WHEN SHOULD I QUIT?
GENDER DIFFERENCES IN EXITING COMPETITIONS

Robin M. Hogarth, Natalia Karelaia, and Carlos Andrés Trujillo

robin.hogarth@upf.edu
natalia.karelaia@insead.edu
cat@adm.uniandes.edu.co

January 20, 2011

* The authors’ names are listed alphabetically. Hogarth: ICREA & Universitat Pompeu Fabra, Barcelona, Spain; Karelaia: INSEAD, Fontainebleau, France; Trujillo: Universidad de los Andes, Bogotá, Colombia. Hogarth is also Affiliate Professor at the Barcelona Graduate School of Economics and thanks the Generalitat of Catalunya for its support. The authors are grateful for comments received from Irina Cojuharenco, Gert Cornelissen, Patricia Funk, Herminia Ibarra, Nagore Iribarri, and four anonymous reviewers for the 2010 Academy of Management meeting. They thank Carolina Barrios Laborda and Catalina Estrada for their help in collecting and coding the data, as well as for their comments during the early stages of this research. The authors also thank RCN TV, and in particular Lucas Khapaz, for granting access to the show recordings.
WHEN SHOULD I QUIT?  
GENDER DIFFERENCES IN EXITING COMPETITIONS

Abstract

We study gender differences in exiting competitive environments by exploiting the “naturalistic experiment” of a TV game show where participants were self-selected and there were no gender-specific constraints or discrimination. In multiple rounds, contestants answer general knowledge questions privately. One participant is eliminated or leaves voluntarily at the end of each round. Women earn 40% less than men and exit the game prematurely at a faster rate, but especially when in a minority. This latter result highlights the importance of structural arrangements in organizations that interact with behavior to maintain “glass ceilings” and explains the differential gender-related risk attitudes observed.

Keywords: competition, gender differences, glass ceilings, minority behaviors, discrimination.

JEL classification: C93, D03, J16, J71.
I. Introduction

Gender equality is important. For too long, society has ignored or suppressed the potential contributions of females in business, the arts, science, and politics, and this has resulted in two major costs. First, many aspirations have been frustrated. Second, society has failed to profit fully from the potential contributions of women.

Despite legislation aimed at gender equality, male-female wage disparities are still large. In the United States, the female-male weekly earnings ratio of full-time workers was about 80% in 2009 (U.S. Bureau of Labor Statistics, 2010a). Moreover, statistics regarding major leadership roles consistently show gender inequality. For example, in the United States, women account for 47% of the workforce (U.S. Bureau of Labor Statistics, 2010b) but only 6.3% of top earners in the Fortune 500 companies (Catalyst, 2009) and 3% of the CEOs (Catalyst, 2010).

In politics, the situation is not much better. Although there are many elected female representatives in different countries, women are still a minority in terms of power and less women than men are heads of state or members of national parliaments. Worldwide, only 19% of local parliament members are women and female presidents over one of the Houses of existing Parliaments are a mere 16% (Inter-Parliamentary Union, 2010).

The most cited reasons for the lack of women in the upper levels of management and politics are preference differences related to child rearing, discrimination, and more recently, gender differences in preferences for performing in competitions. However, contrary to previous research on gender differences concerning self-selection into competitive tasks (e.g., Gneezy, Leonard, & List, 2009; Niederle & Vesterlund, 2007), we report here a novel empirical investigation that examines gender differences in decisions to exit competitions.
To study this issue, we exploit the naturalistic experiment of a TV game show. In this game, contestants earn money by answering general knowledge questions across five rounds and must decide, at the end of each round, whether to stay in the game or leave. If they leave, contestants keep the money earned up to that point; however, if they fail to leave and have the lowest “score” of the remaining contestants, they are expelled from the game and lose all their earnings. The game starts with six contestants and one contestant leaves voluntarily or is expelled at the end of each of five rounds. At the end of five rounds, the overall winner emerges. Note that the rules make no distinctions between men and women. Thus, if gender differences emerge, they are a consequence of abilities, attitudes, and behaviors that contestants bring to the game.

This paper is organized as follows. In the next section, we review briefly the literature on game shows and comment on gender differences in risk attitudes and competitiveness. We then discuss the role of situational and structural factors in creating gender gaps in the work place. This is followed by a detailed description of the game. Next we present the specific questions that we consider and the results of our analysis.

In short, although games start with approximately equal numbers of males and females, women earn less than men and exit the game at a faster rate. In particular, there are more premature voluntary withdrawals by women. Importantly, women leave the game voluntarily more when the proportion of female contestants decreases. By contrast, men’s decisions are not affected by the gender composition of the groups.

We conclude by drawing an analogy between the game and the process by which employees rise through the levels of a corporation except that the game is free of many constraints typically faced by women. Our results therefore imply that one reason for “glass ceilings” lies in women’s own behavior. This, in turn, probably reflects socialization practices, the expectations of gender-specific behavior, and the unequal
situation of genders in societal and organizational structures. Our study suggests two important implications. First, to reduce voluntary turnovers, it is important to retain a sufficiently high proportion of women to men in work groups. Second, obtaining gender diversity at the top of organizations requires promoting and maintaining it at the lower/middle ranks.

II. Game shows

TV game shows have already attracted researchers interested in risky decision making. There are several advantages and disadvantages. For the former, games involve real and, often, substantial payoffs; participants come from a wide, general public; and many choices in the games should depend on explicit probabilistic reasoning. On the other hand, the games can be hard to model analytically, and the public nature of the “entertainment” may have effects on the types of participants recruited as well as their behavior.

Many papers have used game shows to elicit risk attitudes in the presence of substantial payoffs (see, e.g., Beetsma & Schotman, 2001; Blavatsky & Pogrebna, 2008; Gertner, 1993; Metrick, 1995; Post, van den Assem, Baltussen, & Thaler, 2008). An illuminating review by Andersen, Harrison, Lau, and Rutström (2008) indicates methodological difficulties and offers constructive suggestions.

Of special interest is a study by Larkin and Pines (2003) who document significant gender differences. Specifically, females consider going on shows to be more personally risky and have greater concerns about appearing badly in public. Indeed, in some shows women are a distinct minority. There is also evidence of differences in risk attitudes and performance with women both taking less risk than men
(Daghofer, 2007; Mulino, Scheelings, Brooks, & Pfaff, 2006) and earning less (Johnson & Gleason, 2009).

III. Few women at the top: Preference for competition and other explanations

The most cited reasons for the lack of women in leading positions are preferences related to child rearing, discrimination, and gender differences in preferences for competition.

Preferences for child rearing are important. For example, in a study of MBA graduates, Bertrand, Goldin, and Katz (2010) found that the presence of children contributes to less job experience, greater career discontinuities, and shorter work hours for women. In this paper, we focus on the other two explanations.

