Gender Differences in Performance in Competitive Environments: Field Evidence from Professional Tennis Players*

M. Daniele Paserman
Hebrew University
dpaserma@shum.huji.ac.il

August 2006

Abstract

This paper uses data from seven Grand Slam tournaments played between 2005 and 2006 to assess whether men and women respond differently to competitive pressure in a setting with large monetary rewards. In particular, it asks whether the quality of the game (measured by the percentage of unforced errors, winners, and first serves) deteriorates as the stakes become higher. The paper conducts two parallel analyses, one based on aggregate set-level data, and one based on detailed point-by-point data, which is available for a selected subsample of matches in two of the seven tournaments under examination.

The set-level analysis indicates that both men and women perform less well in the final and decisive set of the match. This result is robust to controls for the length of the match and to the inclusion of match and player-specific fixed effects. The drop in performance of women in the decisive set is slightly larger than that of men, but the difference is not statistically significant at conventional levels. On the other hand, the detailed point-by-point analysis reveals that, relative to men, women are substantially more likely to make unforced errors at crucial junctures of the match. Both analyses reveal that the tendency to make unforced errors at crucial stages of the match diminishes with the player's ability.

* PRELIMINARY AND INCOMPLETE. I thank Yaron Aronshtam for outstanding research assistance. All errors are my own.

Recent decades have seen a dramatic increase in female labor force participation rates, and a considerable narrowing of the gender gap. Yet, despite these advances, the gender gap in wages is still substantial, even if one looks at men and women with the same amount of experience and education. At the very high end of the wage distribution, women have found it particularly hard to break the glass ceiling and to make inroads into the upper echelons of management and academia.

Several explanations have been put forth for this phenomenon, ranging from discrimination to differences in preferences, and to differential investments in human capital and on-the-job training. Some recent controversial remarks by former Harvard University President Lawrence Summers have also sparked a lively debate about the reasons for the small number of women in science, engineering, and at the forefront of academic research.

In a recent influential paper, Gneezy, Niederle and Rustichini (2003) have put forward an intriguing hypothesis: women may be less effective than men in competitive environments, even if they are able to perform similarly in a non-competitive environment. Using experimental evidence, they revealed the existence of a significant gender gap in performance in a tournament setting where wages were based on a winner-takes-all principle, while no such gap existed when players were paid according to a piece rate. The reason for this gap is that men's performance increases significantly with the competitiveness of the environment, while women's performance does not. Gneezy et al. also found that the decrease in woman's performance was more pronounced in mixed competitions (women competing against men) than in single-sex competitions (women competing amongst themselves). Similar findings were also obtained by Gneezy and Rustichini (2004), who analyzed the performance of young boys and girls in a race over a short distance. Niederle and Vesterlund (2005), on the other hand, found no gender differences in performance on an arithmetic task under either a non-competitive piece rate compensation scheme, or a competitive tournament scheme. However, they found that men were significantly more likely than women to select into the competitive compensation scheme when given the choice, and that such a choice could not be explained by performance either before or after the entry decision.

In a non-experimental setting, Lavy (2004) found that the gender gap in test scores (which favors girls) on "blind" high school matriculation exams (i.e., exams that are graded by an external committee) is smaller than the gender gap in scores on exams graded internally. Lavy attributes this finding to discrimination on the part of teachers, but one cannot rule out that women underperform because of the increased pressure they face in the blind setting.

In this paper, I complement the existing literature by examining whether men and women respond differently to competitive pressure in a setting with large monetary rewards. Specifically, I focus on professional tennis players and on seven Grand Slam tournaments played between 2005 and 2006. One of the advantages of using sports data is that it is possible to observe detailed measures of productivity or performance. I ask whether players' performance deteriorates as the stakes become higher. A player's performance is measured primarily by looking at the number of unforced errors and winners hit during each match. To the best of my knowledge, this is an improvement over previous research on tennis data, which limited itself to information on the number of games or sets won by the players (Sunde, 2003).

The paper conducts two parallel analyses, one based on aggregate set-level data, and one based on detailed point-by-point data, which is available for a selected subsample of matches in two of the seven tournaments under examination.

The aggregate set-level analysis indicates that both men and women perform less well in the final and decisive set of the match. This result is robust to controls for the length of the match and to the inclusion of match and player-specific fixed effects. I find that the performance of women is slightly inferior to that of men, but the difference is not statistically significant at conventional levels.

