two (the coefficient of correlation is -.05, with a corresponding p-value of 0.65). Less attractive populations of players perform as well as more attractive populations of players.

Two player positions, the first and the last, are of particular interest. The player who is ranked first in terms of earnings must choose one of the others for elimination, and cannot be eliminated himself in that round. The last ranked player is an obvious candidate for elimination. We estimate a conditional logit model for the probability of being in the first or last position conditional on participating in the same episode. 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 serie j has a vector of attributes X_{ij} (gender, age, attractiveness). The conditional logit model has the form:

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

where y_{ij} is an indicator variable which takes value one when the player is in the first (resp. last) position. One important assumption for the validity of the conditional logit estimates is independence of irrelevant alternatives. In the appendix, we discuss this assumption and test for its validity, finding that this assumption is not rejected.

Table 4 reports the results. Columns 1 and 2 show the odds ratios of ending up in the first or last position. Note that the reference value is one, i.e. if the odds ratio is 1, this means that the attribute is irrelevant in determining a player's relative position. An odds ratio below (above) 1 means that the attribute decreases (increases) the probability of ending up in a particular position. None of the coefficients is significantly different from 1, so the likelihood of ending up first or last does not depend on gender, age, or attractiveness.

To investigate further the correlation between the players' attributes and their rank in terms of earnings, we estimate a rank-ordered logit model that explicitly takes account of the ranking of players within a game and specifies this ranking as a function of their relative attributes. The rank-ordered logit specification is a refinement of the conditional logit specification, using information not only on which alternative is "best", but also on the ranking of these alternatives. The results are reported in the third column of Table 4. We find no clear correlation between any of these characteristics and earnings ranking. In particular, there is no evidence that attractive people rank differently from unattractive ones.

Table 4 - Earnings ranking in first round and individual characteristics						
	First position		Last position		Earnings ranking	
	Odds ratios		Odds ratios			
Mean attractiveness	1.16	(.26)	1.33	(.29)	-.08	(.12)
Female	.79	(.21)	.93	(.24)	-.05	(.14)
Age	1.00	(.02)	1.01	(.02)	-.01	(.01)
N. obs.	345		345		345	
Model	Cond. logit		Cond. logit		Rank-ordered logit	

3.2 Decomposing the performance measure

Even if the overall performance of players is uncorrelated with attractiveness, it could be that players differ in how they accumulate earnings. Recall that the earnings in the game depend on the initial capital, the share of capital invested, and the performance in answering questions. Mobius and Rosenblat (2006) find that more attractive people are more confident and appear more productive to others (although they are not). In this context, we would expect highly confident players to be more active in the game in terms of investment and answering questions.

We study the behavior of players in each round, both in regular questions (where the answering decision is taken after players see the question) and in the bonus questions (where the answering decision is linked to the investment decision and is taken before players see the question).

3.2.1 Behavior in the regular questions

Let us now consider how the decision to answer depends upon a player's ability, confidence, and risk aversion. Recall that by being the first to press the buzzer, player i faces a lottery where he gets y_i (his chosen investment level) if he is correct, and gets $-y_i$ if incorrect. Let p_i be the subjective probability assigned by i to his answer being correct.[10] Let x_i be his current

[10] As the player thinks about the question, his subjective probability will evolve over time. The analysis that follows pertains to any instant of time, so the continuation value from not pressing the button ($V_i(x_i)$) includes the option value of waiting, and possibly pressing the buzzer in the future, if no one else presses in interim.

score, and let $V_i(.)$ denote his expected continuation value in the game after this question. It is optimal for player i to answer to the question if:

$$p_i V_i(x_i + y_i) + (1 - p_i) V_i(x_i - y_i) \geq V_i(x_i).$$

That is, i will answer the question if p_i exceeds a critical threshold:

$$p_i \geq \frac{V_i(x_i) - V_i(x_i - y_i)}{V_i(x_i + y_i) - V_i(x_i - y_i)} \equiv \bar{p}_i,$$

where $\bar{p}_i \in (0, 1)$ as long as V_i is strictly increasing in the player's earnings. Notice that the threshold value, $\bar{p}_i$, is larger if the player is more risk averse, i.e. if $V_i(.)$ is more concave. Since we observe the frequency of correct answers for every player, we have an estimate of q_i, the objective probability that a player is correct conditional on answering. That is we have an estimate of $\mathbf{E}(q_i | p_i \geq \bar{p}_i)$. A player who is more risk averse will have a higher threshold value $\bar{p}_i$, and will therefore answer fewer questions but be observed to answer a greater proportion of questions correctly. On the other hand, a player who is more confident – i.e. has a greater value of p_i for a given q_i – will answer more questions and will make more mistakes in his answers. In other words, risk aversion and lower confidence act in very much the same way, in reducing both the number of answers and also the proportion of incorrect answers. On the other hand, if a player is less knowledgeable, and objectively has a lower value of q_i, this will ceteris paribus reduce the number of answers but not raise the proportion of correct answers.