Gender discrimination has been extensively documented both with laboratory and field data. Evidence suggests that women are less likely to be promoted than males, and if promoted, have stronger performance ratings (Lyness & Heilman, 2006). Moreover, even minor gender bias in performance evaluation can trigger dramatic differences at the top levels of organizations. For example, in a computer simulation of an organization with eight hierarchical levels, Martell, Lane, and Emrich (1996) showed that, starting with equal numbers of men and women, a 1% evaluation bias favoring men can lead to almost twice as many men at the top.

The evidence for gender differences in preferences for competition has been mainly experimental. Gneezy, Niederle, and Rustichini (2003) showed that men and women differ in their “drive” to perform well in mixed-sex competitive environments. Men’s performance increases with the competitiveness of the environment, thereby creating a gender gap in performance in winner-takes-all contests. Similarly, Niederle and Vesterlund (2007) provide evidence that, when given the choice, men as opposed to
equally skillful women, self-select more into competitive as opposed to non-competitive compensation schemes.

Recently, Gneezy et al. (2009) emphasized the importance of socialization in gender preferences by comparing preferences for selecting into competitive environments by members of the patriarchal Maasai (in Tanzania) with those of the matrilineal Khasi (in India). Whereas Maasai men chose competitive environments more often than Maasai women, Khasi women chose competitive environments more often than Khasi men. In a related vein, Ahlgren and Johnson (1979) reported that the greatest gender difference in school students’ attitudes towards cooperation versus competition appears in grades 8-10. Female students show consistently more positive attitudes toward cooperation, whereas males favor competition.

From early childhood, individuals absorb social beliefs about the actual and the ideal behavior of women and men. The gender stereotypes learned affect thoughts and motivations, guide behavior (e.g., Martin & Halverson, 1981; Martin, Wood, & Little, 1990), and are themselves obstacles to gender equality in competitive environments. First, behavior congruent with social roles improves self-image (Wood, Christensen, Hebl, & Rothgerber, 1997) thereby explaining the lower willingness of women to behave competitively. Second, perceived incongruence between actual behavior and the female gender role (e.g., “leader” and “woman”) leads to less positive attitudes towards women (Eagly & Karau, 2002). Thus, the punishment anticipated for deviating from expected behaviors reinforces behavior consistent with gender stereotypes.

Studies on negotiations further suggest that in competitive tasks males orient themselves more to the impersonal task of maximizing earnings, while women demonstrate a greater concern for relationships (Rubin & Brow, 1975; Kray & Thompson, 2005). Similarly, women managers appear less inclined than their male
counterparts to define career success in terms of hierarchical and financial progression (Sturges, 1999).

Finally, pressures for maintaining gender identity by conforming to gender stereotypes is greater when gender is salient. For example, Booth and Nolen (2009) provide evidence that girls from single-sex schools develop different attitudes toward risk and competition than girls co-educated with boys. In a controlled experiment, students aged just under 15 from single-sex and coeducational schools were put together in groups, asked to solve some mazes, and then given a choice between tournament-style and competition-free evaluation of their performance. Girls from single-sex schools choose tournaments more than girls from coeducational schools.

Above, we cited three reasons for gender gaps: child rearing, discrimination, and differences in preferences for competition. In our data, all three can be ruled out as explanations. First, child rearing is not in conflict with participating in the game. Second, there are approximately equal numbers of men and women at the beginning of the game and all decisions are taken individually and independently by each contestant. Third, women were aware of the nature of the game before they decided to compete.

### IV. Structural and situational effects

Gender may explain some individual behavior but is often confounded by situational and structural factors (Kanter, 1977a; Ely & Padavic, 2007). For example, highly placed women in organizations are often “tokens” in groups dominated numerically by men. Therefore, conclusions about their behavior can confuse effects of situation and gender. The same can be said about women in lower positions: their behavior inevitably reflects having less power. In fact, literature on gender identity suggests that the juxtaposition of gender and power/status implies that gender is largely
a status characteristic (Ridgeway & Smith-Lovin, 1999). Wagner and Berger (1997) proposed that differences in stereotypical gender tasks are a direct function of status differences between genders or of attempts to cope with these differences.

Kanter (1977b) suggests that the hierarchical arrangements of organizations as well as how social groups differ in relative numerical representation can account for what appear to be “gender phenomena.” That is, power structures and organizational demography are not only the result but also the origin of gender inequalities. For example, Cassirer and Reskin (2000) report that gender differences in career aspirations and attitudes towards promotion can be fully explained by workers’ organizational locations and prior promotions. The fact that men are more likely to occupy positions that can lead to promotion may explain why they attach greater importance to career aspirations. Women respond to lower upward mobility by limiting their aspirations. Importantly, men in positions with low chances of advancement behave similarly (Kanter, 1977b).

“Tokens” face more pressures in organizations than “numerical dominants” (Kanter, 1977a) in that minority achievements are diminished in majority-dominated work groups. For example, Spangler et al. (1978) used data from two law schools with different gender ratios and found that social influences such as performance pressure, social isolation, and role entrapment reduced women’s performance when they were in a small minority.

Women are more sensitive to gender composition than men. For example, Elvira and Cohen (2001) analyzed turnover rates from ten business units of a Fortune 500 financial firm over three years and found that women were less likely to leave the firm voluntarily when more women were employed at their job level. By contrast, the analogous effect did not occur for men.
Tolbert et al. (1995) analyzed turnover in 50 academic departments, and found that the proportion of women had a negligible impact on turnover among male faculty. Literature on tokenism has also consistently reported that unlike women, men do not suffer from the negative effect of being a token (Zimmer, 1988). Thus, being “the few” in a skewed work group has different consequences for men and women. One explanation is that being in the numerical minority is less negative for men precisely because they are accorded more power and social status in society (Konrad, Winter, & Gutek, 1992).

Experimental evidence also suggests that group demography is important in shaping women’s behavior in competitive situations. For example, Gneezy et al. (2003) reported that the gender gap in performance in competitive environments was stronger when women had to compete against men than in single-sex tournaments.

V. The game

We analyzed the first season of a game called “El Jugador” (“The Player”) that was broadcast in Colombia by RCN TV in the first trimester of 2007\(^1\). It is important to emphasize that all episodes of this new game were recorded before any broadcasts were made so that no contestant had any specific knowledge about the game before participating in it. Thus, since the game was new to all contestants, there was no confounding effect of differential knowledge.\(^2\)

Participants were recruited by channel personnel who traveled to major cities in Colombia. Public announcements of a new game were made locally using regular

---

\(^1\) This game is the Colombian version of the TV game “Poker Face” that was broadcast for the first time in the UK in 2006. The original version has never been broadcast in Colombia. Since 2006, the show has been adapted in more than ten countries around the world (http://en.wikipedia.org/wiki/PokerFace_(game_show)).

