

Testing for Discrimination: Evidence from the Game Show *Street Smarts*

Shamena Anwar¹

Abstract

This paper conducts two alternative tests of discrimination in the game show *Street Smarts*, which both exploit the unique setup of the game to determine why discrimination occurs. The results indicate non-black contestants have a lower prior perception of the skill level of blacks relative to non-blacks in answering the average question in the game. When results are stratified by question category, one finds that non-black contestants perceive blacks and non-blacks to have equal ability in answering general knowledge questions. However, they perceive blacks to have a lower ability in answering miscellaneous, entertainment, and slang/common saying questions.

Key words: statistical discrimination; racial prejudice; inaccurate stereotypes.

Classification Code: J7

¹ Contact address for Shamena Anwar: Carnegie Mellon University, Heinz College, 5000 Forbes Ave, Pittsburgh, PA 15213, Office Phone: 412-268-7420, and Email: shamena@andrew.cmu.edu

1 Introduction

Previous literature has shown convincing evidence that discrimination is a factor in the labor market. Bertrand and Mullainathan (2004) sent out identical resumes and found the response rate for resumes with white sounding names received fifty percent more callbacks for interviews. Goldin and Rouse (2000) found that female musicians auditioning for the orchestra were more likely to be hired when their gender was hidden from the judges. Neumark (1996) used an audit study and found women were less likely than equivalently skilled men to get hired in high priced restaurants, but were not discriminated against in medium and low priced restaurants. The drawback to all of these studies, however, is that they cannot determine what type of discrimination is responsible for the hiring disparities. This is important because the effectiveness of policies to reduce discriminatory behavior depends on the type of discrimination.

There are three types of discrimination which can result in the hiring disparities found above: prejudice, and accurate or inaccurate group stereotypes (Fershtman and Gneezy, 2001). If employers have prejudice against certain groups (termed taste discrimination by Becker 1957), they will be less likely to hire members of these groups because they have a distaste for interacting with them. Group stereotypes, on the other hand, are used when employers have limited information about potential applicants, and thus use their prior perceptions (stereotypes) about the average skill level of the group an applicant is from as a proxy for their skill level. If employers' prior perceptions about the average skill level of different groups are different, these stereotypes result in disparities between observationally equivalent individuals from different groups. As the use of stereotypes only has an effect on outcomes if prior perceptions about groups are different, the term stereotype is used synonymously with different prior perceptions throughout this paper. If these prior perceptions are correct, this use of group status is considered efficient and is termed statistical discrimination (see, for example, Phelps 1972, Arrow 1973, Coate and Loury 1993, and Altonji and Blank 1999). However, as Fershtman and Gneezy (2001) showed in the area of trust, it is also possible the prior per-

ceptions evaluators use are incorrect. The primary difficulty in distinguishing between these types of discrimination in the labor market is they all have the same effect—namely, they all result in the discriminated group being less likely to be hired. Several empirical papers have successfully distinguished between the various theories of discrimination in other areas such as trust, negotiation, and racial profiling (Fershtman and Gneezy 2001, List 2004, Knowles, Persico and Todd 2001, and Anwar and Fang 2006). However the identification strategies employed in these studies relied on field experiments and institution specific settings and thus cannot be applied to existing labor market data.¹

Due to these difficulties, this paper examines discriminatory behavior in the second-best environment of a television game show *Street Smarts*.² During *Street Smarts*, contestants earn money by predicting how three random people, known as street savants, answered basic trivia questions. The unique structure of this game show allows us to not only determine if contestants discriminate, but provides insight as to why. Because this game show involves assessing others' skill level with limited information, a key feature present in traditional hiring situations, the results we find here might give us an idea about the discrimination present in the labor market.

A second nice feature of this game show is that because the race of the contestants are observable, one can determine whether the amount and type of discrimination a group faces depends on the group evaluating them. This is important because if both groups happen to discriminate against each other, the aggregate amount will underestimate the true amount of discrimination present. If the discrimination results from statistical discrimination, there will be no difference in contestant behavior, as this requires all contestants to have the same (correct) priors. However, if the discrimination results from incorrect stereotypes or racial prejudice, it is quite plausible that the discrimination results will differ among the contestant groups. Prior perceptions about groups are formed based on interactions with them and it is unlikely that non-blacks have had the same interaction with blacks that other blacks have had. In their model of screening discrimination, Cornell and Welch (1996) put forward the premise that people can evaluate others more accurately when they are of a similar cultural

(racial) group.

Two different tests are used to determine whether contestants discriminate. The first test examines if contestants' predictions of whether particular savants answered questions correctly depends on the savant's race and gender, once all other relevant factors are controlled for. The results show that both black and non-black contestants predict black savants are less likely to answer questions correctly than observationally equivalent non-black savants, although this differential goes away for black contestants later in the game. Because the level of interaction between contestants and savants is not affected by contestants' predictions, it is unlikely taste discrimination plays a role in contestants' decisions. This implies that contestants hold lower prior perceptions about black savants. It is important to note that the questions used in this game show come from a variety of different categories. When the results are broken down by question category, one finds that non-black contestants perceive black savants to have equal ability to non-black savants in answering general knowledge questions, but lower ability in answering miscellaneous, entertainment, and slang/common saying questions. The situation for black contestants is less clear, as the smaller sample size makes it difficult to stratify the results by question category.

The second test used to test for discrimination can determine if contestants hold lower prior perceptions about black savants when we average across all question categories, but cannot determine this within categories. This test examines contestants' choice of savant at the beginning of the third round, when they must select a savant they have to make three future predictions about. Because this choice can potentially be impacted by taste discrimination, this test needs to distinguish between the use of group stereotypes and taste discrimination. This test accomplishes this by exploiting unique features of the game, and can thus provide insight into what situations are necessary in general to be able to determine why discrimination occurs.

The two features of the game that enable one to distinguish between the types of discrimination are: 1) since contestants must select a street savant at the beginning of the third round who they will have to make three future predictions about, they want to select the savant they

perceive as having the most extreme skill level (i.e., the savant most likely to either get the next question right or wrong) and (2) contestants' perception of a savant's skill level when they make this selection depends on both their prior perception about the savant's skill level as well as the savant's performance in answering five questions during the first two rounds of the game.

These two features result in group stereotypes and taste discrimination having different effects on contestants' choice of savant at the beginning of the third round. To see this, consider a simple example with black and non-black savants. If contestants have taste discrimination against black savants they should not want to interact with them. Thus whether one looks among savants that performed poorly in the first two rounds or among savants that performed well, black savants are less likely to be chosen.

If instead contestants discriminate against black savants because they have a low prior perception about them and a high prior perception about non-black savants, a different choice pattern should emerge. Among savants that performed poorly in the first two rounds, black savants are considered more extreme (contestants perceive these low performing black savants to be low skilled overall, and these low performing non-black savants to be average skilled overall). Among savants that perform well the opposite result is obtained, with non-black savants considered more extreme. Since contestants want to choose the savant with the most extreme skill level the following pattern should be observed: among low performing savants, blacks are more likely to be selected, while among high performing savants non-blacks are more likely to be selected. Thus, if stereotypes are being used, the effect of a savant's race on their likelihood of being selected depends on their performance in the first two rounds, while with taste discrimination it does not. Explicit empirical support is shown for the key assumptions necessary for this identification strategy to be valid.

This identification strategy is used to test for different prior beliefs in a descriptive manner and within the context of a formal structural discrete choice model. The results of both the descriptive and structural analysis confirm the findings of the first test used. Namely, there is evidence that non-black contestants have lower prior perceptions about black savants' ability to

answer the average question. There is no evidence of discrimination among black contestants. As this identification strategy is applied at a later point in the game, this conclusion is also consistent with the first test, which found that later in the game black contestants do not discriminate.

The organization of this paper is as follows: Section 2 explains how *Street Smarts* is played, Sections 3 and 4 set up and carry out the two alternative tests for discrimination, and Section 5 discusses why the nature of the game makes it difficult to determine whether contestants are using accurate or inaccurate stereotypes. Finally, Section 6 concludes with a discussion of the external validity of these results.

2 Background on *Street Smarts*

Street Smarts is a game show in which two contestants compete to most accurately predict whether a person will correctly answer a random trivia question. Prior to the show, the host goes out to popular locations all over the U.S. and talks separately to three random people, termed street savants. The savants, who are selected by producers, are sometimes from the cities they are interviewed in, but oftentimes are just vacationing there. Each of the savants are asked the same set of trivia questions. These trivia questions are all designed to have the same (moderately easy) difficulty level but they span many different categories, including general knowledge, entertainment, sports and more. Table 1 shows the six different question categories, as well as some examples of each.³ The two contestants playing the game then have to predict whether these street savants answered the questions right or wrong. The game consists of four rounds, with each round having a slightly different structure.

At the start of the show, the host introduces the two contestants and viewers can observe their race and gender. The contestants are then introduced to the savants they will be making predictions about by being shown a short interview between the host and each of the three street savants. Contestants typically learn the occupation of each of the savants during the course of the interview, and can observe their race and gender.

