

Win at All Costs or Lose Gracefully in High-Stakes Competition? Gender Differences in Professional Tennis

Abstract

This paper examines line-call challenges by male and female professional tennis players in major tournaments around the world. In terms of utilization rates, we find that the genders behave similarly. Nevertheless, we do detect some intriguing gender differences in these challenges. First, male players' challenges are more likely to be provoked by those of their opponents. More importantly, at tiebreaks, females are more likely to reverse an umpire's unfavorable call, while males make relatively more unsuccessful challenges. Furthermore, we find that men are a lot more likely to make "embarrassing" line-call challenges at tiebreaks and offenses (i.e., when the shot lands at the opponent's side of the tennis court) than women. These significant gender differences suggest that women particularly diverge from men at crucial junctures of the match such as tiebreaks. Differences in factors such as risk-aversion, overconfidence, pride, shame and strategic signalling behavior might help us to explain these gender-difference findings in line-call challenges.

JEL Classification: J16

Keywords: Gender differences, professional tennis, high stakes, line-call challenges.

1 Introduction

Since the middle of the 20th century, women’s labor force participation rates in many developed countries have been converging rapidly with those of men. This period has also witnessed a remarkable narrowing of the gender pay gap, particularly since the 1970s. Despite these improvements in the labor market prospects, women still lag behind men significantly in average wages and salaries, even after accounting for differences in socioeconomic characteristics, such as education and experience.¹ In the upper tail of the skill distribution, many studies have verified that, in the U.S., women still constitute only a negligible percentage of the top executives in major corporate firms, politicians, public and government administrators, academicians (including tenured faculty at leading research institutions), conductors of philharmonic orchestras, and top surgeons.² Because a large proportion of the gender gap in these outcomes remained unexplained after accounting for socioeconomic characteristics, there is now a growing body of literature involving laboratory experiments which examine the extent to which the gender gap in pay and under-representation in high-profile jobs are due to innate or culturally/socially-determined gender differences in certain personality traits, such as risk aversion, overconfidence, and competitiveness.³

These studies have broadened researchers’ horizons regarding gender differences in fundamental traits and attitudes toward risk. One of the most prominent papers in this literature is by Niederle and Vesterlund (2007), who find no gender differences in performance on an arithmetic task under either a non-competitive piece rate compensation scheme or a competitive tournament scheme. However, the authors do find that men are significantly more likely than women to self-select into the competitive compensation scheme, with such a choice being unable to be explained by their performance either before or after the entry decision.⁴ Perhaps

¹According to the U.S. Census data, the labor force participation gap went down from 60% to slightly more than 10% (<http://stats.bls.gov/data>). Also, women’s average pay rates, which were about 55% of men’s earnings in the early 1970s, are still less than 80% of those of men (<http://www.census.gov/ftp/pub/hhes/income/histinc/histinctb.html>).

²Bertrand and Hallock (2010) note that only 2.5% of the highest paid executives at top corporate firms are women. A similar under-representation of women (3%) was also found among CEOs in Fortune 500 companies in 2009 (<http://money.cnn.com/magazines/fortune/fortune500/2009/womenceos>). Regarding politicians, public and government administrators, see Krook and Childs (2010). Women are also found to be underrepresented among university faculty in science and engineering (<http://www.nsf.gov/statistics/infbrief/nsf08308/nsf08308.pdf>), conductors (<http://www.infoplease.com/ipea/A0106174.html>), and top surgeons (<http://nymag.com/bestdoctors>).

³See, among others, Booth and Nolen (2012a,b); Croson and Gneezy (2009); Gneezy et al. (2009); Eckel and Grossman (2008); Dohmen et al. (2010); Gneezy and Rustichini (2004); and Gneezy et al. (2003).

⁴Campbell (2002) cites numerous psychology studies which suggest that men are more competitive than

the result from that paper that has been emphasized least in the literature is that “about 57 percent of the original gender effect can be accounted for by general differences in overconfidence and risk and feedback aversion while the residual ‘competitive’ component is 43 percent” (Niederle and Vesterlund, 2007, p. 1096). Thus, even a personal trait such as competitiveness may involve underlying gender differences in other personal traits such as confidence and risk aversion.

Despite these elaborate experimental results which have been obtained in an attempt to identify gender differences in various important personal traits, there are hardly any laboratory or field experiments which involve sufficiently large stakes and high levels of competition among their participants. Would these significant gender differences in personal traits which have been measured in experimental labs still persist in a real setting with large monetary rewards? In other words, is it possible that men and women would respond differently to competitive pressure which involves high stakes?

In considering this question, this paper adds to the existing literature by examining a particular gender-balanced competitive profession with high stakes, namely professional tennis. The particular activity which we consider within that profession is the line-call challenges made by highly-ranked male and female professional tennis players.

In the existing literature, there have been two particular papers utilizing data from the game of professional tennis which are closely related to the current study. With a motivation similar to ours, Paserman (2010) investigates gender differences by using stroke-by-stroke data from seven Grand Slam tournaments between 2006 and 2007. He specifically examines the probabilities of men and women hitting winning shots and making unforced errors. In terms of gender differences, he finds that men and women are similar in adopting a more conservative and less aggressive playing strategy. Nevertheless, he also finds that the odds of making an unforced error relative to hitting a winner fall for women, while remaining constant for men. On the other hand, the second paper, by Abramitzky et al. (2012), examines line-call challenges in the game of tennis, as in our paper, but with a different focus. Excluding female players from their study entirely, Abramitzky et al. (2012) find that male players are more likely to challenge when the stakes are higher and the option value of challenging is lower. They also find that the challenging behaviors of male players are consistent with their model’s optimal challenging

women by self-selection, as boys spend most of their time at competitive games and girls select activities that are less competitive and more cooperative (see also Brizendine, 2006).

strategy. Our work aims to complement these existing studies by examining both male and female players and focusing on their different behaviors in line-call challenge decisions.

In March 2006, a player line-call challenge system was introduced in leading high-stakes professional tennis tournaments around the world. Under the new system, called the *Hawk-Eye* system, players are allowed to appeal the line judges' calls which are upheld by the umpire (we will refer to such line calls as the umpire's calls unless otherwise noted). If the challenge is successful, a player is able to reverse the umpire's line call; if the challenge is unsuccessful, the player loses one of his limited number of line-call challenge opportunities for the set. Thus, challenging a call is a risk-taking behavior, since an incorrect challenge may deprive one of the opportunity to challenge a line call later in the set; in other words, the player loses the option value of a future line-call challenge.

In this context, tennis line-call challenges are risk-involving decisions taken in a very well-defined setup. However, they are unlike the risky behaviors which have been the focus of many recent experimental studies in the literature on risky choices and risk aversion. In lab and field experimental studies, probabilities are set explicitly (referred to as "objective probability lotteries" by Croson and Gneezy, 2009), and focus on estimating the curvature of a utility function. On the other hand, in the real world, probabilities should be evaluated subjectively (Tversky and Kahneman, 1974). Moreover, in reality, risk-taking individuals often incur some reputation costs and shame, especially when the consequences of such behaviors turn out to be negative.

Note that the stakes in professional tennis are extremely high relative to the levels of pay in laboratory experiments or the stakes involved in field experiments; for instance, in the final match at a typical ATP Masters (men's) tournament or a typical WTA Tier I (women's) tournament, which rarely lasts more than a couple of hours, the winner earns roughly a million dollars, while the runner-up earns half that amount. In addition, laboratory experiments typically employ college students who have not yet become specialized professionals, even though it has been found that professionals are behaviorally different from the general population (Palacios-Huerta and Volij, 2008).⁵

Not only is singles tennis a high-stake game, it also involves two highly-skilled professional

⁵Antonovics et al. (2009, p. 152) conclude that "care must be taken when translating results regarding gender and competition from low-stakes laboratory experiments to other environments".

players in a highly competitive setting.⁶ The professional tennis players who play at the highest levels are those who have also accumulated high-level skills in competitiveness and risk management. As a result, some of these players make more than \$35 million a year, including the prize money and sponsorship income.⁷ In addition, these players compete at high-prize-money tournaments frequently, and have ample opportunities to rectify their rookie mistakes over time.

While discrimination or anticipated discrimination may cause women to shy away from certain occupations, gender discrimination, in terms of pay,⁸ participation rates or the difficulty of time commitments, is likely to be less of an issue in tennis.⁹ Men's and women's tennis tournaments have the same draw sizes and match formats,¹⁰ and men and women receive the same amounts of prize money and similar sponsorship incomes. Consequently, the numbers of high-profile male and female players are similar.

Clearly, any gender differences in line-call challenges in professional tennis due to potential differences in personal traits such as confidence and risk aversion cannot explain differences in labor market participation or pay between men and women in professional tennis, as they can only explain within-gender pay differences. In other professions (where gender differences

⁶These players have typically spent their childhood and youth in honing their physical, technical and competitive skills with the aim of becoming professional players (even junior players typically practice more than four hours a day by attending high-caliber tennis academies, resorting to home-schooling during their mandatory schooling years in order to achieve their high aspirations regarding playing professional tennis). In the U.S. alone, in ten age groups for each gender, tens of thousands of competitive junior tennis players compete every week in more than 15 sections and 100 regions; nationally ranked players travel from tournament to tournament almost every week, all over the country. Only a few of these players will ever make the professional ranking, and only a fraction will last more than a few years. Those who remain in professional tennis are not only the most skilled ones physically and technically, they are also the strongest mentally, and are trained to make very quick decisions involving risk (an attribute which they share, not only with other sports professionals, but also with many surgeons, CEOs, stock market brokers, and active-duty police officers).

⁷See the Forbes Celebrity Income rankings: the top tennis players, both male and female, are among the highest earning celebrities in the world.

⁸At any combined tournament where both males and females take part they get the same amount of prize money for the level that they progress throughout the tournament. Obviously, the overall career prize might be different between males and females depending on factors such as the number of years they have been playing as professional players (hence, the age distribution between males and females) and whether they also play doubles or not.

