Ambivalent sexism in a competitive environment. Evidence from a game show.

Natalia Starzykowska, Michał Krawczyk

Faculty of Economic Sciences, University of Warsaw

Corresponding author: nstarzykowska@wne.uw.edu.pl

Abstract: Sexism and gender discrimination remain to be likely contributors to females' lower wages and slower professional advancement. We use a novel dataset of decisions made by participants of the Ten to One TV show. During the game, contestants nominate next person to answer a question. Being nominated reduces one's probability of eventually winning the game, so a general tendency to nominate one gender more often than the other signifies taste-based discrimination against this gender. Having analyzed over 6000 decisions from 117 episodes we find taste-based discrimination favoring females, which, however, nearly disappears when we focus on highly-competent targets.

Keywords: ambivalent sexism, backlash effect, gender bias, gender discrimination

Introduction

Studies such as Weichselbaumer and Winter-Ebmer (2005) show that a substantial gender gap persists in the labor market. While a part of it is often ascribed to some form of discrimination against women (see Romei and Ruggieri,2014, for a review), such practices are often difficult to identify using existing field data. This is partly due to possible unobservable productivity differences. It is only rarely the case that these can be controlled for, as in the Goldin and Rouse's (2000) analysis of the impact of blind vs. non-blind orchestra auditions on the fraction of job offers made to women.

Even if a bias can be established, it is typically difficult to distinguish between information-based discrimination (believing, typically incorrectly, that members of group X are more competent than members of group Y) and taste-based discrimination (caring more about the well-being of members of group X than members of group Y or willing to interact with Xs rather than Ys). This is also true for field experiments on discrimination, the correspondence studies being the most prevalent sort. In their extensive review of this literature, Bertrand and Duflo (2017) note that "while field experiments have been overall successful at documenting that discrimination exists, they have (with a few exceptions) struggled with linking the patterns of discrimination to a specific theory". Yet this feat can be highly important, in particular when choosing between possible remedies, as discussed later. The identification difficulties may be overcome in laboratory experiments such as Albrecht et al. (2013), but such studies also have obvious limitations, including homogeneity of the sample, possible experimenter effect, the artificiality of tasks, and low stakes.

Some innovative methods seek to avoid these constraints while maintaining lab-like control. One of the new approaches involves the use of data from TV shows (List, 2006). Because the game is observable and follows well-defined rules, it is much easier to quantify participants' decisions and tell if they show gender bias. Importantly, unlike in the labor market, the researcher generally knows the information upon which the decision was based. In comparison to typical lab experiments, samples are much more diversified

(especially in terms of age, education level, employment status, and place of residence) and the stakes tend to be much higher.

The present study also follows this path, by exploring decisions made by participants of a game show *Ten to One* (*Jeden z dziesięciu* in Polish), a close relative of the British show *Fifteen to One*. Specifically, we look for a gender bias in the way the contestants nominate the next person to answer a question. Such a decision reduces the nominee's chances of surviving in the game and winning the prize (and posts on relevant players' forums show that they are keenly aware of that). Thus a systematic preference for nominating one gender rather than the other can be interpreted as a manifestation of taste-based discrimination. At the same time, contestants would ideally like to eliminate the strong opponents, but nominating them may be risky. By comparing the circumstances in which strong opponents tend to be nominated to the circumstances in which *males* tend to be nominated we can tell if most participants associate the male gender with higher performance.

Moreover, we can explore the interaction between the gender of a potential nominee and their performance record. This line of investigation is inspired by the literatures on ambivalent sexism (Glick & Fiske, 2001) and backlash against agentic women (Rudman & Glick, 1999). Both of them are rooted in the observation that males are expected to be competitive leaders and bread-earners, while females are expected to be sensitive care-takers. Those who embrace conventional gender roles tend to be rewarded, the others being punished. Such patterns have been observed e.g. in the education process (Dresden et al., 2018; Smith & Gayles, 2017) and in experiments (Otterbacher et al., (2017). In the context of our game, a hypothesis arises that among poorly performing contestants, females may be treated better (nominated less often) than males, so that they are not challenged and do not have to challenge back. However, once a female becomes a successful competitor, she might be perceived as behaving like a stereotypical male would have, resulting in backlash/hostile sexism. Thus, females' probability of being nominated will grow faster than that of males as their performance improves.

To the best of our knowledge, this game format has not been subject to academic work so far. One of its advantages is that such decisions are plenty, about 60 per episode in our sample, which increases the statistical power to identify inequality. Related to this, the participants are forced to make their nomination decisions very quickly. Some studies suggest that this could strengthen (implicit) discrimination (Bertrand et al., 2005; Price & Wolfers, 2010), giving a better understanding of what the participants are naturally inclined to do. Related to this, unlike in *The Weakest Link* the contestants do *not directly vote to eliminate* anyone from the show, so again they may be less inclined to hide their prejudice. Another attractive feature of the game is that it has been broadcasted in unchanged format for more than 20 years now. We can therefore compare the intensity of gender discrimination in the early 90s to the present situation. As Poland has undergone very substantial political, economic, and societal changes in the period, our study can offer some test whether (post-communist) transition affects patterns of gender discrimination.

We find participants' choices to be driven by taste-based discrimination against men and backlash towards agentic women. This is generally consistent with the notion of ambivalent sexism – females are "protected" but only as long as they do not take the initiative in their hands.

Neither of these tendencies seems to have changed much over the analyzed period. In the following section we review relevant literature, including studies using television game shows. Then we discuss the rules of *Ten to One* and provide our predictions concerning participants' decisions. Next, we examine our data set and the identification strategy. Finally, we report the empirical results and discuss some of the lessons they teach us.

Literature review

This project is most closely related to the strand of literature that explores discrimination in game shows. In his path-breaking paper, Levitt (2004) investigated the decisions of participants of *The Weakest Link*, in which it pays to vote to eliminate weak competitors at the beginning of the game and strong competitors

towards the end. Thus discrimination based on beliefs (but not based on taste) would predict phase-dependent biases. He found some information-based discrimination against Hispanics and taste-based discrimination against the elderly, but no gender discrimination. Likewise, Anwar (2012) observed that non-black participants of *Street Smarts* underestimated the ability of black "savants" to correctly answer the questions (in some categories), while gender played no role. Van den Assem et al. (2012) found no trace of (taste-based) gender discrimination in the show *Golden Balls*. By contrast, Atanasov and Dana (2013) reported same-sex favoritism in *the One Bid* game of *The Price is Right* show. Similarly, Antonovics et al. (2005), who used a different method to analyze *The Weakest Link*, found evidence of taste-based voting of women against men. Finally, Wall (2011) reported information-based discrimination against female contestants in the reality show *Survivor*.

Because we track the extent of gender effect over time, our study also relates to the literature studying how gender discrimination changes with an economic transition (of post-communist countries). Official communist propaganda emphasized gender equality and treated labor as a privilege and duty of all adult citizens. However, while women were often better educated, traditionally male occupations such as those in the mining industry were generally considered prestigious and paid better. Moreover, the highest managerial and partisan positions were hardly accessible for women.