Table 5 - Initial capital and share invested in each round of the game
OLS estimates

	Initial capital		Share invested					
	(1)		(2)		(3)		(4)	
			Round 1		Round 2		Round 3	
Attractiveness	.609	(.805)	-.032	(.020)	-.024	(.028)	.009	(.034)
Female	-.138	(1.089)	-.011	(.025)	-.041	(.032)	-.058	(.038)
Age	.096	(.061)	-.004	(.002)***	-.009	(.002)***	-.006	(.002)***
Constant	43.05	(4.70)***	.940	(.096)***	.993	(.132)***	.845	(.165)***
Adjusted R^2	.32		.025		.059		.044	
N. obs.	345		345		276		207	

Notes: Standard errors are clustered by episode. Dummies are included for episodes that had a slightly different setup for the first regression (col. 1) (see also footnote 5). Significance levels: *: 10 percent, **: 5 percent, ***: 1 percent.

We decompose the overall performance in the game into these different components and examine whether there is a correlation between these decisions and attractiveness.

The first decision of players regards their initial capital and the level of investment for the 10 regular questions. Table 5 shows OLS estimates for the level of initial capital (regression (1)) and the share invested in each round (regressions (2)-(4)). We find no clear correlation between these two variables and attractiveness. The only variable affecting the share invested is age, with older players investing slightly less.

The second decision is whether to answer or not, given the share invested. Since each question can be answered by one player at most, the decisions to answer are mutually exclusive. We investigate whether some characteristics (such as attractiveness and gender) make a player more likely to be the one answering the question. We estimate a conditional logit model, where the probability of answering is a function of the characteristics of the player answering *relative* to the characteristics of the players competing with him. Then we investigate whether these characteristics affect the probability of answering correctly, conditional on answering.

Table 6 - Probability of answering and correct answer (conditional on answering)
Odds ratios (Conditional logit estimates)

	Round 1		Round 2		Round 3	
	Answer	Correct	Answer	Correct	Answer	Correct
	Cond. logit	Logit	Cond. logit	Logit	Cond. logit	Logit
Attractiveness	.98	.85	.99	1.27	1.02	1.03
	(.07)	(.16)	(.08)	(.26)	(.09)	(.19)
Age	1.00	.99	.99	1.01	1.01**	1.01
	(.01)	(.01)	(.01)	(.02)	(.01)	(.01)
Female	.69***	1.15	.81**	.61*	.75**	.88
	(.07)	(.31)	(.09)	(.17)	(.09)	(.23)
Share invested	.74*	.96	.94	1.24	.96	.94
	(.14)	(.46)	(.16)	(.58)	(.18)	(.40)
Chose the topic	1.44***	1.08	1.64***	1.26	1.85***	1.77*
	(.20)	(.32)	(.22)	(.44)	(.25)	(.56)
N. obs.	3450	580	2760	536	2070	484

Notes: logit estimates include a random effect for each participant.
Significance levels: *: 10 percent, **: 5 percent, ***: 1 percent.

Table 6 reports the results, and we find no systematic difference in behavior with respect to attractiveness. This is in contrast with the findings of Mobius and Rosenblat (2006), who find, in an experimental setting, that more attractive players are also more self-confident. If more attractive players were more confident, they should be more likely to answer and, conditional on answering, perform worse. We do find a systematic difference according to gender: Women are much less likely to answer a question. This is what you would expect if women are more risk averse or less confident. However, conditional on answering, they actually do not perform better than men, what you would expect with risk aversion or lack of confidence. In the second round, we even find that they are significantly less likely to answer correctly. Overall, 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.[11] Finally, we included a dummy variable identifying the player who chose the topic of the questions. In the first forty episodes, each round of questions was in the same topic, and the topic was chosen by the winner of the last round (or the player with the highest capital in the first round). We find that the participant who has chosen the topic of the questions in the round is more likely to answer. Note that this variable captures partly an ability effect. This is not true in the first round, where the player choosing the topic is the one with the highest initial capital.[12]

3.2.2 Behavior in the bonus questions

The last determinant of earnings is the bonus question at the end of each round. Players decide on their investment *before* they see the question and the player with the highest investment gets to answer the question. The bonus questions are generally more subtle and arguably more difficult than the regular questions. It is therefore not surprising that the success rate in answering the bonus question is much lower (slightly above 50%) than for the regular questions (above 75%).

Answering the bonus question has strong elements of team production. In rounds 1 and 2, the capital for all players who progress to the next round, $t+1$, is set equal to the earnings of the player with the greatest earnings in round t. It is plausible to assume that a player with larger earnings is at least as likely to answer correctly as a player with smaller earnings. This implies that the expected starting capital for the next round is maximized if the first player answers the question and invests all her earnings. In case of an incorrect answer, the starting capital in the next round will then be equal to the earnings of player 2. If players are relatively risk averse, player 2 should

[11] In this context, the finding of Gneezy et al. (2003) that women do less well than men in a competitive environment, is relevant. We also did not find any evidence that the topics of the questions could explain gender differences in performance.

[12] These specifications do not correct for the possible correlation in answering probabilities across questions from the same episode, so that the standard errors could be underestimated. We have obtained very similar results with a Dirichlet specification (using the Stata command 'dirmul' proposed by Guimaraes and Lindrooth (2005)).

always invest her entire capital and answer the question. Intuitively, if the second player answers, this provides greater insurance since the earnings of the first ranked player become the fall-back option. Irrespective of risk preferences, from a team point of view, no lower ranked player should ever answer.