\(^2\) In fact, the game has only been broadcast in Colombia twice: once, the initial version that we analyze, and secondly in a celebrity version.
media, and the channel personnel interviewed those who attended the recruitment meetings. There was no attempt to have a random sample of the Colombian population. Instead, the organizers set out to select a pool of contestants who would be literate, vocal, heterogeneous, with at least a middle level income, inclined to take risks, and who would feel relatively comfortable in front of the camera (all characteristics were subjectively evaluated by the organizers).

Among people who responded to the recruiting announcements, there were approximately 60% men and 40% women. The organizers then recruited more female players by approaching potential contestants at commercial malls. Of the 216 selected contestants, 102 (47%) were women. Based on the selection criteria, the female contestants on the show were probably more extraverted and risk-seeking than the average women in Colombia. Contestants varied in age between 18 and 65 (median of 30) and, in terms of occupation, could be classified as students (21%), professionals (69%), and others (10%). These proportions were almost identical for men and women.

The game was based on answering general knowledge questions. Importantly, the questions covered multiple diverse domains with overall no male or female content bias. A content analysis of a random sample of 112 questions revealed the following distribution of questions between domains: 17% history, 13% language/vocabulary skills, 7% sports, 7% soap operas and movies, 6% geography, 6% literature, 6% science, 4% popular music, 4% basic math skills, and 30% distributed between smaller domains (e.g., celebrities, arts, nature, architecture, popular culture and others).³

We distinguish between three sets of games. The first were “regular” games broadcast each day from Monday through Saturday for six consecutive weeks. Thus there were 36 different regular games. The winner of each regular game (Mondays

³ Examples of the questions used can be obtained from the authors.
through Saturdays) participated in a “Sunday” game with the other winners from the same week. Thus there were 6 different Sunday games. Finally, the six winners of these Sunday games played in a grand final game at the end of the season. Given that the stakes were a lot higher in the single, grand final game, we limit our analysis in what follows to the 36 regular and six Sunday games (exceptions are explicitly indicated).

The organizers of the game arranged groups of six contestants trying to maximize the variety of profiles in each group, as well as to balance the numbers of men and women. Of the 36 groups in the regular games, 26 were composed of three men and three women, eight groups included four men and two women, and the two remaining groups included two men and four women.

Prior to starting each game, there is a meeting of the six contestants where they introduce themselves to each other and may engage in strategic behavior. For example, during this presentation they may tell the truth or lie. (In the broadcasts, it is shown on-screen whether they are lying or not). Contestants are also asked privately how “strong” they perceived the others to be based solely on that first encounter.

Each game consists of five rounds of general knowledge questions (eight questions in the first round and five in each of the subsequent rounds). From the perspective of each contestant, all questions are asked publicly but answered privately using an individual screen. Correct answers are rewarded with monetary values that increase across rounds.

After each round, contestants are ranked according to the money they have accumulated up to that point in the game. Also, after each round they have to decide whether to stay or withdraw from the game. If a contestant withdraws, s/he can keep her/his accumulated prize money. If nobody withdraws, the last contestant in the rankings is expelled and loses all of her/his accumulated prize money. Ties are broken
using the time contestants took to answer. (Faster is better). The decision to withdraw is made by pushing a special button.

If several contestants manifested their decision to withdraw by pushing the withdrawal button, only the player who was the first to do so actually leaves the game. Contestants who were slower in pushing the button stay in the game and face no immediate material consequences for their decision to withdraw. In the analyses that follow, we pool together all decisions to withdraw – i.e., all instances of pushing the decision button, both resulting in actual exits and not. All exceptions are explicitly indicated.

Contestants are allowed and encouraged to interact publicly after each round of questions in order to gain strategic advantages (e.g., by bluffing). The host asks questions about performance and opinions and contestants may say whatever they want. There is no sure way for the participants to figure out who was lying and who was not. There were many such interactions that took a significant amount of game time. We were able, therefore, to observe and code types of behavior for all participants in each round. The following are some examples of contestant interactions:

“I did great in this round. Questions were very simple!”

“I feel extremely confident.”

“You (addressing other contestant) seem very upset; you should leave before losing everything, because I am doing great!”

“You clearly seem to ignore all these topics, leave while you can.”

“Be sure that I am not withdrawing, believe me, you better push that button.”

Finally, the game show host gives the contestants a 10-second warning before the withdrawal button needs to be pushed. During these ten seconds, the contestants can also behave strategically by addressing comments to their competitors.
During the game participants don’t know their own rank. The only feedback each receives is her/his number of correctly answered questions during the last round played and the total money that he/she had accumulated.

Each round ends with a private debriefing interview with the contestant that withdrew or was expelled in which the rankings are revealed and impressions about performance relative to others are discussed.

The financial stakes of the game are high. Across rounds, the rewards for correct answers vary from 75 USD (round 1) to 375 USD (round 5). Questions were selected by the organizers to be approximately equal in difficulty across rounds.

Two researchers coded all the data by watching replays of all 43 episodes of the game show. Most of the data were objective and thus used directly (e.g., numbers of correctly answered questions, accumulated gains, and so on). Apart from these quantitative data, we coded the behavior of contestants during the time allowed for interactions (i.e., prior to the withdrawal decisions). In particular, two coders watched the interaction episodes and independently coded how active the contestants were during the interactions on a scale from 0 (“not vocal”, i.e., the contestant stays silent) to 3 (“very vocal”, i.e., the contestant is actively engaged in persuading the others, talks a lot, or dominates the conversation). The subjective judgments of the two coders were highly correlated. The initial agreement was 98%. When the coders disagreed, they watched the relevant episodes again, and reached an agreement by consensus.

**VI. Main questions**

Our main question is whether equally skillful women and men quit competitive environments at the same rate. In a context free of discriminatory practices and preferences for special considerations such as child rearing, does the gender gap
disappear? We start the games with an approximately equal gender ratio and so the null hypothesis is no differences in the gender ratio as the game progresses across rounds.

If the null hypothesis is rejected, what factors could explain any ensuing gender gap? We focus on understanding whether women who self selected into the competition quit the game earlier than men. As noted above, experimental evidence suggests that when given a choice, women as opposed to equally skilful men, self-select less into competitive environments (Niederle & Vesterlund, 2007). Our data provide a unique opportunity to analyze the behavior of women who have already entered a competitive environment. Do they stay in the competition as long as men?

We also ask what structural and situational factors can explain differences in the behavior of men and women. One important factor is the proportion of women relative to men in organizational units. We hypothesize that men’s and women’s decisions to leave the game will be differentially affected by the gender composition of the groups. Specifically, men should be less sensitive to the gender ratio. That is, whereas we expect women to be more likely to exit the game as the proportion of remaining females falls, we do not expect an analogous effect for men. Since the gender compositions of the groups change every round as a consequence of eliminations and withdrawals, our data allow us to test this hypothesis. The hypothesis, we note, reflects the literature summarized above that documented the greater power and status accorded to men in society as well as the fact that men’s experiences of being in the numerical minority are less negative than those of women (see, e.g., Konrad, Winter, & Gutek, 1992).