Unlike in the West, the nineties were characterized by a decline in female labor force participation (Goraus & Tyrowicz, 2013). Findings concerning the gender wage gap and its unexplained component (often interpreted in terms of discrimination) are more mixed, depending on specific data coverage and methods used. In particular, Newell and Reilly (2001) concluded that the gap remained quite stable throughout the nineties, Brainerd (2000) reported diminishing gap and discrimination in former non-Soviet Eastern Bloc countries and the opposite in Ukraine and Russia. By contrast, Münich et al. (2005) found an increase in gender discrimination in the Czech Republic in the same period, and Adamchik and Bedi (2003) no change in the gender wage gap in Poland (in the years 1993-1997).

The game

The Polish version of *Fifteen to One,* locally known as *Jeden z dziesięciu* is a highly successful game show, on-air continuously since 1994. The rules are essentially analogous to the UK original, except that only ten contestants are involved. Standing behind randomly allocated, consecutively numbered lecterns forming a semicircle, they are asked several trivia quiz questions. Each incorrect answer (or no answer at all within three seconds) means losing one of the "lives"; a contestant with no lives left is eliminated from the show. The game proceeds in three rounds. The first one involves no strategic decisions – contestants are asked two questions each and have to answer at least one of them correctly to proceed to Round 2, where they retain two or three of their initially assigned lives.

Round 2 starts with Contestants 1, 2, 3, etc. each being asked one question. However, as soon the first correct answer is given, often by Contestant 1 (or C1 for short) already, the players start nominating one another to answer the next question. The contestant who gave the last correct answer is always the one to nominate. For example, if C2 nominates C4 and she fails to answer, C2 nominates again (possibly C4 again if she is still in the game). Round 2 ends when only three players are left standing.

These three start Round 3 with three lives each, no matter how many they had left after Round 2. The first questions in this round are answered by whoever is the first one to push their buzzer upon hearing the question. Players start nominating only once one of them has answered three questions correctly. In this round, unlike in Round 2, contestants sometimes nominate themselves,[1] because a correct answer yields 20 points after self-nomination (provided at least one other contestant is still alive) and only 10 points otherwise. Players also receive a small tie-breaking bonus equal to the number of chances they had after

[1]The game goes back to the buzzer mode after a self-nominated contestant answers incorrectly (unless only one contestant is left – in this case she will always get all the remaining questions, earning 10 points per correct answer). Nominations start again after the first correct answer in such a case.

Round 2 plus the number of chances they have after Round 3. The round ends after 40 questions or (more often) as soon as all contestants lose all their lives. In the former case, the high scorer among the survivors (not necessarily having the largest number of lives left) is the winner. If all contestants have been eliminated, the last survivor is the winner (even if his or her final score is lower than that of somebody eliminated previously). The winner typically earns 3,000 PLN (ca. 700 euro), which is just shy of mean gross salary, and a weekly stay in a luxurious hotel. The number of points the winner ends up with is also relevant, because only 10 top scorers in a series of 25 consecutive shows qualify to the Grand Finale, which follows the same rules, except that the value of the prizes is as high as 50,000 PLN.

Predictions

Close to 60 nomination decisions are made on average in an episode in our sample, with up to nine possible actions (corresponding to remaining players) in each of them. On top of that, there are even more resolutions of risk (corresponding to players trying to answer the questions). The game tree is thus gargantuan. Moreover, unknown (but partially revealed during the game) individual probabilities of answering correctly should be taken into account when deciding whom to nominate. Finally, one should note that players are often indifferent in the sense that one of two or more hitherto behaviorally indistinguishable players must be nominated. This means that in all probability numerous equilibria survive even subtle refinements suitable for such a dynamic game with incomplete information and that players are very unlikely to be able to tell how these equilibria look like. Attempting a complete game-theoretic analysis therefore appears futile. However, this does not preclude suggesting some features of seemingly reasonable moves and inferring something about players' tastes and beliefs.

Given the rules of the game, being nominated during Round 2 is clearly an unfortunate occurrence, as it results in a risk of losing a life.[2] Therefore, other things being equal, stronger taste-based discrimination against any specific group is expected to result in its member being nominated more often.

For the information-based discrimination, the picture is much more complex. Our identification strategy relies on the assumption that willingness to nominate a strong rather than weak contestant may depend on the current strategic position of the nominating contestant. Again, because the game is so complex, we do not make a priori assumptions about the direction of this link. Suppose for example that contestants currently in a strong strategic position (who have performed relatively well so far) tend to nominate strong opponents (with high performance record) and that they also tend to nominate men. This would suggest that they associate the male gender with high performance. The same conclusion would follow if they tended to nominate weak opponents and women, while the opposite would be concluded for the two remaining patterns of findings (contestants in a strong position nominating strong and female opponents or weak and male opponents).

Taste-based discrimination against women would thus mean that they are disproportionally often nominated throughout the game (or at least in Round 2). Information-based discrimination against women

---

[2]This is not always true in Round 3, as players earn points for correct answers (and more points if they self-nominate while an opponent is alive) and the game ends after 40 questions. Thus when a contestant is nominated (and answers correctly, but this is typically the case, as poor players rarely survive Rounds 1 and 2), she earns points and gains the right to self-nominate, which may be necessary to win the episode (when another player survives till the end) and to qualify for the Grand Finale. However, because typically only one player survives till the end and the expected payoff in the Grand Finale of a player who is just able to reach it is low anyway, being nominated is, again, unprofitable in a majority of cases also in Round 3.

would mean that the pattern of interaction between the current strategic situations of the nominating contestant and female gender of the potential nominee is the same as the pattern of interaction between the current strategic situations of the nominating contestant and poor past performance of the potential nominee (at least in Round 2).

The dataset and methods of analysis

Ideally, the past episodes would be obtained from the broadcaster. Unfortunately, this turned out to be impossible. We thus used various Internet sources. We have not verified if they were allowed to distribute the files. Downloading and streaming from unauthorized sources is not illegal in Poland (let alone when done for research purposes only) and in our view, it is obviously justified from an ethical viewpoint in these particular circumstances. One disadvantage of this is that only some episodes could be found. In particular, we wish we could find more episodes from the 1990s. On the other hand, it is very hard to think of a reason why episodes characterized by a specific pattern of nominations (say: women being nominated particularly often) were to be more likely to be available than other episodes. In this sense, we believe we have a random selection with the time of broadcasting being the only variable that significantly affected the probability that an episode is included in the sample. Table 1 shows summary statistics, broken by the period in which the program was aired.

The low fraction of women stands out. It is not completely clear what are its causes. The questions are generally not considered (strongly) male-oriented and the show has solid viewership among both genders. While statistics are not available, imbalance among participants generally reflects the imbalance among applicants. It could be that the time pressure and the competitiveness involved in the show discourage some women. Somewhat related activities such as gambling, playing video games, and partaking in sports competitions also tend to be more popular among males. Still, this feature of our data is regrettable, given our focus on gender effects as it reduces statistical power. One may also wonder whether the gender composition itself affects nomination patterns. Even if this is the case, it should be noted that females are

also strongly underrepresented in other economically and socially important environments, including top management, many well-paying professions at large (particularly the IT industry), and protestant clergy, to name some examples. In this sense, even if gender composition plays a major role, it may not render extrapolating our results to other interesting settings less meaningful. What is more, even if females who have participated in that game show differ significantly from those who have not, they are likely to not differ much from those females, who take part in other male-dominated environments. Both our female contestants and females working in fields dominated by males self-select themselves to be present in a given environment. Therefore we believe that to some extent those game contestants can be treated as an analogy present in a number of outside of the game settings.