During Round 1 the host reads a question and contestants are told that one of the three street savants answered the question right and the other two got the question wrong. The contestants must simultaneously predict which street savant got the question right. Contestants are not told the answer to the trivia question before they make their prediction. Contestants then observe how each street savant answered the question by viewing the videotaped clip of the host asking the savant the particular question. There are three questions in this round and each right prediction by a contestant is worth \$100. There are no penalties for wrong answers.

During Round 2 the host picks two of the three street savants and then reads a question. Contestants are told that one of the two savants got the question right and the other got the question wrong. Contestants must simultaneously predict which of the two savants got the question wrong, and are then shown how each of the two savants answered the question. There are three questions in this round, and for each question a different combination of two of the three street savants is chosen. Each question is worth \$200 with no penalties for wrong answers.

By the start of Round 3 contestants have seen each savant answer five questions. Contestants must now each choose a savant that they have to make predictions about during this round. The two contestants cannot choose the same savant, so the player trailing at the start of the round chooses first, and the other contestant chooses their savant from the remaining two. The contestant that chose first is then told a question, and has to predict whether the savant they picked answered the question right or wrong. They are then shown how their savant answered the question. The other contestant is then told a different question and must predict how their savant answered this question. This process repeats until both contestants have predicted how their savants answered three questions. Each correct prediction is worth \$300 in this round with no penalties for wrong answers.

In Round 4 the host reads a question to the contestants. Each contestant then picks a street savant and predicts if they answered the question right or wrong and wagers an amount of money that cannot exceed what they have currently earned in the game. If the contestant

predicts the savant's answer correctly they add the amount of money they wagered to their total. If the contestant incorrectly predicts the savant's answer they lose the amount of money they wagered from their total. This concludes the game. The player with the most money wins and keeps their win total, which can be anywhere between \$1 and \$4600, and the losing player gets nothing.

It is important to note that producers have complete control over how this game is put together. Each of the savants has been asked numerous questions before the game. Only ten of these answers are included in the show and it is producers who choose which answers and the order in which these answers appear. This opens up the possibility that producers could manipulate the game to try and trick contestants. The implications this could have on the identification strategy is discussed in Section 4.1.2.

Data for 299 episodes were collected, resulting in a total of 598 contestants and 897 street savants.⁴ Table 2 provides information about the demographic characteristics of both the contestants and savants. Race was subjectively classified by the author into one of the five groups listed. Individuals that were difficult to identify were classified in the other/unknown group. 63% of the contestants are white, 22% are black, 7% are Hispanic, 6% are Asian, and 2% are unknown. For the savants, 59% are white, 24% are black, 8% are Hispanic, 6% are Asian, and 3% are unknown.⁵ Both genders are, for the most part, equally represented in the game: for contestants, 52% are female and 48% are male, while for savants 55% are female, and 45% are male.

While there is not explicit information on contestants and savants' education level, their occupation level can be used as an indicator of whether they went to college or not.⁶ The percentages of college educated people are very similar across both contestants and savants, with 24% of the contestants and 23% of the savants being college educated.⁷ A higher percentage of blacks are college educated than whites across both contestants and savants. While male contestants tend to be more educated than female contestants, female savants are more educated than male savants, although the differences are not large.

Due to the small number of contestants that fall into the Hispanic, Asian and unknown

racial categories, these groups are henceforth combined with whites to form the group non-blacks.

3 Do Contestants Discriminate?

Table 3 shows how contestants' predictions about which savants answered questions correctly in Rounds 1-3 depend on the characteristics of the savant.⁸ As the factors that should affect contestants' decisions can be controlled for, if the decision depends on the race or gender of the savant, it will indicate contestants are discriminating against certain groups. Furthermore, because contestants' decisions do not affect the level of interaction between contestants and savants, taste discrimination is unlikely to occur in this setting, implying that racial/gender prediction differentials are likely driven by statistical discrimination.⁹

The structure of Rounds 1 and 2 is different than Round 3, which necessitates evaluating them separately. Specifically in Rounds 1 (and 2) only one of the three (two) savants answered the question correctly, and contestants must choose which one. To determine if contestants' prediction depends on the group status of savants, a discrete choice model is estimated. The utility contestants get from predicting savant i answered correctly is defined as:

$$U_i = X_i\beta + \varepsilon_i \tag{1}$$

where X_i includes the savant's previous proportion of questions answered correctly, and whether the savant is college educated, black, and/or female.¹⁰ ε_i is an error term that picks up contestants unobserved utility for selecting a particular savant. Each ε_i is assumed to be distributed i.i.d. extreme value, resulting in the probability that contestants select savant i being specified as:

$$P_i = \frac{e^{X_i\beta}}{\sum_k e^{X_k\beta}} \tag{2}$$

where k is the number of savants contestants can choose between. Maximum likelihood

estimation is used to find the vector of β coefficients that make contestants actual choice of savant as likely as possible. Marginal effects are shown in the table to make the coefficients easier to interpret. Because contestants choose between the same group of savants six times in the first two rounds, the errors might be correlated across questions for a particular contestant. To account for this, standard errors are clustered at the contestant level.

The results shown in Column 1 of Table 3 indicate that contestants are significantly less likely to predict that black savants answered the question correctly in the first two rounds, but that their predictions do not depend on the gender of the savant. Columns 3 and 4 separate out the results by contestant race and show that both non-black and black contestants behave in the same way. Specifically, non-black contestants perceive black savants to be 4% less likely to answer a question correctly than non-black savants, while for black contestants this differential is 6%. Together, the results imply that all contestants discriminate against black savants in the early rounds.¹¹ No evidence of gender discrimination was found when the results were separated out by contestant gender.

In Round 3 the structure changes, as contestants only predict whether a particular savant answered a question correctly. Columns 5-8 of Table 3 present the results from an OLS regression where the dependent variable is an indicator for whether the contestant predicted the savant answered the question correctly.¹² Standard errors are again clustered at the contestant level to take into account that contestants make a prediction about the same savant three times. The results for all contestants mirror those found in Rounds 1 and 2 as one finds contestants continue to discriminate against black savants. However, when the results are separated out by contestant race, one finds that non-black contestants believe that black savants are 7% less likely to answer questions correctly, but black contestants view these groups as equivalent. Thus it seems the aggregate result is caused by non-black contestants only, as black contestants do not discriminate anymore.

The above analysis groups all question types together, which implicitly assumes that the categories of the questions do not matter to contestants when they make their predictions. One potential issue with this is that because contestants hear the question before they have

to make their prediction, they will rely on how the savant has answered previous questions in that category, as opposed to how they have answered previous questions overall. This would imply the proper control to use would be the category-specific proportion of questions answered correctly, as opposed to the overall proportion that was used in the main analysis in Table 3. Columns 2 and 6 of Table 3 investigate this possibility by including both of these variables as controls.¹³ Interestingly, one finds that contestants do not rely on the category-specific proportion when making their predictions, and instead use the overall proportion correct. Thus, the previous specification used with overall proportion correct is valid.

Another issue with ignoring the categories of questions is that contestants' prior perceptions about savants' ability to answer questions may be category-specific. This means the results in Table 3 will reflect contestants' average prior perception about savants across all categories; however, it might be more interesting to examine the category-specific priors, especially since the question types are so different. Table 4 stratifies the Round 1 and 2 results from Table 3 by question category. All specifications include controls for overall proportion correct and education level, but the results are not shown due to brevity reasons.¹⁴ Due to sample size, only the four most prominent question categories were used. Non-black contestants view black and non-black savants as equivalent in answering general knowledge questions, but view black savants as less skilled in answering miscellaneous questions, entertainment questions, and slang/common saying questions. Thus the overall lower prior perception non-black contestants have of black savants found in Table 3 is only caused by the latter three categories. The potential external relevance of these category-specific results will be discussed in Section 6. The situation for black contestants is not as clear—all categories have a similar negative coefficient for black savants, implying black contestants have a lower perception about black savants in all categories. However, none of these coefficients are statistically significant, potentially due to the small sample sizes in these regressions. With respect to gender, neither non-black or black contestants view males and females any differently across the categories.

4 An Alternative Test of Discrimination

This section develops an alternative test of discrimination which relies on examining contestants' choice of savant at the beginning of the third round. This additional test is helpful for a couple reasons. First, it strengthens the overall conclusion if this test, which uses a completely different identification strategy and outcome measure, provides the same result as the first test. Second, this test involves examining a choice which potentially can be affected by taste discrimination. This requires the test be able to distinguish between the use of group stereotypes and taste discrimination, which previous discrimination literature has shown to be difficult to accomplish. Because the identification strategy used here capitalizes off the unique incentive structure present in this game show, it provides insight in how to distinguish between these two types of discrimination more generally.

Throughout this section it is assumed for simplicity that the specific categories of questions do not matter. As will be discussed at the end of Section 4.1.2 the main implication this has is that we will be identifying contestants' prior perceptions averaged across all categories (akin to the results in Table 3), as opposed to identifying category-specific perceptions. Because the evidence of discrimination in the previous section was found primarily along racial lines, this section focuses exclusively on testing for racial discrimination.