In practice, players 1 and 2 often invest less than their entire earnings and often one of the other players gets to answer the question. From a team perspective this is inefficient, resulting in an expected monetary loss per round of around 15 percent of earnings – since earnings are effectively compounded across rounds, the overall losses are even larger.[13] This can be understood from an individual perspective. For instance, if the first ranked player invests her entire earnings, she will end up last in the event of an incorrect answer and becomes a prime candidate for elimination. Thus investment decisions by first or second ranked players are likely to be risk-averse, even though team production considerations would prescribe risk-taking. Similarly, individuals who are low ranked may invest aggressively in order to improve their ranking so possibly securing their own position in the next round.[14] Figure 3 confirms this. It shows the cumulative distribution of shares invested for the first and last players, where we see that the latter distribution first-order stochastic dominates that of the former.

Turning to the role of player preferences or psychological characteristics, we would expect low risk-averse and overconfident players to invest relatively more in the bonus question, and be relatively less successful in answering.

[13] We compute this (conservative) estimate as follows. If a player other than player 1 invests, then the benchmark efficient investment level is one where where the second player answers, and invests her entire capital – which corresponds to the risk averse choice. In the case that player 1 invests, we compute the efficiency level relative to player 1 investing all his capital. Compounding over rounds arises since lower earnings in any round reduces investments and earnings in later rounds.

[14] This kind of inefficiency is similar to that arising in many organizations, where individuals who are favorably placed for promotion are likely to advocate safe projects, whereas dark-horses are inclined to lobby for more risky projects.

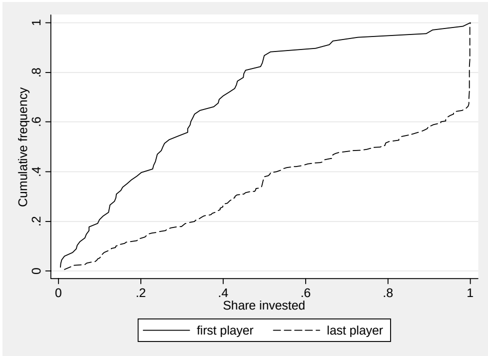

Figure 3: Cumulative distributions of shares invested for highest and lowest ranked players.

Players with a higher ability should also be prepared to invest more. Table 7 reports the results for the share of earnings invested in the bonus question, and the probability of answering correctly. We find that women invest a relatively smaller share. On the other hand, attractiveness and age are uncorrelated with the share invested. Turning to the probability of giving a correct answer, we find no significant effect of gender, age, attractiveness or share invested. The results confirm our previous findings regarding attractiveness and gender. We find no evidence of a difference in self-confidence, risk-aversion or true ability according to attractiveness. On the other hand, women are less likely to answer both regular and bonus questions but do not perform better conditional on answering, which means that the reason why they invest less in the bonus question has probably more to do with ability than with risk-aversion or lack of confidence.

Table 7 - Investment and performance in the bonus question				
	Share invested		Correct answer	
	OLS estimates		Probit estimates marginal effects	
	(1)		(2)	
Attractiveness	.02	(.02)	.01	(.05)
Female	-.06	(.02)***	.09	(.08)
Age	.00	(.00)	.00	(.00)
Earnings (x 1000)	-.02	(.01)*	-.01	(.03)
Dummy lowest earnings	.32	(.03)***	-.13	(.10)
Round 2	-.12	(.02)***	.05	(.09)
Round 3	-.05	(.03)	.11	(.10)
Constant	.36	(.09)***		
N	782		207	

Notes: Standard errors are clustered by episode, pooled sample including data from the three rounds. The share invested is unavailable for players with zero earnings.
Significance levels: *: 10 percent, **: 5 percent, ***: 1 percent.

3.3 Beauty and cooperativeness

Players can only materialize their gains in this game after a final stage, where the two remaining players play a prisoner's dilemma game and decide simultaneously to share the accumulated money or not. A companion paper conducts 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.

We augment this analysis by including the player's own beauty rating and the opponent's beauty rating as explanatory variables. The results are reported in Table 8. We find no correlation between beauty and cooperative behavior. Attractive players are no more (or less) likely to share – indeed, the coefficient is very close to 0.[15] Attractive opponents are also no more (or less)

[15] The coefficients are potentially biased if lead players use private signals to select

likely to induce sharing behavior from their opponents. This is interesting – although our overall results suggest that players obtain consumption value from having attractive co-players, they are no more likely to share with attractive players than with others.

Table 8 - Attractiveness and probability of sharing
Bivariate probit estimates - Marginal effects

	(1)		(2)	
Own attractiveness	-.03	(.07)	-.04	(.06)
Opponent's attractiveness	-	-	-.01	(.11)
Age	.00	(.01)	.00	(.01)
Female	.19	(.09)**	.21	(.10)**
Contribution to prize money (%)	-.72	(.31)**	-.82	(.35)**
Total gains (x €1,000)	.03	(.01)**	.03	(.01)***
N. obs.	138		138	

* significant at 10% level, ** 5% level and *** 1% level;(1) and (2) are bivariate probit estimates, standard errors clustered by episode; the marginal effects are computed at the means of the lead player's characteristics.

It is possible that attractive people are *perceived* as being more cooperative even though they are not really so, so that selection decisions are driven by incorrect beliefs.[16] While we do not observe the beliefs of participants on the show, we used the following experimental strategy to infer what beliefs might be. We showed a sample of games to independent subjects, asking them to predict the probability that the players would share. The sample of 6 episodes is stratified to be representative of all episodes in terms of some key respects, including gender, percentage sharing, and median stakes. Subjects were rewarded based on the accuracy of their predictions, using a quadratic scoring rule, thus giving them strict incentives to predict correctly.

players. In that case, the sample of unattractive players may be unusually cooperative. We address the selection bias in detail in the companion paper, and find no evidence of a bias.