VII. Results

As noted above, we report data for six weeks of the show involving 36 regular games and six Sunday games. Of the 216 players, 90 were expelled and 155 decided to
leave the game voluntarily (i.e., pushed the decision button). Of these 155 decisions, 125 resulted in actual withdrawals. The mean payoff that contestants took home was 2,618 USD, with a median of 625 USD. The winner of the final game (a man) earned 29,625 USD. We report general results on the behavior of all contestants first and then describe gender-specific results.

Withdrawals. Table I presents the number of people who decided to withdraw in each round (aggregated across all games). The decision to withdraw was taken 155 times in the coded shows.

Do players decide to withdraw too often or too seldom? If all players decide to stay in, the base-rate probabilities for being expelled vary by round: 1/6 for round 1; 1/5 for round 2; 1/4 for round 3; 1/3 for round 4; and ½ for round 5 (last column in Table I). These probabilities can be used to find the expected number of withdrawals (lower panel of Table I). All in all, if players were correctly calibrated, one player should decide to exit each round, so that there would be 36 exits per round in 36 regular games, and 6 exits per round in 6 Sunday games.

In fact, in the 36 regular games too few people chose to exit, e.g., only 9 and 23 in rounds 1 and 2. Across all rounds of the regular games, there were 108 decisions to withdraw, while the expected number of withdrawals is 180. However, in the 6 Sunday games, participants attempted to exit the game more often than the base-rate probabilities suggest: there were 42 decisions to exit against 30 expected exits. One explanation for the apparent increase in risk aversion is that accumulated gains were much larger in Sunday as opposed to the regular games (details are provided further
below). In addition, on the Sundays contestants were playing the game for a second time and knew that they were facing other winners of regular games.

Winners. Twenty-five men and 11 women won in the 36 regular games. Of these, 5 men and 1 woman won in Sunday games.

The future winners-to-be already did better than other contestants in the first rounds of regular games. Table II details the gains of the eventual winners and other players in the five rounds of regular games. The advantage of performing well early in the game is demonstrated by the fact that the difference between the payoffs of the eventual winners and others is statistically significant for all but the fourth rounds.4

We next report our main gender-specific results.

Result 1: There is a significant gender gap in earnings.

Women earned significantly less than men: mean earnings were 3,390 USD for men and 1,756 USD for women (52% relative to men).5 The maximum earnings among women were 23,550 USD (79% of the male maximum). The median earnings among men were also significantly larger: 936 USD for men vs. 375 USD for women (40% relative to the men’s median).

Result 2: The proportion of women decreases significantly at later stages of the game.

Of the 216 contestants in regular games, 102 were female (47%). Figure I shows that the proportion of female contestants decreased dramatically as the game advanced. That is, in the last round of regular games, 50 of 72 contestants were males (69%). In the last round of Sunday games, there were ten men (83%) and two women.

---

4 This difference is particularly striking when it is recalled that payment for correct answers increased by 75 USD each round.
5 The difference is statistically significant, Wilcoxon rank-sum test \( z = 2.29, p = 0.02, n = 216 \).
We next analyze possible reasons for the decreasing female/male ratio among contestants. First, it is possible that women were forced out earlier than men from the game due to inferior performance. Second, it is possible that more women withdrew voluntarily from the game than men.

**Result 3: In regular games, more women are expelled than men.**

We first analyzed performance across rounds at the individual level. Specifically, for every contestant, we averaged across rounds of regular and Sunday games the percentage of questions s/he answered correctly – see Figure II. Median individual performance was 57.6% for men and 51.4% for women; mean individual performance was 56.0% and 50.5%, respectively; the difference between the means is statistically significant (Wilcoxon rank-sum test, $z = 2.395 \ p = 0.003, \ n = 216$). In addition, the range of individual performance was greater for women than men. The percentage of questions answered correctly by the worst individual performer was 12.5% among women and 25% among men. The best individual performances for women and men were, respectively, 77.5% and 81.9% correct answers.

We next analyzed the data on expulsions. Across all games, 47 women and 43 men who had the lowest performance at the end of rounds were expelled from the game. Figure III illustrates the percentage of expelled contestants by game and round, separately for males and females. The percentage of expelled contestants is larger for

---

注：在第一轮的常规游戏中，两男一女都表现得比前几轮都要好。具体来说，在这些轮次中，中位数表现分别为75.0%和62.5%；均值个体表现分别为69.8%和63.8%，分别；两组间的差异具有统计学意义（Wilcoxon rank-sum test, $z = 2.338 \ p = 0.019, \ n = 216$)。
women than men in the first four rounds of the regular game. Gains that male and female contestants accumulated at the end of each round (before any expulsion was made effective) confirm the pattern of expulsions (Table III). Female contestants had lower accumulated gains than male contestants in all rounds of regular games. However, in Sunday games there was no significant difference between the accumulated gains of men and women who remained in the game.

Result 4: *The rate of voluntary withdrawals is larger for women than for equally skilled men.*

Across all games, 68 women and 87 men decided to leave the game voluntarily. Figure IV shows the rates of withdrawals (both resulting and not resulting in actual exits) by gender, game, and round. The proportion of contestants who voluntarily withdraw from the game is consistently larger among females. Moreover, the gap increases in Sunday games. In regular games, 6% of female as opposed to 3% of male contestants decided to withdraw from the first round. The withdrawal rate increased to 12% among men and 23% among women in the third round, and to 28% among men and 36% among women in the last round. In Sunday games, 32% of men vs. 64% of women decided to leave the game in the first round and 30% of men vs. 50% of women in the last round.

Did women exit the game voluntarily because, having accumulated less gains than men, they anticipated the risk of being expelled? To answer this question, we fitted a population-averaged logistic model of exit decisions (made in regular and Sunday
games) with gender (female coded as 1) and accumulated gains at the moment of the decision as predictors. Indicator variables for rounds and the type of game (regular vs. Sunday) were additionally included. This specification illuminates how, within a given moment of the game, the behavior of women differed from that of men who had the same amount of accumulated gains (and thus the same probability of being expelled from the game). Standard errors were adjusted for the correlation across repeated observations. The results are presented in Table IV (Model 1). The model showed a significant effect of gender ($p<0.05$). The probability of withdrawal was 6.5 percentage points larger for women than men ($p<0.05$ also for the marginal effect of gender in Model 1). Figure V depicts predicted probabilities of withdrawal as a function of accumulated gains, separately for men and women (other predictors were fixed at their mean values). The figure illustrates that gender differences in exists were especially pronounced at lower levels of accumulated gains and tended to dissipate when the gains grew larger.