4.1 Identification Strategy

There are two features about contestants' choice of savant at the beginning of the third round that enable one to distinguish between group stereotypes and taste discrimination. The first feature is that contestants' selection of savant is based on how extreme the savant's skill level is. Recall that during the third round contestants have to make predictions about how their savant answers three questions. Thus, they should want to choose the savant whose future answers are easiest to predict. This implies they choose the savant who they feel is either most likely to get the next question right or most likely to get the next question wrong. This feature of selecting someone based on the extremeness of their skill level is very unique to this

game show.¹⁵

The second feature of this decision is that a contestant's perception of a savant's skill level when they make this selection should depend on both their prior perception of the savant's skill level, as well as the savant's performance in answering five questions during the first two rounds of the game. The combination of these two features cause different priors and taste discrimination to have different effects on a contestant's choice of savant in the third round.

4.1.1 Identification in a Simple Example

Suppose contestants taste discriminate against black savants. Consider the situation at the beginning of the third round where contestants must choose between a black and non-black savant who performed equivalently in the first two rounds. These two savants are considered equally skilled, because contestants' prior perceptions about the two groups are also the same. However, racially prejudiced contestants can have an aversion to interacting with black savants, and thus are more likely to select the non-black savant. Recall that contestants' monetary gain during this round will only be tied to the savant they choose and thus they only have to pay attention to the videotaped answers their chosen savant gives. Because this interaction is minimal, we are unlikely to find taste discrimination here. However, because it is still possible that contestants will want to avoid even this minimal interaction, it is important to use a method that can control for any potential taste discrimination. If taste discrimination is present, this pattern of choosing the non-black savant over an equivalently performing black savant will always occur, no matter whether the non-black and black savant both performed poorly or both performed well in the first two rounds. This selection pattern can be summarized in the following table:

	Group contestants are most likely to select
high performance	non-blacks
low performance	non-blacks

Now consider the situation where there is no taste discrimination, but instead contestants

have different priors about black and non-black savants. Specifically, they have a very low prior perception about black savants, and a very high prior perception about non-black savants. Contestants' perception of the savant's skill level at the beginning of the third round is an average of their prior perception and the savant's performance. Thus a low performing black savant is considered low skilled overall, while a low performing non-black savant is considered average skilled overall. Since contestants want to select the savant with the most extreme skill level, black savants are more likely to be selected among low performing savants. Among high performing savants the opposite result is obtained: black savants are considered average skilled, while non-black savants are considered highly skilled, resulting in non-black savants being more likely to be selected. Thus, with different prior beliefs the relationship between contestants' selection decision and savant performance can be summarized in the following way:

	Group contestants are most likely to select
high performance	non-black
low performance	black

By comparing the selection pattern under different prior beliefs and taste discrimination the following idea should be clear: with different prior beliefs the effect a savant's group status has on their likelihood of being selected depends on their past performance, while with taste discrimination this effect is independent of past performance. This is the key distinction that makes it possible to distinguish between these two types of discrimination. The unique feature of choosing the player with the most extreme skill level can be contrasted with traditional hiring decisions in the labor market, where employers typically select the applicant with the highest skill level. Examining the table for different prior beliefs, one can see that this would imply that among savants who perform poorly, non-black savants would be more likely to be selected (since an average skill level is greater than a low skill level). Thus the choice pattern

is the same under both different prior beliefs and taste discrimination, making it difficult to distinguish between them in a labor market scenario.

4.1.2 Formal Analysis of the Identification Method

Since the above example is quite specific, it is important to formalize what is required for this identification strategy to work in general. Assume contestants view each question a savant answers as a Bernoulli trial, where $p_i \in (0, 1)$ denotes the true probability that savant i gets each question correct. This requires that contestants do not view questions as being category-specific. The implications of this assumption will be further discussed later this section. Contestants do not know a savant's true p_i and thus have to use their perception of it.

Define $\mu_B, \mu_N \in (0, 1)$ as the prior perceptions contestants have about the average skill level of black and non-black savants, respectively. This can be thought of as the probability they feel a savant from a particular group correctly answers a question at the beginning of the game (before they have seen them answer any questions). Define $\widehat{p}_{B,i}$ ($\widehat{p}_{N,i}$) as the contestant's perception of a black (non-black) savant's true skill level at the beginning of the third round. Assuming this is just a weighted average of their prior perception and the savant's performance in the first two rounds, $\widehat{p}_{B,i}$ and $\widehat{p}_{N,i}$ can be written as follows:

$$\widehat{p}_{B,i} = (\sigma)\mu_B + (1 - \sigma)x_i \tag{3}$$

$$\widehat{p}_{N,i} = (\sigma)\mu_N + (1 - \sigma)x_i \tag{4}$$

where $x_i \in (0, 1)$ is the proportion of questions savant i answered correctly in the first two rounds, and $\sigma \in (0, 1)$ denotes how much weight is put on the prior.¹⁶

The utility a contestant gets from selecting a savant at the beginning of the third round can be defined as a function of the extremeness of the savant's skill level as well as any taste

discrimination they may have:

$$U_{B,i} = \max(\widehat{p}_{B,i}, 1 - \widehat{p}_{B,i}) + \theta + \varepsilon_i \quad (5)$$

$$U_{N,i} = \max(\widehat{p}_{N,i}, 1 - \widehat{p}_{N,i}) + \varepsilon_i \quad (6)$$

where ε is a random disturbance term with mean 0. In this model, different prior beliefs mean contestants perceive $\mu_B \neq \mu_N$. This affects the utility contestants get from selecting each savant through the \widehat{p}_i term. On the other hand, if contestants have a taste preference towards black or non-black savants, they get more utility from picking them. Because this personal taste should not affect their perception of the savant's skill level, \widehat{p}_i , it enters into the utility function separately. The parameter θ measures this taste preference. If θ is negative, contestants have a taste preference towards non-black savants, while if θ is positive they have a taste preference towards black savants. Note that a taste preference for non-black savants can pick up two things: (1) contestants' personal prejudice against black savants or (2) contestants' preference to select non-black savants because they feel they can predict their answers better. For ease of exposition θ is referred to as the taste discrimination parameter throughout this section. However, the complexity of this taste parameter is taken into account when discussing the empirical results. Regardless, this taste parameter will control for any taste discrimination present, allowing for a clean identification of the use of different priors.

The key to the identification strategy is that taste discrimination and different prior beliefs cause the expected utility differential between black and non-black savants to change differently as performance level x increases. This is evident by examining the different cases that can arise.

First, consider the case where there is taste discrimination against black savants, but prior beliefs are the same. This means $\mu_B = \mu_N$ and $\theta < 0$. Figure 1A shows the expected utility contestants get from selecting black and non-black savants over the different performance levels when $\mu_B = \mu_N = .6$, $\sigma = .5$, and $\theta = -.2$. One should see that the expected utility differential

between black and non-black savants is constant, no matter what the performance level. From looking at the utility specifications it should be evident that this always happens, no matter what the values of $\mu_B = \mu_N, \sigma$ and θ .

Now consider the case where there are lower prior beliefs about black savants, but no taste discrimination. This means $\mu_B < \mu_N$ and $\theta = 0$. In this case two situations can result depending on the values of μ_B, μ_N and σ . In the first situation the difference in expected utility contestants get from black and non-black savants depends on the savant's performance level. Figure 1B shows an example of this, where $\mu_B = .2, \mu_N = .8, \sigma = .5$, and $\theta = 0$. These parameters correspond to the different prior beliefs example presented in Section 4.1.1. One can see that the expected utility differential between black and non-black savants depends on the performance level, which results in black savants being more likely to be selected at low performance levels and non-black savants more likely to be selected at high performance levels.¹⁷

Different prior beliefs can also result in a different situation, however, where the expected utility differential between black and non-black savants is constant, no matter what the performance. This is the same pattern that taste discrimination results in. Figure 1C shows an example of this where $\mu_B = .65, \mu_N = .85, \sigma = .8$, and $\theta = 0$. This latter result is problematic because if this pattern is observed empirically one cannot determine if it is caused by different prior beliefs or taste discrimination. (Note that the pattern in Figure 1C can also be generated by the parameters $\mu_B = \mu_N = .85, \sigma = .8$, and $\theta = -.16$.) Thus the only way to empirically distinguish different prior beliefs from taste discrimination is if different prior beliefs always result in the first situation.

There are two cases where different prior beliefs do not result in the first situation:

$$\text{Case 1: } \widehat{p}_B > .5 \text{ and } \widehat{p}_N > .5 \text{ at } x = 0$$

$$\text{Case 2: } \widehat{p}_B < .5 \text{ and } \widehat{p}_N < .5 \text{ at } x = 1$$

The scenario represented in the first case is what is shown in Figure 1C. The scenario represented in the second case would result in two parallel downward sloping lines. In either of these situations the max function is not necessary in the utility functions of black and non-black savants.¹⁸ Because this function is precisely what causes expected utility differentials between black and non-black savants to depend on performance level when different prior beliefs are present, it means that when the max function is not important one cannot distinguish between different prior beliefs and taste discrimination. Luckily, both the descriptive and structural analyses indicate the data does not reflect either of these situations.

In order to use the identification strategy outlined above, several key assumptions built into the model must hold. First, contestants must choose the savant with the most extreme skill level. Empirical support for this assumption is shown when the model is structurally estimated in Section 4.3.2.