[16] There is some evidence in favor of such incorrect beliefs in trust experiments – see Andreoni and Petrie (2005) and Eckel and Wilson (2004). However, the evidence is somewhat mixed and indirect.

We refer to a companion paper for a detailed exposition and analysis of the results of this experiment.

The results are reported in Table 9. Because the sample is limited in size, dummies are used for the variables attractiveness, age, and stakes. We find that our subjects do not believe that attractive people are more cooperative – on the contrary, they predict above average attractive people to be slightly *less* cooperative, although the effect is small.

Table 9: Predicted probability of sharing
OLS estimates

Attractive (above average)	-.06	(.02)***
Age above median (34)	.02	(.02)
Female	.05	(.03)*
Contribution to prize money over 30%	-.11	(.03)***
Total gains above median (x €1,000)	-.09	(.03)***
Dummy for second player	.03	(.02)
N. obs.	1056	

Significance levels: *: 10 percent, **: 5 percent, ***: 1 percent.

Our findings are on the whole consistent with experimental studies on the relation between attractiveness and cooperation, since the findings of this literature are mixed. Mulford et al. (1998) study a prisoner's dilemma game where subjects have an outside option. Subjects are more likely to opt in and play more cooperatively against opponents they regard as attractive. Andreoni and Petrie (2005) study a public goods game, and find a beauty premium when contributions are private, but this disappears in the treatment with publicly observable contributions. They attribute this to subjects expecting more from attractive players. Eckel and Wilson (2004) study a trust game and find that attractive players receive less, both in the role of first mover (second movers return less) and in the role of second mover (first movers send less). In an ultimatum game study, Solnick and Schweitzer (1999) find that proposers offer more to attractive opponents, but are also more likely to reject offers from attractive proposers.

4 Beauty and selection

We now study the elimination / selection decision, having established that there is no objective reason to discriminate in favor of attractive players either on the grounds of performance or because they are more cooperative. Thus any bias towards attractive 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 et al. (2005)) are often made by majority voting, involving all the participants remaining at that stage. Majority voting games are plagued by multiple equilibria, and this becomes even more of a problem in a dynamic context. 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. In other words, given the presence of multiple equilibria in voting, and the strategic 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 the context of 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.

As figure 1 in the introduction shows, beauty plays an important role in the elimination decision – the average attractiveness of players increases over the rounds, and the attractiveness of the eliminated player in any round is lower on average than that of those that survive. Other summary statistics confirm this picture. If a player is average-looking (i.e. 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 (see Table 10).

Table 10 - Attractiveness and survival, by round

	% reaching round 2	% reaching round 3	% reaching final
Attractive	83	62	51
Average-looking	81	62	40
Unattractive	72	52	31

Note: Attractive (unattractive) is more than one standard deviation above (below) the mean.

We investigate in more detail the role of physical attractiveness in the selection decision by the lead player by estimating a conditional logit model, where the dependent variable indicates whether the player was eliminated (1) or not (0). We start by focusing on the first round, where the sample is largest and no selection has taken place. Later on we discuss results for other rounds.

The results are shown in Table 11. Column (1) is a benchmark specification controlling for age, gender, score ranking and the measure of attractiveness. The odds ratio corresponding to attractiveness is below unity and significant at the 10% level. Thus, attractive players are significantly less likely to be eliminated. In columns (2) and (3) we use dummies for the most and least attractive player within an episode, excluding the player who is in the position to eliminate. We find that the most attractive player is equally likely to be eliminated as his or her counterparts who are average-looking. The least attractive player, on the other hand, is significantly more likely to be eliminated.[17]

[17] This asymmetry, between the positive effect of being attractive and the negative effect of being unattractive, is also found by Hamermesh and Biddle (1994).

Table 11 - Probability of being eliminated at the end of the first round Conditional logit estimates (odds ratios)								
	(1)		(2)		(3)		(4)	
Attractiveness								
Mean attractiveness	.66	(.16)*						
Most attractive			1.21	(.42)				
Least attractive			1.91	(.51)**	1.78	(.51)**	1.77	(.51)**
Performance								
Fourth highest	.42	(.13)***	.47	(.15)**	.47	(.15)**	.48	(.16)**
Third highest	.30	(.10)***	.35	(.13)***	.34	(.12)***	.36	(.15)**
Second highest	.18	(.08)***	.17	(.08)***	.17	(.08)***	.19	(.10)***
% correct answers							.77	(.39)
Behavioral								
% capital invested							.73	(.53)
number of answers							.97	(.13)
Female	.73	(.23)	.71	(.22)	.71	(.22)	.71	(.22)
Age	1.00	(.02)	1.00	(.02)	1.00	(.02)	1.00	(.02)
N. obs.	276		276		276		276	

Significance levels: *: 10 percent, **: 5 percent, ***: 1 percent.