The detailed point-by-point data, available for a subset of matches played at the 2006 French Open and Wimbledon tournaments, allows me to create a less coarse measure of the decisiveness of each point. Following Klaassen and Magnus (2001), I define the importance of each point as the difference in the probability of winning the entire match as a result of winning or losing the current point. This importance variable evolves very non-linearly over the course of a match, generating an abundance of useful variation that can be used for estimation. I find that while men's performance does not vary much depending on the importance of the point, women's performance deteriorates significantly as points become more important. About 31 percent of men's points end in unforced errors, regardless of their placement in the distribution of the importance variable. For women, on the other hand, about 33 percent of points in the bottom quartile of the importance distribution end in unforced errors; the percentage of unforced errors rises to more than 40 percent for points in the top quartile of the importance distribution. These differences persist and are statistically significant even in a multinomial logit model that controls for players' ability, tournament, set number, and individual match fixed effects.

The rest of the paper is structures as follows. The next section introduces some basic concepts in the game of tennis. Section 2

1. Tennis: basic concepts

Rules. Tennis is a game played by two players ("singles") or two teams of players ("doubles"). The players stand on opposite sides of a net and strike a ball in turns with a stringed racket. Their objective is to score *points* by striking the ball within a delimited field of play (the *court*) and out of the reach of the opponent. The player who first puts the ball in play is designated the *server*, while the player who is ready to return the serve is called the *receiver*. *A* tennis *match* usually comprises an odd number of *sets* (three or five). A *set* consists of a number of *games*, and games, in turn, consist of *points*. The match winner is the player who wins more than half of the sets. The match ends as soon as this winning condition is met. A set consists of a sequence of games played with service alternating between games, ending when the count of games won meets certain criteria. Typically, a player wins a set when he wins at least six games and at least two games more than his opponent. It has become common, however, to play a "tiebreaker" when each player has won six games. A *game* consists of a sequence of points played with the same player serving, and is won by the first player to have won at least four points and at least two points more than his opponent.

Typology of points. A point is lost when one of the players fails to make a legal return of the ball. This can happen in a number of ways: an *unforced error* is an error in a service or return shot that cannot be attributed to any factor other than poor judgment by the player; a *winner* is a forcing shot that cannot be reached by the opponent and wins the point; all other points are *forced errors*, errors in a return shot that were forced by the opponent.

Playing Surface. There are three main types of courts, depending on the materials used for the court surface. Each surface provides a difference in the speed and bounce of the ball, which in turn can affect the level of play of the individual players. Some players are notoriously better on certain surfaces than on others. Clay courts are considered "slow", meaning that the balls lose speed as they hit the court and bounce relatively high, making it more difficult for a player to hit an unreturnable shot, called a *winner*. Hard courts and grass courts are "fast" surfaces, where fast, low bounces keep rallies short, and powerful, hard-serving and hard-hitting players have an advantage. Grass courts add an additional variable, with bounces depending on how healthy the grass is and how recently it has been mowed.

Tournaments. Tournaments are organized by gender and number of players. The four most important tournaments on the professional tennis circuit are the *Grand Slam* tournaments: the Australian Open, played in January in Melbourne on a hard court surface; the French Open, played in May-June in Paris on clay courts; Wimbledon, played in June-July in London on grass courts; and the U.S. Open, played in August-September in New York on hard courts. The Grand Slam tournaments have 128 entrants, organized in a predetermined *draw* of 64 matches: the winner of a match advances to the next round, while the loser exits the tournament. The winner of the tournament is the player who wins seven consecutive matches. The prizes for reaching each round are set in advance: typically prizes double at each successive round in the latter rounds, while they increase somewhat less steeply in the earlier rounds. Women matches are played on a "best-of-three" basis (i.e., the first player who wins two sets wins the match), while for men, contrary to other tournaments on the professional tour, all matches starting from the first round are played on a "best-of-five" basis.

Ranking of players. The tournament draw is set up so that the best players are kept apart until the final rounds through a seeding mechanism. The best 32 players in the tournament are seeded and cannot play each other in the first two rounds. The seeding is determined by the weekly Entry Ranking of players by the ATP (the professional male players' association) and the WTA (the professional female players' association). The weekly ranking takes into account all results obtained in professional and satellite tournaments over the past 52 weeks.