We would have liked to classify decisions as “correct” or “incorrect” ex-ante but this proved infeasible given the available data. We therefore classified decisions ex-post. Thus, a decision was correct if a contestant was indeed ranked last and would be

\footnote{The estimates of population-averaged (or marginal) specifications are interpreted as group differences, i.e., as effects on the average person (Diggle et al., 2004). Since we are interested in how exit decisions differed between men and women, these specifications seem more appropriate than alternative random-effect specifications. Although the latter have the advantage of accounting for natural heterogeneity among subjects in terms of unmeasured characteristics, their estimates are interpreted as within-subject effects and thus seem less appropriate to assess the effect of gender. We nevertheless checked the results of random-effect specifications; these yielded worse fit (as evidenced by lower $\chi^2$), but qualitatively the same results (not reported here).}

\footnote{The marginal effects of all indicators are positive and significant ($p<0.01$), in line with the fact that there were more withdrawals in later rounds and in Sunday as opposed to regular games. The negative marginal effect of accumulated gains in this model suggests that within each round, contestants with larger accumulated gains were less likely to withdraw – 12 percentage points less for each additional 1,000 USD of accumulated gains.}
expelled. Incorrect decisions were premature withdrawals, i.e., taken by contestants who were not ranked last in a given round. Across all games, there were 72 correct and 83 incorrect withdrawal decisions. Among women who decided to exit the game, 30 did so correctly while 38 took incorrect decisions. As for men who decided to leave the game voluntarily, 42 were right in doing so, while 45 decided to abandon the game prematurely. These data, along with the number of withdrawals resulting in actual exits are summarized in Table V. The data on expulsions are also added for reference.

To understand the effect of gender on correct vs. incorrect decisions to withdraw, we fitted separate logistic models of correct and incorrect exit decisions in regular and Sunday games with gender and accumulated gain as predictors (population-averaged models, standard errors adjusted for repeated observations). The results from the models are presented in Table IV (Models 2 and 4). The models showed that within each round, the probability of correctly withdrawing from the game was the same for men and women with the same accumulated gains (p=0.68). On the other hand, the probability of incorrect withdrawals was 4.4 percentage points higher for women than men (p=0.04).

Result 5: Incorrect withdrawals among women increase when the proportion of women in the group decreases.

We next analyzed how gender composition of groups affected contestants’ decisions to withdraw. For each group and round, we coded the proportion of female contestants in the group. For example, for a group with 2 women and 3 men the new variable was coded as 2/5=0.4. We next included this new variable and the interaction between the proportion of women and the indicator for female in the logit models of incorrect exits. The results are presented in Table IV – see Models 3 and 5. The results showed that gender composition mattered for premature withdrawals: the interaction
term in Model 5 was significant (p=0.03). Moreover, the two gender composition variables significantly improved the goodness of fit of the model, adding 26% to the \( \chi^2 \) of Model 4 \( (p=0.001) \). Interestingly, adding the gender composition variables to the model of incorrect exits does not cancel the direct effect of gender \( (p<0.01 \text{ in Model 5}) \), implying that women were overall more likely to withdraw incorrectly than men.

The significant interaction of gender composition and the indicator for female in Model 5 suggests that when deciding to withdraw, women but not men were sensitive to the gender composition of groups. The probability of women prematurely leaving the game is larger when the relative number of women in the group is smaller. Figure VI depicts the predicted probabilities of incorrect withdrawals for men and women (as implied by Model 5), as a function of the proportion of women in the group at the moment of the decision to exit. When about 65% of the group is composed of women, the probability of incorrect withdrawal is approximately the same for women and men – about 0.09 or 9%. When there are 20% women in a group, the probability of voluntary premature withdrawal by a woman is 0.18. When there are 50% women in the group, the probability of incorrect exit is 0.11, a decrease of 0.07. The probability of incorrect withdrawal further drops to 0.07 for a woman making decisions in a group with 80% of women, a decrease of 0.04.

\[ \text{Insert Figure VI about here} \]

As to men, the probability of exiting the game prematurely increases as the proportion of women in the group increases. However, the effect is not as strong as that for women. Indeed, logistic models similar to Model 5 fit separately on the data corresponding to men and women confirmed that men’s incorrect decisions to exit did not vary with the gender composition of groups (coefficient of 1.33, \( z=1.45, \ p=0.15, \)
507 observations, n=114, $\chi^2(7)=42.70$), while women exited more when the proportion of men in the group increased (coefficient of -1.79, $z=-2.08$, $p=0.04$, 333 observations, n=102, $\chi^2(7)=20.53$). Additional analyses – variations of logistic models explaining women’s premature withdrawals – revealed a threshold effect. Specifically, women were less likely to withdraw prematurely when there were at least two other women in their group (coefficient for the dummy indicating at least two more women in the group is -1.26, $z=-2.85$, $p=0.004$, 333 observations, n=102, $\chi^2(7)=20.86$). The presence of one more woman in the group was not sufficient to reduce significantly the probability of premature dropouts by female contestants (coefficient for the dummy indicating one more women in the group – vs. no other female contestants – is 0.54, $z=0.95$, $p=0.34$, 168 observations, n=72, $\chi^2(7)=14.77$).

One possible explanation for women exiting more in groups with more men is that they may anticipate performing worse in such circumstances. To test this hypothesis, we analyzed whether group composition affected the number of correctly answered questions. We fitted population-averaged linear models of the number of correct answers separately for men and women. Standard errors were corrected for multiple observations on individuals. The independent variables were the proportion of women in the group, the type of game (regular vs. Sunday), and indicators for rounds. The models revealed that women indeed performed better in groups with more female contestants: for each additional 20% of women in the group, women’s performance increased by, on average, 0.2 correctly answered questions ($z=2.11$, $p=0.04$, 333 observations).

For each group and round except the first rounds, we also coded indicator variables of whether in the previous round a woman (1) decided to leave the game voluntarily; and (2) was expelled from the game. Further analyses of logistic models revealed that these variables could not account for the women’s withdrawal behavior. That is, we did not find any evidence of women mimicking the behavior of their female peers in the previous rounds.
observations, n=102, $\chi^2(6)=334.36$). By contrast, men’s performance did not differ in groups with differential gender ratios ($z=0.76$, $p=0.45$, 507 observations, n=114, $\chi^2(6)=277.08$).

Does the enhanced performance of female contestants in groups with more women explain the lower rate of premature dropouts among women in such circumstances? To answer this question, we included the proportion of correctly answered questions in the last round as an additional predictor in Model 5 (Table IV). While the effect of the new predictor was significant (coefficient of -1.28, $z=-2.26$, $p=0.02$, $\chi^2(10)=63.71$), its inclusion altered neither the effect of gender (coefficient of 1.77, $z=2.88$, $p=0.004$) nor the effect of the interaction between gender and the gender composition (coefficient of -2.93, $z=-2.21$, $p=0.03$). Thus, the higher rate of premature withdrawal among women when competing against more men is robust and cannot be explained by women’s lower performance in such circumstances.