The second assumption is that contestants use a savant's past performance as a predictor of future performance. While this seems plausible, this may not happen if producers manipulate the game and contestants are aware of it. Prior to the actual in-studio game, savants are asked numerous questions. From these, producers select which ten questions to use in the game, as well as the order in which they are shown. If producers manipulate the order of the questions such that savant performance in the first two rounds is completely unrelated to their performance in the third round, contestants aware of this will not put any weight on x_i .

To determine whether or not contestants believe past performance is positively related to future performance, an indicator for whether the contestant predicted the savant answered a question correctly in the third round can be regressed on the savant's previous performance, education level, race and gender. These regression results were previously presented in Column 5 of Table 3. The coefficient on past performance is positive and strongly significant, indicating contestants believe past performance is an important predictor of future performance. If one instead regressed whether or not the savant actually answered the question correctly on these same regressors (results not shown), the coefficient on past performance is negative and significant, indicating past performance is strongly negatively related to future performance.

Together these results imply producers do manipulate the game, but contestants are not aware of it. The only consequence of this is that contestants do a poor job of predicting savants' answers. This is not important because the objective is not to determine how accurate contestants' perceptions of savants are—it is to identify their actual perceptions.¹⁹

The third key assumption made throughout this section is that contestants do not view the questions as being category-specific, and thus choose the savant with the most extreme skill level averaged *across* all categories. However, because contestants learn the questions before they make their predictions about savant answers, it is also possible they will choose the savant that is the most extreme *within* each category. Consider a case where there are only two types of questions—half are general knowledge and half are miscellaneous. Suppose a savant was perceived to do poorly at general knowledge questions (based both on the prior perception of them in this category, as well as how they answered these types of questions in the first two rounds), and was perceived to do well at miscellaneous questions. If contestants do not consider categories, this savant would not be chosen because they are not extreme when we average across all categories. However, if categories are important, this savant is clearly extreme within each category, and thus would be ideal to choose. This alternative selection pattern will clearly affect the identification strategy used above.

To determine the validity of the current identification strategy two things need to be examined: first, do contestants use overall past performance or category-specific past performance when making their predictions? Second, do they have different prior perceptions regarding savants for different categories? The results from Columns 2 and 6 of Table 3 answer the first question, as they show contestants only consider the overall previous proportion correct, and not the category specific percentage, when making predictions. This implies the only measure of past performance the contestant considers when choosing a savant is their overall past performance, and thus the part of the identification strategy that relies on overall past performance should continue to hold.

The results from Table 4 answer the second question, as they clearly show contestants have different prior perceptions about savants for different categories. The only implication this

should have is on the interpretation of the results. Instead of assuming all questions are in one category (i.e. the questions are successive Bernoulli trials) and we are identifying contestants' perceptions about savants' ability to answer this one type of question, we need to explicitly recognize that questions come from different categories and this method allows us to identify prior perceptions about savants' ability averaged across all categories.²⁰ To see this, note that among savants that perform poorly in the first two rounds, for any category that contestants have a low prior perception about them, they will be perceived to perform extremely bad in that category. For any category where contestants have a high prior, they will be perceived to perform average in that category. (This takes into account the fact that contestants use overall past performance when determining how savants will perform in a particular category.) If contestants want to choose the savant that is the most extreme in every category, they will pick the savant they have the lowest prior about in the most categories. This savant will also be the one that has the lowest average skill level when we average across all categories. Thus at the low performing end, choosing the savant that is the most extreme in every category is synonymous with choosing the savant with the lowest average skill level across all questions. Repeating this analysis at the high performing end, one will find that choosing the savant that is the most extreme in every category is synonymous with choosing the savant with the highest average skill level across all questions. Thus the choice pattern at the low and high performing ends will allow us to identify contestants' perceptions about particular groups of savants' ability to answer the average question, although it cannot determine the category-specific perceptions. This will be taken into account when the results are interpreted.

4.2 Descriptive Analysis

The methodology outlined in the previous section shows different prior beliefs result in the differential in expected utility contestants receive from different groups of savants depending on the savant's past performance level in a particular way. This implies a simple descriptive test of different prior beliefs can be carried out by examining how the probability a member of a particular group is selected at the beginning of the third round depends on their performance

level. Essentially, this treats the probability a group member is selected as a proxy for the expected utility contestants get from selecting members of that group. Furthermore, the shape of the graph should enable one to determine which group is perceived as having the lower prior skill level.

Figure 2A shows, for non-black contestants, how the probability of a savant in a particular racial group being selected in the third round depends on their performance in the first two rounds. Due to the paucity of observations, proportion correct is condensed from six to three categories. The trend line for blacks indicates that 48% of the black savants that answer between zero to twenty percent of their questions correctly in the first two rounds are selected at the beginning of the third round. This percentage drops to 31% among black savants that answer between eighty to one-hundred percent of their questions correctly.²¹ The patterns for black and non-black savants clearly show that non-black contestants have different prior beliefs about the two groups, as the probability differential between the two groups changes a lot over the different performance levels. The probability differentials at the low and high performance levels are statistically different from each other at the 10% level.²² (The results are not more significant due to the relatively small number of observations at the extreme performance levels.) The downward sloping shape for blacks is consistent with $\widehat{p}_B < .5$ at $x = 0$. The upward sloping curve for non-blacks is consistent with $\widehat{p}_N > .5$ at $x = 0$. Together, this means non-black contestants perceive $\mu_B < \mu_N$. This is strong evidence that non-black contestants have lower prior perceptions about black savants ability to answer the average question in the game.

Figure 2B shows these results for black contestants. Recall that the analysis done in Section 3 implied that black contestants no longer discriminate against black savants by this point in the game, so one would expect two identical lines. The actual results are unclear, as the picture for black savants does not match a pattern the model would predict. This might be partly due to the small number of observations at the high performance level. Another reason it is difficult to interpret the results for black contestants is likely caused by the fact that in this simple descriptive test of the model, the probability a given person is selected

is measured as whether or not they were chosen in the third round. This does not control for the characteristics of the other two savants that were available to be chosen, nor does it control for any of the other characteristics about the savant besides their race. To control for the information that contestants have at the time they make their decision the next section estimates a structural discrete choice model of a contestant's choice of savant at the beginning of the third round.

4.3 Structural Analysis

4.3.1 Structural Model and Simulation Results

The goal of the structural estimation is to estimate the parameters of a model similar to the one outlined in Section 4.1.2. That model allows savants to belong to one of two groups, but savants can differ by their education level and race. To take this into account, and allow non-black and black contestants to have different prior beliefs about both non-black and black savants, contestant j 's prior perception about savant i can be written as:

$$\begin{aligned} \mu_{ij} = & \omega_1 + \omega_2 \cdot COLLEGE_i + \omega_3 \cdot BLACK_j \cdot BLACK_i + \\ & \omega_4 \cdot BLACK_j \cdot NON - BLK_i + \omega_5 \cdot NON - BLK_j \cdot BLACK_i \end{aligned} \quad (7)$$

where $0 \leq \mu_{ij} \leq 1$. $COLLEGE_i$ indicates savant i has attended or is currently attending college. The race variables are indicators that pick up the racial match of the contestant and the savant. For example, $BLACK_j \cdot BLACK_i$ takes the value of 1 if both contestant j and savant i are black. The parameter ω_1 represents the prior perception non-black contestants have of non-college educated, non-black savants. Recall that this prior perception can be interpreted as the probability with which contestants think a savant will answer the average question correctly at the beginning of the game. The parameter ω_5 picks up the difference between non-black contestants' prior perception of black savants and non-black savants. If $\omega_5 < 0$, non-black contestants have lower priors of black savants than non-black savants.

Similarly, finding $\omega_3 < \omega_4$ implies black contestants have lower priors of black savants than non-black savants. Note that comparing ω_3 with ω_5 does not make a good test for racial stereotypes because black contestants could be systematically different than non-black contestants. If black contestants tend to have higher ratings of everyone than non-blacks do, finding evidence that black savants were perceived better by black contestants than by non-black contestants would be due to this and not racial stereotypes.

Contestant j 's perception of savant i 's skill level at the beginning of the third round can be written as:

$$\widehat{p}_{ij} = (\sigma) \mu_{ij} + (1 - \sigma) x_i \quad (8)$$

The utility contestant j gets from selecting savant i is a function of both the extremeness of the savant's skill level and any taste preference they may have towards the savant. The utility function defined below is the simplest way to include both of these requirements:

$$U_{ij} = \gamma \cdot \max(\widehat{p}_{ij}, 1 - \widehat{p}_{ij}) + \theta_1 \cdot BLACK_j_BLACK_i + \theta_2 NON - BLK_j_NON - BLK_i + \epsilon_{ij} \quad (9)$$

where γ is a scale parameter and the θ 's are taste parameters. $\theta_1 > 0$ indicates black contestants have a taste preference for choosing black savants over non-black savants. Likewise if $\theta_2 > 0$ non-black contestants have a taste preference for selecting non-black savants over black savants. This model is consistent with how contestant utility is defined in Section 4.1.2. However, Monte Carlo simulations could not find the true parameter values, most likely because the utility function is discontinuous.²³ To keep a similar framework but make the utility specification a continuous function, one can redefine it as:

$$U_{ij} = \gamma \cdot (\widehat{p}_{ij} - \frac{1}{2})^2 + \theta_1 \cdot BLACK_j_BLACK_i + \theta_2 NON - BLK_j_NON - BLK_i + \epsilon_{ij} \quad (10)$$

Note that this function still requires savants with the most extreme skill levels to generate

the highest utility, which means the identification strategy presented in Section 4.1 still applies. The main difference with this new function is it also requires utility to increase exponentially as the savant's skill level becomes more extreme, rather than linearly. This means the utility differential between two savants that have skill levels \hat{p} of .85 and .75, respectively, is larger than the utility differential between two savants that have skill levels of .65 and .55, respectively. This should not make a huge difference in the results.