The least attractive player is almost twice as likely to be eliminated at the end of the first round as compared to any other player, a substantial effect. As we will show, this result is stable over different specifications. Note that age and gender are irrelevant in the selection decision. Also, the score ranking is a very good predictor of elimination: the player with the lowest score (the reference category) is more than 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 second highest score. Finally, controls for behavior during the game do not change the results and do not matter as such in the selection decision (column (4)). Less attractive players are discriminated against, for reasons that are uncorrelated with their performance or behavior during the game.

4.1 Discrimination over rounds

One explanation for the discrimination taking place in the first round is that players have very little information about each other. They had relatively little time to get to know each other and to learn about each other's ability. With so little information, perhaps they retreat to attractiveness to select one player over another.

If this is the reason, we expect discrimination to disappear over the rounds when more information becomes available. Although there does not seem to be a lot of discrimination going on in the second round, the discrimination taking place in the third round is very similar to what happens in the first round. The least attractive players are again very likely to be eliminated with an odds ratio comparable to the coefficient we find for the first round (see Table 12). So discrimination does not disappear with learning about player ability.

Table 12 - Probability of being eliminated in 3rd round
Conditional logit estimates (odds ratios)

	(1)		(2)		(3)		(4)	
Mean attractiveness	.58	(.18)*	.64	(.21)	-	-	-	-
Least attractive player	-	-	-	-	1.89	(.57)**	1.76	(.55)*
Ranked last	.32	(.10)***	.33	(.10)***	.33	(.10)***	.34***	(.10)
Age	-		1.02	(.03)	-	-	1.03	(.03)
Gender	-		1.15	(.44)	-	-	1.14	(.45)
N. obs.	138		138		138		138	

Significance levels: *: 10 percent, **: 5 percent, ***: 1 percent.

4.2 Heterogeneity in discriminatory behavior

We now explore the nature of discrimination in greater detail. We first investigate the relationship between the "ugliness penalty" and performance in the game. Is being less attractive particularly a disadvantage when having performed less well? We ran a separate regression for the sample of players who have the lowest earnings and compare these results to the remaining sample. The results are striking: The player who has the lowest score but is not the least attractive of that round, is about twice as likely to be eliminated as the others. In shows where the player with the lowest score is also the least attractive player, he or she is about 6 times as likely to be eliminated as the other players (see Table 13).

Table 13 - Probability of being eliminated in first round
Conditional logit estimates - Odds ratios

	(1)		(2)	
Lowest score	2.39	(.66)***	-	
Lowest score & player least attractive	-		6.26	(3.39)***
N. obs.	276		64	

Significance levels: *: 10 percent, **: 5 percent, ***: 1 percent.

We now investigate the relation between discrimination and gender. Do men care more about looks than women do? Do people care more about the looks of the opposite sex? Table 14 reports separate regressions for male and female lead players. We find that women do discriminate more against the least attractive player than men do, i.e. women seem to care more about looks than men do. The difference in coefficients between male and female lead players is not significant though. Columns (3) and (4) investigate the elimination decision when the lead and least attractive player are of opposite sex on the one hand, and of the same sex on the other hand. We find that discrimination is indeed greater when the candidates for elimination are of the opposite sex. The difference in coefficients is significant at the 10% level. This could explain why women discriminate more, since the least attractive player in the first round is more likely to be a man (men being on average less attractive than women). We have found almost identical results for the last selection round, where the discrimination against the least attractive player took place in games where the lead player and the least attractive player were of opposite sex (results not reported for the sake of brevity). Since in the last round, most lead players are men, the discrimination plays mostly against women.

Overall, these results reinforce the idea that beauty has a consumption value, and that this is the main reason why people discriminate against less attractive players.

Table 14: Discrimination and beauty of the lead player (first round)
Conditional logit estimates (odds ratios)

	Lead player female		Lead player male		Lead player and least attractive - opposite sex		Lead player and least attractive - same sex	
	(1)		(2)		(3)		(4)	
Fourth highest score	.24	(.18)*	.50	(.18)*	.36	(.19)**	.53	(.23)
Third highest score	.24	(.19)*	.39	(.16)**	.51	(.25)	.18	(.11)***
Second highest score	.21	(.15)**	.13	(.08)***	.31	(.16)**	.05	(.05)***
Least attractive	4.55	(3.11)**	1.39	(.48)	2.80	(1.19)**	1.10	(.53)
Age	.95	(.04)	1.01	(.02)	.98	(.02)	1.03	(.03)
Gender	.93	(.67)	.70	(.24)	.68	(.31)	.56	(.27)
N. obs	84		192		140		136	
Test equality coefficients for the least attractive (1) = (2) P-value (3) = (4) P value	.10				.16			

Notes: Equality of coefficients is tested with a generalized Hausman test
Significance levels: *: 10 percent, **: 5 percent, ***: 1 percent.

4.3 The price of beauty

The game show is such that the stakes are substantial. 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. We now do a back-of-the-envelope calculation of the price that they pay by eliminating the least attractive player.

We can identify 13 cases where the lead player eliminates the least attractive player in the third round, even though this player does not have the lowest score. These players have average earnings of around €750, while the players who are chosen instead have average earnings of only €400. Hence, by eliminating the least attractive players in these cases, the lead player diminishes the prize money E by €350 on average. This is a lower bound since we are only looking at costs associated with discrimination in the third round, and not earlier rounds, since a precise imputation of financial costs in earlier rounds is more difficult.[18]

5 Gender differences

It is worthwhile contrasting our results on beauty with differences between players based on gender. Overall, women earn less than men, earning €269 on average, as compared to €428. As with attractiveness, we can decompose the gender gap in earnings into different components. In terms of performance, we found that women were less likely to answer a question, and conditional on answering were no more likely to answer correctly (in round 3, they are actually less likely to answer correctly). The lower performance of women could be due to the competitive character of the game show — Gneezy et al. (2003) present experimental evidence showing that women perform worse in competitive environments, especially when they compete against men.