How symmetric is the effect of gender ratio on women’s behavior? We have found that women who decided incorrectly to withdraw from the game did so more when the gender ratio was skewed toward men in their groups. Could a similar argument be made about women who incorrectly decided to stay in the game and consequentially were expelled? Does the probability of women being expelled increase with the proportion of women in the group? To answer this question, we focused on the data on expulsions and correct exits, and fitted a logistic model, similar to those presented above, with the dependent variable “expulsion.” The results are presented in Table IV – Model 6. Neither the gender ratio nor the interaction between gender and the gender ratio are significant in this model ($p=0.16$ and 0.46, respectively). Thus, the effect of group composition is not symmetric in our data and only applies to women
who perform relatively well and yet still decide to leave the game prematurely— and especially when competing against fewer women than men.

**Result 6:** *Female winners-to-be are as active in the interactions as male contestants.*

Our more qualitative data revealed that female contestants engaged less in strategic end-of-round interactions (women were “very vocal” in 43% of the interactions versus 59% for men; proportion test z=3.17, p<0.01). However, female winners-to-be were as active in the interactions as male contestants. In particular, female winners-to-be were especially active in 67% of the interactions of the regular games (versus 37% for female non-winners; proportion test z=-3.09, p<0.01), while male future winners were very active in 58% of the interactions (versus 60% for male non-winners; proportion test z=0.20, p=0.84).\(^\text{10}\)

This behavioral difference between female winners-to-be and non-winners contributed to the overall difference between winners-to-be and non-winners. In particular, across both genders, future winners were more active in the end-of-round interactions of the regular games. Winners-to-be talked “a lot” in 61% of the interactions compared to 49% for other contestants (proportion test z=-2.06, p<0.05).

As to persuading others to withdraw from the game, male contestants employed this strategy marginally more than their female counterparts – in 10% vs. 6% of the interactions. The difference, however, is not statistically significant (proportion test z=0.60, p=0.55). Within-gender comparisons showed that across rounds, winners-to-be did not use this strategy more than less successful contestants.

We next examined whether the contestants were similarly vocal in the groups differing by gender composition. We fitted logistic models analogous to those presented above with the dependent variable “very vocal” (0/1). First, we started with two

\(^{10}\)The difference between 67% for female winners-to-be and 59% for all male contestants is not statistically significant (proportion test z=0.92, p=0.36).
predictors revealed to be important by the analyses of proportions: gender (1 for female), “winner-to-be” (1 for eventual winners), and the interaction between the two predictors. Consistent with the results on proportions, the logistic model showed that overall women were less vocal than men (coefficient of -0.91, \( z=-4.15, p<0.001 \)), except for female winners-to-be (interaction of 1.02, \( z=2.82, p=0.05 \)), 860 observations, \( n=216 \), \( \chi^2(3)=21.35^{11} \). Second, we added the proportion of female contestants as an additional predictor to the model. The amended model did not affect significantly the coefficients of the three other predictors, and yielded goodness of fit that was not significantly different from the baseline model (\( \chi^2(4)=21.50, p=0.70 \) for the difference between the models). Similar separate analyses of how vocal female and male participants were consistently revealed across multiple specifications no significant effect of gender composition. Thus, the gender ratio did not affect the contestants’ propensity to be active in the post-round interactions.\(^{12}\)

VIII. General Discussion

In this paper, we present novel empirical evidence on gender differences in competitive environments. While previous research documented that women are less willing than men to enter competition (e.g., Niederle & Vesterlund, 2007), our results show that attitudes toward staying in competition differ between genders. Specifically,

\(^{11}\)The result is robust to the inclusion of additional predictors in the model, such as dummies for rounds and Sunday games, accumulated gains, and the proportion of correctly answered questions in the last round.

\(^{12}\)We also analyzed whether the contestants’ decisions to leave the game were affected by how vocal their competitors were. For this effect, for each group, round, and participant we coded the number of very vocal players, excluding self. In addition, we coded separately the number of very vocal men and women (again, excluding self). We analyzed the effect of these variables in the logistic models of exit similar to those presented in Table IV. Across various specifications, none of these variables had any predictive power to explain either correct or incorrect exits, or women’s or men’s decisions. Strategic posturing in this game was not persuasive enough to modify the behavior of competitors.
we find that, despite starting with approximately equal numbers, the proportion of women drops dramatically as the game advances.

Our results suggest that women may be less prevalent in some competitive environments because they exit prematurely. In our game, women withdraw voluntarily more from competition than equally skilled men. Moreover, these premature exits are exacerbated when the proportion of women in a group is low. By contrast, men’s decisions are not affected by the gender composition of the groups.

We also found a difference in performance. Women gave fewer correct answers than men and this, together with the premature withdrawals of skilled women, eliminated most women from later stages of the tournament. There are at least two possible explanations. First, women might experience more anxiety when performing in public (Larkin & Pines, 2003). Second, although we detected no overall gender bias in the questions, “stereotype threat” – or the risk of confirming a negative stereotype of one’s group (Steele & Aronson, 1995; Spencer, Steele, & Quinn, 1999) – could have led to poorer performance. Moreover, this threat is exacerbated when the distinctive characteristics of the group are salient. For example, women tend to do worse on math tests as the proportion of men in the group increases (Inzlicht & Ben Zeev, 2000). Future research should address the question of whether skilled women stay longer in competitive environments when the stereotype threat is reduced, for example, by self-affirmation (Martens, Johns, Greenberg, & Schimel, 2006).

At a qualitative level, winning contestants engaged in more strategic bantering (to persuade others to quit) and men did this more than women. In addition, while women engaged less in strategic end-of-round interactions, female winners-to-be were as active in the interactions as the males. This suggests that women who won regular
games became a self-selecting group who behaved more competitively and that “glass ceilings” are not inevitable.

One way to interpret the game is as a metaphor for executives climbing a corporate hierarchy where promotions depend on both relative ability and the willingness to compete with others. At the outset, the corporation recruits approximately the same number of male and female executives of equal ability. Periodically, these executives receive information about their performance and must decide whether this justifies continuing. In other words, the rounds of the game are similar to levels in a corporation where workers have to decide whether to withdraw from competing with their peers or to attempt to climb further. And, similar to corporations, if the “weak” don’t choose to leave, they are eliminated. At the same time, as in corporate life, the game allows for strategic posturing by competitors.

However, the game differs crucially in two respects from its corporate counterpart. First, there is no equivalent for gender-specific “constraints” such as child rearing for female participants. Second, since performance is assessed objectively in the game, there is no bias in evaluation. These differences are critical because it means that the glass ceiling we observed can be explained neither by women choosing alternative activities nor by assessors evaluating women less favorably than men.

It is tempting to suggest that our results simply reflect greater risk aversion on the part of women. However, although there is much evidence of such differences (Croson & Gneezy, 2009), these disappear when account is taken of profession, knowledge, and experience (Dwyer, Gilkeson, & List, 2002). For example, no gender differences in risk attitudes were found among entrepreneurs (Birley, 1989), fixed-income mutual fund managers (Atkinson, Baird, & Frye, 2003), managers of large companies (Adams & Funk, 2009), and people undergoing managerial education.
Self-selection and experience are the most likely explanations for these exceptions (Croson & Gneezy, 2009).