2. Set level analysis: data and summary statistics

I collected data on the seven Grand Slam tournaments played between the 2005 Australian Open and the 2006 Wimbledon tournament. The data were collected between January and July 2006 from the official web sites of the tournaments. One advantage of focusing on the Grand Slam tournaments is the uniformity of the available statistics, kept by IBM. The web sites record detailed match statistics broken down by set for every match played in six of the seven tournaments; for the 2005 U.S. Open, the full statistics are available for every match from the third round onwards, and for selected matches played in the first two rounds. I have information on the final score in the set, the number of points played in the set, the length of the set, and a number of statistics on the performance of both players: the total number of first serves and the number of first serves that were in, the number of aces, the number of double faults, the number of unforced errors, the number of points won on the first and on the second serve, the number of break points and the number of break points won, the number of net approaches and the number of net approaches won, and the total number of points won in the set. In addition, I also recorded the players' 52-week ranking at the beginning of the tournament, from which I calculate each player's ability rating (Klaassen and Magnus, 2001) as $Rating = 8 - log_2(Rank)$. Klaassen and Magnus justify the use of this variable as a smoothed version of the expected round to be reached by a player of a given rank: for example, the number 1 ranked player in the world is expected to win all matches, and therefore reach round 8 (i.e., will win the tournament). This variable also has three additional advantages: first, the distribution of this variable is less skewed than the distribution of rank, and it explains about twice the variance in the percentage of points won than the simple rank; second, it captures the fact that the difference in ability between the number 1 and the number 2 ranked players is probably greater than the difference between players ranked 101 and 102; finally, it has the advantage of taking on higher values for better players, a desirable feature for a measure of ability.

From the raw data, I constructed five key indicators of performance for every player and set: the percentage of unforced errors out of total points played, the percentage of winners, the percentage of forced opponent errors,[1] the percentage of first serves that were in, and the percentage of net approaches. The first three indicators all clearly measure the quality of the player's game: a player who hits many winners and few unforced errors, and forces many opponent errors, will on average win a high percentage of points. The first serve percentage is also widely cited by tennis analysts and commentators as an important indicator of performance, and it is a good predictor of the likelihood of winning a point on one's serve: the percentage of points won by the server when the first serve is in is about 70 percent for

[1] The number of forced opponent errors is calculated by subtracting the number of winners and the number of opponent forced errors from the total number of points won by a player in a set.

men and 62 percent for women, as compared to about 51 percent and 46 percent on one's second serve. As for the percentage of net approaches, it is not clear *a priori* whether it reflects a high quality or low quality game, but we include it mostly as an indicator of the riskiness of the player's game strategy. Note that also the other indicators tell us something about risk as well as quality: a more aggressive strategy will generally lead to more winners and forced opponent errors, but it may also generate more unforced errors, as the players attempts for more risky shots. However, for both men and women, the partial correlation (after having netted out own and opponent rating) between unforced errors on one side and winners, forced opponent errors and first serves on the other is negative. This suggests that, on aggregate, all the variables are more a reflection of performance than of risk. Players who are tired and lack concentration and effort make more unforced errors, hit fewer winners, force fewer opponent errors, and have a lower first serve percentage.

Summary statistics for these variables are presented in Table 1, broken down by gender and tournament. For the analysis, I keep all matches with available detailed statistics, but I drop matches that were abandoned before the end because of one player's injury. This leaves me with a sample of 6,018 player-set observations in 827 matches for men, and 3,816 player-set observations in 829 matches for women. The summary statistics show that there are important differences in the typology of the game, both across genders and across tournaments. These differences will probably not come as a surprise to even the casual tennis observer. Men hit on average fewer unforced errors than women, more winners and force more opponent errors. There are however large differences across tournaments: on the slow clay court surface of the French Open, it is much harder to hit winners and the percentage of unforced errors soars for both men and women. At the opposite extreme, the fast grass court surface of Wimbledon leads players to hit slightly more winners and to force more opponent errors, while unforced errors decrease. The hard court surfaces of the Australian Open and the U.S. Open stand somewhere in the middle. The percentage of first serves is similar across genders, even though there is some variation across tournaments, again with faster surfaces inducing higher first serve percentages for both men and women. Men approach the net much more frequently than women, and more so on fast surfaces.