The last three columns of Table 1 are based on contestants' short introductions which generally provide incomplete information. The trend observed in the last two variables can probably at least partly be explained by urbanization and dramatic expansion of the sector of higher education in the relevant period. Contestants virtually never report their age but judging by their looks and their reported occupational status, it varies considerably, with standard deviation likely exceeding 15 years. Most post-2000 episodes come from the year 2011, they were thus recorded nearly 15 years after those from the nineties.

Table 1: summary statistics

	#episodes	#nomination decisions	%women	%with university education	%living in a big city	%student
1990s	35	1936	12.29	18.86	40.29	12.00
2000s	82	4707	13.54	22.56	50.85	22.80
total	117	6643	13.16	21.45	46.50	19.57

Overall, 55.59% of questions were answered correctly. The median success rates were 56.23% for males and 51.14% for females, a significant difference as verified by a Mann-Whitney test (p=0.037).[3]

This difference makes it more cumbersome to identify unwarranted information-based discrimination. Indeed, of two contestants with an identical record, the man can typically be rationally expected to perform slightly better in the future because of the difference in the base rate. We thus proceed as follows. First, using each player's performance in the entire episode we calculate kernel density estimate of the player-specific probability of answering a question correctly. We do so separately for males and females, gender being the only variable that is significant in terms of performance observed in the game. Using thus obtained prior distributions—one for males and one for females—we apply the Bayes rule to calculate, given each player's performance showed so far, his or her probability of answering the next question correctly (*posterior pot.*). For example, a woman who has so far answered 9 questions correctly and 4 questions incorrectly may be expected to answer any question with a probability of 63.42%, compared to 64.51% for a man with a 9-4 record. As a proxy for the nominating player's strategic situation, we simply use the percent of his or her correct answers up to the moment of decision in question, a variable we call *past % correct own.*

We are interested in nomination decisions, we thus disregard the first round, early questions of Round 2 (before the first correct answer) and of Round 3 (before the third correct answer of the same contestant) as well as the final questions of Round 3 when only one contestant remains. For the remaining questions, we focus on the variable *nominated*, which takes the value of 1 if a *potentially nominated* contestant (often abbreviated to *pot.*) was actually nominated and 0 otherwise.

---

[3]Other variables like education level, graduated faculty, and size of the city of origin of a player were not significant predictors of performance.

To account for repeated decisions and possibly heterogeneity of decision-makers, we run a mixed logit model (Revelt & Train, 1998), which allows estimating individual-specific coefficients. Let $n = 1, \dots 1170$ denote a decision-maker facing at time $t = 1, \dots, T_n$ a choice between J_{nt} alternatives (potential nominees), J_{nt} ranging between 1 and 9. T_n may vary between decision-makers, as by design of the game they may face a different number of choice occasions. The utility that a decision-maker n derives from choosing potential nominee j on choice occasion t is assumed to be

$U_{njt} = \beta_n' * x_{njt} + \varepsilon_{njt}$,

where β_n is a vector of individual-specific coefficients and x_{njt} is a vector of observed attributes relating to decision-maker n and potential nominee j on choice occasion t, while ε_{njt} is a random term assumed to be independently and identically distributed, following the Type 1 extreme value distribution. The density of β is denoted as $f(\beta|\theta)$, where θ are the parameters of the distribution. The probability of contestant n choosing contestant i on choice occasion t is given by:

$$L_{nit}(\beta_n) = \frac{exp(\beta_n' * x_{nit})}{\sum_{j=1}^{J} \exp(\beta_n' * x_{njt})}$$

which is the conditional logit formula (McFadden, 1974). The probability of the observed sequence of choices is given by

$$S_n(\beta_n) = \prod_{t=1}^{T} L_{ni(n,t)t}(\beta_n)$$

Where $i(n,t)$ denotes the contestant chosen by contestant n on choice occasion t. The unconditional probability of the observed sequence of choices is the conditional probability integrated over the distribution of β:

$$P_n(\theta) = \int S_n(\beta) f(\beta|\theta) d\beta = \int \prod_{t=1}^{T} L_{ni(n,t)t}(\beta_n) f(\beta|\theta) d\beta$$

The unconditional probability is thus a weighted average of a product of logit formulas evaluated at different values of β, with the weights given by the density f. This specification is general because it allows

fitting models with both individual-specific and alternative-specific explanatory variables (Hole, 2007). The probability that a contestant will be nominated in a given choice situation is based on the estimation of the following model:

$$probability\ of\ choosing\ contestant_{nit}(\beta_n) = \frac{[\![exp[\beta_n' * (Var\_info_{nit} + Var\_taste_{nit} + Var\_other_{\ nit})]]\!]}{\left[\!\left[\sum_{j=1}^{J} exp[\beta_n' * (Var\_info_{nit} + Var\_taste_{nit} + Var\_other_{nit})]\right]\!\right]}$$

Where *Var_info* denotes the group of variables allowing to identify information-based discrimination, while *Var_taste* concerns taste-based discrimination and *Var_other* contains all other control variables.

All models have been estimated using Stata 16 native function *cmxtmixlogit* designed for mixed logit model incorporating panel data, which is essentially the case for our dataset. We use noconstant specification of the model, as by the design of the game, alternatives present in our choice sets have to be treated as non-labeled. That means that nominating Player 1 (i.e. choosing that alternative) is not the same thing in two different episodes. Unfortunately, native functions of Stata 16 do not support calculation of marginal effects for non-labeled alternatives. To overcome that difficulty, we first verified that marginal effects for alternatives are not significant, which is consistent with the notion that the 1-10 positions are assigned randomly in each episode and contestants are unaffected by these numbers per se when nominating. Then, marginal effects were estimated for each of the possible outcomes and at the final stage an average of obtained values was calculated. That allows aggregating information on marginal effects and standard errors for each of the variables. The advantage of this approach is that, unlike in the case of manual computations of marginal effects as suggested by Hole (2007), we can obtain standard errors without bootstrap, which was shown to be inefficient in terms of time required to perform that procedure. The procedure applied also limits the risk of human-made error, as the averaged results are directly resulting from outcomes reported by native Stata function *margins*. All continuous variables have been normalized to N(0, 1) in order to provide comparability of variables originating from different scales.

We cluster standard errors on the level of the episode, to deal with the problem of potential non-independence of observations within an episode. Because in Round 2 self-nominating is suicidal and utmost rare, we drop the cases when *potentially nominated=nominating* and simultaneously *round*=2. Because the strategic circumstances are different in Round 3 compared to Round 2, as described before, we also run separate regressions for these two rounds.

Identifying belief-based discrimination requires us to construct a measure of contestant's relative strategic situation (strong vs. weak). To make it orthogonal to contestant's characteristics, we use the following:

$$relative\ measure_i = past\ performance\ of\ player\ i\ - overall\ performance\ of\ player\ i$$

Past performance of player i is expressed as percentage of correct answers which decision-maker has given up to the particular moment *t* in the game. The *overall performance of player i* is the percentage of correct answers which particular decision-maker have given in the entire episode.

If we now observe the choices to be systematically different when this measure is high compared to when it is low, it will mean that the difference is due to current strategic situation of a contestant, not their time-invariant characteristics (such as being a high vs. low performer in general).

