ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 624 January 2007

Is Beauty only Skin-Deep ? Disentangling the Beauty Premium on a Game Show

Michele Belot, V Bhaskar and Jeroen van de Ven

Is Beauty only Skin-deep? Disentangling the Beauty Premium on a Game Show*

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

January 11, 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 than 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.

Keywords: beauty premium, discrimination.

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

*We are very grateful 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%. There are several competing explanations for this premium. 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 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, and as Becker (1957) pointed out, discriminators must be willing to pay a price. 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 generalized phenomenon.

This paper analyzes participant behavior on a television game show in order to disentangle these components of the beauty premium.[3] Specifically, we can ask: are attractive people more productive? 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

[1] See Mobius and Rosenblat (2006) for experimental evidence for this hypothesis.

[2] Reverse causality is to some extent addressed in Biddle and Hamermesh (1998), since they use attractiveness as measured by photographs taken at law school as the explanatory variable. It is also not a problem in experimental or quasi-experimental studies provided one controls for income.

[3] List (2006) and Levitt (2004) also use gameshows to study discrimination on other dimensions (such as race, gender, and age).

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 – most notably gender – 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.

Our key findings are as follows. Overall, there is a substantial beauty premium, with attractive people earning €300 more than unattractive ones. This 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, an attractive player is significantly less likely to be chosen than a less attractive one. These results support a "consumption value" basis for the beauty premium in our context. Indeed, 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.

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. 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 four 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. At the end of 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 are depicted in Table 1. These monetary payoffs correspond to generalized prisoner's dilemma, where G is a weakly dominant strategy.[4]

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

Table 1: Monetary payoffs

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

[4]Players share 43 percent of of the time, suggesting that non-monetary motivations are important. A detailed analysis of behavior in the prisoner's dilemma stage and its consistency with models of social preferences can be found in Belot et al. (2006). The prisoner's dilemma stage is similar to that in the game show *Friend or Foe* (see List (2004, 2006), Kalist (2005) and Oberholzer-Gee et al. (2004)); however, the overall game is different.

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 who loses his entire capital 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.[5]

2.2 Measure of beauty

We construct a measure of beauty based on independent ratings. Each of the 345 participants in the game show was rated on a scale from 1 (very unattractive) to 7 (very attractive) by approximately 10 raters (5 males and 5 females). 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.[6] There is a high degree of concurrence on

[5]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.

[6]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

attractiveness across raters, as measured by Cronbach's alpha for all subsamples of judges who rated the same sample of participants. The Cronbach's alpha ranged from 0.70 to 0.85, showing high agreement. Table 2 reports some summary statistics of the ratings. Contrary to what one might expect, the average participant is rated slightly below the population average of 4 (the mean of all ratings is equal to 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.

Table 2: 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 one standard deviation above the mean earn roughly €300 more than those whose attractiveness is one standard deviation below the mean.[7] Given the many determinants of payoffs in the show, this premium could be attributable to a variety of factors. We now study each of these factors in turn.

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, we find the same results.

[7] This earnings gap translates into a 100 percent premium for the attractive players compared to unattractive players. However, this high percentage is largely due to the asymmetry in earnings in this game: those who are eliminated have zero earnings.

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 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 1 plots the total number of questions answered correctly in the first round against the average attractiveness of participants in the show. We find no correlation between the two. Less attractive populations of players perform equally 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 has a vector of attributes X_j (gender, age, attractiveness). The conditional logit

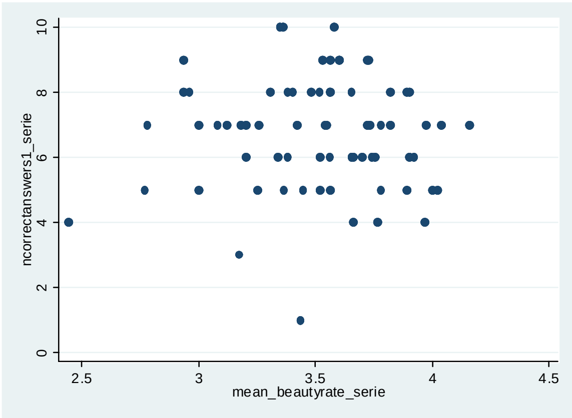

Figure 1: Mean beauty rate of players per episode and average number of correct answers.