⁹It is still true, however, that female players may face additional costs in starting a family, and may have career interruptions due to responsibilities of child bearing and rearing, unless they delay these until the end of their tennis career. Moreover, the level of competition may be somewhat different in men's and women's matches: as will be elaborated later, Table 1 reveals that there are more three-set matches, upsets and tiebreaks per set in men's competitions.

¹⁰Professional singles tennis involves groups of two women or two men competing against each other, as per Niederle and Vesterlund (2007, p. 1068), for instance, where "groups of two women and two men perform a real task". As a matter of fact, there is no Olympic-level or professional sport in which a woman competes against a man individually.

in such personal traits can cause labor market participation or pay differences), the overall extent of the differences may also involve convoluted and indirect effects. For example, less risk-averse male medical residents may specialize in conducting risky but well-paying spinal-cord surgeries, whereas very risk-averse female medical residents may choose to specialize in conducting virtually risk-free but low-paying cataract eye surgeries. Thus, any personal-trait gender differences that can potentially cause gender differences in line-call challenges in professional tennis will only have stand-alone effects, without any such convoluted indirect effects. As such, the measurement of such pure stand-alone gender differences via professional tennis line-call challenges data will be a novel contribution of this paper.

In order to establish a sound background for our empirical analysis, we begin by using a simple theoretical framework under uncertainty which considers an individual line-call challenge as a risky choice according to a threshold probability. Our model establishes that players' line-call challenge utilization and challenge success rates must be inversely related for all rational players, whether male or female. As we will discuss in detail, our empirical results using data from 480 ATP Masters and WTA Tier I matches around the world, and 1,484 individual line-call challenges in these matches, confirm this theoretical prediction to a large extent.

We find that the genders are very similar in terms of line-call challenge utilization rates. This is not very surprising, given that these players are highly-skilled professionals participating in high-stake sports. More specifically, the utilization rates of both male and female players are significantly higher the higher the total number of games played in the set, and increase significantly with the proportion of games won by the opponent. However, the latter finding is more robust for males than females. Utilization rates are also significantly higher in the third and second sets than in the first set. Although these relationships seem stronger for male players, the gender differences are not statistically significant. While a higher self-ranking tends to increase male utilization rates significantly, female utilization rates decrease significantly as the opponent player's ranking increases. Nevertheless, these gender differences in utilization rates again turn out to be statistically negligible.

In terms of success rate results, the major difference between males and females appears in the case of a tiebreak. At tiebreaks, females are more likely to make a correct challenge while males are more likely to make an incorrect challenge, and this gender difference is statistically significant. The results are similar in the case of a challenge made in an offense (i.e., when the

shot lands at the opponent’s side of the tennis court), but the difference is less significant.

Our further analysis indicates that men are a lot more likely to make “embarrassing” line-call challenges at tiebreaks and offenses than women. Moreover, a higher-ranked male player is more likely to make an embarrassing challenge, while a higher-ranked female is less likely to make an embarrassing challenge. Again, this difference is also statistically significant. Finally, while the propensity for making embarrassing challenges increases with the opponent’s ranking in male games, we do not observe any such statistically significant correlation in female games. Along with the previously mentioned results, this finding might be interpreted as indicating that men’s line-call challenges are more likely to be provoked by their opponents.

Given these findings, it might be fair to say that, at least when it really matters the most (i.e., at tiebreaks), male players try to win at all costs, while female players accept losing more gracefully when they are losing the match, rather than making embarrassing line-call challenges. In a sense, this result extends the Niederle-Vesterlund finding to a domain in which neither men nor women choose to shy away from competition: even if it involves very high stakes, women are less likely to try to win the competition at all costs than men. Gender differences in factors such as overconfidence, pride, shame and strategic signalling behavior turn out to be instrumental in providing plausible explanations for our main gender-difference findings in line-call challenges.

The remainder of the paper proceeds as follows. Section 2 develops a simple theoretical model, from which we derive some empirical implications. Section 3 presents our empirical methodology and describes the data. Section 4 presents the results. In Section 5, we provide some plausible interpretations for our few but significant gender-difference findings. Section 6 concludes.

2 A Simple Model and Empirical Implications

2.1 Individual Challenge Decisions

In a singles tennis tournament, two players play a match in a round, and the winner advances to the next round (or wins the tournament, if that is the final match). In a match, the prize for the winner is $W > 0$, that for the loser is $L \geq 0$, and $W > L$.¹¹ A player’s utility function is $U(\cdot)$, and the player (‘he’) maximizes the expected utility (EU):

¹¹ W is the minimum prize guaranteed at the next round plus the option value of advancing to the higher rounds. L is simply the predetermined prize at the current round.

$$EU = \pi U(W) + (1 - \pi)U(L), \tag{1}$$

where π is the probability of winning the match, given the umpire’s current call.

Under the player line-call challenge system, after a play, if the umpire’s call is not in his favor, the player decides whether or not to challenge the call. If he challenges the line call, the umpire resorts to a review of the spot where the ball landed, as recorded by a set of overhead cameras, called the Hawk-Eye system.¹² The animated trajectory of the ball is also shown to viewers, so that they find out, not only whether the line-call challenge (simply “challenge” hereafter) is successful, but also exactly where the ball landed.¹³ Since the player is given a limited number of chances to utilize the system (two unsuccessful challenges per set without a tiebreak), challenging is a risk-taking behavior. Furthermore, it is a risk-taking behavior which may potentially involve shame or embarrassment, since it occurs in front of a large number of viewers.¹⁴

The player should be assessing the probability of winning the match, conditional on the outcome of the review. Let $\pi^H \in [0, 1]$ denote the match-winning probability and $(1 - \pi^H)$ the match-losing probability when the umpire’s call is reversed. Also, let $\pi^L \in [0, 1]$ denote the match-winning probability and $(1 - \pi^L)$ the match-losing probability when the player fails to reverse the call, and consequently, loses one (perhaps the last) opportunity to challenge. In most cases, $\pi^L \leq \pi \leq \pi^H$, although π^L could be greater than π if an unsuccessful challenge stalls the momentum of the opponent or gives a breather to the player who challenges the call.

We assume that the umpire behaves like a non-strategic robot with error. In other words, for a player in a given match, the probability with which the umpire makes a wrong call, p , is drawn randomly from a distribution $f(\cdot)$. We assume that the distribution may depend on factors such as the player’s average ball speed, but it does not vary according to the match’s trajectory, as could be relevant if a particular player was losing the match badly and the umpire was trying to make the match more competitive by favoring that player with his calls.

¹²The Hawk-Eye technology is not used in tournaments played on clay, where a ball leaves a mark on the court and its landing location can easily be verified by the umpire.

¹³As was noted by Abramitzky et al. (2012), the Hawk-Eye verdict is a statistical prediction, and lab testing indicates a mean prediction error of 3.6 mm. Given the small magnitude of a potential error and lack of availability of any better option, the Hawk-Eye decision is final, and is treated as the ‘truth’.

¹⁴“As the animated ball retraces the shot trajectory, all eyes transfix on the screen to see if it was in or out, and by how much” (New York Times, September 10, 2010). In a recent experiment, Charness and Rustichini (2011) found that the degree of cooperative behavior changes depending upon the type of audience. The same may hold for risk-taking behavior as well.

Under this setup, the player will make a challenge if and only if:

$$\begin{aligned} & p\{\pi^H U(W) + (1 - \pi^H)U(L) + b\} + (1 - p)\{\pi^L U(W) + (1 - \pi^L)U(L) - c\} \\ & \geq \pi U(W) + (1 - \pi)U(L). \end{aligned}$$

Note that we have a disutility of $c \geq 0$ associated with an unsuccessful challenge; likewise, we have an additional utility of $b \geq 0$ associated with a successful challenge. Part of b can be viewed as the “pride” resulting from a successful challenge, and part of c as the “shame” following an unsuccessful challenge.¹⁵ The other part of b can be viewed as the benefit of either stalling the opponent’s momentum (and thus irritating the opponent) or obtaining a breather for the player who challenges the call. Likewise, the other part of c can be viewed as the cost of stopping his own momentum.

Normalizing $U(L)$ to 0 without loss of generality, we find that there exists a threshold probability, p^* :

$$p^* = \frac{(\pi - \pi^L)U(W) + c}{((\pi^H - \pi) + (\pi - \pi^L))U(W) + b + c}. \quad (2)$$

This threshold depends upon five factors: (i) $U(W)$, the utility of winning the match; (ii) $(\pi^H - \pi)$, the expected increase in the match-winning probability if the umpire’s call is reversed; (iii) $(\pi - \pi^L)$, the expected increase in the match-losing probability if the umpire’s call is challenged but not reversed, and thus the option value of another challenge opportunity is lost;¹⁶ (iv) the disutility from an unsuccessful challenge, c ; and (v) the extra utility from a successful challenge, b . Note that $(\pi^H - \pi)$ and $(\pi - \pi^L)$ pertain to one’s overconfidence, which can be defined as the tendency to overestimate the chance of obtaining a favorable outcome — and to underestimate the chance of obtaining an unfavorable outcome — as a result of a presumptuous belief in one’s abilities.¹⁷ In this case, the favorable outcome is winning

¹⁵Note that part of b corresponds to Atkinson’s (1957, p. 360) “pride in accomplishment” and part of c to his “shame upon failure.” It is well established in the Psychology literature that failure — especially in public — is embarrassing. “For example, sporting performances typically take place in the public arena and, as such, mistakes and failures expose athletes (and possibly their parents) to others’ judgment and negative evaluation and to feeling shame” (Sagar and Lavalley, 2010, p. 178).

¹⁶In that sense, players’ decisions to save their challenges for more crucial moments towards the end can be interpreted as a form of precautionary saving.