3. Set-level Analysis: Results

I first investigate whether performance decreases systematically as the stakes increase, i.e. as the match reaches its crucial stages. I start by using the individual set as the unit of analysis, and take the combined performance of both players together as our dependent variable. I adopt a particularly simple indicator for the level of the stakes: a dummy indicator for whether the set in question is the decisive set of the match, i.e., the third set in women's matches, and the fifth set in men's matches. Therefore, letting $FINALSET_{mt}$ be a dummy indicator for whether set t in match m is the final and decisive set in the match, and stacking all the tournaments together, I run regressions of the following form, separately for men and women:

$$Performance_{mt} = \beta_0 + \beta_1 FINALSET_{mt} + \gamma' X_{mt} + c_m + u_{mt} \,,$$

where the performance variable represents the total percentage of shots of a given type played by *both* players (e.g., the percentage of unforced errors by both players out of the total number of points played in a set). The set of control variables includes a full set of tournament dummies interacted with the two players' rating, the tournament round, and the cumulative number of points played up to the beginning of the set, to capture any effects of fatigue on performance.[2] The fixed effect c_m captures any residual unobserved factor that may affect performance in match m, such as the weather, the physical health of a player, the way the two players match up against one another, and so on.

[2] Ideally, we would have wanted to control for the cumulative duration (in minutes) of the match up to the beginning of the set. Unfortunately, data on match duration are unavailable for one of the tournaments (Wimbledon 2005), so we proxy for duration with the length of the match in points. The correlation between match duration (in minutes) and match length is above 0.95 for both genders.

I estimate the equation both with and without match fixed effects. In the specification without fixed effects the identification of the coefficient of interest comes from variation in the performance variable both between and within players and matches. To account for potential within-match correlation in the error terms, standard errors are adjusted for clustering at the match level. In the specification with match fixed effects, identification comes exclusively from variation in a player's performance *within* a match, i.e., whether the quality of the player's game deteriorates in the final set relative to the previous sets. Clearly, in this specification, all the fixed player characteristics drop out of the equation, and only the final set dummy and the cumulative match length remain in the equation.

I first estimate the equation separately for the two genders, and I then stack all the data, add interactions between the female dummy and all the explanatory variables. In this stacked specification, the interaction between the final set indicator and the female dummy represents the difference between women and men in the final set effect.

The results are presented in Table 2. Each row presents results for a different dependent variable. For both men and women, performance as measured by either unforced errors or winners deteriorates in the final set. On the other hand, the percentage of first serves does not change in the final set for either men or women. The percentage of unforced errors rises by between 1.25 percent and 1.63 percent for men, and by about 2.5 percent for women. For both men and women, the percentage of winners drops by between 1.1 and 1.4 percent. These effects are not very large – the standard deviation in the percentage of unforced errors and the percentage of winners are about 12 and 8 percentage points respectively – but the effects are estimated fairly precisely. The drop in performance for women is always statistically significant at the 10 percent level, while the coefficients for men are statistically significant only in the fixed effects specification. The women's drop in performance is larger than that of men, but the differences are never statistically significant at conventional significance levels.

In Table 3 I replicate the above analysis using the player-set as the unit of observation. The performance variable represents the percentage of shots of a given type played by each player, and the fixed effects are now player-match fixed effects. This specification doubles the number of observations, and allows to relate directly performance to player ability, but one has to use caution in making statistical inference since two observations from the same set are clearly not independent. Standard errors are robust to clustering at the player-match level.

The first three rows mirror the results of the previous table, both in terms of the magnitude and the significance of the coefficients. The fourth row of the table uses as dependent variable an aggregate performance index that attempts to combine into a single measure the different aspects of a player's performance. This index is constructed by first running a regression of the percentage of points won in a set on the percentage of unforced errors, the percentage of winners, the percentage of forced opponent errors, the percentage of first serves, and the percentage of net approaches, separately by gender and tournament; then, a raw performance index is calculated by simply taking the predicted values from the above regressions; finally, the standardized performance index is calculated by rescaling the raw index so that it has mean 50 and standard deviation 10. Hence, a one-unit increase in the standardized performance index represents a one-tenth of a standard deviation increase in performance.[3] The results confirm and reinforce the findings described previously. Women's performance declines significantly in the final and decisive set, by almost one-fifth of a standard deviation. The decline in men's performance in the final set is smaller, and only statistically significant at the 10 percent level. The difference between genders is less than one-tenth of a standard deviation, and is not statistically significant.