model has the form:

$$p(y_j = 1) = \frac{\exp(\beta' X_j)}{\sum_j \exp(\beta' X_j)} \qquad \text{for } j = 1, ...5,$$

where y_j is a dichotomous variable indicating whether the player is in first (resp. last) position or not ($y_j = \{1, 0\}$). One important assumption for the validity of the conditional logit estimates is the assumption of independence of irrelevant alternatives. In the appendix, we discuss this assumption and test for its validity. We found that the modelling assumptions were not rejected.

Table 3 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 3. We find no clear correlation between any of these characteristics and score ranking. In particular, there is no evidence that attractive people rank any different from unattractive ones.

Table 3 - 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)	-.06	(.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 et. al. (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.[8] Let x_i be his current 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. 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

[8] 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.

lower value of q_i, this will ceteris paribus reduce the number of answers but not raise the proportion of correct answers.

Table 4 - 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.
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 4 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. As we discussed earlier, since each question can only be answered by one player at most, the decisions to answer are mutually exclusive: As soon as one player has pushed the buzzer, the others cannot answer the same question. We investigate whether some characteristics (such as attractiveness and gender) make a player more likely to be the one answering the question or not. We estimate a conditional logit model, where the probability of answering is defined as 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 matter for the probability of answering correctly, conditional on answering.

Table 5 - 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	.96	.90	.94	1.15	1.02	1.07
	(.08)	(.11)	(.09)	(.19)	(.11)	(.15)
female	.70	1.12	.81	.67	.74	.88
	(.07)***	(.40)	(.09)*	(.11)**	(.09)**	(.28)
age	1.00	.99	1.00	1.00	1.01	1.01
	(.01)	(.01)	(.01)	(.01)	(.01)*	(.01)
share invested	.65	1.03	1.10	1.33	.81	.79
	(.14)**	(.44)	(.26)	(.37)	(.22)	(.27)
pseudo R^2	.006	.003	.003	.008	.005	.003
N. obs.	3450	580	2760	536	2070	484

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

Table 5 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. Here their probability of answering seems even smaller, although the effect is not significant. 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.[9]

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 than the regular questions, and are probably more difficult. It is therefore not surprising that the success rate in answering the bonus question is much lower than for the regular questions, only slightly more than 50% are answered correctly, against more than 75% for the regular questions.

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 answers, 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 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

[9] In this context, the finding of Gneezy et al. (2003) that women do less well than men in a competitive environment, is relevant.

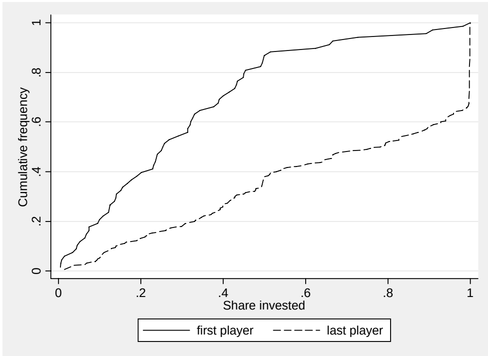

Figure 2: Cumulative distributions of shares invested and earnings' position

per round of around 15 percent of earnings – since earnings are effectively compounded across rounds, the overall losses are even larger.[10] 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.[11] Figure 2 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,

[10] 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.

[11] 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.

we would expect low risk-averse and overconfident players to invest relatively more in the bonus question, and be relatively less successful in answering. Players with a higher ability should also be prepared to invest more. Table 6 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 6 - 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 (Belot et. al. (2006)) 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 7. 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.[12] Attractive opponents are also no more (or less) likely to induce sharing behavior from their opponents.