While theoretically all nine of these parameters ($\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \theta_1, \theta_2, \gamma$ and σ) should be identified, both Monte Carlo simulations and the actual estimation indicate that γ and σ cannot be separately estimated. Significantly different values of γ and σ resulted in the same log likelihood. To resolve this it is necessary to fix either γ or σ . Since γ is a very non-intuitive parameter it is difficult to know a reasonable value(s) to fix this parameter at. On the other hand, σ is a more intuitive parameter that only takes on a small range of values. Furthermore, simple descriptive statistics can give a good idea as to the value of this parameter. In the regression results in the fifth column of Table 3 the coefficient on the proportion correct in rounds 1 and 2 is about .27. This means if the proportion correct in the first two rounds is increased from 0 to 1, the probability that contestants predict that savants get the question correct increases by .27. The difference between contestants' perception of two savants (say 1 and 2) who are equivalent in all characteristics except their performance level at the beginning of the third round is:

$$\hat{p}_1 - \hat{p}_2 = (1 - \sigma)(x_1 - x_2) \tag{11}$$

The regression results imply that if $(x_1 - x_2) = 1$, $\hat{p}_1 - \hat{p}_2 = .27$. This implies $\sigma = .73$. As this estimate seems slightly high, values of $\sigma = .5, .65$, and $.75$ are tried. As long as the results are robust over these reasonable values one can be confident that even if the true σ was known the results would not be that different.

Assuming ϵ_{ij} is distributed i.i.d. extreme value, the probability that contestant j chooses savant i is²⁴:

$$P_{ij} = \frac{e^{V_{ij}}}{\sum_k e^{V_{kj}}} \quad (12)$$

where V_{ij} is the observed portion of the utility in (10). Maximum likelihood estimation is used to find the parameters $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \theta_1, \theta_2$, and γ that make contestants' actual choice of savant at the beginning of the third round as likely as possible, subject to $0 \leq \mu_{ij} \leq 1, \forall i, j$.

The data set includes 598 observations (in each of the 299 shows, each of the two contestants select a savant). To determine how well the estimation method used uncovers the true parameters in a sample of this size, Monte Carlo simulations were performed on the specification in (10). 1000 simulated data sets of 598 observations each were created. Each of the 1000 data sets was created by randomly drawing with replacement 598 observations from the original data set of 598 observations. Specified parameters are used, along with an extreme value distributed error term, to determine contestants' choice of savant in these simulated data sets. The estimated parameters were found for each of the 1000 data sets. The distribution of these estimates should provide a reasonable idea about what the sampling distribution of each of the parameter estimators looks like.

Column 1 of Table 5 shows what the true parameters are. Columns 2 and 3 report the mean and standard deviation of the sampling distributions for each of the parameters. The estimator does a relatively good job of estimating the constant ω_1 , the taste parameters θ_1 and θ_2 , and γ . However, the estimators of ω_2, ω_3 and ω_5 are biased towards 0. The estimator of ω_4 is also biased towards zero, but does not show up here because the parameter it is estimating is actually zero.²⁵ This needs to be taken into account when interpreting the actual estimation results presented in the next section.

To determine if the estimation method is consistent, the simulations were redone for 1000 data sets of 15000 observations each. These results are shown in Columns 4 and 5. The estimation method works much better with this larger sample size, as the parameter estimates are closer to the true parameters and the standard deviation of the sampling distributions have decreased for the most part. Together these results imply that for most of the ω parameters,

the estimator used is biased but consistent.

4.3.2 Structural Results

Table 6, Column 1 presents the structural estimation results of the model specified in (10). The model estimated in Column 2 is just a slightly different version of this main specification, as it condenses the two taste parameters into one *RACE_MATCH* variable, which takes the value of 1 if the contestant and the savant are the same race and 0 otherwise. In both specifications, σ is assumed to be .75. The model was also estimated with $\sigma = .5$ and $\sigma = .65$, but the results are not shown for brevity reasons. While the magnitudes do change over the σ specifications, the sign and statistical significance does not change. In order to make sure the results are robust to the model specification, the interpretation of the results mainly focus on the sign and significance of the parameters.

Before examining the race parameters it is useful to check whether the other results seem realistic. Most importantly, the parameter γ reveals whether utility depends on the extremeness of a savant’s skill level. The results show γ is positive and significant, indicating that the more extreme a savant’s skill level, the more utility a contestant gets from selecting them. This provides explicit empirical support to the assumption made in Section 4.1 that contestants select the savant with the most extreme skill level. The parameter ω_2 reveals how much contestants’ prior perception of a savant changes when the savant is college educated. As one might expect, this parameter is always positive, although it is not significant.²⁶

Recall that for the specifications presented in Table 6, $\omega_1 + \omega_2 \cdot COLLEGE_i + \omega_3 \cdot BLACK_j - BLACK_i + \omega_4 \cdot BLACK_j - NON - BLK_i + \omega_5 \cdot NON - BLK_j - BLACK_i$ is contestant j ’s perception of savant i ’s probability of answering the average question correctly at the beginning of the game. The estimate of ω_5 in Column 1 implies that at the beginning of the game, non-black contestants perceive that black savants are 8% less likely to get the average question correct than non-black savants. While this result is not significant in Column 1, it is significant in Column 2 when the number of parameters is decreased. Furthermore, Monte Carlo simulations showed a strong tendency for these ω estimators to be biased towards zero,

indicating it is likely that these estimates are underestimating the true parameters in absolute value. The estimates of ω_3 and ω_4 reveal what is happening among black contestants. While there is a decent differential in magnitude in both specifications, neither of these differentials are statistically significant. (It is difficult to know the effect of the biased estimator on these results, as both could have been underestimated, resulting in an unbiased estimator finding the same insignificance that was found here.) This implies black contestants have equivalent perceptions of the average skill level of black and non-black savants. None of the taste parameters have a significant impact on contestants' choices.

Overall, the results found here are consistent with both the descriptive analysis results and the results from the first test for discrimination. All three sets of results reveal that non-blacks have lower prior perceptions about black savants ability to answer the average question. Although this second test used cannot determine category-specific perceptions, the first test showed that this overall result is likely driven by non-black contestants perceiving blacks have lower ability in answering miscellaneous, entertainment and slang/common saying questions. For black contestants, both tests imply that black contestants do not discriminate at this later point in the game.

5 Statistical Discrimination versus Inaccurate Stereotypes

The empirical evidence indicates that, at least in the early rounds, all contestants have a lower prior perception of blacks than of non-blacks in answering the average question. A natural question to ask is whether this lower prior is justified. If it is, contestants are using statistical discrimination. However, if it is not, contestants are basing their decisions on inaccurate stereotypes, resulting in inefficient choices.

Earlier it was discussed that the potential manipulation of the game should not affect the ability to identify contestants' true perceptions, because contestants do not seem to know the game is manipulated. Unfortunately, the fact that producers can manipulate the game does make it difficult to answer whether the perceptions contestants use are accurate. First, the

savants selected to be on the show can be non-random, so whether or not the perceptions are accurate for the savants in the game will not reveal if they would be accurate for the general population. Second, as mentioned previously, savants are asked numerous questions and producers pick ten of these that make the show. To make the game more difficult for contestants (and viewers), it might be in producers best interest to choose questions such that no observable savant information, such as race, gender, or education level, has any predictive power over whether the savant answers the question correctly. This potential manipulation means that while we can determine if contestants have incorrect priors about savants *in the game*, we cannot determine the accuracy of these priors *outside the game*. From an external validity standpoint, the latter is the question we want to answer.

For interest's sake, Table 7 shows how actual savant performance in Rounds 1-3 depends on their characteristics. Column 1 estimates a discrete choice model similar to that estimated in the first column of Table 3, except now the choice indicates which savant answered the question correctly in Rounds 1 and 2. Marginal effects are shown in the table. Columns 2-5 break out these results by question type. Column 6 estimates a model similar to that estimated in the Column 5 of Table 3, except now the dependent variable indicates whether the contestant actually got the question correct in Round 3. The results indicate that, for the most part, savant characteristics have no bearing on their performance in the game, either overall or within each question type. This implies contestants are using inaccurate racial stereotypes within the game, as they incorrectly perceive black savants to be of lower skill in answering the average question than non-black savants. Whether or not these stereotypes would be accurate in the real world is not a question this game can answer.²⁷

6 Conclusion

In the past it has proven quite difficult to empirically distinguish between taste discrimination and group stereotypes in situations involving skill assessment, such as hiring. This is primarily because both types of discrimination can lead to the same outcome, where the affected group is

less likely to be hired. This paper finds an alternative situation, the game show *Street Smarts*, where discrimination against blacks occurs, but this problem of determining why doesn't arise.