In order to establish if questions asked during the game are, by design, easier for one gender than the other, 112 participants of laboratory experiment have been asked, after facing each question, if in their opinion it was relatively easier for males, females or equally difficult for both genders. Each participant evaluated 84 questions and each question was evaluated by 16 or 32 participants. In total we evaluated a sample of 504 questions. In 71.72% of cases a question was described as being equally difficult for males and females. In 14.24% of cases participants stated that it was rather easier for males and in 4% of cases as definitely easier for males. We believe that such result allows to state that the design of analyzed game-show is not gender biased.

Results

To investigate robustness of our findings, we run a number of specifications. Model (1) only includes the basic variables identifying taste-based and belief-based discrimination and same-sex favoritism. Each of these effects is tested for stability over time (i.e. we test if there are changes between observation from 1994-1999 and those from 2000 and later). Additionally, that specification allows (in the model estimated on the entire dataset) for the possibility that the effect of the potential nominee being female differs between rounds and that male and female nominators react differently to the number of chances of the potential nominee. The probability that the potentially nominated player answers the question correctly, as it can be calculated based on his or her gender and past performance (also in interaction with the past performance of the nominating player) is included here. Model (2) additionally tests whether that probability has a different effect for female nominees. It also verifies if a strategic situation of the decision makers has the same effect depending on the gender of the nominee.

Model (3) verifies whether males and females differ in terms of probability of self-nominating (only on round 3 and the entire dataset, as the extremely rare, suicidal cases of self-nomination in round 2 are discarded). Also, it verifies whether the number of chances held by the potential nominee has an impact on the probability of being chosen (and if this effect is stronger in same-sex dyads). Model (4) controls if the willingness to nominate a player holding his or her last chance differs between male and female contestants. It also allows for the willingness (potentially different for men and women) to nominate a player back immediately after having been nominated.

Model (5) studies the effect of proximity between the decision-maker and a potential target in terms of the number of chances left. Additionally, that specification controls whether the nominations are affected by the information concerning the level of education of male and female contestants and their academic major. It also controls for the experience of being nominated by a given player in the past (except for the case of direct nomination), also in interaction with gender. Model (6) allows for taste-based discrimination

based on information concerning the size of the city of origin of a potentially nominated player and whether the level of education affects the probability of being nominated differently for males and females.

To facilitate the presentation of the estimates for our numerous explanatory variables, we divide them by the type of discrimination. Table 2 includes the variables that allow investigating taste-based discrimination and Table 3 is dedicated to information-based one. In Appendix B we show results for remaining (control) variables, most of which come out insignificant. Entries in all the tables are based on the same specifications, e.g. model (1) includes variables from *potential is female* through *same-sex in* 90's visible in Table 2, *relative measure own* posteriori pot.* through *relative measure own* fem. pot.* from Table 3 etc. All the tables show *marginal effects* rather than the mean coefficient for a given variable, followed by the standard error and the *p* value. For example, the value -0.065 for *potential is female* in Table 2 means that (if we ignore contestant's other characteristics) the probability of being nominated is by 6.5 percentage points lower for women than for men.

Table 2: Marginal effects for taste-based discrimination hypothesis.

Variable \ Specification*	(1)	(2)	(3)	(4)	(5)	(6)
ROUND 2						
potential is female	-0.065 (0.013) **0.000**	-0.132 (0.030) **0.000**	-0.141 (0.033) **0.000**	-0.137 (0.033) **0.000**	-0.124 (0.039) **0.001**	-0.124 (0.038) **0.001**
potential is female in 90's	-0.021 (0.027) **0.431**	-0.016 (0.030) **0.602**	-0.012 (0.030) **0.692**	-0.011 (0.030) **0.000**	0.004 (0.032) **0.888**	0.000 (0.030) **0.990**
same sex	-0.062 (0.014) **0.000**	-0.063 (0.014) **0.000**	-0.061 (0.014) **0.000**	-0.062 (0.014) **0.000**	-0.050 (0.015) **0.001**	-0.048 (0.015) **0.002**
same sex in 90's	-0.026 (0.028) **0.356**	-0.028 (0.028) **0.315**	-0.030 (0.029) **0.299**	-0.026 (0.028) **0.352**	-0.012 (0.030) **0.697**	-0.014 (0.029) **0.622**
N	34888	34888	34888	34888	34888	34888

ROUND 3						
potential is female	-0.036 (0.010) **0.000**	-0.157 (0.054) **0.004**	-0.055 (0.042) **0.191**	-0.047 (0.039) **0.238**	-0.042 (0.037) **0.254**	-0.030 (0.037) **0.410**
potential is female in 90's	0.008 (0.024) **0.755**	-0.008 (0.025) **0.734**	0.011 (0.021) **0.603**	0.012 (0.021) **0.238**	0.016 (0.019) **0.411**	0.029 (0.017) **0.092**
same sex	-0.081 (0.009) **0.000**	-0.083 (0.010) **0.000**	-0.014 (0.009) **0.128**	-0.013 (0.008) **0.131**	-0.009 (0.007) **0.204**	-0.010 (0.005) **0.068**
same sex in 90's	0.025 (0.029) **0.394**	0.016 (0.029) **0.581**	0.028 (0.024) **0.257**	0.026 (0.023) **0.245**	0.026 (0.021) **0.208**	0.026 (0.021) **0.209**
N	6779	6779	6779	6779	6779	6779
ENTIRE DATASET						
potential is female	-0.195 (0.027) **0.000**	-0.261 (0.032) **0.000**	-0.117 (0.031) **0.000**	-0.110 (0.031) **0.000**	-0.118 (0.032) **0.000**	-0.121 (0.032) **0.000**
potential is female in 90's	0.023 (0.064) **0.717**	0.048 (0.065) **0.463**	0.087 (0.056) **0.120**	0.083 (0.055) **0.000**	0.091 (0.060) **0.130**	0.083 (0.060) **0.164**
same sex	-0.083 (0.001) **0.000**	-0.084 (0.010) **0.000**	-0.012 (0.008) **0.144**	-0.013 (0.008) **0.123**	0.024 (0.022) **0.285**	-0.013 (0.010) **0.187**
same sex in 90's	-0.004 (0.026) **0.880**	-0.007 (0.026) **0.787**	0.020 (0.021) **0.340**	0.021 (0.022) **0.324**	-0.006 (0.004) **0.094**	0.022 (0.023) **0.333**
potential is female * round	0.055 (0.010) **0.000**	0.051 (0.011) **0.000**	0.023 (0.012) **0.036**	0.020 (0.011) **0.063**	0.026 (0.012) **0.033**	0.025 (0.012) **0.041**
N	41667	41667	41667	41667	41667	41667

* see Tables 3, B1, and B2 for the differences between specifications. For each independent variable marginal effects, (standard errors) and **confidence levels** were given.