Table 7 - Attractiveness and probability of sharing
Probit estimates

	(1)		(2)	
Own attractiveness	-.01	(.07)	-.04	(.06)
Opponent's attractiveness	-	-	.09	(.06)
Age	.00	(.01)	.00	(.01)
Female	.16	(.10)*	.18	(.10)*
Contribution to prize money (%)	-.55	(.22)**	-.58	(.23)**
Total gains (x 1,000)	.04	(.01)**	.04	(.01)***
N. obs.	138		138	

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

It is possible that attractive people are perceived as being more cooperative even though they are not really so.[13] They may expect unattractive

[12] Our coefficients are potentially biased if lead players use private signals to select 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 (see Belot et. al., 2006).

[13] 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.

people to be less cooperative, and therefore disfavor them in the elimination rounds. We do not observe the beliefs of participants on the show. However, we showed a sample of games to independent subjects, asking them to predict the probability that the players would share. We find that observers do not believe that attractive people are more cooperative – on the contrary, they predict above average attractive people to be slightly *less* cooperative, although this is not significant.

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)) 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 others 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

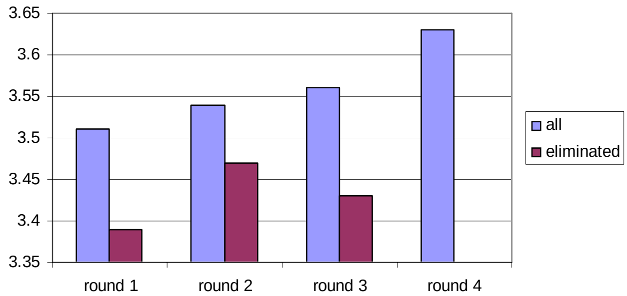

Figure 3: Average beauty per round of all players and eliminated players.

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.

Figure 3 illustrates the importance of beauty in the elimination decision. It shows the average attractiveness of remaining players in rounds one to three, and also plots the attractiveness of the two players in the final stage. Average attractiveness increases steadily over rounds one to three, and the most attractive player is the one who is chosen to be in the final, at the end of round three. 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 has probability only 0.31 (see Table 8).

Table 8 - Attractiveness and probability of playing 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 the role of physical attractiveness in the selection decision by the lead player in more detail by estimating a conditional logit model, where the dependent variable indicates whether the player was eliminated (1) or not (0).

The results are shown in Table 9. 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 the average-looking players. The least attractive player, on the other hand, is significantly more likely to be eliminated. The effect is substantial: The least attractive player is almost twice as likely to be eliminated at the end of the first round than any other player. 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 (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.

Table 9 - Probability of being eliminated at the end of the first round Conditional logit estimates (odds ratios)								
	(1)		(2)		(3)		(4)	
Attractiveness	.66	(.16)*						
Mean attractiveness		(.23)						
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.

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 10).

Table 10 - Probability of being eliminated in 3d round Conditional logit estimates (odds ratios)								
	(1)		(2)		(3)		(4)	
ranked last	.32	(.10)***	.33	(.10)***	.33	(.10)***	.34***	(.10)
mean attractiveness	.58	(.18)*	.64	(.21)	-	-	-	-
least attractive player	-	-	-	-	1.89	(.57)**	1.76	(.55)*
age	-		1.02	(.03)	-	-	1.03	(.03)
gender	-		1.15	(.44)	-	-	1.14	(.45)
N	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 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 three times more 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 6.6 times as likely to be eliminated as the other players.

Table 11 - Probability of being eliminated in first round Conditional logit estimates - Odds ratios				
	(1)		(2)	
lowest score	2.96***	(.72)	-	
lowest score & player least attractive	-		6.6	(3.56)***
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 12 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 more present among players of opposite sex. The difference in coefficients is significant at the 10% level. Altogether, these results seem to reinforce the idea that beauty has a consumption value, and that this is the main reason why people discriminate against less attractive players.