One possible explanation for the final set effect is that it simply reflects fatigue: as the match progresses and enters the final set, players are obviously more tired, and hence are

[3] Details on the calculation of the standardized performance index are available from the author upon request.

more likely to make errors and less likely to hit winners. It should be remembered, however, that in all the regressions I control for the cumulative number of points prior to the beginning of the set, in practice a proxy measure of fatigue. Moreover, the coefficients on this variable are almost always positive and significant: if anything, performance quality increases with the duration of the match, maybe because it takes some time for players to hit their stride. In addition, the inclusion of an interaction between the final set dummy and the length of the set turns out to be insignificant: the drop in performance in the final set is not related to the duration of the match, contrary to the predictions of a fatigue-based explanation.[4]

These results indicate that both men and women perform less well in high-stakes situations; however, there is only weak evidence to suggest that women cope less well with pressure. The analysis, however, is limited because of the coarseness of our measure of pressure. There are many critical junctions in a match that occur well before the final and decisive set. A break point in the latter stages of an evenly fought early set can be more decisive for the fate of a match than a point in the early stages of the final set. Therefore, I now move to the analysis of point-by-point data, where I will be able to construct a more refined measure of the level of the stakes at each stage of the match.

4. Point-by-point data: description and the importance variable

For the last two tournaments in the sample, the 2006 French Open and Wimbledon tournaments, I was able to collect detailed point-by-point data for a selected number of matches that were played on the main championship courts and were covered by IBM's Point Tracker technology. For every single point played in these matches, I recorded who won the point, whether the first serve was in, who hit the last shot, the way the point ended (winner, unforced error, forced error, ace, double fault), and the score of the match. This data is available for a total of 123 matches: 37 men's matches and 35 women's matches at the 2006

[4] The results are essentially unchanged if one controls for the duration of the match prior to the beginning of each set in minutes, for those tournaments in which this information is available.

French Open, and 31 men's matches and 20 women's matches at Wimbledon 2006. Altogether, I have data for more than 22,000 points that were played in these matches.

One of the key objectives of the analysis using point-by-point data is to construct a measure of the importance of each point. Following Morris (1977), and Klaassen and Magnus (2001), I define the importance of a point as the probability that player 1 wins the match conditional on him or her winning the current point minus the probability that player 1 wins the match conditional on him or her losing the current point. It is immediate to see that the importance of a point from the perspective of player 2 is exactly identical to the importance of a point from the perspective of player 1. To calculate the importance of each point, I assume that in every match there is an associated fixed probability of each player winning a point, which depends on the gender, the playing surface, the two players' ability ratings and on the identity of the server. These fixed probabilities are calculated using the full 2005 French Open and Wimbledon data by running regressions of the proportion of points won by the server (the receiver) in each match as a function of the rating of the two players, separately by tournament and gender. These probabilities are then fed into a computer program that takes into account the structure of a match in a Grand Slam tournament, and calculates recursively, for every pairing of players, the probability of winning the match at every possible stage. From this it is then possible to calculate the importance of each point.

Table 4 presents summary statistics for the importance variable, separately by tournament and gender. The mean of the importance variable is 0.0231 for men, and 0.0288 for women, reflecting the fact on average points in the women's game are more important, given that matches are played in the "best-of-three" sets format, rather than "best-of-five." The distribution of the importance variable is heavily skewed to the right, indicating that most points played in a tennis match have relatively little potential to significantly affect the fate of a match. Figures 1 and 2 show the evolution of the importance variable over the course of the four finals played in the 2006 French Open and Wimbledon tournaments. Note how the importance variable evolves in a very nonlinear fashion: importance tends to rise towards the latter stages of each set, but only if the set is evenly fought. There are a number of clusters of high importance points even in the early sets and in the latter stages of the late sets. Most of the spikes in importance are associated with break points. This is particularly true in the men's tournament at Wimbledon, where the fast playing surface means break points are relatively rare, and hence can change the direction of a match substantially.

Summing up, there is substantial variation in the importance measure both across matches and within matches, meaning that one should be able to detect variation in performance that depends on the degree of competitive pressure.