While differences in risk aversion might have played a role, we do not have measures of the contestants’ risk attitudes. However, gender differences in risk attitudes do not explain why women behaved in a less risk-averse manner in the groups involving more women. Nor do they explain why men were more active than women in interactions between competitors, and at the same time female-winners-to-be were as active as men. In our view, attributing the observed gender gap to greater risk aversion on the part of women is not satisfactory. It does not illuminate (a) why women behave as if more risk-averse in these circumstances, nor (b) what could be done to change behavior. For example, the fact that women exited less when there were more women present could perhaps be explained by similarity attraction or the notion that individuals are attracted to and prefer to interact with similar others (Byrne, 1971; Tolbert et al., 1999).

The observed gender gap is also consistent with “expectation states” theory. This argues that individuals use status beliefs to organize their interaction in goal-oriented settings (Berger et al., 1977). Thus, gender is salient in settings when it differentiates the actors – i.e., in a mixed-sex context. In such settings, gender status beliefs shape – often unconsciously – expectations for competence and behaviors in a self-fulfilling way (Ridgeway & Smith-Lovin, 1999). Even in the context of the game, free of gender differences in hierarchical roles and status relationships, implicit gender status beliefs that participants brought with them might have induced behavior consistent with their everyday experience thereby recreating the external gender system. Also consistent with expectation states theory is the idea that men and women could have different expectations in terms of earnings and that these would affect behavior.
That is, if women expected to earn less, they would be more motivated to leave the game earlier than men. However, we lack the data on expectations to test this hypothesis.

Expectation states theory further predicts that in mixed-sex settings with gender-neutral tasks, performance expectations triggered by gender status beliefs make men participate more, speak up more often, defend their views, and overall display more confident and assertive behavior (Ridgeway, 2001). Low-status members behave according to their status — taking a more reactive role (Wagner & Berger, 1997). Several studies have shown that *ceteris paribus* men talk more when performing gender-neutral tasks in mixed-sex groups (e.g., Dovidio, Brown, Heltman, Ellyson, & Keating, 1988) as was the case in our data.

Importantly, competence expectations based on gender status assumptions mediate men’s tendency to speak more and be more active overall (Wood & Karten, 1986). When competence expectations for men and women are the same, gender differences disappear. In our data, women who performed relatively well in the task (i.e., the future winners) were as active as male participants. Evidence of competence from performance feedback might have modified the gender status beliefs of these women, thereby reducing their adherence to stereotypical female/low-power behavior.

Drawing parallels between the structure of the game and that of a corporate hierarchy, our results imply that the effects of group membership (here: male and female) are cumulative in nature. Thus, any female disadvantage in being promoted to the top level (or winning the competition) has its origin in what happens early in the process. For example, in our data both higher rates of expulsion and voluntary exits for women at earlier stages of the game (i.e., at “lower levels” of the hierarchy) meant that, in the competition for the “top spots,” women were a minority.
This insight has important ramifications for promoting gender diversity in organizations. In particular, it’s no use bemoaning the lack of good female candidates for senior positions. Instead, it is important to commit resources to maintaining and promoting women at the lower and middle levels so that the ultimate pool of senior female candidates can be as large as that of men. Importantly, we find that women exit competitive environments less when more women participate in the contest. It is relative numbers that are important in shaping labor outcomes of disadvantaged groups. An illuminating task for future research, therefore, would be to simulate effects of different distributions of status and performance differences on gender ratios across organizational hierarchies. How does the initial distribution of talent interact with different factors to mitigate or promote glass ceilings (cf. Martell & Emrich, 1996)?

Our work represents an opportunistic attempt to illuminate an important social problem by capitalizing on a “natural experiment.” It is therefore appropriate to assess its strengths and weaknesses. On the positive side, we note, first, that all contestants volunteered to participate in what they knew was a competitive game with a strong public dimension. This is not dissimilar to starting work in a large corporation although, of course, the game involves even wider dissemination of performance data. Second, payoffs were substantial. Third, as noted above, both gender-specific constraints and discrimination can be ruled out as alternative explanations.

At the same time, the field nature of this natural experiment suggests care in interpreting results. First, there were small differences in ability (as measured by performance) between men and women in the regular games. Whereas we find – controlling statistically – that women still exited the game voluntarily at a higher rate than men, an improved experimental design would have involved equating men and women on a number of dimensions before the games even started. Since we are
measuring performance, however, controls should not concern just knowledge about the questions asked but also comfort levels in terms of performing in public.

Second, one can question attempts to generalize from performance on general knowledge questions. Clearly, this is a large inferential jump and, as is well known, jobs can have characteristics that are gender related or stereotypical (e.g., nursing). A further consideration is the lack of opportunities for learning to improve performance during the game which, once again, might be gender-related for different types of tasks in naturally occurring environments. Future research should also specifically target the types of interactions between men and women. Does this lead to accentuating or mitigating effects of gender status differences that exist in society?

All these disadvantages inherent in our research clearly define questions for further investigation using both more focused experimental designs and other field studies. Assuming, however, that our quasi-experimental results are valid, we believe that the critical question for future investigations lies in identifying specific strategies for changing the structure of organizations in order to modify existing gender status beliefs and thus prevent skilful women from imposing their own glass ceilings. To be effective, any affirmative action aimed at increasing the relative number of women in organizations must be combined with attention to how prevailing gender stereotypes and structural arrangements condition the behavior of all involved.
REFERENCES


Table I: Actual and expected number of withdrawals

<table>
<thead>
<tr>
<th>Round</th>
<th>Actual number of withdrawals*</th>
<th>Per round</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Total number of participants)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Regular games</td>
<td>Sunday games</td>
<td>The Final</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(216)</td>
<td>(36)</td>
<td>(6)</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>(180)</td>
<td>(30)</td>
<td>(5)</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(144)</td>
<td>(24)</td>
<td>(4)</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(108)</td>
<td>(18)</td>
<td>(3)</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(72)</td>
<td>(12)</td>
<td>(2)</td>
</tr>
<tr>
<td>Total</td>
<td>108</td>
<td>42</td>
<td>5</td>
</tr>
</tbody>
</table>

*Expected number of withdrawals

<table>
<thead>
<tr>
<th>Per round</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>6</td>
</tr>
<tr>
<td>180</td>
<td>30</td>
</tr>
</tbody>
</table>

*including all decisions to withdraw whether resulting in actual exits or not
Table II: Mean accumulated gains to winners-to-be and other players after different rounds in regular games

<table>
<thead>
<tr>
<th>Round</th>
<th>Wins Gains</th>
<th>Wins n</th>
<th>Others Gains</th>
<th>Others n</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular games</td>
<td>1</td>
<td>471</td>
<td>36</td>
<td>388</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>913</td>
<td>36</td>
<td>784</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1,506</td>
<td>36</td>
<td>1,227</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2,423</td>
<td>36</td>
<td>2,268</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3,631</td>
<td>36</td>
<td>3,244</td>
<td>36</td>
</tr>
</tbody>
</table>