The results from two different tests of discrimination indicate that non-black contestants perceive that black savants have a lower ability to answer the average question in the game than non-black savants. However, when these results are separated out by question category, one finds this average differential is actually driven by non-black contestants having lower perceptions of black savants' ability to answer miscellaneous, entertainment, and slang/common saying questions. Their perception of savants' ability to answer general knowledge questions is racially neutral. To determine the external validity of these results it is important to think about which of these question types most closely reflect the skills employers might look for. The only question type that is likely to be of any relevance for potential employers is general knowledge questions, because as the examples in Table 1 show, they are a decent measure of general intelligence. The fact that non-black contestants view black and non-black savants equivalently in this regard is promising as it could potentially imply that employers may not statistically discriminate. However due to the very specific nature of this game show, the external validity of these results should be interpreted with caution.

Unlike the results for statistical discrimination, the lack of taste discrimination found in *Street Smarts* are not directly applicable to the labor market. While there was no evidence that contestants decisions reflected racial prejudice, the level of interaction between contestants and savants is much less than employers and employees in a typical firm setting. Determining the presence of taste discrimination in the labor market remains an important topic for future research.

At the start it was said it is difficult to distinguish between types of discrimination in the labor market with currently available data. However, this game show provides insight into what type of data we might want to collect. The unique feature about this game show is that contestants make decisions based on how extreme a savant's skill level is. While it is unlikely that this feature would be found in a labor market setting, there are situations where the opposite feature might be found. Namely, there are many entry-level positions in which

certain applicants are over-qualified and only apply to have a safety net or because they cannot find a better position. Employers will typically not want to hire these individuals out of fear they will either not take the job or quickly leave when a better opportunity arises. Thus in these situations, employers will choose to hire the medium skilled applicants. The same identification arguments that were presented within the context of the game show would be valid in this setting also. Specifically, one could imagine running a study similar in vein to Bertrand and Mullainathan (2004) whereby fake resumes were sent to entry level positions. For each of the employers, one could send resumes for identical black and white pairs that are either under-qualified for the job, exactly qualified, or over-qualified, and then measure the call-back rates. If taste discrimination against blacks is present, one should find that the call-back rate for black applicants is always lower than for white applicants whether we look within under-qualified workers or within over-qualified workers. However, if statistical discrimination against black workers is present, one would expect to find a different pattern. Namely, among under-qualified workers, white applicants should be more likely to be selected, while among over-qualified workers, black applicants should be more likely to be selected. Situations like this could potentially be exploited in future research.

Notes

¹Oettinger (1996) and Altonji and Pierret (2001) develop models of statistical discrimination that make predictions about how the black-white wage gap should change over time within a firm. However, neither model explicitly considers how other types of discrimination (i.e. taste discrimination and inaccurate group stereotypes) might affect this wage gap over time.

²The innovative approach of distinguishing between theories of discrimination using game show data was proposed in two previous papers simultaneously written about the game show *Weakest Link* by Antonovics, Arcidiacono, and Walsh (2005) and Levitt (2004). Several other studies have used game shows to test for things besides discrimination. For example, Beetsma and Schotman (2001) measure risk attitudes in the game show *Lingo*, Berk, Hughson,

and Vandezande (1996) test whether observed bidding patterns in *The Price is Right* can be explained by rational decision theory, Metrick (1995) uses data from *Jeopardy* to analyze behavior under uncertainty and the ability of players to choose strategic best responses, and List (2006) studies factors affecting cooperation in the prisoners dilemma setting present in *Friend or Foe*.

³The questions in the game do not have official categories. The categorization in this table is a subjective classification. Roughly 40% of the questions asked are "general knowledge", 20% are "entertainment", 5% are "sports", 10% are "slang", 7% are "childrens' interests", and 18% are "miscellaneous". The category "miscellaneous" groups together about eight different categories, each of which comprises less than 4% of the total amount of questions.

⁴*Street Smarts* went on for five seasons, beginning in Fall 2000 and finishing in Spring 2005. There were about 800 episodes produced in total. The 299 episodes used in this paper include all of the episodes from the fifth season (recorded from a local television station), and random episodes from each of the previous four seasons (recorded from the cable television station *Game Show Network* which shows reruns of *Street Smarts*).

⁵In 30 of the 299 episodes, two of the three savants are black. None of the episodes have three black savants.

⁶The Occupational Outlook Handbook was used to determine this in the following way: if an individual's occupation requires a college degree or if the individual says they are currently in college, they are classified as college educated. Individuals with occupations that do not require a college degree, or whose occupations are unknown, are classified as non-college educated.

⁷Contestants and savants do not state their age during the show, but rough visual classifications indicate that 89% of the contestants and 75% of the savants are less than 35 years old. As this classification is somewhat subjective this variable is not included in the subsequent analysis.

⁸Round 4 data was not used since contestants have to make three simultaneous decisions in this round.

⁹Contestants are typically shown all of the savants' taped responses, regardless of who they

choose.

¹⁰For the first question answered by savants previous proportion correct is not defined, and thus was set at .5 for all savants.

¹¹Specifications were also run which included an interaction term between *black* and *college*. The interaction term was not significant, implying the race differential is not mitigated by education level.

¹²OLS regression results are shown to make the coefficient interpretations simpler. Results from a logit regression produced similar results.

¹³This specification only includes questions for which there has already been a question asked in that category. This results in many observations being lost.

¹⁴Due to the significantly smaller number of observations in Round 3, this round was excluded from this analysis.

¹⁵One might think instead contestants would pick the savant who they have correctly predicted the most answers for in the first two rounds. However, due to the structure of the game where contestants are simultaneously predicting one savant answered the question correctly and the others answered it incorrectly, each contestant should have an equal success rate across all three savants.

¹⁶This updating structure, which implicitly assumes the signal is normally distributed, is used to keep the model simple. Originally, the updating structure assumed the prior distribution to be beta distributed, and the signal to be binomially distributed, but the model proved too difficult to estimate.

¹⁷When the utility differential between black and non-black savants changes over performance levels it sometimes results in a switch in which group is more likely to be selected (as it does in this example) and sometimes it doesn't. However, whether or not a switch takes place is not important—all that matters is that the utility differential changes over the different performance levels.

¹⁸In Case 1, equivalent figures would result if the utility functions are specified as:

$$U_{B,i} = \widehat{p}_{B,i} + \theta + \varepsilon_i$$

$$U_{N,i} = \widehat{p}_{N,i} + \varepsilon_i$$

¹⁹Related to this, one might also worry that producers select questions in such a manner that savants who would be perceived as highly skilled perform relatively poorly during the game. If contestants realized this they would have no incentive to base their decisions on their true perception of the savant's skill level. However, if this were the case variables like savant's education level should not have a positive effect on predicted performance. The results in Table 3 show this is not the case—contestants perceive education to be an important positive predictor of savants' performance, implying a contestant's true prior perception of the savant's skill level is being identified.

²⁰This is actually a weighted average across all categories, where the prevalence of particular categories of questions determines the weights.

²¹If a savant is selected by the first contestant they are represented in the data set once and coded as being selected. If a savant is not selected by the first contestant but is selected by the second, they are represented in the data set twice and coded once as not being selected and once as being selected. Finally if a savant is not selected by either contestant, they are represented in the data set twice and coded both times as not being selected.

²²Statistical significance of the results is determined by running the following regression:

$$\begin{aligned} choice_i = & \beta_0 + \beta_1 \cdot black_i + \beta_2 \cdot medium_i + \beta_3 \cdot high_i + \\ & \beta_4 \cdot black_i * medium_i + \beta_5 \cdot black_i * high_i + \varepsilon_i \end{aligned}$$

where *choice* is an indicator for whether the savant is chosen, *black* is an indicator for whether the savant is black, *medium* is an indicator for whether the savant's performance level is .4 or .6, and *high* is an indicator for whether performance is .8 or 1. If β_5 is different than 0, it

means the probability differential between selecting black and non-black savants at the high end is statistically different than the differential at the low end. OLS results of this regression (not shown) find this coefficient to be significant at the 10% level.

²³To optimize a discontinuous function, the Nelder-Mead Simplex Search algorithm was used. Monte Carlo simulations showed that this algorithm was not searching very far away from the initial values and is thus not able to find the true parameter values.

²⁴The ε_{ij} 's are likely to reflect a contestant's positive or negative taste for qualities about the savant that are unobservable to the researcher, such as their personality. The i.i.d. extreme value assumption requires that a contestant's perception of a savant's personality is uncorrelated with their perception of another savant's personality. It also requires that two contestants' perceptions of a particular savant's personality are uncorrelated.

²⁵Further simulation results imply the finite sample bias results from there being very few savants at the extreme performance levels (as evidenced in Figure 2). The estimator does a much better job when the majority of savants have extreme performance levels (either 0 or 1).