¹⁷Many researchers in economics and psychology have used the terms ‘optimism’ and ‘overconfidence’ interchangeably, despite their fundamental difference. Optimism can be defined as the tendency to overrate the chance of a favorable outcome and to underrate the chance of an unfavorable outcome. Both behaviors are very prevalent: “Dozens of studies show that people generally overrate the chance of good events, underrate the chance of bad events and are generally overconfident about their relative skills or prospects. For example, 90 percent of American drivers in one study thought they ranked in the top half of their demographic group in driving skill” (Camerer, 1997).

the match and the unfavorable outcome is losing the match; the ability in which the players presumptuously believe is their ability to make a successful challenge. Thus, an overconfident player will overestimate $(\pi^H - \pi)$ and underestimate $(\pi - \pi^L)$, due to his presumptuous belief in his ability to make a correct challenge.

Except for $U(W)$, which is constant for a player within a match, the remaining four factors (i.e., $(\pi^H - \pi)$, $(\pi - \pi^L)$, b , and c) may vary over the course of the match. They should be evaluated subjectively based on the situations within the match.

2.2 Challenge Utilization and Success Rates

In our simple model, the individual player’s decision-making process is summarized simply by the threshold probability. However, one empirical difficulty is that the threshold probability is not observable. Therefore, in our empirical analysis, we will examine two aggregate statistics of challenging behavior which are informative about the threshold probability.

First, we examine the *challenge utilization rate* per set. Each player is allowed to use as many challenges as possible in each set until they have accumulated two “wrong” challenges (plus a third challenge when the set score is 6-6 and the set reaches a tiebreak). Overall, we observe more challenges from more accurate or luckier players. In order to disentangle a player’s intention to utilize his available challenges from his ability to detect the umpire’s wrong calls (or his luck), we top-code the number of challenges to two (or three for a tiebreak set) in each set. Then, we obtain the challenge utilization rate by dividing the top-coded number of challenges by the number of challenge opportunities available in the set. If p^* is the threshold probability in the set, then the challenge utilization rate for the set is as follows:

$$r \equiv \min\left\{n \int_{p^*}^1 f(p)dp, 2 + TB\right\} / (2 + TB), \quad (3)$$

where n denotes the total number of plays within the set in which the umpire’s judgements are against the player, and TB is one if there is a tiebreak in the set and zero otherwise. Note that the utilization rate depends not only on p^* , but also on n , TB , and f . After removing the effects of n , TB and f , the remaining variation in the challenge utilization rates across players and matches should be attributable to differences in p^* . Furthermore, as can be seen from Equation (2), after accounting for differences in $U(W)$ across players and matches, the variation in p^* is attributable to differences in the factors, π , π^H , π^L , c , and b .

Second, we examine the probability of the player successfully reversing the umpire's call, i.e., the *challenge success rate*:

$$s \equiv E(p|p \geq p^*) = \int_{p^*}^1 pf(p|p \geq p^*)dp. \quad (4)$$

The success rate depends upon p^* and $f(\cdot)$. As with the utilization rate, after accounting for differences in f and $U(W)$ across players and matches, the variation in p^* should be attributed to differences in the factors, π , π^H , π^L , c , and b .

Our empirical analysis in the next section is based on the simple model. The parameter in which we are chiefly interested is the threshold probability, p^* , since it can indicate how players assess their winning probabilities, as well as the disutility from failure in a challenge and the utility from success. Ideally, we want to compare the thresholds of men and women. That is, we want to test whether $E(p^*|Male, X) = E(p^*|Female, X)$, where X is the vector of control variables representing players' characteristics and the contexts in which individual challenges occur. However, it is not possible to identify the threshold per se, and therefore, we choose some observable variables that are expected to influence the threshold. For instance, we hypothesize that a certain variable x affects the threshold probability. Suppose that we estimate the impacts of x on r (utilization rate) and s (success rate), i.e., $\frac{\partial r}{\partial x}$ and $\frac{\partial s}{\partial x}$, for each gender separately. Because $\frac{\partial r}{\partial p^*} \leq 0$ and $\frac{\partial s}{\partial p^*} \geq 0$, we have the following result:

$$sign\left\{\frac{\partial p^*}{\partial x}\right\} = sign\left\{-\frac{\partial r}{\partial x}\right\} = sign\left\{\frac{\partial s}{\partial x}\right\}. \quad (5)$$

This result means that we can test gender differences in terms of the impact of x on the threshold. We can infer that there is a gender difference in the threshold if we find that, for example, a variable increases the challenge success rate for women but not for men. Generally, if the signs of the marginal effects differ between men and women, this at least implies that the threshold is determined differently by gender.

In our approach, we are able, not only to examine how x influences the threshold probability, but also to check the validity of the model by examining whether $\frac{\partial r}{\partial x}$ and $\frac{\partial s}{\partial x}$ have opposite signs for males and females.

However, one significant limitation of our approach is that that we cannot identify the exact channels through which x affects the threshold probability. For example, an increase in the threshold probability due to a decrease in $(\pi^H - \pi)$ is observationally equivalent to that due to an increase in c (and likewise, an increase in $(\pi^H - \pi)$ is observationally equivalent to that

due to an increase in b , etc.). Still, we believe that our findings will teach us how individuals make risky choices in competitive environments with extremely high stakes. We will explore possible explanations in Section 5.

3 Estimation and Data

3.1 Estimation Methods

The main empirical implication of our model in the previous section is that we can infer the determinants of the threshold probability from an examination of the determinants of the challenge utilization and success rates. This suggests that we should estimate the equations determining the utilization and success rates.

First, we estimate a Tobit model with two-sided censoring for the challenge utilization rate (corresponding to r in Equation (3)) for each gender, separately.

$$r_{ism}^* = \beta_0 + \beta_1 \tilde{n} + \beta_2 X_{ism} + \alpha_{im} + U_{ism} \quad (6)$$

$$r_{ism} = \begin{cases} 0 & \text{if } r_{ism}^* < 0 \\ r_{ism}^* & \text{if } 0 \leq r_{ism}^* \leq 1 \\ 1 & \text{if } 1 < r_{ism}^* \end{cases}$$

where the subscripts i , s and m represent the individual player i , set s and match m . We interpret r_{ism}^* as the true utilization rate, representing the player's true willingness to utilize challenges without any constraint, while r_{ism} is the observed utilization rate. Note that the utilization rate is defined at the set level, since the frequency of challenges is limited by the set. The fixed-effect Tobit (FE-Tobit) model is estimated using the estimator developed by Honoré (1992), who extended Powell's (1986) idea of a symmetrically trimmed least squares estimator. In terms of a robustness check, we also estimate random effect Tobit (RE-Tobit) models.

Based on the results from the previous section, we include not only the total number of games played in the set, but also the proportion of games won by the opponent in the set. Intuitively, one could imagine the utilization rates in a particular set being positively correlated with the total number of games played in that set. The 'relative number of games won by the opponent' variable proxies the number of the umpire's calls that have not been in favor of the player, n in Equation (3). The vector X_{ism} includes x s that we hypothesize will affect the threshold probability: the logarithm of the player's self-ATP ranking, the logarithm of the

opponent’s ATP ranking,¹⁸ a dummy variable which equals one if the player won the first set and the current set is the second set, a dummy variable which equals one if the player lost the first set and the current set is the second set, a dummy variable which equals one if the current set is the third set, and a dummy variable which equals one if there is a tiebreak in the current set. In addition, taking 2006 as the base year, two year dummies for 2007 and 2008 are included in order to capture any time trend in the utilization of challenges.

Lastly, the variable α_{im} is the player-by-match fixed effect (FE). Recall that we want to infer how a certain variable, representing the player’s or the match’s characteristics, affects p^* . However, as can be seen in Equation (2), p^* also depends on $U(W)$, which is fixed for a given player in a particular match. We remove this confounding effect by including the player-by-match FE. It is also our key identification assumption for estimating $\frac{\partial r}{\partial x}$ consistently in Equation (5) (and later $\frac{\partial s}{\partial x}$) that the distribution of the probability in which the umpire makes a wrong judgement is constant within a match for a player. This assumption is valid unless the umpire is affected by the situational factors of a match.

Note that the player-by-match fixed effect should absorb any umpire-specific effects. Also, the fixed effect should capture any individual-player-specific effects. For the sake of robustness, we tried to include the match- and player-specific FEs separately, but the results were similar.

For the success rate, we use challenge-level data, and estimate the following binary dependent variable model separately for each gender:

$$s_{kism} = 1\{\gamma_0 + \gamma_1 Z_{kism} + \delta_{im} + V_{kism} > 0\}, \quad (7)$$

where s_{kism} indicates whether the k th challenge by individual i in set s of match m is successful. Thus, the expected value of the dependent variable for a player in a match corresponds to s in Equation (4).

As in the case of utilization estimations, the vector of control variables, Z_{kism} , includes the logarithms of both player’s rankings, together with some general characteristics of the set in which the particular challenge was made: a dummy variable which equals one if the player won the first set and the current set is the second set, a dummy variable which equals one if the player lost the first set and the current set is the second set, and a dummy variable which

¹⁸Following the same course as Abramitzky et al. (2012), we took the natural logarithms of players’ rankings when controlling for differences in abilities or in the degree of competition in the match. However, our results remain unchanged when we include the rankings linearly, without taking their natural logarithms. Please note that we can only include these ranking measures in the random effects estimations, not in fixed effects ones, as they do not vary in a given match.

equals one if the current set is the third set. For the most part, the challenge-specific control variables follow the spirit of Abramitzky et al. (2012), and include the values of the following when the challenge was made: total number of games played, proportion of games won by the opponent, a dummy variable indicating whether the point ended naturally or the player stopped during the play, a dummy variable indicating whether the game score was a deuce, a dummy variable indicating whether the game score was 40-0 or 0-40, time elapsed since the start of the match, a dummy variable indicating a tiebreak game, an indicator of whether the shot landed on the opponent’s side of the tennis court, and an indicator of whether the challenge was the second one in the set.¹⁹ In addition, where appropriate, two year dummies for 2007 and 2008 are included in order to capture any time trend in the success rates.