Table 12: 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)**	.50	(.24)	.18	(.11)***
Second highest score	.21	(.15)**	.13	(.08)***	.29	(.15)**	.05	(.05)***
Least attractive	4.55	(3.11)**	1.39	(.48)	2.75	(1.17)**	1.10	(.53)
Age	.95	(.04)	1.01	(.02)	.97	(.02)	1.03	(.03)
Gender	.93	(.67)	.70	(.24)	.69	(.32)	.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 are willing to pay to eliminate 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 likely to be a lower bound because other unattractive players may already have been eliminated in earlier rounds, who might otherwise have pushed up the score further.[14]

5 Concluding comments

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 attempted to disentangle its components. Biddle and Hamermesh (1998) analyze a sample of lawyers, and find a premium irrespective of their area of expertise and also including the self-employed. They argue that the most plausible explanation is taste-based discrimination by clients. Mocam 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 to disentangle attractiveness from ability. There is often no precise measure of productivity, so it is 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 fundraisers, and Pfann et. al. (2000) find that companies with better looking executives have higher revenues.

[14] It might be that lead players expect those with a lower score to be more cooperative, for which there is evidence (Belot et. al., 2006). 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.

The beauty premium is also replicated in experimental studies. Mobius and Rosenblat (2006) present evidence showing that attractive people are more confident, and this may increase an observer's estimate of their ability. In their setup "employers" have to estimate the productivity of workers (i.e. the number of mazes they can solve). Most striking is their finding that attractive people are estimated to have higher productivity even when their interaction with the employer is only oral, not visual. They attribute this to the self-confidence of attractive workers. We find no evidence that beautiful people are more confident, since they answer the same number of questions and invest the same amount of money. One key difference between our set up and theirs is that individuals get repeated public feedback in the performance phase, on their success in answering quiz questions. This might prevent them from sustaining higher levels of self-confidence.

In relation to this literature, the novelty of our paper has been to use data from a game show in order to shed light on the sources of the beauty premium. The game show format has several advantages, insofar as it allows us to disentangle the role of different factors that could lead to the premium. However, there are important caveats. People might behave differently on national television, and may hesitate to discriminate, leading us to underestimate discrimination. On the other hand, the interaction time is relatively short, and the participants are not well trained to select the best person. Personnel managers might have more experience and discriminate less. A translation of our results to the labor market should therefore be made with caution.

Our research is also complementary to the studies by List (2006) and Levitt (2004). They use game shows to study discrimination based on public characteristics such as race, sex or age (they do not consider attractiveness). They find some evidence of taste-based discrimination against older players. List (2004) studies the game show *Friend or Foe?,* which has a similar final stage as *Shafted* but with differences in the selection stage. Most importantly, teams are formed prior to the trivia rounds so that players are uninformed about the ability of their partners. This uncertainty about performance makes it harder to isolate taste based discrimination from incorrect perceptions on productivity. *The Weakest Link*, studied by Levitt (2004), has the disadvantage of a complicated optimal strategy during the trivia rounds. The strategic incentives change over the rounds: at first, it is best to keep the best performing players in the game to increase the prize money, but later on this switches because they will be the competitor in the final. This is further complicated because elimination is via majority voting, which implies that there are multiple equilibria.

Our findings can be related to experimental work on the relation between attractiveness and behavior in prisoner's dilemma type games. Mulford et al. (1998)[15] 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. In the treatment with observable contributions, expectations are not born out and others reduce their contributions accordingly, turning the beauty premium into a beauty penalty. 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) offer more to attractive opponents, but also appear to demand more from them as they more often reject offers from attractive proposers. Overall, these experimental results are somewhat contradictory.

To return to the question posed by the title of this paper, we find that beauty is indeed only skin-deep, and has no implications for a person's

[15] In an early contribution, Kahn et al. (1971) study a version of the repeated prisoner's dilemma game.

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. Since this positive selection is unrelated to performance, it seems to reflect consumption value considerations on the part of the other players in the game.

References

[1] 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.

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

[3] Belot, M., V. Bhaskar and J. van de Ven (2006), A Public Dilemma: Cooperation with Large Stakes and a Large Audience, Working Paper 221, ELSE, University College London.

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

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

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

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

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

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

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

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

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

[13] 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.

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

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

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

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

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

[19] 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.

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

[21] 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.

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

[23] 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.

6 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.[16]

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 substitutes 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.

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.[17] The results

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

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

are reported are reported in Tables A1-A3.[18] 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		

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

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