Both in Rounds 2 and initial specifications for Round 3 we find evidence of taste-based discrimination against male contestants, as the estimate on *potential is female* is significantly below zero. For Round 2, the effect is robust between all tested specifications and remains negative within an entire confidence interval. The same is true for the entire sample and these findings are very robust. However, the attitude

toward females changes as the game progresses, as indicated by the effect of the interaction of *potential is female* and *round* which is positive and significant in all models. From the estimations based on the entire dataset, we can therefore learn that taste-based gender discrimination in Round 3 is still present, although it is weaker than in Round 2. In estimations run solely on Round 3 that effect could not have been accounted for properly due to the low number of females at this stage of the game. From estimation on entire dataset we observe that the effect of nominee's gender is strong enough to hold. As the interaction with the 90's is in general not significant, there is no reason to reject the null hypothesis of no change in the discriminatory patterns during the period under scrutiny.

Same-sex favoritism is statistically significant for all of the specifications for Round 2 but it is less robust in Round 3 (again, we have too few observations for women there) or the entire dataset. These estimates are lower than for *potential is female,* implying that indeed both genders discriminate against females. There is no evidence for that tendency to change over time as the effect for *the same sex in the '90s* is not significant in almost any of the specifications.

Table 3: Marginal effects for information-based discrimination hypothesis.

Variable \ Specification	(1)	(2)	(3)	(4)	(5)	(6)
ROUND 2						
relative measure own* posteriori pot.	-0.120 (0.009) **0.000**	-0.119 (0.009) **0.000**	-0.129 (0.010) **0.000**	-0.130 (0.010) **0.000**	-0.142 (0.011) **0.000**	-0.142 (0.011) **0.000**
posteriori pot.	0.015 (0.002) **0.000**	0.014 (0.002) **0.000**	0.004 (0.002) **0.060**	0.003 (0.002) **0.161**	0.004 (0.003) **0.133**	0.004 (0.003) **0.169**
posteriori pot. * fem. pot.		0.021 (0.008) **0.014**	0.025 (0.010) **0.013**	0.023 (0.010) **0.019**	0.020 (0.012) **0.117**	0.020 (0.012) **0.109**
relative measure own* fem. pot.		0.003 (0.003) **0.246**	0.004 (0.003) **0.173**	0.004 (0.003) **0.184**	0.003 (0.003) **0.274**	0.004 (0.003) **0.212**
chances_pot=2			0.018 (0.004) **0.000**	0.015 (0.004) **0.000**	0.019 (0.005) **0.000**	0.019 (0.005) **0.000**

chances_pot=3			0.049 (0.006) **0.000**	0.046 (0.006) **0.000**	0.056 (0.008) **0.000**	0.056 (0.008) **0.000**
chances_pot=2 and same sex			-0.007 (0.009) **0.489**	-0.005 (0.009) **0.620**	-0.002 (0.011) **0.836**	-0.002 (0.011) **0.826**
chances_pot=3 and same sex			-0.012 (0.012) **0.318**	-0.011 (0.012) **0.344**	-0.009 (0.014) **0.539**	-0.009 (0.015) **0.532**
fem. nominates * potential has last chance				-0.021 (0.014) **0.129**	-0.016 (0.015) **0.288**	-0.016 (0.015) **0.292**
abs. chances diff.					-0.005 (0.003) **0.118**	-0.005 (0.003) **0.114**
N	34888	34888	34888	34888	34888	34888
ROUND 3						
relative measure own* posteriori pot.	-0.248 (0.062) **0.000**	-0.254 (0.061) **0.000**	-0.168 (0.043) **0.000**	-0.171 (0.042) **0.000**	-0.158 (0.039) **0.000**	-0.160 (0.039) **0.000**
posteriori pot.	0.013 (0.003) **0.000**	0.009 (0.004) **0.023**	0.010 (0.003) **0.002**	0.008 (0.003) **0.019**	0.005 (0.003) **0.155**	0.005 (0.003) **0.113**
posteriori pot. * fem. pot.		0.036 (0.016) **0.022**	0.016 (0.013) **0.210**	0.013 (0.013) **0.313**	0.017 (0.013) **0.214**	0.004 (0.014) **0.784**
relative measure own* fem. pot.		0.003 (0.008) **0.686**	0.003 (0.006) **0.652**	0.002 (0.006) **0.801**	0.002 (0.006) **0.678**	0.002 (0.006) **0.684**
chances_pot=2			0.014 (0.004) **0.002**	0.010 (0.004) **0.021**	0.007 (0.004) **0.096**	0.007 (0.004) **0.073**
chances_pot=3			0.027 (0.005) **0.000**	0.021 (0.005) **0.000**	0.017 (0.005) **0.000**	0.018 (0.005) **0.000**
chance_pot=2 and same sex			-0.004 (0.010) **0.684**	0.001 (0.009) **0.917**	-0.002 (0.009) **0.809**	0.000 (0.009) **0.964**
chances_pot=3 and same sex			-0.008 (0.011) **0.477**	-0.005 (0.011) **0.644**	-0.009 (0.011) **0.446**	-0.005 (0.011) **0.686**
self nominated			-0.079 (0.007) **0.000**	-0.058 (0.009) **0.000**	-0.048 (0.012) **0.000**	-0.048 (0.012) **0.000**
female nominates herself			-0.040	-0.049	-0.056	-0.055

			(0.032) **0.214**	(0.033) **0.136**	(0.032) **0.080**	(0.027) **0.039**
fem. nominates * potential has last chance				-0.003 (0.009) **0.764**	-0.002 (0.010) **0.876**	-0.001 (0.009) **0.936**
abs. chances diff.					-0.007 (0.003) **0.017**	-0.006 (0.003) **0.023**
N	6779	6779	6779	6779	6779	6779
ENTIRE DATASET						
relative measure own* posteriori pot.	-0.115 (0.009) **0.000**	-0.114 (0.009) **0.000**	-0.111 (0.009) **0.000**	-0.111 (0.009) **0.000**	-0.112 (0.009) **0.000**	-0.112 (0.009) **0.000**
posteriori pot.	0.014 (0.001) **0.000**	0.013 (0.001) **0.000**	0.009 (0.002) **0.000**	0.008 (0.002) **0.000**	0.007 (0.002) **0.000**	0.007 (0.002) **0.000**
posteriori pot. * fem. pot.		0.023 (0.006) **0.000**	0.014 (0.007) **0.051**	0.013 (0.007) **0.052**	0.012 (0.008) **0.115**	0.011 (0.008) **0.151**
relative measure own* fem. pot.		0.003 (0.002) **0.201**	0.004 (0.002) **0.098**	0.003 (0.002) **0.104**	0.003 (0.002) **0.135**	0.003 (0.002) **0.121**
chances_pot=2			0.016 (0.003) **0.000**	0.014 (0.004) **0.000**	0.011 (0.004) **0.002**	0.011 (0.004) **0.002**
chances_pot=3			0.042 (0.005) **0.000**	0.040 (0.005) **0.000**	0.037 (0.005) **0.000**	0.037 (0.006) **0.000**
chances_pot=2 and same sex			-0.006 (0.007) **0.370**	-0.005 (0.006) **0.453**	-0.004 (0.007) **0.531**	-0.004 (0.007) **0.607**
chances_pot=3 and same sex			-0.011 (0.008) **0.181**	-0.011 (0.008) **0.176**	-0.008 (0.009) **0.365**	-0.007 (0.009) **0.418**
self nominated			-0.156 (0.012) **0.000**	-0.147 (0.013) **0.000**	-0.141 (-0.063) **0.000**	-0.141 (0.013) **0.000**
female nominates herself			-0.036 (0.031) **0.242**	-0.047 (0.033) **0.153**	-0.056 (0.037) **0.126**	-0.054 (0.037) **0.138**
fem. nominates * potential has last chance				-0.011 (0.009) **0.247**	-0.010 (0.010) **0.311**	-0.009 (0.010) **0.328**
abs. chances diff.					-0.011 (0.002)	-0.011 (0.002)

					0.000	**0.000**
N	41667	41667	41667	41667	41667	41667

For each independent variable marginal effects, (standard errors) and **confidence levels** were given.