Like α_{im} in the utilization rate equation, the variable δ_{im} is the player-by-match FE. As in the utilization rate equation, we interpret the coefficients for the variables in Z_{kism} as the effects of c , π^H and π^L on the threshold probability, under the assumption that the FE plays a role of controlling for $U(W)$ and f . Due to the presence of the fixed effect, we use a fixed effect Logit model (FE-Logit) to estimate Equation (4), under the assumption that V_{kism} follows the logistic distribution. Again, we also estimate the random effect Logit (RE-Logit) models for success rates in terms of robustness checks.

3.2 Data and Descriptive Statistics

In March 2006, starting with Miami’s Sony Ericsson Open, the professional men’s and women’s tennis bodies, the Association of Tennis Professionals (ATP) and the Women’s Tennis Association (WTA), decided to give players a chance to challenge some calls by the chair umpires (see Appendix A for more details on the challenge system). We focus mainly on men’s Masters’ and women’s Tier I events, since these tournaments involve both men and women playing matches which are the best (two) out of three sets.²⁰ Our data set covers all such tournaments in 2006–2008 for which the ATP and WTA had challenge information. Some women’s tournaments, especially those in 2006, did not use the challenge system or keep compiled detailed challenge

¹⁹Abramitzky et al. (2012) use the sort of information mentioned above to construct two proxy measures for the importance and option value of a specific challenge, and focus mainly on the relationship between these two measures and the probability of the challenge being successful. Rather than taking such an approach, we simply look at the direct relationships between our control variables and the challenge success rates, and examine the gender differences in those direct relationships.

²⁰In Grand Slam tournaments, on the other hand, men play matches which are the best (three) out of five sets, while women play matches which are the best (two) out of three sets. Thus, including Grand Slam events would prevent direct comparisons between men and women.

records; thus, we were not able to include those events in our analysis. As a result, most of the Tier I women's tournaments in 2006 were not included in our analysis.

The final data set we use includes 331 men's matches and 149 women's matches. The data are constructed from the challenge review reports provided directly to the authors by ATP and WTA. Figure 1 presents an example report. We had to drop some tournaments because some crucial variables, such as scores, had not been recorded in the relevant reports. We also dropped four matches from which players retired, so that the matches were not completed.

Table 1 presents summary statistics of some match characteristics. Overall, it seems that men's and women's matches are comparable. However, we find that men's matches are slightly more competitive than those of women. Men's matches are more likely to go to the third and deciding set, more likely to be an upset, and tend to have more tiebreaks. Also, the probability of winning the match conditional on winning the first set is slightly lower for men.²¹

Another interesting difference is found in the correlation between the numbers of challenges between two players. The correlation is significantly positive in men's matches (0.16), while it is smaller and statistically insignificant in women's matches (0.05). Two players' challenges may be correlated because they are judged by the same umpire. However, if this were the case, there would be no reason why the correlation would only exist for men. One possible explanation of the gender difference is that, for male players, challenges are sometimes a response to the opponent's challenges.

For both men and women, the average number of challenges per set is significantly below two: 0.31 for men and 0.33 for women. That is, neither male nor female players utilized all available challenges in a set fully.²² This is somewhat surprising, since the opportunity cost of using a challenge drops considerably as a set approaches its end, finally becoming virtually zero at the last play that ends the set. Thus, low utilization rates suggest that there is a stigma associated with a wrong or unnecessary challenge, and/or that there is a strong incentive to save a challenge for a critical moment, which is likely to occur at the end of a set.²³

²¹As an additional exercise, we create a measure of the competitive imbalance as the sum of the squared between-player differences in the number of games won within sets, normalized by the number of sets. According to this measure, we find that men's matches are more competitively balanced than women's, a difference which is statistically significant.

²²Abramitzky et al. (2012) argue that, even though there is a big difference between the utilization rates and the number of available challenges, male players' actual behaviors are close to the optimal challenging strategy described by their model. However, they do not consider female players' challenging strategy.

²³As we will show later, utilization rates are significantly higher for both male and female players in the third set than in the first or second sets. Players' decisions to save some of their challenges for more crucial moments

The average success rate of challenges is somewhat higher for men: men’s success rate is about 40% while women’s is about 33%, but they are not statistically different. The average ball-line distance is similar for both genders, about 37 mm.

4 Estimation Results

4.1 Challenge Utilization and Success Rates

We present the utilization rate estimation results in Table 2. We estimate the determinants of the utilization rates of male and female players separately using both the RE and FE approaches. For each approach, we also estimated the utilization rate equation using the pooled sample of male and female players with all explanatory variables interacted with the gender dummy, in terms of estimating the coefficient differences across genders and their standard errors. These differences and corresponding standard errors for the RE and FE methods are presented in columns (3) and (6), respectively. The success rate estimation results are presented in Table 3 in a similar fashion.

Overall, the results seem to support the validity of our model, where an individual challenging decision is described as a threshold strategy. We find that when the challenge utilization rate is higher, the success rate tends to be lower, and vice versa. Although there are a few exceptions, the sign of $\frac{\partial r}{\partial x}$ turns out to be opposite to that of $\frac{\partial s}{\partial x}$ in most cases, especially for male players.

The results regarding the utilization rate in columns (1) and (2) of Table 2 are quite similar for male and female players. First, as the number of total games played in a set increases, challenge utilization rates increase significantly for both genders in both the RE and FE models. However, this robust positive correlation is clearly intuitive and not surprising: the larger the number of games played, the greater the chance of a player challenging an umpire call. The results also indicate that, keeping all else constant, particularly the total number of games played, a player tends to use more challenges as the opponent wins relatively more games in a set. Again, this association is not surprising if we consider the variable as a proxy for the number of unfavorable judgements. Equation (3) showed that, the more judgements there are against a player, the more likely the player is to challenge. Nevertheless, this finding is more robust for male players than for females, because the estimated coefficient is significant around the end of a set can be interpreted as a form of precautionary saving.

in both the RE and FE models for males, while it is only significant in the RE model for females. Utilization rates are also significantly higher in the second and third sets than in the first set.²⁴ More specifically, it turns out that a player who has lost the first set will tend to use more challenges in the second set. This effect is significant for men in both the RE and FE models, and for women in the RE model. This is in contrast to when the player has won the first set, in which case there is no significant change in the utilization rate of the second set, except for the RE model of male players. Regardless of their gender, all players challenge more during the third and deciding set. These relationships are robust in both the FE and RE models, though they seem stronger for male players. In addition, RE estimations suggest that, while a higher self-ranking tends to increase male utilization rates significantly, female utilization rates decrease significantly as the opponent player's ranking increases.

As was mentioned above, columns (3) and (6) of Table 2 represent the coefficient differences between male and female players and the standard errors of those differences for RE and FE models, respectively. Given the similarities between male and female challenging strategies, as suggested by separate estimation results, it is not surprising that the coefficient differences for all variables turn out to be statistically insignificant in both the RE and FE estimation results. This exercise also confirms the similarity of the challenging behaviors of male and female players in a statistical sense.

Next, we turn our attention to the success rate results presented in Table 3. Note that our set of explanatory variables is richer in the success rate estimations, as we are able to control for challenge-specific variables by using individual challenge data. The findings indicate that a challenge made closer to the end of a set is less likely to be successful. This is suggested by the negative coefficient estimated on the total number of games played in the set up to the time of the challenge. The coefficients are significant for both male and female players in the FE model, but turn out to be insignificant in the RE one, although they have the same negative sign. Looking at columns (3) and (6), however, suggests that there is not a statistically significant gender difference with respect to this correlation between the total number of games played at the time of a challenge and its success probability. We also notice from Table 3 that a female player is more likely to make an unsuccessful challenge when the opponent has a

²⁴These findings are in line with those of Abramitzky et al. (2012), who found that players are more likely to challenge when the point is more important or the stakes are higher. One can easily imagine that sets that are closer to the end of a match or points that are closer to the end of a set are more likely to be important.

relatively higher ranking, but we do not observe any such significant relationship for male games. Another factor which seems to be important in explaining challenge success rates is the type of challenge; that is, whether the challenge point ended naturally or the player stopped during the play turns out to be significant, particularly for males' success rates. A player's challenge is more likely to be unsuccessful when the point ends naturally than when the player stopped. Again, the gender difference with respect to this factor is not significant in either the RE or FE models. In terms of success rate results, the major difference between males and females appears in the case of a tiebreak. At tiebreaks, females are more likely to make a correct challenge, while males are more likely to make an incorrect challenge, and this gender difference is statistically significant. The results are similar in the case of a challenge made in an offense, but the difference is less significant. These findings are particularly noticeable in the RE model. The FE model still suggests the same conclusions in terms of their direction, but they are less apparent in a statistical sense. Finally, a second challenge is more likely to be successful than the first challenge for both male and female players, especially in the FE models. This makes sense intuitively, as a player would be expected to be more careful in utilizing his second challenge if it were potentially his last.

4.2 Embarrassing Challenges

In this subsection, we focus on somewhat extreme challenges. If a challenge is wrong by more than 50 mm, we call it an "embarrassing challenge" here.²⁵ Figure 2 is the histogram of ball-line distances. For wrong challenges, the distances are given a negative sign. It is quite surprising that some wrong challenges were made when the distance was quite long, for example, more than 200 mm. In fact, it seems that wrong challenges have larger distances. The histogram looks asymmetric and slightly skewed to the left, as more than 60% of the overall challenges are unsuccessful. This suggests that, although many of the embarrassing challenges are simply players' unintentional mistakes, some players may be exploiting challenges strategically. Comparing the frequencies of making such bad calls by gender would be interesting for the purpose of our paper. Certainly a player is more likely to make such a challenge when he or she is far away from the ball. However, it is also possible that players may make such bad calls deliberately, in spite of the fact that the success rate is low, in order to slow down the opponent's

²⁵The critical distance for embarrassing challenges is as big as the diameter of the tennis ball itself, and is chosen arbitrarily. The results below are robust to the choice of the critical distance.

momentum or to obtain rest.²⁶ Either way, these challenges may involve some strategic behavior in terms of increasing match winning probabilities. Lastly, embarrassing challenges might be a consequence of emotional reactions to some intensively competitive high-stake moments.