Despite the fact that women perform worse, they are as likely to reach the final round as men – 35% of the participants on the show are women, while 37% of the players in the final round of the game are female. One explanation for discrimination in favour of women in the selection decision is their greater cooperativeness, since women are more likely to share in the final round — 55% of women share against only 36% of men. However, this does not seem to be the explanation since women are no more likely to be selected than men in the *last* round (holding earnings constant), when the lead player selects the player with whom he will play the final round and

[18] It might be that lead players expect those with a lower score to be more cooperative, for which there is evidence (in our companion paper). But this doesn't explain why the player with a lower score is rarely chosen to play the final if he is is the least attractive player.

where one may expect cooperativeness considerations to be dominant.[19] It is more plausible that women are chosen despite their lower performance due to gender balance considerations (since women are relatively scarce in the show)[20] or positive discrimination (since the selection decision is public).

Womens' greater cooperativeness widens the gender gap in earnings. Indeed, since women are much more likely to share in the final round, and since the opponent does not share any more when facing a woman, the difference in take-home earnings across gender is larger than between individual scores.

Overall, we find no evidence of negative discrimination against women. The gender earnings gap is mainly driven by differences in performance and cooperativeness. Perhaps the fact that the selection decision is public prevents gender discrimination - since gender is an objective and obvious characteristic in contrast to attractiveness, which is more subtle.

6 Interpretation: Insidious Discrimination?

Following the work of Hamermesh and Biddle (1994), who find a beauty premium in the labor market in a variety of occupations, several papers have replicated its findings, and also attempted to disentangle the components of the premium. Biddle and Hamermesh (1998) analyze a sample of lawyers, and find a premium irrespective of their area of expertise, including the self-employed. They argue that the most plausible explanation is taste-based discrimination by clients. Mocan and Tekin (2006) find that unattractive people sort into criminal activity due to the existence of a beauty premium on the legal labor market.

One difficulty with most empirical studies is in disentangling attractiveness from ability. There is often no precise measure of productivity, so it is

[19] Our companion paper provides a comprehensive analysis of the determinants of the sharing decision in the final stage, and its implications for how selection decisions are made just before this.

[20] Some participants explicitly mentioned "gender balance" as a consideration while making the selection decision in the second round.

hard to establish whether the premium is due to productivity differences or discrimination. Such productivity effects are sometimes present. Landry et al. (2006) find that attractive female solicitors are more productive fund-raisers, and Pfann et al. (2000) find that companies with better looking executives have higher revenues.

Disentangling the beauty premium is easier in experimental studies. Mobius and Rosenblat (2006) demonstrate, in an experimental setting, that "employers" have have higher estimates of the productivity of more attractive individuals, even though they are given independent evidence on productivity. An intriguing finding is that attractive people 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 self-confidence of attractive workers. There are two important differences between our set up and theirs. First, after making the estimates, there is no interaction in their experimental set up between subjects who are assigned the role of "employers" and those assigned the role of "workers", so that consumption value considerations may be naturally less important. Second, in our set up, individuals get repeated public feedback in the performance phase, on their success in answering quiz questions. Thus it is compelling that we find that discrimination in favour of attractive people does not disappear over the rounds. We also find no evidence that beautiful people are more confident, since they answer the same number of questions and invest the same amount of money.

There is a large literature on discrimination on the basis of race or gender (Altonji and Black (1999) provide a useful survey). One difficulty in labor market studies is in establishing discrimination — see, for example, the criticisms of Heckman (1998). Bertrand and Mullainathan (2005) conduct an field experiment which circumvents these criticisms and shows convincing evidence of employer discrimination against Afro-American names. However, it remains an open issue whether such discrimination is statistical or taste based, since employee performance on the job is not observable.

The main contribution of our paper is that we are able to shed light on the sources of the beauty premium. We show that the premium is not due to superior performance, or more aggressive behavior in answering questions. Nor is it due to differences in cooperativeness between attractive and unattractive players. Indeed, a striking finding is that attractiveness is orthogonal to all measures of performance or behavior, and only affects the elimination decision. We conclude therefore that the beauty premium arises in our setting due to consumption value considerations, and constitutes a form of taste based discrimination.

One concern with our study is external validity, especially if the participants on the show are not representative of the population. The producer of the show told us that no explicit criteria were used in recruiting participants. The participants have a diverse background, and somewhat surprisingly, it does not seem that the show attracts especially good looking people.[21] For the purpose of our study (examining discrimination), it seems that one should expect less discrimination, since participants are subject to public scrutiny.

The phenomenon we find appears to be a form of *insidious discrimination,* where the persons being discriminated against (or third parties) are not aware of this discrimination. Indeed, it may be possible that the discriminator is also not aware that he or she is discriminating. One might expect that participants on a TV show would be reluctant to discriminate on the basis of race or gender, since their behavior is subject to public scrutiny. This is consistent with the findings of other studies of discrimination based on TV shows (see List (2006), Antonovics et al. (2005), and Levitt (2004)), which find no evidence of discrimination based on race, gender or ethnicity, but some evidence of discrimination against older players (these papers do not examine the role of beauty).