Turning now to variables allowing identification of information-based discrimination, from Table 3 we conclude that seemingly strong contestants are slightly more likely to be nominated, as the marginal effect for *posteriori pot.* is positive. However, in both rounds that effect becomes insignificant in some models if we analyze data derived from particular rounds separately. Only in estimations on the entire dataset that effect remains strongly significant for all model specifications. Interestingly, in Round 2, we see evidence that strong females are treated differently than males. The effect for the interaction of variable *posteriori pot.* and gender of the potential player is significant or nearly significant, depending on the specification. High-performing females are nominated *relatively* often, a result is consistent with previously mentioned literature which concerned the backlash against agentic women.

A question thus arises if this interaction leads, on balance, to discrimination against high-performing females in absolute terms. This is not the case. For example, the marginal effect for a male contestant to nominate a male opponent whose *posteriori_pot* is better than that of 90% of all players is 13.2%, considerably higher than 6.6% for an equally well performing woman.

Concerning the effect of the relative strategic situation, contestant are more likely to nominate a weak opponent when their own situation is good (*relative measure own * posteriori pot.*). This effect is highly significant in all the specifications, thus allowing us to compare this interaction to that with the gender of the nominee (*relative measure own * female pot.*). Here, the estimates are not statistically different from zero, a finding speaking against belief-based discrimination. In other words, obtained results suggest that females competence is judged by their performance, not automatically assumed to be inferior.

We sought to compare our findings with contestants' self-reported strategies. We surveyed over 100 former participants of the show using an on-line questionnaire. Sadly, out of 59 who claimed that they have made it to Round 2, only 27 respondents decided to reveal their strategy. These were predominantly

fairly simple, the most common involving eliminating weak players as soon as possible. This seems quite different from our findings, possible reasons including the sample being small and self-selected; player not willing to reveal their strategies; players following gut feelings rather than a well-articulated strategy. Then again, in isolated cases the mechanisms we observe in the data were also explicitly formulated as "I was aiming to nominate relatively strong contestants, who would be a threat in Round 3".

Conclusions

We believe that our data set offers an excellent opportunity to identify the main types of gender discrimination in a real-life setting involving a fierce, high-stakes contest. To be sure, there are some inherent limitations. In particular, arguably the game creates an artificial environment, rather different from daily professional life. Yet, to perform well in the show one needs broad knowledge and the ability to focus and perform well under stress and time pressure while competing with others. Such qualities are also valued in the labor market, especially in the highly competitive corporate environment. It may thus be hoped that patterns identified here will generalize to other settings. Moreover, while our focal nomination decision may not have an immediate, direct equivalent in the world of business or politics or the academia, it bears some resemblance e.g. to assigning unwanted, unrewarding tasks (such as organizing a conference), a process in which some groups (e.g. women) may be discriminated against.

The selection of participants is also an issue. Those applying to take part do not represent a random sample of the society (and then they have to pass a screening involving some trivia questions similar to those appearing in the show and wait for their turn). Of particular concern from our viewpoint is that the vast majority of the contestants are male. Obviously, this reduces statistical power to identify any systematic gender effects. Luckily, it turns out that at least some of them are strong enough to show up anyway (partly thanks to the large number of observations we have). One may also wonder whether gender composition per se affects discriminatory patterns. We have run some regressions accounting for the number of women on the episode and it did not seem to make a difference. However, we only have limited statistical

power to test for such effects as there is little variation in the starting composition, for example as many as 51.3% of episodes in our sample started with but one female and 30.8% of episodes started with two. The negative result may also be associated with the fact that players take into account the fraction of women in the whole population of participants, not in the particular episode in which they happen to partake. For example, it could be that females tend to be "spared" because there are so few of them on *Ten to One* anyway. Then again, such a gender imbalance may not severely restrict external validity, because other important environments are also strongly dominated by men, as mentioned before. Nevertheless, in future research, it would be desirable to also have more gender-balanced samples (as well as any variation at all on the racial dimension for example).

In general, we observe taste-based discrimination against males. One way of understanding this pattern is in terms of "ambivalent sexism" (Glick & Fiske, 1996) under which women are perceived as weaker (also mentally) than men and thus deserving protection (Glick & Fiske, 2001; Reilly et al., 2017). The tendency to nominate males more often could be understood as a manifestation of a chivalry effect: perceiving females as weak and vulnerable, males express protective attitudes towards typical females. That is basically what literature calls benevolent sexism (Glick & Fiske, 2001; Reilly et al., 2017). According to Naurin et al. (2019), it is more prevalent in societies displaying a relatively low level of gender equality, such as Poland is one of them XXX bib ref, nie przypis pls[4].

However, as soon as females prove themselves not to be inferior, the protective behavior vanishes and their manifestation of competencies triggers stronger repercussions than it does for males. In our context, this indeed translates into a preference for letting rather underperforming females remain in the show.

[4] Accordingly to Gender Equality Index 2020 Poland with 55.8 out of 100 possible points ranks 24th in the European Union on the Gender Equality Index. For more information see : https://eige.europa.eu/publications/gender-equality-index-2020-poland

That observation is in accordance with the concept of hostile sexism (Glick & Fiske, 1996, 2001), whereby competent females are more likely to experience backlash or, in our case, nominations which may exclude them from the show. Both of reported results are robust across specifications and over time.

References

Adamchik, V. A., & Bedi, A. S. (2003). Gender pay differentials during the transition in Poland. *Economics of Transition*, *11*(4), 697–726.

Albrecht, K., von Essen, E., Fliessbach, K., & Falk, A. (2013). The influence of status on satisfaction with relative rewards. *Frontiers in Psychology*, *4*, 804.

Antonovics, K., Arcidiacono, P., & Walsh, R. (2005). Games and discrimination lessons from The Weakest Link. *Journal of Human Resources*, *40*(4), 918–947.

Anwar, S. (2012). Testing for discrimination: Evidence from the game show Street Smarts. *Journal of Economic Behavior & Organization*, *81*(1), 268–285.

Atanasov, P., & Dana, J. (2013). The price is sexist: taste based discrimination by contestants on the price is right. Citeseer.

Bertrand, M., Chugh, D., & Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, *95*(2), 94–98.

Bertrand, M., & Duflo, E. (2017). Field experiments on discrimination. *Handbook of Economic Field Experiments*, *1*, 309–393.

Brainerd, E. (2000). Women in transition: Changes in gender wage differentials in Eastern Europe and the former Soviet Union. *ILR Review*, *54*(1), 138–162.

Dresden, B. E., Dresden, A. Y., Ridge, R. D., & Yamawaki, N. (2018). No girls allowed: women in male-dominated majors experience increased gender harassment and bias. *Psychological Reports*, *121*(3), 459–474.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*(3), 491.

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, *56*(2), 109.