A descriptive look at the data suggests that men and women have the same ratio of embarrassing challenges: of each gender’s challenges, 17% are embarrassing challenges.²⁷ However, in tiebreaks, i.e., when the set is at stake, men make many more of them (34%), while women make many fewer of them (9%). Men also tend to make more embarrassing challenges when the ball lands on the other side of the net, i.e., when it is rather risky to challenge since the player’s view is impeded.

The estimation results for embarrassing challenges are presented in Table 4. Consistent with the descriptive statistics, when controlling for other factors in a regression setting, men are still a lot more likely to make embarrassing line-call challenges at tiebreaks and offenses than women. Moreover, the gender differences in regard to these variables turn out to be statistically significant, particularly in RE estimations. From Table 4, we also note that a higher-ranked male player is more likely to make an embarrassing challenge, while a higher-ranked female is less likely to make an embarrassing challenge. This gender difference in challenging behavior is again statistically significant. Finally, while the propensity to make an embarrassing challenge increases with the opponent’s ranking in male games, we do not observe such a statistically significant correlation in female games. This finding also supports the previously mentioned idea that men’s line-call challenges are more likely to be provoked by their opponents.

5 Interpretations of Gender Differences

Our empirical analysis above provides some evidence in support of our simple model that describes players’ challenges as choices made according to a threshold probability. As we mentioned earlier, these choices involve a risk-taking behavior, because each player has a limited number of challenge rights in a given set. As one can imagine, it might be crucial to

²⁶Abramitzky et al. (2012) argue that this aspect of challenging behavior is unlikely to be common driver of observed challenges for men, as the distance distribution would be a lot more asymmetric if such were the case, with a much fatter left tail. Nevertheless, given that we still observe a relatively asymmetric distribution of challenge distances, we cannot exclude such a possible contributor of a challenging strategy entirely.

²⁷As men typically hit balls harder than women, one would expect that it is more difficult for male players as well as umpires to judge where the ball lands, and this will increase the chances of an “embarrassing” challenge for males. However, these descriptive statistics do not seem to suggest such a difference between males and females.

save some of these challenges for the most critical points, in terms of maximizing individuals' match winning probabilities. Thus, when deciding to make a challenge, a player must weigh the risks involved and choose the best timing to utilize these limited number of opportunities.

The results indicate that male and female players exhibit many similarities in their use of line-call challenges, which is not surprising, since they are all top-notch professionals and play high-stake games. Still, we do find that challenge decisions are made somewhat differently by male and female players. Recall that, in our model, the threshold probability is determined by factors parameterized by b , c , $(\pi^H - \pi)$, and $(\pi - \pi^L)$. In this section, we will explore possible interpretations of our findings about gender differences within the framework of our optimal decision making model.

Croson and Gneezy (2009) list three potential explanations for gender differences in any decision making which involves risk-taking: (1) 'emotions', (2) 'overconfidence', and (3) 'perception of risk'. Clearly, in our framework, b and c pertain directly to 'emotions', via pride and shame, and partially to 'perception of risk'. In addition, $(\pi^H - \pi)$ and $(\pi - \pi^L)$ pertain to 'overconfidence' directly, and to 'perception of risk' to some extent.

We have already noted that an overconfident player would overestimate $(\pi^H - \pi)$ and underestimate $(\pi - \pi^L)$ when making a challenge decision, due to his presumptuous belief in his tennis playing ability. Previous studies on overconfidence have consistently shown that individuals overestimate the precision of their knowledge (Alpert and Raiffa, 1982). In addition, overconfidence tends to be more severe for difficult tasks (Fischhoff et al., 1977). Even our world-class professional players can be subject to a cognitive bias due to overconfidence. Based on a sample of about 3,000 entrepreneurs, Cooper et al. (1988) found significant levels of overconfidence in these entrepreneurs' self-perceptions of their business abilities. About 81% of them perceived their own likelihood of success with these abilities to be at least 70%, while a third viewed their odds of success as 100%. Meanwhile, they had much less favorable outlooks on other businesses which were comparable to theirs.²⁸

Various psychological studies have found that men tend to be more overconfident than women (e.g., Lichtenstein et al., 1982; Beyer, 1990; and Beyer and Bowden, 1997).²⁹ According

²⁸Such overconfidence has been observed in the behavior of professionals in many different fields. Clinical psychologists (Oskamp, 1965), physicians and nurses (Christensen-Szalanski and Bushyhead, 1981; Baumann et al., 1991), investment bankers (Stael von Holstein, 1972), engineers (Kidd, 1970), entrepreneurs (Cooper et al., 1988), lawyers (Wagenaar and Keren, 1986), negotiators (Neale and Bazerman, 1990), and managers (Russo and Schoemaker, 1992) have all been observed to exhibit overconfidence in their judgments.

²⁹Barber and Odean (2001) show that men trade more excessively in financial markets than women, as a

to Niederle and Vesterlund (2007, p. 1087), “while 75 percent of the men think they are best in their group of four, only 43 percent of the women hold this belief”.

These factors might potentially help explain the gender differences that we observe in this paper. It is interesting to note that the gender differences among professional players are salient mainly at competitively-intense high-stakes moments. We observe the two most notable gender differences during tiebreaks. First, the results on the success rate for female players are opposite to those for male players. The success rate is lower than usual for men in tiebreaks, whereas it is higher than usual for women, despite very similar utilization rates by the genders. Second, in tiebreaks, men make more embarrassing challenges than usual and women make fewer embarrassing challenges than usual (in spite of the fact that the two genders have equal ratios of embarrassing challenges to all challenges). In both cases, the gender differences are statistically significant.

One possible explanation for both of these stark gender differences in tiebreaks could be that women play more conservatively at such times. That is, women may become relatively more risk averse than men in intensely-competitive high-stake moments. This is consistent with the finding of Paserman (2010) that the odds of making an unforced error relative to hitting a winner at crucial junctures of the match fall for women, while they remain constant for men.³⁰

Another potential explanation for the significant gender differences in tiebreaks could be that the overconfidence of male players leads them to make more risky challenges, while the lower level of overconfidence of female players leads them to make fewer risky and embarrassing challenges in similar situations.³¹

Our theoretical approach assumes that, in part, players with a lower c will feel much less shame for a bad challenge, and players with a higher b will feel much more pride for a good challenge. However, what about the observed differences in challenging behaviors between men and women in relation to these parameters? Are there any potential differences between

consequence.

³⁰It is generally found in the experimental literature that men are more risk-taking than women, and perform better in a competitive environment (see Eckel and Grossman, 2008, for a summary of related results). However, there are exceptions; for example, Schubert et al. (1999) find that men are more risk averse than women when lotteries are framed as losses rather than gains.

³¹At a given point in a match, *ceteris paribus*, a less confident player will perceive $(\pi - \pi^L)$ as high and $(\pi^H - \pi)$ as low, and will not challenge a call unless he perceives the probability of the umpire making a wrong call, p , to be sufficiently high. In the same situation, a more confident player will perceive $(\pi - \pi^L)$ as low and $(\pi^H - \pi)$ as high, and will be willing to challenge a call even if he perceives the probability p to be low.

men and women with respect to their level of concern over making a mistake or their fear of experiencing shame?

There is some evidence that women tend to take negative feedback more seriously than men (see for instance Roberts and Nolen-Hoeksema, 1989). Second, due to their more emotional nature, women may view a negative signal as an indicator of their self-worth, rather than simply of their one-time performance on a task. Dweck (2000) suggests that women may fall into “confidence traps” more easily than men. In addition, Brizendine (2006) reports that girls’ well-developed brain circuits (for gathering meaning from faces and responding to unspoken cues in others) push them to comprehend the social approval of others very early. These findings of the differences between genders also seem useful in providing potential explanations of why women might take a more conservative approach in their playing strategy in tiebreaks, and make fewer embarrassing challenges than men.

The fact that the challenge utilization rate is very low for both genders implies that all tennis players have concerns about making unsuccessful challenges. Otherwise, all of the remaining challenges of every player could easily be used as strategic ploys to at least change the momentum of the game by irritating the opponent, or to obtain a breather for the player who challenges the call, whenever possible. This is hardly the case for female players, and is only the case for male players when a lot is at stake in the match, such as during tiebreaks, which suggests that the extent of such concerns differs for men and women.

At high-stake moments in a match, such as tiebreaks, the embarrassment of a strategic challenge (i.e., a challenge with virtually no chance of success) may be overcome by its strategic benefits, such as stalling the opponent’s momentum and getting a breather. In addition, such a strategic advantage may even have a signaling value, with the challenging player’s signal being, “I’ll keep fighting at all costs until the last point of the match and won’t allow any easy victory, without caring much about shame and embarrassment.” Our gender difference findings indicated that a male player is relatively more likely to send such a strategic signal. A player who sends such a signal of resilience may hope to deter both current and future opponents to some extent, and as such, to gain a mental edge over them, however small.³² Our finding of

³²In Charness and Rustichini’s (2011) experiment, two groups in two rooms participated in each session, where these groups could be both all-male, both all-female, or one all-male and one all-female. Each member of each group played the Prisoners’ Dilemma game once with an audience of the same group and once with audience of the other group. Males cooperate substantially less often when they are observed by their peer group, while females cooperate substantially more often. Thus, both males and females wish to signal to their in-group peers: males wish to signal their formidableness, while females wish to signal their cooperativeness.

male player’s challenges being more likely to be provoked by those of their opponents can be considered along these lines as well.