²⁶It was shown in Section 4.1.2 that if both \widehat{p}_B and \widehat{p}_N are greater than .5 at $x = 0$ or both are less than .5 at $x = 1$, the model will not do a good job of distinguishing between taste discrimination and different priors. To check if this is a problem we can first compare \widehat{p} for blacks (\widehat{p}_B) versus non-blacks (\widehat{p}_N) for non-black contestants in Table 6, Column 1. At $x = 0$, $\widehat{p}_N = .75(.508) + .25(0) = .381$, while $\widehat{p}_B = .75(.508 - .08) + .25(0) = .321$. At $x = 1$, $\widehat{p}_N = .75(.508) + .25(1) = .631$, while $\widehat{p}_B = .75(.508 - .08) + .25(1) = .571$. These results imply there should not be a problem distinguishing between the two types of discrimination. Similar exercises can be done for black contestants.

²⁷In analysis not shown, whether or not a prediction is correct in Round 3 was stratified by contestant/savant race combinations. No evidence was found that contestants do a better job of predicting answers for savants of their own race

References

- Altonji, Joseph G., and Rebecca M. Blank. 1999. Race and Gender in the Labor Market. In *Handbook of Labor Economics*, ed. David Card and Orley Ashenfelter. Amsterdam: Elsevier Science.
- Altonji, Joseph G., and Charles R. Pierret. 2001. Employer Learning and Statistical Discrimination. *Quarterly Journal of Economics* 116: 293-312.
- Antonovics, Kate, Peter Arcidiacono, and Randall Walsh. 2005. Games and Discrimination: Lessons from 'The Weakest Link'. *Journal of Human Resources* 40: 918-47.
- Anwar, Shamena and Hanming Fang. 2006. An Alternative Test of Racial Profiling in Motor Vehicle Searches: Theory and Evidence. *American Economic Review* 96: 127-151.
- Arrow, Kenneth. 1973. The Theory of Discrimination. In *Discrimination in Labor Markets*, ed. Orley Ashenfelter and Albert Rees. Princeton: Princeton University Press.
- Becker, Gary S. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Beetsma, Roel, and Peter C. Schotman. 2001. Measuring Risk Attitudes in a Natural Experiment: Data from the Television Game Show *Lingo*. *Economic Journal* 111: 821-848.
- Berk, Jonathan B., Eric Hughson, and Kirk Vandezande. 1996. The Price Is Right, but Are the Bids? An Investigation of Rational Decision Theory. *American Economic Review* 86: 954-970.
- Bertrand, Marianne and Sendhil Mullainathan. 2004. Are Emily and Brendan More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review* 94: 991-1013.
- Coate, Stephen and Loury, Glenn. 1993. Will Affirmative-Action Policies Eliminate Negative Stereotypes? *American Economic Review* 83: 1220-1240.

- Cornell, Bradford, and Ivo Welch. 1996. Culture, Information, and Screening Discrimination. *Journal of Political Economy* 104: 542-571.
- Fershtman, Chaim and Uri Gneezy. 2001. Discrimination in a Segmented Society. *Quarterly Journal of Economics* 116: 351-377.
- Goldin, Claudia and Cecilia Rouse. 2000. Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians. *American Economic Review* 90: 715-741.
- Levitt, Steven D. 2004. Testing Theories of Discrimination: Evidence from *Weakest Link*. *Journal of Law and Economics* 47: 431-452.
- List, John A. 2004. The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field. *Quarterly Journal of Economics* 119: 49-89.
- List, John A. 2006. Friend or Foe? A Natural Experiment of the Prisoner's Dilemma. *Review of Economics and Statistics* 88: 463-471.
- Metrick, Andrew. 1995. A Natural Experiment in *Jeopardy!* *American Economic Review* 85: 240-253.
- Neumark, David. 1996. Sex Discrimination in Restaurant Hiring: An Audit Study. *Quarterly Journal of Economics* 111: 915-41.
- Oettinger, Gerald S. 1996. Statistical Discrimination and the Early Career Evolution of the Black-White Wage Gap. *Journal of Labor Economics* 14: 52-78.
- Phelps, Edmund. 1972. A Statistical Theory of Racism and Sexism. *American Economic Review* 62: 659-61.
- Sackett, Paul R., Cathy L. Z. Dubois. 1991. Rater-Ratee Race Effects on Performance Evaluation: Challenging Meta-Analytic Conclusions. *Journal of Applied Psychology* 76: 873-877.
- Stoll, Michael A., Steven Raphael, and Harry J. Holzer. 2004. Black Job Applicants and the Hiring Officer's Race. *Industrial and Labor Relations Review* 57: 267-287.

Table 1: Examples of Questions Listed by Category

Category	Examples of Questions
general knowledge	What state was founded by William Penn? What letter is silent in the word autumn? In what U.S. city would you find the gateway arch? Who was the U.S. president during the Iran Contra affair? If your friend is 5'4" and you are ten inches taller, how tall are you?
entertainment	What is James Bond's numeric code name? The show "Nip/Tuck" is about people in what profession? Rapper Marshall Mathers is better known as whom? What character did Michael Richards play on "Seinfeld"? What three part series of mob movies was directed by Francis Ford Coppola?
Sports	What sport do the LA Clippers play? In baseball what is a grand slam? Major sports teams the Astros, the Texans, and the Rockets all call what city home? What college sport is played every winter in the Sugar Bowl? Tony Hawke is a legend in what sport?
slang/common sayings	In slang, what happened if you blew a fuse? Complete the saying: looks can be what? According to the saying, every cloud has a what? If your voice is hoarse, you are said to have what animal in your throat? In slang what does it mean if someone's got your back?
childrens' interests	In the popular childrens books, Christopher Robbins is friends with Winnie the who? What food did Hansel and Gretel leave on the ground as a path to lead them home? In the nursery rhyme what is the London Bridge doing? Grranimals is a popular line of kids what? On Sesame Street, what color is Grover's Fur?
miscellaneous	What color is pesto sauce? What store's motto is "always low prices, always"? What type of alcohol is mixed with orange juice to make a screwdriver? If you're addicted to Manolo Blahniks what accessory are you addicted to? In order to play Yahtzee you need five what? What tasty herb is believed to ward off vampires? According to the Bible, what should you do to someone who trespasses? What University is nicknamed the Fighting Irish?

Table 2: Demographic Characteristics of Contestants and Savants

		Contestants		Savants	
		Total %	% College-Educated	Total %	% College Educated
race	white	62.7%	22.4%	59.4%	23.1%
	black	22.4%	26.9%	24.1%	24.1%
	Hispanic	6.5%	23.1%	7.9%	22.5%
	Asian	6.4%	34.2%	5.7%	23.5%
	other/unknown	2.0%	16.7%	2.9%	26.9%
gender	male	48.2%	27.4%	44.9%	21.1%
	female	51.8%	21.0%	55.1%	25.3%
sample size		598	144	897	210

Table 3: The Effect of Savant Characteristics on Contestant Predictions

	Rounds 1 & 2				Round 3			
	Discrete Choice Model--Marginal Effects Shown				OLS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	all contestants	all contestants	non-black cont.	black cont.	all contestants	all contestants	non-black cont.	black cont.
<i>previous prop. correct</i>	***.043 (.015)	***.091 (.034)	** .035 (.016)	** .066 (.033)	***.270 (.061)	***.286 (.087)	***.304 (.069)	.167 (.127)
<i>previous category-specific prop. correct</i>		.008 (.024)				-.026 (.043)		
<i>college</i>	***.059 (.013)	.065 (.021)	***.056 (.015)	** .064 (.028)	.042 (.026)	.036 (.034)	*.051 (.030)	.023 (.057)
<i>black</i>	***-.044 (.013)	*-.036 (.021)	***-.037 (.015)	**-.061 (.026)	**-.053 (.025)	*-.053 (.031)	**-.065 (.030)	-.012 (.052)
<i>female</i>	.004 (.010)	.014 (.016)	.005 (.011)	.002 (.019)	-.038 (.022)	*-.047 (.027)	**-.050 (.025)	-.007 (.048)
<i>constant</i>					***.412 (.031)	.413 (.038)	***.408 (.034)	***.422 (.072)
R^2					0.0141	0.0138	0.0186	0.0043
observations	8952	3326	6975	1977	1779	1234	1383	396

Note: Standard errors clustered at the contestant level are in parentheses.

Columns 1-4 show the marginal effects from conditional logit regressions.

*** denotes a parameter is significant at the 1% level.

** denotes a parameter is significant at the 5% level.

* denotes a parameter is significant at the 10% level.

Table 4: The Effect of Savant Characteristics on Contestant Predictions in Rounds 1 and 2, Stratified by Contestant Race and Question Category

		general knowledge questions	miscellaneous questions	entertainment questions	slang/common saying questions
non-black contestants	<i>black</i>	.002 (.022)	*-.060 (.035)	***-.094 (.030)	*-.085 (.046)
	<i>female</i>	.017 (.017)	.015 (.029)	-.004 (.027)	.021 (.035)
	Observations	2715	1127	1503	700
black contestants	<i>black</i>	-.061 (.048)	-.046 (.056)	-.072 (.053)	-.070 (.066)
	<i>female</i>	.069 (.049)	-.042 (.048)	-.025 (.041)	.034 (.061)
	Observations	698	317	487	176

Note: Standard errors clustered at the contestant level are in parentheses. All regressions also include controls for past proportion of questions correct, as well as education level. The results show the marginal effects from conditional logit regressions.

*** denotes a parameter is significant at the 1% level.

** denotes a parameter is significant at the 5% level.

* denotes a parameter is significant at the 10% level.