Notes: z: Two-sample Wilcoxon rank-sum test
*** p<0.001; ** p<0.01; * p<0.05; + p<0.10.
Table III: Mean accumulated gains in regular and Sunday games

<table>
<thead>
<tr>
<th>Round</th>
<th>Men Gains</th>
<th>Men n</th>
<th>Women Gains</th>
<th>Women n</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular games</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>419</td>
<td>114</td>
<td>383</td>
<td>102</td>
<td>2.34*</td>
</tr>
<tr>
<td>2</td>
<td>842</td>
<td>98</td>
<td>772</td>
<td>82</td>
<td>2.22*</td>
</tr>
<tr>
<td>3</td>
<td>1,439</td>
<td>83</td>
<td>1,297</td>
<td>61</td>
<td>2.09*</td>
</tr>
<tr>
<td>4</td>
<td>2,394</td>
<td>71</td>
<td>2,177</td>
<td>37</td>
<td>1.70+</td>
</tr>
<tr>
<td>5</td>
<td>3,659</td>
<td>50</td>
<td>2,935</td>
<td>22</td>
<td>3.42***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sunday games</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8,481</td>
<td>25</td>
<td>8,339</td>
<td>11</td>
<td>0.96</td>
</tr>
<tr>
<td>2</td>
<td>9,414</td>
<td>23</td>
<td>9,568</td>
<td>7</td>
<td>-0.69</td>
</tr>
<tr>
<td>3</td>
<td>10,512</td>
<td>19</td>
<td>11,130</td>
<td>5</td>
<td>-1.35</td>
</tr>
<tr>
<td>4</td>
<td>12,045</td>
<td>14</td>
<td>11,950</td>
<td>4</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>14,650</td>
<td>10</td>
<td>14,825</td>
<td>2</td>
<td>-0.11</td>
</tr>
</tbody>
</table>

Notes:  
z: Two-sample Wilcoxon rank-sum test  
*** p<0.001; ** p<0.01; * p<0.05; + p<0.10.
### Table IV: Logistic models of voluntary exits.

<table>
<thead>
<tr>
<th>Model:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable:</td>
<td>Exit</td>
<td>Correct exit</td>
<td>Correct exit</td>
<td>Incorrect exit</td>
<td>Incorrect exit</td>
<td>Expulsion (vs. correct exit)</td>
</tr>
<tr>
<td>Independent variables:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-2.72***</td>
<td>-3.81***</td>
<td>-4.09***</td>
<td>-3.16***</td>
<td>-3.67***</td>
<td>2.16**</td>
</tr>
<tr>
<td>(0.25)</td>
<td>(0.46)</td>
<td>(0.59)</td>
<td>(0.27)</td>
<td>(0.45)</td>
<td>(0.67)</td>
<td></td>
</tr>
<tr>
<td>Female (0/1)</td>
<td>0.48*</td>
<td>0.13</td>
<td>-0.35</td>
<td>0.55*</td>
<td>1.80**</td>
<td>1.62+</td>
</tr>
<tr>
<td>(0.23)</td>
<td>(0.31)</td>
<td>(0.91)</td>
<td>(0.26)</td>
<td>(0.60)</td>
<td>(0.95)</td>
<td></td>
</tr>
<tr>
<td>Accumulated gains, in $1000</td>
<td>-0.88***</td>
<td>-1.25***</td>
<td>-1.30***</td>
<td>-0.35+</td>
<td>-0.32+</td>
<td>0.64**</td>
</tr>
<tr>
<td>(0.20)</td>
<td>(0.29)</td>
<td>(0.28)</td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.23)</td>
<td></td>
</tr>
<tr>
<td>Sunday game (0/1)</td>
<td>9.48***</td>
<td>12.24***</td>
<td>12.75***</td>
<td>4.52**</td>
<td>4.42**</td>
<td>-7.73***</td>
</tr>
<tr>
<td>(1.81)</td>
<td>(2.48)</td>
<td>(2.40)</td>
<td>(1.74)</td>
<td>(1.69)</td>
<td>(2.32)</td>
<td></td>
</tr>
<tr>
<td>Indicator variables for rounds 2-5</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Proportion of women</td>
<td>0.85</td>
<td>1.16</td>
<td>-1.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.95)</td>
<td>(0.88)</td>
<td>(1.25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of women*Female</td>
<td>0.42</td>
<td>-2.85*</td>
<td>-1.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1.69)</td>
<td>(1.29)</td>
<td>(1.96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model $\chi^2$</td>
<td>90.89</td>
<td>65.50</td>
<td>73.11</td>
<td>46.53</td>
<td>58.61</td>
<td>34.56</td>
</tr>
<tr>
<td>Number of observations</td>
<td>840</td>
<td>840</td>
<td>840</td>
<td>840</td>
<td>840</td>
<td>158</td>
</tr>
<tr>
<td>Number of participants</td>
<td>216</td>
<td>216</td>
<td>216</td>
<td>216</td>
<td>216</td>
<td>148</td>
</tr>
</tbody>
</table>

Notes: Population-averaged models were fitted. Standard errors, reported in parentheses, corrected for correlation across repeated observations on individuals.

*** p<0.001; ** p<0.01; * p<0.05; + p<0.10.
Table V: Number of withdrawals and expulsions by gender.

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winner of the big final game</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Expulsions</td>
<td>43</td>
<td>47</td>
<td>90</td>
</tr>
<tr>
<td>Decisions to exit:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>42</td>
<td>30</td>
<td>72</td>
</tr>
<tr>
<td>Incorrect</td>
<td>45</td>
<td>38</td>
<td>83</td>
</tr>
<tr>
<td>Total</td>
<td>87</td>
<td>68</td>
<td>155</td>
</tr>
<tr>
<td>Decisions to exit resulting in actual exits:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>37</td>
<td>23</td>
<td>60</td>
</tr>
<tr>
<td>Incorrect</td>
<td>33</td>
<td>32</td>
<td>65</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>55</td>
<td>125</td>
</tr>
<tr>
<td>Total number of contestants</td>
<td>114</td>
<td>102</td>
<td>216</td>
</tr>
</tbody>
</table>
Figure I: Numbers of men and women in the game
Figure II: Percentage of correct answers – by gender
Figure III: Percentage of contestants expelled from the game – by gender
Figure IV: Percentage of contestants who decided to withdraw from the game – by gender
Figure V: Predicted probabilities of withdrawal – by gender

Note: Plots are based on the results presented in Table IV (Model 1).
Figure VI: Predicted probabilities of incorrect decision to withdraw – by gender and group composition.

Note: Plots are based on the results presented in Table IV (Model 5). The probability of withdrawal includes the main effect of gender (0 for men and 1 for women), the intercept between gender and the proportion of women in the group, and the effect of accumulated gains, estimated at its mean level.