Nevertheless, no kind of strategic ploy is carried out by all players at all times. Evidently, even the most strategic players who play such mind games must resort to such strategic ploys only when it is worth it, namely when they have a chance to turn things around, such as in tiebreaks.

Although we do not discuss them in detail here, similar explanations might apply to the additional significant gender differences observed in our empirical analysis above.

6 Conclusions

Tennis line-call challenges provide a suitable context in which to investigate the origins of gender differences in a competitive setting with large monetary rewards. Given the nature and design of the tennis market, gender discrimination in terms of pay or participation rates on the demand side are unlikely to happen. This makes professional tennis an attractive setting for investigating whether men and women respond differently to competitive pressures. Unlike in an experimental setting, the prize money gained by winning matches is reasonably large. Thus, by examining the behavior of individuals in such a real setting, we aim to make a significant contribution to the existing literature on gender differences.

To summarize our results, we observe that the genders are very similar in terms of their line-call challenge utilization rates. Specifically, the utilization rates of both male and female players are significantly higher the higher the total number of games played in the set, and increase significantly with the proportion of games won by the opponent. Intuitively, utilization rates are also significantly higher in the second and third sets than in the first set. While a higher self-ranking tends to increase male utilization rates significantly, female utilization rates decrease significantly as the opposing player’s ranking increases. Nevertheless, all of these gender differences in utilization rates turn out to be statistically insignificant.

In terms of success rates, the major differences between males and females appear in challenges made in tiebreaks. In tiebreaks, females are more likely to make correct challenges, while males are more likely to make an incorrect challenges, and this gender difference is statistically significant. The results are also similar in the case of a challenge made in an offense, but the

Applying their intuition to our context, one may be able to guess that male and female tennis players may perhaps have behaved similarly to each other if no one else had been watching them.

difference is less significant.

Furthermore, men are a lot more likely to make “embarrassing” line-call challenges in tiebreaks and offenses than women. While a higher-ranked male player is more likely to make an embarrassing challenge, a higher-ranked female is less likely to do so. Finally, the propensity to make an embarrassing challenge increases with the opponent’s ranking in male games, but we do not observe any such statistically significant correlation in female games. When coupled with the descriptive statistics of the data, this finding might suggest that men’s line-call challenges are more likely to be provoked by their opponents.

The major gender differences found at crucial moments of the match, such as tiebreaks, suggest that male players try to win at all costs, while female players accept losing more gracefully. In that sense, this finding extends the Niederle-Vesterlund finding to a domain in which neither men nor women are choosing to shy away from competition: even if it involves very high stakes, women are less likely to try to win the competition at all costs than men. Differences in factors such as risk-aversion, overconfidence, pride, shame and strategic signalling behavior are some of the potential explanations of our main gender-difference findings in line-call challenges.

Finally, our paper studies the risk-taking decisions of two highly-skilled professionals in a high-stakes “interactive environment”, in which their risk-taking trajectories are shaped by the constantly changing score and momentum of the match. In addition, these professionals make their decisions “in front of others.” We are not aware of any previous study on that has looked at risk-taking in an “interactive environment” and “in front of others.” The limited literature on other topics where the audience is a factor, such as the study by Charness and Rustichini (2011), suggests that people do not always behave in the way that would be expected when faced with an audience. Given that many individuals do not make their risk-taking decisions in isolation or seclusion, further research into risk-taking behaviors in “interactive environments” and “in front of others” will be significant and fruitful.

References

- Abramitzky, Ran, Liran Einav, Shimon Kolkowitz, and Roy Mill (2012), “On the Optimality of Line Call Challenges in Professional Tennis.” *International Economic Review*, 53, 939-963.

- Alpert, Marc, and Howard Raiffa (1982), "A Progress Report on the Training of Probability Assessors," in *Judgment Under Uncertainty: Heuristics and Biases*, eds. D. Kahneman, P. Slovic, and A. Tversky, 294-305. Cambridge University Press, Cambridge and New York.
- Antonovics, Kate, Peter Arcidiacono, and Randall Walsh (2009), "The Effects of Gender Interactions in the Lab and in the Field." *Review of Economics and Statistics*, 91, 152-162.
- Atkinson, John W. (1957), "Motivational Determinants of Risk-taking Behavior." *Psychological Review*, 64, 359-372.
- Barber, Brad M. and Terrance Odean (2001), "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment." *Quarterly Journal of Economics*, 116, 261-292.
- Baumann, Andrea O., Deber, Raisa B., and Gail G. Thompson (1991), "Overconfidence Among Physicians and Nurses: The 'Micro-Certainty, Macro-Uncertainty' Phenomenon." *Social Science and Medicine*, 32, 167-174.
- Bertrand, Marianne, and Kevin F. Hallock (2010), "The Gender Gap in Top Corporate Jobs." *Industrial and Labor Relations Review*, 55, 3-21.
- Beyer, Sylvia (1990), "Gender Differences in the Accuracy of Self-Evaluations of Performance." *Journal of Personality and Social Psychology*, 59, 960-970.
- Beyer, Sylvia and Edward M. Bowden (1997), "Gender Differences in Self-Perceptions: Convergent Evidence from Three Measures of Accuracy and Bias." *Personality and Social Psychology Bulletin*, 23, 157-172.
- Booth, Alison, and Patrick Nolen (2012a), "Gender Differences in Risk Behaviour: Does Nurture Matter?" *Economic Journal*, 558, F56-F78.
- Booth, Alison, and Patrick Nolen (2012b), "Choosing to Compete: How Different Are Girls and Boys?" *Journal of Economic Behavior and Organization*, 81, 542-555.
- Brizendine, Louann (2006), *The Female Brain*. Random House, New York.

- Camerer, Colin (1997), "Progress in Behavioral Game Theory." *Journal of Economic Perspectives*, 11, 167-188.
- Campbell, Anne (2002), *A Mind of Her Own: The Evolutionary Psychology of Women*. Oxford, UK: Oxford University Press.
- Charness, Gary, and Aldo Rustichini (2011), "Gender Differences in Cooperation with Group Membership." *Games and Economic Behavior*, 72, 77-85.
- Christensen-Szalanski, Jay J., and James B. Bushyhead (1981), "Physicians' Use of Probabilistic Information in a Real Clinical Setting." *Journal of Experimental Psychology: Human Perception and Performance*, 7, 928-935.
- Cooper, Arnold C., Carolyn Y. Woo, and William C. Dunkelberg (1988), "Entrepreneurs' Perceived Chances for Success." *Journal of Business Venturing*, 3, 97-108.
- Croson, Rachel, and Uri Gneezy (2009), "Gender Differences in Preferences." *Journal of Economic Literature*, 47, 1-27.
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jurgen Schupp, and Gert G. Wagner (2010), "Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey." *Journal of the European Economic Association*, 3, 522-550.
- Dweck, Carol S. (2000), *Self-Theories: Their Role in Motivation, Personality, and Development*. Philadelphia, PA: Psychology Press.
- Eckel, Catherine C., and Philip J. Grossman (2008), "Men, Women and Risk Aversion: Experimental Evidence." In *Handbook of Experimental Economics Results*, Volume 1, ed. Charles Plott and Vernon Smith, 1061-1073. Elsevier, New York.
- Fischhoff, Baruch, Paul Slovic, and Sarah Lichtenstein (1977), "Knowing with Certainty: The Appropriateness of Extreme Confidence." *Journal of Experimental Psychology: Human Perception and Performance*, 3, 552-564.
- Gneezy, Uri, Muriel Niederle, and Aldo Rustichini (2003), "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics*, 118, 1049-1074.

- Gneezy, Uri, and Aldo Rustichini (2004), "Gender and Competition at a Young Age." *American Economic Review*, 94, 377-381.
- Gneezy Uri, Kenneth L. Leonard, John A. List (2009), "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society." *Econometrica*, 77, 1637-1664.
- Honoré, Bo E. (1992), "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects." *Econometrica*, 60, 533-565.
- Kidd, John B. (1970), "The Utilization of Subjective Probabilities in Production Planning." *Acta Psychologica*, 34, 338-347.
- Krook, Mona Lena, and Sarah Childs (2010), *Women, Gender, and Politics: A Reader*. Oxford University Press, New York.
- Lichtenstein, Sarah, Baruch Fischhoff, and Lawrence Phillips (1982), "Calibration and Probabilities: The State of the Art to 1980," in *Judgment under Uncertainty: Heuristics and Biases*, Daniel Kahneman, Paul Slovic, and Amos Tversky, eds. New York: Cambridge University Press.
- Neale, Margaret A. and Max H. Bazerman (1990), *Cognition and Rationality in Negotiation*. The Free Press, New York.
- Niederle, Muriel, and Lise Vesterlund (2007), "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, 122, 1067-1101.
- Oskamp, Stuart (1965), "Overconfidence in Case Study Judgments." *Journal of Consulting Psychology*, 29, 261-265.
- Palacios-Huerta, Ignacio, and Oscar Volij (2008), "Experientia Docet: Professional Play Minimax in Laboratory Experiments." *Econometrica*, 76, 71-115.
- Paserman, Danielle (2010), "Gender Differences in Performance in Competitive Environments? Evidence from Professional Tennis Players." mimeo, Boston University.
- Powell, James L. (1986), "Symmetrically Trimmed Least Squares Estimation for Tobit Models." *Econometrica*, 54, 1435-1460.