[21] Harrison and List (2004) have an extensive discussion on the external validity of game show data.

We believe that insidious discrimination differs from well known forms of discrimination, on the basis of race, sex or ethnicity, and is well worth further study. Apart from beauty, height is one characteristic on the basis of which there may be insidious discrimination. There is evidence from the labor market that taller individuals obtain higher wages. Interestingly, Persico et al. (2004) find that a labor market reward to height during adolescense. They interpret this as due to the greater participation in social activities of individuals during adolescense.

A related phenomenon, which has been identified by psychologists is *implicit discrimination*. Greenwald et al. (1998) have developed the Implicit Association Test, where the subject's speed of response in categorizing words or faces is used to measure an individual's implicit bias against, for example, African-Americans. Bertrand et al. (2005) discuss the implications of such implicit discrimination — this is unintentional and the discriminator may be unaware of it, although individuals are very much aware of the *possibility* of such discrimination. Insidious discrimination has some common features, since the discriminator may be unaware of his or her discrimination, and may not even be aware of the possibility of such discrimination; however, this discrimination is reflected in explicit behavior. While there is no IAT with beauty as one of the categories, it noteworthy that researchers have found an implicit bias against older people and those who are overweight (see https://implicit.harvard.edu/implicit/demo/selectatest.jsp for details of the available tests).

We believe that insidious discrimination deserves further examination from economic theorists as well as empirical economists. For example, it is clear that standard models of statistical discrimination (Arrow, 1973; Phelps, 1972) cannot deliver insidious discrimination. Statistical discrimination arises since discriminated groups believe that their group status will be used as an informative signal, thereby reducing their incentives to undertake costly investment. If unattractive people are unaware that they will be discriminated against, and indeed the group itself is not clearly defined, this mechanism cannot work for insidious discrimination. Nevertheless, it seems likely that insidious discrimination can give rise to a phenomenon similar to statistical discrimination, although this arises due to individual feedback rather than group identity. An unattractive individual who is repeatedly given negative feedback will revise downwards her estimate of her own ability, thereby reducing incentives for effort. This in turn will induce a relationship between attractiveness and productivity. We raise these issues for further research.

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, even from a narrow monetary standpoint. Attractive players earn a substantial 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 beauty premium is therefore a form of discrimination. 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, both discriminators and the discriminated against may be unaware that it occurs. It is likely that insidious discrimination has 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

[1] Altonji, J.G., and R.M. Black (1999), Race and Gender in the Labor Market, in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Amsterdam: North Holland.

[2] Andreoni, J. and R. Petrie (2004), Beauty, Gender and Stereotypes: Evidence from Laboratory Experiments, University of Wisconsin, Madison, Department of Economics Working Paper 2004-06.

[3] Antonovics, K., P. Arcidiacono and R. Walsh (2005), Games and Discrimination: Lessons From The Weakest Link, *Journal of Human Resources* 40(4), pp. 918-947.

[4] Arrow, K.J. (1973), The Theory of Discrimination, in O.Ashenfelter and A. Rees (eds), *Discrimination in Labor Markets*, Princeton, NJ: Princeton University Press.

[5] Becker, G. (1957), *The Economics of Discrimination*, Chicago University Press, Chicago.

[6] Bertrand, M., D. Chugh and S. Mullainathan (2005), Implicit Discrimination, *American Economic Review Papers and Proceedings,* 95(2), 94–98.

[7] Biddle, J. and D. Hamermesh (1998), Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre, *Journal of Labor Economics* 16(1), pp. 172–201.

[8] Eckel, C. and R. Wilson (2004), Detecting Trustworthiness: Does Beauty Confound Intuition?, mimeo.

[9] Gabriel, M., J. Critelli and S. Jullana (1994), Narcissistic Illusions in Self-Evaluations of Intelligence and Attractiveness, *Journal of Personality* 62(1), 143–155.

[10] Gertner, R. (1993), Game Shows and Economic Behavior: Risk-Taking on 'Card Sharks', *Quarterly Journal of Economics* 108(2), 507-522.

[11] Gneezy, U., M. Niederle and A. Rustichini (2003), Performance in competitive environments: Gender differences, *Quarterly Journal of Economics* 118(3), 1049-1074.

[12] Greenwald., A. G., D. McGhee and J.L.K. Schwartz (1998), Measuring Individual Differences in Implicit Cognition: The Implicit Association Test, *Journal of Personality and Social Psychology*, 74 (6), 1464-80.

[13] Guimaraes, P. and R. Lindrooth (2005), Dirichlet-Multinomial Regression, Economics Working Paper Archive at WUSTL, Econometrics, no. 0509001.

[14] Hamermesh, D. and J. Biddle (1994), Beauty and the Labor Market, *American Economic Review* 84(5), 1174–1194.

[15] Harrison, G. and J. List (2004), Field Experiments, *Journal of Economic Literature* 27, 1009-1055.

[16] Heckman, J. (1998), Detecting Discrimination, *Journal of Economic Perspectives,* 12(2), 101-116.

[17] Kalist, D. (2004), Data from the Television Game Show Friend or Foe?, *Journal of Education Statistics* 12(3).