Table 5: Monte Carlo Simulation Results

	specified parameter	598 observations		15000 observations	
		average estimated parameter	st.deviation of estimated parameter	average estimated parameter	st. deviation of estimated parameter
ω_1 (<i>constant</i>)	.5	.503	.122	.524	.105
ω_2 (<i>COLLEGE_i</i>)	.2	.114	.225	.128	.228
ω_3 (<i>BLACK_j_BLACK_i</i>)	.2	.074	.217	.177	.099
ω_4 (<i>BLACK_j_NON-BLK_i</i>)	0	-.002	.149	.002	.042
ω_5 (<i>NON-BLK_j_BLACK_i</i>)	-.2	-.096	.229	-.176	.160
θ_1 (<i>BLACK_j_BLACK_i</i>)	.5	.622	.503	.524	.136
θ_2 (<i>NON-BLK_j_NON-BLK_i</i>)	.5	.504	.351	.498	.084
Γ	8	9.932	16.843	7.093	3.578

Note: σ was set at .75 for both simulations run. 1000 unique data sets were created for each of the two simulations.

Table 6: Structural Estimation Results for Model with Contestant Race Controls

	(1)	(2)
ω_1 (<i>constant</i>)	***.508 (.036)	***0.476 (.054)
ω_2 (<i>COLLEGE_i</i>)	.0643 (.056)	0.102 (.071)
ω_3 (<i>BLACK_i_BLACK_i</i>)	.104 (.108)	***0.264 (.087)
ω_4 (<i>BLACK_i_NON-BLK_i</i>)	.007 (.060)	0.038 (.097)
ω_5 (<i>NON-BLK_i_BLACK_i</i>)	-.080 (.070)	*-0.122 (.066)
θ_1 (<i>BLACK_j_BLACK_i</i>)	.467 (.399)	---
θ_2 (<i>NON-BLK_j_NON-BLK_i</i>)	-.010 (.228)	---
θ_3 (<i>RACE-MATCH_{ij}</i>)	---	.100 (.147)
Γ	***29.8029 (4.08)	***16.213 (3.08)
log-likelihood	-528.7895	-528.831
Observations	598	598

Note: Standard errors are in parentheses. $\sigma=.75$ in all three specifications.

*** denotes a parameter is significant at the 1% level.

** denotes a parameter is significant at the 5% level.

* denotes a parameter is significant at the 10% level.

Table 7: Actual Performance of Savants in Rounds 1-3, Overall and by Question Category

	Rounds 1 and 2					Round 3
	Discrete Choice Model--Marginal Effects Shown					OLS
	(1)	(2)	(3)	(4)	(5)	(6)
	All Questions	general knowledge questions	miscellaneous questions	entertainment questions	slang/common saying questions	All Questions
<i>constant</i>						***.505 (.014)
<i>college</i>	.036 (.015)	.031 (.030)	*.076 (.042)	.027 (.042)	**-.124 (.054)	-.022 (.021)
<i>black</i>	-.013 (.014)	-.025 (.027)	.020 (.039)	-.009 (.035)	.008 (.056)	.010 (.019)
<i>female</i>	.015 (.011)	.022 (.021)	.002 (.034)	-.003 (.030)	.006 (.045)	-.023 (.017)
observations	4434	1676	717	992	438	1779

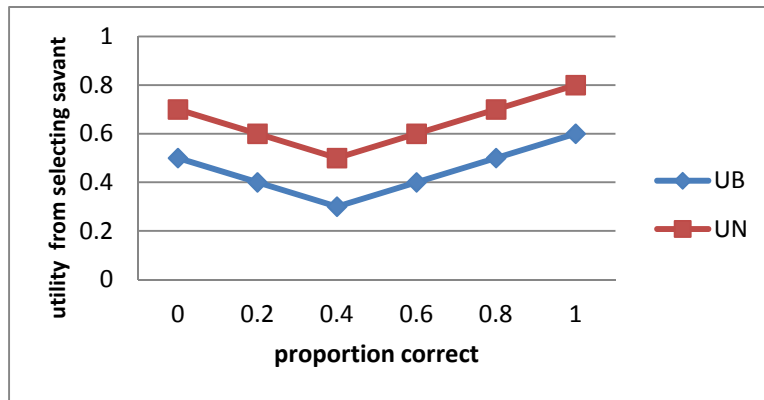
Note: The standard errors for Columns 1-5 are clustered at the game level. The standard errors for Column 6 are clustered at the contestant level. Columns 1-5 show the marginal effects from conditional logit regressions.

*** denotes a parameter is significant at the 1% level.

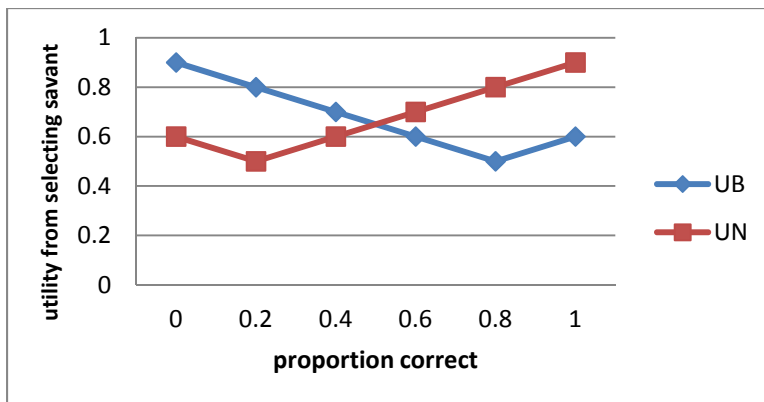
** denotes a parameter is significant at the 5% level.

* denotes a parameter is significant at the 10% level.

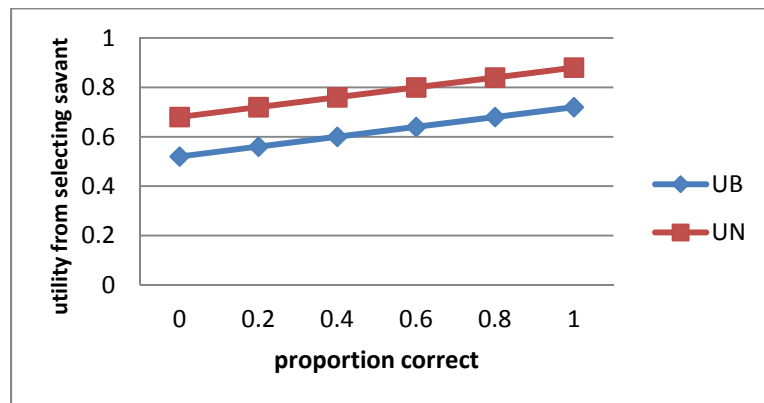
Figure 1: Examples of Taste Discrimination and Different Prior Perceptions



A: Taste Discrimination

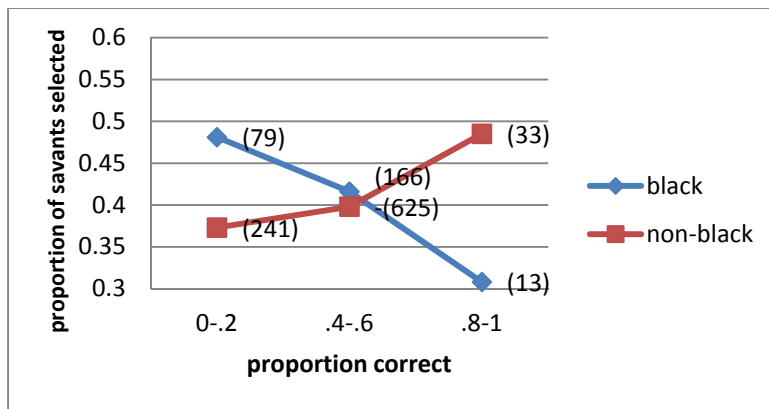


B: Different Prior Perceptions

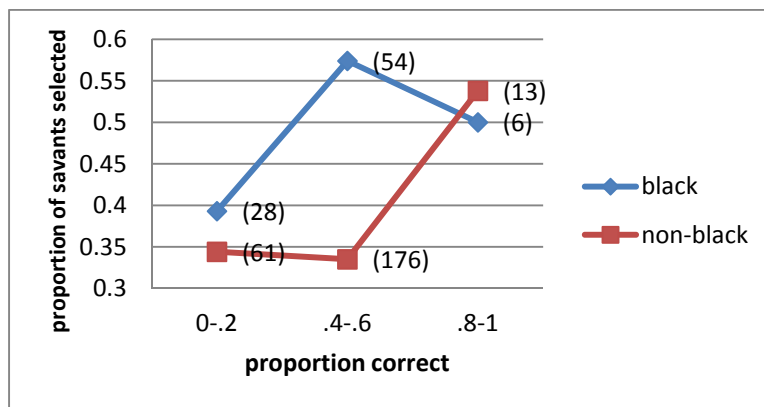


C: Taste Discrimination or Different Prior Perceptions

Figure 2: How the Relationship Between Savant Race and Selection Depends on Performance



A. Non-Black Contestants



B. Black Contestants

Note: Sample sizes are in parentheses.