- Roberts, Tomi-Ann, and Susan Nolen-Hoeksema (1989), "Sex Differences in Reactions to Evaluative Feedback." *Sex Roles*, 21, 725–747.
- Russo, Edward J. and Paul J.H. Schoemaker (1992), "Managing Overconfidence." *Sloan Management Review*, 33, 7-17.
- Sagar, Sam S. and David Lavalley (2010), "The Developmental Origins of Fear of Failure in Adolescent Athletes: Examining Parental Practices." *Psychology of Sport and Exercise*, 11, 177-187.
- Schubert, Renate, Martin Brown, Matthias Gysler, and Hans Wolfgang Brachinger (1999), "Financial Decision-Making: Are Women Really More Risk-Averse?" *American Economic Review*, 89, 381-385.
- Stael von Holstein, Carl-Axel. S. (1972), "Probabilistic Forecasting: An Experiment Related to the Stock Market." *Organizational Behavior and Human Performance*, 8, 139-158.
- Tversky, Amos, and Daniel Kahneman (1974), "Judgement under Uncertainty: Heuristics and Biases." *Science*, 185, 1124-1131.
- Wagenaar, Willem, and Gideon B. Keren (1986), "Does the Expert Know? The Reliability of Predictions and Confidence Ratings of Experts," in *Intelligent Decision Support in Process Environments*, eds. E. Hollnagel, G. Maneini, and D. Woods, 87-107. Springer, Berlin.

A Singles Tennis Rules

Singles tennis is a game played by two players of the same gender who stand on opposite sides of a net on a well-defined court and take turns to strike a ball using a stringed racket. Each player’s objective is to score points by striking the ball to the other player’s side of the court and out of reach of their opponent. A point ends when a player either fails to strike a ball that has touched his side of the court, or strikes the ball, but hits it out of the court or into the net.

A tennis match comprises an odd number of sets — three, in all tournaments in our data set. A set consists of a number of games. A game is a sequence of points played with the same player serving until the game is won by one or other of the players. The match is won by the player who wins more than half of the sets (i.e., two sets, in the tournaments we consider here). Typically, a player wins a set when he wins at least six games and at least two more games than his opponent. In all of the tournaments that we consider, players play a “tiebreak” if the score in games is tied at 6-6; this is won by the player who first reaches seven points and at least two more points than his opponent (if the score is still tied at 6-6 in a tiebreak, then a player can win that set by winning two more points than his opponent). A game is won by the first player to have won at least four points and at least two more points than his opponent.

Starting in the second part of March 2006, in the Miami Masters and Tier I combined tournament, many major tennis tournaments adopted a line-call challenge system involving an electronic review technology, the Hawk-Eye system, which allows a player to challenge an umpire’s line calls in a point.³³ Accordingly, a player is able to challenge both “out” calls by the umpire and shots by the opponent that the umpire assumes to be “in” but the player believes were actually out. For a ball to be called “out”, it must be 100% out of the tennis court, i.e., it should not touch any portion of any line. A player may use an unlimited number of challenges in a set, provided that he is not incorrect more than a limited number of times (two, in all tournaments in our data set). One additional challenge opportunity is given if a set reaches a tiebreak. Unused challenge opportunities in a set do not carry over to the following set, if any.

³³Umpires, like professional tennis players, are seasoned professionals in their field. Before being selected to umpire ATP Masters or WTA Tier I matches, they have to umpire at lower-level tournaments for many years. The variation in their quality is considered very minimal, as is evident from the fact that all of them, not a small subset, have umpired finals matches at various top-level tournaments.

Figure 1. Example Challenge Report

	EAST WEST BANK CLASSIC <small>Tournament Name</small>	LOS ANGELES <small>City</small>
	SANIA MIRZA vs ALEXSANDRA WOZNIAK	1 <small>Round</small>
CRISTINA ROMERO <small>Chair Umpire</small>	CHRIS WILSON <small>Review Official</small>	AUG 6, 2007 <small>Date</small>

PLAYER 1: SANIA MIRZA

CHALLENGE RECORD - PLAYER 1

SET 1										SET 2										SET 3									
2 1 + TB										2 1 + TB										2 1 + TB									
Set #	Line	Call	O	Hawk	Dist. (mm)		Challenges			Challenge Type		Score		Match Time	Notes														
			✓		in	out				Point-end	Player-stop																		
1	LB	in out		in out	18		W	L	NA	✓	✓	1-1 15-30	8																
2	RB	in out	✓	in out	23		W	L	NA	✓		1-0 15-30	34																
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				

PLAYER 2: ALEXSANDRA WOZNIAK

CHALLENGE RECORD - PLAYER 2

SET 1										SET 2										SET 3									
2 1 + TB										2 1 + TB										2 1 + TB									
Set #	Line	Call	O	Hawk	Dist. (mm)		Challenge			Challenge Type		Score		Match Time	Notes														
			✓		in	out				Point-end	Player-stop																		
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				
		in out		in out			W	L	NA																				

Legend

- LF: Left Far
- LB: Left Base
- LS: Left Service
- LC: Left Centre Serve
- RF: Right Far
- RB: Right Base
- RS: Right Service
- RC: Right Centre Serve

MIRZA <small>Player 1</small>	WOZNIAK <small>Player 2</small>	<small>Match</small>
Challenges Won 1	Won 0	Won 1
Challenges Lost 1	Lost 0	Lost 1
Total 2	Total 0	Total 2

Match Winner: SANIA MIRZA

Score: 6, 3, 6

[Signature]
 Review Official Signature

Figure 2. Ball-Line Distance

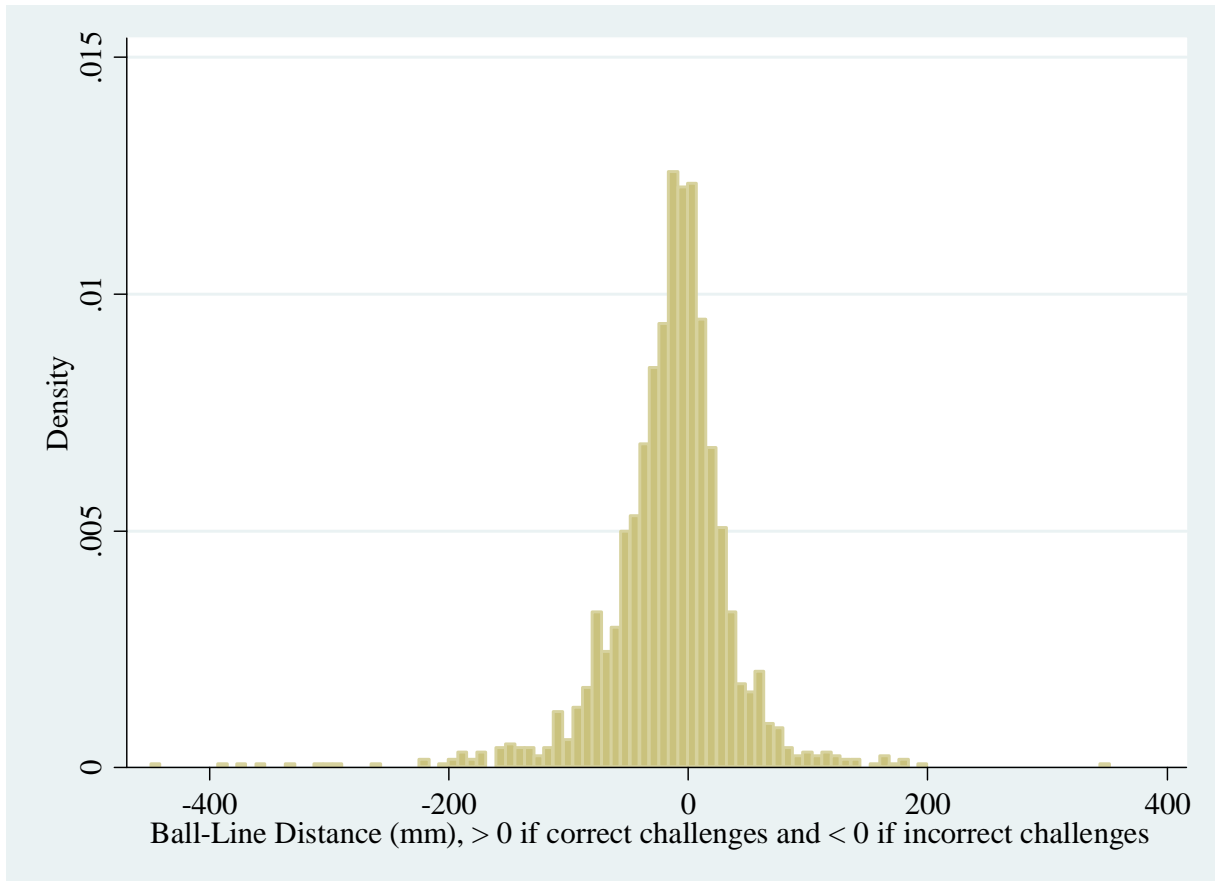


Table 1. Match and Challenge Characteristics

	Men	Women
Three-set matches	0.3172 (0.4661)	0.302 (0.4607)
Upsets	0.3625 (0.4815)	0.2886 (0.4546)
Tie-break per set	0.2729 (0.3071)	0.2159 (0.2653)
Probability of winning match after first set won	0.8512 (0.3570)	0.8861 (0.3197)
Challenge utilization rates	0.3137 (0.2191)	0.3345 (0.2327)
Correlation in challenges between opponents	0.1635 [p = 0.003]	0.0517 [p = 0.532]
Ball-line distance (mm)	36.97 (43.58)	36.81 (43.28)
Challenge success rates	0.3983 (0.4898)	0.3252 (0.4691)
Number of matches	331	149
Number of sets	767	343
Number of individual players	113	91
Number of challenges	1,057	427

Notes: Standard deviations are presented in parentheses. The p-value is in brackets. *Three-set matches* and *Upsets* are dummy variables.