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of" blind" auditions on female musicians. *American Economic Review*, *90*(4), 715–741.

Goraus, K., & Tyrowicz, J. (2013). The Goodwill Effect? Female Access to the Labor Market Over Transition: A Multicountry Analysis. *Faculty of Economic Sciences Working Paper*, *19*.

Hole, A. R. (2007). Fitting mixed logit models by using maximum simulated likelihood. *The Stata Journal*, *7*(3), 388–401.

Levitt, S. D. (2004). Testing theories of discrimination: evidence from Weakest Link. *The Journal of Law and Economics*, *47*(2), 431–452.

List, J. A. (2006). Friend or foe? A natural experiment of the prisoner's dilemma. *The Review of Economics and Statistics*, *88*(3), 463–471.

McFadden, D. (1974). Frontiers in Econometrics, chapter Conditional logit analysis of qualitative choice behavior. Academic Press New York, NY, USA.

Münich, D., Svejnar, J., & Terrell, K. (2005). Is women's human capital valued more by markets than by planners? *Journal of Comparative Economics*, *33*(2), 278–299.

Naurin, N. D., Naurin, E., & Alexander, A. (2019). Gender stereotyping and chivalry in international negotiations. A survey experiment in the Council of the European Union. *International Organization*, *73*(2), 469–488.

Newell, A., & Reilly, B. (2001). The gender pay gap in the transition from communism: some empirical evidence. *Economic Systems*, *25*(4), 287–304.

Otterbacher, J., Bates, J., & Clough, P. (2017). Competent men and warm women: Gender stereotypes and backlash in image search results. *Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems*, 6620–6631.

Price, J., & Wolfers, J. (2010). Racial discrimination among NBA referees. *The Quarterly Journal of Economics*, *125*(4), 1859–1887.

Reilly, E. D., Rackley, K. R., & Awad, G. H. (2017). Perceptions of male and female STEM aptitude: The moderating effect of benevolent and hostile sexism. *Journal of Career Development*, *44*(2), 159–173.

Revelt, D., & Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics*, *80*(4), 647–657.

Romei, A., & Ruggieri, S. (2014). A multidisciplinary survey on discrimination analysis. *The Knowledge Engineering Review*, *29*(5), 582–638.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, *77*(5), 1004.

Smith, K. N., & Gayles, J. G. (2017). " Setting Up for the Next Big Thing": Undergraduate Women Engineering Students' Postbaccalaureate Career Decisions. *Journal of College Student Development*, *58*(8), 1201–1217.

Van den Assem, M. J., Van Dolder, D., & Thaler, R. H. (2012). Split or steal? Cooperative behavior when the stakes are large. *Management Science*, *58*(1), 2–20.

Wall, G. (2011). Outwit, outplay, outcast? Sex discrimination in voting behaviour in the reality television show Survivor. *New Zealand Economic Papers*, *45*(1–2), 183–193.

Weichselbaumer, D., & Winter-Ebmer, R. (2005). A meta-analysis of the international gender wage gap. *Journal of Economic Surveys*, *19*(3), 479–511.

Appendix A. Definitions of variables

potential is female denotes sex of potentially nominated player: 1 if the potentially nominated contestant is female, 0 otherwise;

potential is female in '90s represents the interaction between sex of the potentially nominated player and the fact that the episode aired in the '90s: 1 if potentially nominated is female and the episode is from '90s, 0 otherwise;

same-sex: 1 if the potentially nominated and the nominating player are of the same sex, 0 otherwise;

same-sex in the '90s represents the interaction between *same-sex* and the fact that the episode aired in the '90s: 1 if the potentially nominated and the nominating player are of the same sex and episode is from '90s, 0 otherwise;

*potential is female * round* represents the interaction between the sex of the potentially nominated player and the round: 2 if potentially nominated is female and the *round*==2, 3 if a potentially nominated player is female and *round*==3, 0 otherwise;

closest contestant (on the left): 1 if the potentially nominated player is the closest surviving contestant to the left of the nominating player, 0 otherwise;

closest contestant (on the right): 1 if the potentially nominated player is the closest surviving contestant to the right of the nominating player, 0 otherwise;

distance from nominating to potential denotes the distance from the nominating player to the potentially nominated (absolute value of the difference of numbers of these two players);

potential from big city: 1 if the potentially nominated player comes from a city of 100,000 or more inhabitants, 0 otherwise.

potential from mid-sized city: 1 if the potentially nominated player comes from a city of 20,000-100,000 inhabitants;

posteriori pot denotes the posteriori expectation of correctness of answers for the potentially nominated player. It is a characteristic estimated separately for each moment of the game and player based on his or her or performance hitherto and the prior gender-specific distribution;

relative measure own posteriori pot.* stands for the interaction of relative measure of situation of the nominating player and *posteriori pot;*

*posteriori pot. * fem. pot.* represents the interaction of *posteriori pot* with the sex of the potentially nominated player, equal to *posteriori pot.* for female nominees, 0 for males;

relative measure own fem. pot.* is the interaction of relative measure of situation of the nominating player with the sex of the potentially nominated player, equal to *relative measure own* for female nominees, 0 for males;

chances_pot=2 is a dummy variable indicating that the potentially nominated player has exactly 2 chances;

chances_pot=3 is a dummy variable indicating that the potentially nominated player has exactly 3 chances;

*chances_pot=2 * fem.pot* is the interaction of *chances=2* with the sex of the potentially nominated player (1 if the potentially nominated is female and has 2 chances, 0 otherwise);

*chances_pot=3 * fem.pot* is the interaction of *chances=3* with the gender of the potentially nominated player (1 if potentially nominated is female and has 3 chances, 0 otherwise);

chances_pot=2 and same-sex: the number of chances of the potentially dominated if the latter and the nominator are both male or both female, 0 otherwise;

chances_pot=3 and same-sex: the number of chances of the potentially dominated if the latter and the nominator are both male or both female, 0 otherwise;

self-nominated is a dummy variable denoting that the nominator and the potential nominee is the same person;

female nominates herself is the interaction of *self-nominated* with the sex of the nominator;

*fem. nominates * potential has last chance*: 1 if a female nominates and the potentially nominated has 1 chance, 0 otherwise);

abs. chances diff.: *the* absolute value of the difference of the number of chances between the nominator and the potential nominee;

instant revenge takes the value of 1 if the potentially nominated player has nominated the current nominator to answer the previous question, 0 otherwise;

female instant revenge is the interaction of *instant revenge* with the sex of the nominator: *instant revenge* if she is female, 0 otherwise;

revenge later: 1 if the nominator has ever been nominated by the potentially nominated player in the past, 0 otherwise;

female takes revenge later: *revenge later* if the nominating player is female, 0 otherwise;

major in ____ dummies encode the academic major of the potentially nominated contestant

___ *edu.* dummies encode the education level of the potentially nominated contestant

Appendix B. Estimates for control variables

Table B1: Reciprocal motives (marginal effects)

Variable \ Specification	4	5	6
ROUND 2			
instant revenge	0,030 (0,004) **0,000**	0,012 (0,008) **0,163**	0,012 (0,008) **0,159**
female instant revenge	-0,037 (0,010) **0,000**	-0,032 (0,012) **0,007**	-0,032 (0,012) **0,007**
revenge taken later		0,002 (0,008) **0,742**	0,002 (0,007) **0,757**
female takes revenge later		-0,023 (0,012) **0,066**	-0,023 (0,012) **0,054**
N	34888	34888	34888