[18] Levitt, S. (2004), Testing Theories of Discrimination: Evidence from Weakest Link, *Journal of Law and Economics* 47, 431–452

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

[20] List, J. (2004), Young, Selfish and Male: Field Evidence of Social Preferences, *Economic Journal* 114, 121-49.

[21] List, J. (2006), Friend or Foe? A Natural Experiment of the Prisoner's Dilemma, *NBER Working Paper* 12097.

[22] McFadden, D. (1974), Conditional Logit Analysis of Qualitative Choice Behavior, in *Frontiers in Econometrics*, edited by Paul Zarembka, New York: Academic Press.

[23] Mocan, N. and E. Tekin (2006), Ugly Criminals, *NBER Working Paper* 12019.

[24] Mobius, M. and T. Rosenblat (2006), Why Beauty Matters, *American Economic Review* 96(1), 222-235.

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

[26] Myers, G.D., 1996, *Social Psychology,* New York: McGraw Hill.

[27] Oberholzer-Gee, F., J. Waldfogel and M. White (2004), Friend or Foe? Coordination, Cooperation, and Learning in High-Stakes Games, mimeo, Harvard Business School.

[28] Persico, N., A. Postlewaite and D. Silverman (2004), The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height, *Journal of Political Economy* 112, 1019–1053.

[29] Pfann, G. J. Biddle, D. Hamermesh and C. Bosman (2000), Business Success and Businesses' Beauty Capital, *Economics Letters* 67(2), 201–207.

[30] Phelps, E.S., (1972), The Statistical Theory of Discrimination, *American Economic Review* 62(4), 659-661.

[31] Solnick, S. and M. Schweitzer (1999), The Influence of Physical Attractiveness and Gender on Ultimatum Game Decisions, *Oganizational Behavior and Human Decision Process* 79(3), 199–215.

7 Appendix

We use a conditional logit specification in three different contexts: (1) Probability of being first or last, (2) Probability of answering a question, (3) Probability of being eliminated. The conditional logit specification estimates the probability that one alternative is chosen (realized) among a set of possible alternatives, as a function of the attributes of the choices.

For example, suppose that the 'ability' of a player i, z_i, can be written as:

$$z_i = \beta X_i + u_i$$

where X is a vector of attributes such as attractiveness and u is a random component. Suppose there are two players A and B. Player A answers a particular question if $z_A > z_B$. The probability that player A answers the question is equal to:

$$\begin{aligned} P(z_A > z_B) &= P(\beta X_A + u_A > \beta X_B + u_B) \\ &= P(u_A - u_B > \beta X_B - \beta X_A) \end{aligned}$$

McFadden (1974) showed that $u_A - u_B$ follows a logistic distribution if the individual errors u_A, u_B are independent and follow a type 1 Extreme Value distribution.[22]

One important assumption is that the error terms are independent across alternatives. This implies that the odds ratio of choosing alternative j over alternative k does not depend on the other alternatives available. This is the well-known *Independence of Irrelevant Alternatives* (IIA) assumption. This assumption is likely to be violated if there exists an alternative C that is a closer substitute to A than to B. In this case, the odds ratio of A versus B will increase if C is eliminated from the choice set. This assumption may be even more problematic in the context of the lead player's choice of which player is to be eliminated – if we model this as the lead player choosing the subset of *continuing players* that has greatest value to her, then complementarities across player types could arise, leading to a violation of IIA.

[22] This distributional assumption conveniently links the random utility model to the logistic model.

Testing for the independence of irrelevant alternatives

We test for the validity of the IIA assumption as follows. Since the choice set varies across episodes, we need to label the alternatives in a meaningful way to test the IIA assumption. Since our main coefficient of interest regards attractiveness, we labeled the alternatives according to their attractiveness ranking. We then excluded each alternative at a time and tested (with a Hausman test) the stability of coefficients across specifications.[23] The results are reported in Tables A1-A3.[24] Overall, the IIA assumption is not rejected at the 5% level. Even in the cases where the p-value is small (between 5 and 15%), the exclusion of an alternative never affected the conclusion regarding the attractiveness coefficient: Attractiveness remained not significant and the odds ratio remained close to 1 in all specifications.

Table A1 - Probability of being first or last in first round
Hausman test - P-value equality of coefficients

Alternative excluded	First	Last
$\text{player}_i = 1$	.63	.98
$\text{player}_i = 2$	.93	.23
$\text{player}_i = 3$	.59	.33
$\text{player}_i = 4$	.13	.97
$\text{player}_i = 5$	.56	.051

Table A2 -Probability of answering in first round
Hausman test - P-value equality of coefficients

Alternative excluded	First	Second	Third
$\text{player}_i = 1$	.29	.12	.98
$\text{player}_i = 2$	.77	.11	.11
$\text{player}_i = 3$	.34	.06	.11
$\text{player}_i = 4$	.11	.27	
$\text{player}_i = 5$	.25		

[23] We experimented with different ways of labeling and found very similar results (the IIA assumption is generally not rejected)

[24] Note that the IIA assumption is not invoked in the selection decision in the third round as there are only two alternatives left.

Table A3 -Probability of being eliminated in first round	
Alternative excluded	Hausman test - P-value equality of coefficients
$\text{player}_i = 1$	1.00
$\text{player}_i = 2$	.67
$\text{player}_i = 3$	.94
$\text{player}_i = 4$	.32