Table 2. Challenge Utilization by Gender

	RE-Tobit			FE-Tobit		
	Men	Women	Difference	Men	Women	Difference
Total number of games played in the set	0.073*** (0.016)	0.084*** (0.020)	-0.008 (0.025)	0.055*** (0.017)	0.060*** (0.017)	-0.005 (0.024)
Proportion of games won by the opponent in the set	0.235** (0.117)	0.278* (0.158)	-0.029 (0.192)	0.339** (0.133)	0.086 (0.160)	0.253 (0.208)
Logarithm of player's self ATP ranking	0.033** (0.016)	-0.012 (0.025)	0.044 (0.029)			
Logarithm of opponent player's ATP ranking	0.003 (0.016)	-0.044* (0.025)	0.046 (0.029)			
Second set after first set won	0.099* (0.055)	0.086 (0.088)	0.018 (0.101)	0.064 (0.046)	0.082 (0.080)	-0.018 (0.093)
Second set after first set lost	0.155*** (0.055)	0.192** (0.087)	-0.029 (0.100)	0.145*** (0.049)	0.113 (0.081)	0.031 (0.095)
Third set	0.271*** (0.063)	0.189* (0.104)	0.091 (0.118)	0.188*** (0.052)	0.149* (0.079)	0.040 (0.095)
Tiebreak set	-0.052 (0.078)	-0.136 (0.136)	0.080 (0.152)	-0.127 (0.082)	-0.175 (0.124)	0.048 (0.149)
Year Dummies	Yes	Yes		NA	NA	
Observations	1,522	680		1,522	680	

Notes: Robust standard errors are presented in parentheses.

Table 3. Success Rates by Gender

	RE-Logit			FE-Logit		
	Men	Women	Difference	Men	Women	Difference
Total number of games played in the set when the challenge made	-0.047 (0.028) <i>-0.011</i>	-0.000 (0.047) <i>0.000</i>	-0.046 (0.055) <i>-0.011</i>	-0.104** (0.045) <i>-0.012</i>	-0.114* (0.069) <i>-0.025</i>	0.010 (0.083) <i>0.001</i>
Proportion of games won by the opponent when the challenge made	0.066 (0.273) <i>0.016</i>	-0.208 (0.413) <i>-0.045</i>	0.274 (0.495) <i>0.064</i>	0.020 (0.459) <i>0.002</i>	0.100 (0.622) <i>0.022</i>	-0.080 (0.773) <i>-0.012</i>
Logarithm of player's self ATP ranking	-0.035 (0.054) <i>-0.008</i>	0.040 (0.082) <i>0.009</i>	-0.075 (0.098) <i>-0.018</i>			
Logarithm of opponent player's ATP ranking	-0.048 (0.054) <i>-0.011</i>	-0.168** (0.083) <i>-0.037</i>	0.120 (0.099) <i>0.028</i>			
Whether the point naturally ended or the player stopped	-1.081*** (0.383) <i>-0.263</i>	-0.802** (0.319) <i>-0.187</i>	-0.279 (0.499) <i>-0.066</i>	-0.988* (0.547) <i>-0.159</i>	-0.750 (0.516) <i>-0.174</i>	-0.238 (0.752) <i>-0.035</i>
Whether the game score was a deuce when the challenge made	-0.176 (0.556) <i>-0.041</i>	-0.252 (1.196) <i>-0.052</i>	0.076 (1.319) <i>0.018</i>	0.221 (0.736) <i>0.028</i>	0.831 (1.471) <i>0.199</i>	-0.610 (1.645) <i>-0.072</i>
Whether the game score was 40-0 or 0-40 when the challenge made	0.543 (0.403) <i>0.134</i>	0.976 (0.621) <i>0.235</i>	-0.433 (0.740) <i>-0.095</i>	0.126 (0.520) <i>0.016</i>	0.812 (1.396) <i>0.195</i>	-0.686 (1.490) <i>-0.079</i>
Elapsed time since the start of the match when the challenge made	0.000 (0.004) <i>0.000</i>	0.004 (0.006) <i>0.001</i>	-0.003 (0.007) <i>-0.001</i>	-0.005 (0.008) <i>-0.001</i>	0.003 (0.006) <i>0.001</i>	-0.007 (0.010) <i>-0.001</i>
Second set after first set won	-0.054 (0.265) <i>-0.013</i>	-0.136 (0.410) <i>-0.029</i>	0.083 (0.489) <i>0.020</i>	-0.066 (0.450) <i>-0.008</i>	-0.581 (0.658) <i>-0.117</i>	0.514 (0.797) <i>0.083</i>
Second set after first set lost	-0.254 (0.254) <i>-0.060</i>	-0.023 (0.389) <i>-0.005</i>	-0.231 (0.465) <i>-0.053</i>	-0.023 (0.457) <i>-0.003</i>	0.252 (0.509) <i>0.056</i>	-0.275 (0.684) <i>-0.037</i>

Third set	-0.451 (0.441)	-0.692 (0.638)	0.241 (0.775)	0.135 (0.780)	-0.386 (0.827)	0.521 (1.137)
	<i>-0.103</i>	<i>-0.136</i>	<i>0.058</i>	<i>0.017</i>	<i>-0.080</i>	<i>0.084</i>
Tiebreak	-0.245 (0.339)	1.799* (0.922)	-2.044** (0.982)	0.258 (0.477)	1.768 (1.087)	-1.510 (1.187)
	<i>-0.057</i>	<i>0.421</i>	<i>-0.322</i>	<i>0.033</i>	<i>0.415</i>	<i>-0.140</i>
Offense	-0.281** (0.134)	0.104 (0.257)	-0.386 (0.290)	-0.272 (0.200)	-0.100 (0.432)	-0.172 (0.476)
	<i>-0.067</i>	<i>0.023</i>	<i>-0.089</i>	<i>-0.032</i>	<i>-0.022</i>	<i>-0.024</i>
Second challenge	0.235 (0.214)	-0.196 (0.347)	0.431 (0.408)	1.113*** -0.269	1.154** (0.517)	-0.042 (0.583)
	<i>0.057</i>	<i>-0.042</i>	<i>0.104</i>	<i>0.173</i>	<i>0.275</i>	<i>-0.006</i>
Year Dummies	Yes	Yes		NA	NA	
Observations	1005	411		556	192	

Notes: Robust standard errors are presented in parentheses. Marginal effects, evaluated under the assumption that individual effects are zero, are presented in italics.

Table 4. Embarrassing Challenges

	RE-Logit			FE-Logit		
	Men	Women	Difference	Men	Women	Difference
Total number of games played in the set when the challenge made	0.052 (0.038) <i>0.007</i>	0.084 (0.067) <i>0.009</i>	-0.019 (0.072) <i>-0.002</i>	0.090 (0.057) <i>0.010</i>	0.222* (0.132) <i>0.028</i>	-0.133 (0.144) <i>-0.015</i>
Proportion of games won by the opponent when the challenge made	0.120 (0.414) <i>0.015</i>	-0.661 (0.606) <i>-0.075</i>	0.658 (0.697) <i>0.081</i>	0.000 (0.677) <i>0.000</i>	-1.330 (0.840) <i>-0.168</i>	1.331 (1.079) <i>0.150</i>
Logarithm of player's self ATP ranking	0.139* (0.071) <i>0.018</i>	-0.201 (0.122) <i>-0.023</i>	0.322** (0.129) <i>0.040</i>			
Logarithm of opponent player's ATP ranking	0.140* (0.072) <i>0.018</i>	0.017 (0.123) <i>0.002</i>	0.116 (0.131) <i>0.014</i>			
Whether the point naturally ended or the player stopped	0.994 (0.633) <i>0.090</i>	0.416 (0.493) <i>0.043</i>	0.599 (0.787) <i>0.069</i>	0.422 (1.271) <i>0.053</i>	1.494* (0.886) <i>0.255</i>	-1.071 (1.550) <i>-0.102</i>
Elapsed time since the start of the match when the challenge made	0.005 (0.006) <i>0.001</i>	-0.000 (0.009) <i>0.000</i>	0.006 (0.010) <i>0.001</i>	0.016 (0.010) <i>0.002</i>	-0.009 (0.023) <i>-0.001</i>	0.025 (0.025) <i>0.003</i>
Second set after first set won	-0.512 (0.360) <i>-0.058</i>	0.363 (0.576) <i>0.045</i>	-0.855 (0.637) <i>-0.085</i>	-0.624 (0.593) <i>-0.078</i>	2.213 (1.476) <i>0.170</i>	-2.838* (1.591) <i>-0.538</i>
Second set after first set lost	-0.098 (0.332) <i>-0.012</i>	-0.035 (0.586) <i>-0.004</i>	-0.060 (0.630) <i>-0.007</i>	0.026 (0.592) <i>0.003</i>	0.088 (1.257) <i>0.011</i>	-0.062 (1.389) <i>-0.007</i>
Third set	-0.362 (0.565) <i>-0.042</i>	-0.225 (0.963) <i>-0.024</i>	-0.181 (1.047) <i>-0.021</i>	-1.107 (1.000) <i>-0.154</i>	1.542 (2.261) <i>0.139</i>	-2.649 (2.472) <i>-0.497</i>
Tiebreak	0.582* (0.349) <i>0.087</i>	-1.431 (1.243) <i>-0.098</i>	2.064* (1.225) <i>0.410</i>	0.602 (0.556) <i>0.055</i>	-0.852 (1.364) <i>-0.140</i>	1.455 (1.473) <i>0.106</i>

Offense	0.716***	0.087	0.627*	0.536*	0.220	0.316
	(0.182)	(0.359)	(0.376)	(0.285)	(0.569)	(0.636)
	<i>0.092</i>	<i>0.010</i>	<i>0.084</i>	<i>0.059</i>	<i>0.028</i>	<i>0.035</i>
Second challenge	0.302	0.200	-0.006	-0.034	-0.510	0.476
	(0.268)	(0.440)	(0.483)	(0.338)	(0.560)	(0.654)
	<i>0.041</i>	<i>0.024</i>	<i>-0.001</i>	<i>-0.004</i>	<i>-0.071</i>	<i>0.047</i>
Year Dummies	Yes	Yes		NA	NA	
Observations	1,005	411		338	119	

Notes: Robust standard errors are presented in parentheses. Marginal effects, evaluated under the assumption that individual effects are zero, are presented in italics.