ROUND 3			
instant revenge	0,031 (0,006) **0,000**	0,016 (0,005) **0,002**	0,017 (0,005) **0,001**
female instant revenge	-0,025 (0,013) **0,051**	-0,021 (0,013) **0,101**	-0,021 (0,012) **0,080**
revenge taken later		0,035 (0,006) **0,000**	0,033 (0,006) **0,000**
female takes revenge later		-0,017 (0,016) **0,288**	-0,015 (0,014) **0,299**
N	6779	6779	6779
ENTIRE DATASET			
instant revenge	0,017 (0,004) **0,000**	0,013 (0,005) **0,008**	0,013 (0,005) **0,006**
female instant revenge	-0,022 (0,009) **0,015**	-0,018 (0,009) **0,055**	-0,018 (0,009) **0,056**
revenge taken later		0,001 (0,004) **0,860**	0,000 (0,004) **0,924**
female takes revenge later		-0,006 (0,008) **0,495**	-0,006 (0,008) **0,484**
N	41667	41667	41667

For each independent variable marginal effects, (standard errors) and **confidence levels** were given.

Each specification for Round 3 and the entire dataset indicates that the players are willing to nominate back, directly after having been nominated. According to the estimates for the entire dataset, it is the female contestants that are less likely to take such revenge. The estimate for the delayed revenge is positive in Round 3, and the interaction with the gender of the nominating player shows that that tendency is not affected by the gender of the decision-maker. However, in Round 2 we can observe that being nominated by a player in the past makes a female contestant, at a later stage of the game, even less likely to nominate that player.

Table B2: Other control variables (marginal effects)

Variable \ Specification	6	7
ROUND 2		
major in social sciences	0,008 (0,006) **0,159**	0,009 (0,006) **0,159**
major in natural sciences	-0,001 (0,007) **0,887**	0,000 (0,007) **0,946**
major in humanities	0,012 (0,006) **0,041**	0,013 (0,006) **0,020**
major in health sciences	0,006 (0,010) **0,516**	0,009 (0,010) **0,380**
major in mathematics etc.	0,003 (0,007) **0,710**	0,004 (0,007) **0,584**
other major	-0,014 (0,016) **0,374**	-0,011 (0,017) **0,541**
secondary edu.	0,000 (0,006) **0,982**	0,002 (0,005) **0,763**
vocational edu.	0,000 (0,006) **0,978**	-0,003 (0,006) **0,685**
university edu.	0,001 (0,005) **0,858**	-0,001 (0,006) **0,917**
secondary edu. * fem.pot		-0,017 (0,014) **0,229**
vocational edu. * fem.pot		0,044 (0,023) **0,059**
university edu. * fem.pot		0,011 (0,011) **0,331**
closest contestant (on the left)	0,027 (0,006) **0,000**	0,026 (0,006) **0,000**
closest contestant (on the right)	0,004 (0,006) **0,508**	0,004 (0,006) **0,501**
	0,019	0,019

distance from nominating to potential	(0,001) **0,000**	(0,001) **0,000**
potential from big city		0,004 (0,004) **0,380**
potential from mid-sized city		-0,002 (0,005) **0,687**
N	34888	34888
ROUND 3		
major in social sciences	-0,001 (0,006) **0,960**	0,000 (0,006) **0,960**
major in natural sciences	0,017 (0,009) **0,060**	0,019 (0,009) **0,033**
major in humanities	0,004 (0,005) **0,507**	0,003 (0,005) **0,566**
major in health sciences	0,005 (0,011) **0,660**	0,003 (0,009) **0,777**
major in mathematics etc.	-0,002 (0,006) **0,790**	-0,004 (0,007) **0,570**
other major	0,016 (0,009) **0,073**	0,016 (0,008) **0,052**
secondary edu.	0,007 (0,006) **0,244**	0,001 (0,006) **0,813**
vocational edu.	-0,002 (0,005) **0,691**	-0,007 (0,006) **0,180**
university edu.	-0,006 (0,005) **0,277**	-0,010 (0,005) **0,043**
secondary edu. * fem.pot		0,033 (0,011) **0,004**
vocational edu. * fem.pot		0,056 (0,017) **0,001**
university edu. * fem.pot		0,034 (0,010)

		0,000
closest contestant (on the left)	0,012 (0,004) **0,009**	0,012 (0,005) **0,007**
closest contestant (on the right)	0,001 (0,006) **0,887**	0,001 (0,006) **0,855**
distance from nominating to potential	0,000 (0,001) **0,819**	0,000 (0,001) **0,915**
potential from big city		-0,006 (0,004) **0,113**
potential from mid-sized city		0,001 (0,004) **0,771**
N	6779	6779
ENTIRE DATASET		
major in social sciences	0,004 (0,004) **0,343**	0,004 (0,004) **0,343**
major in natural sciences	0,003 (0,005) **0,584**	0,003 (0,005) **0,477**
major in humanities	0,007 (0,004) **0,041**	0,007 (0,004) **0,047**
major in health sciences	0,002 (0,009) **0,863**	0,001 (0,009) **0,892**
major in mathematics etc.	0,001 (0,005) **0,789**	0,002 (0,005) **0,657**
other major	-0,001 (0,008) **0,913**	0,001 (0,009) **0,941**
secondary edu.	-0,001 (0,004) **0,735**	-0,001 (0,004) **0,777**
vocational edu.	-0,003 (0,004) **0,420**	-0,006 (0,004) **0,148**
university edu.	-0,004 (0,004) **0,239**	-0,006 (0,004) **0,106**

secondary edu. * fem.pot		0,000 (0,007) **0,973**
vocational edu. * fem.pot		0,040 (0,016) 0,011
university edu. * fem.pot		0,016 (0,008) **0,030**
closest contestant (on the left)	-0,006 (0,004) **0,855**	-0,007 (0,004) **0,855**
closest contestant (on the right)	-0,020 (0,004) **0,000**	-0,020 (0,004) **0,000**
distance from nominating to potential	0,006 (0,001) **0,000**	0,006 (0,001) **0,000**
potential from big city		0,001 (0,003) **0,665**
potential from mid-sized city		-0,001 (0,003) **0,750**
N	41667	41667

For each independent variable marginal effects, (standard errors) and **confidence levels** were given.

Generally speaking, measures of physical proximity have a plausible effect: contestants often nominate their nearest surviving neighbor on the left (the next clock-wise, often bearing a number equal to nominator's number plus one). However, they are also more likely (in Round 2) to nominate an opponent far away from self and thus presumably psychologically distant, as well as comfortably visible without the need to turn the head.

Demographic variables other than gender are limited in terms of their impact and are slightly different between the rounds. In Round 2 only major in humanities was statistically significant and had a positive marginal effect. While in Round 3 we see quite a similar effect for major in natural sciences or majors other

than humanities, health sciences or mathematics. For the estimation based on an entire dataset, the effect reported for Round 2 is predominant.

Males' level of education plays no role, but in Round 2 we observe a positive effect for vocational education of a female potential nominee. In the case of estimations on an entire dataset, we see that both vocational and university education have a positive impact on the probability of being nominated.