5. Point-by-point analysis: results

Table 5 the typology of points played by men and women, split by quartiles of the importance variable. The results are fairly striking. For men, there appears to be no systematic pattern in the typology of points by importance quartile: the percentage of unforced errors hovers between 30 and 31 percent, regardless of the importance of the point. About three fifths of these unforced errors are made by the server, and the remainder by the receiver. On the other hand, the typology of points in women's matches is strongly affected by the importance variable. In the bottom quartile of the importance distribution, point types are distributed fairly uniformly, and this distribution is not too dissimilar from that of men. However, as the importance of the points grows, women commit a growing number of unforced errors, with the percentage of winners and forced errors falling, especially the latter. In the top quartile of the importance distribution, the percentage of unforced errors reaches nearly 41 percent, nearly eight percentage points higher than what it was in the bottom quartile.

At a first glance, these results suggest that men and women react differently to increases in competitive pressure, with women exhibiting a lower level of performance as the stakes become higher. These results could of course be due to composition effects: maybe the more important points are disproportionately more likely to involve low-ranked players (who are more likely to commit unforced errors), or are more likely to be played at the French Open, where unforced errors are more frequent. To address these concerns, I proceed to a multinomial logit analysis for the typology of the point.

Specifically, define Y_{im} as the outcome of point i in match m. Y_{im} can take on three possible values: 1 – forced error, 2 – winner, 3 – unforced error. I estimate the following multinomial logit model:

$$P\left(Y_{im}=k\right)=\frac{\exp\left(\alpha_k+\beta_k IMP_{im}+\gamma_k{}' X_{im}+c_{mk}\right)}{1+\sum_{k'=2}^{3}\exp\left(\alpha_{k'}+\beta_{k'} IMP_{im}+\gamma_{k'}{}' X_{ik'}+c_{mk}\right)}, \qquad k=2,3.$$

The main coefficient of interest is β, the coefficient on the importance variable. The set of control variables includes the server rating, the receiver rating, set dummies, a Wimbledon dummy, and the interaction of the Wimbledon dummy with all of preceding variables. The base category is forced errors: this implies that β represents the increase in the log odds of unforced errors or winners relative to forced errors when the importance variable increases by one unit. In addition to this basic specification, I also estimate two additional models (quadratic and piecewise constant) to detect potential nonlinearities in the effect of importance. The model is estimated with and without match fixed effects. In the model without match fixed effects, identification comes from variation in importance both between and within matches, while inclusion of fixed effects implies that the parameters of interest are identified solely off the variation in the importance variable within matches.

The results of the multinomial logit analysis are shown in Table 6. The top panel presents the coefficients on the importance variables in the three models (linear, quadratic and piecewise constant) for the "winners" equation, and the bottom panel presents the coefficient for the "unforced errors" equation.

In the linear model, none of the coefficients on the importance variable are statistically significant, although the pattern of signs suggests that women are more likely to make unforced errors as importance grows. The quadratic equation reveals that there are some meaningful non-linearities in the effect of importance that were overlooked in the linear specification. For men, both the linear and the quadratic terms in the unforced error equation are insignificant, while for women both coefficients are significant at the 5 percent level, even after controlling for match fixed effects. The positive coefficient on the linear term and the negative coefficient on the quadratic term imply that the log odds of unforced errors relative to forced errors increases as importance grows, but at a declining pace. The turning point in the quadratic equation occurs at importance levels between 0.11 and 0.13, beyond the top 95th percentile of the importance distribution. The gender differences in the coefficients of the "unforced error" equation are statistically significant at the 10 percent level. The coefficients in the "winners" equation for women have a similar pattern, implying that the odds of a point ending in a winner also increase with importance. However, the coefficients are smaller and less significant than in the "unforced errors" equation, and the difference between men and women is not statistically significant.

The piecewise constant specification confirms and reinforces the results of the quadratic specification. In particular, we find that there is a significant difference between genders in the propensity to make unforced error in the top quartile of the importance distribution. This difference stems from an increased propensity to make errors by women in the no fixed-effects specification, and from a decreased propensity by men in when fixed match effects are controlled for. By contrast, there appears to be no significant relationship between the propensity to hit winners and the importance quartile for both genders.

Summing up, these results indicate that, even after controlling for player characteristics, set dummies, and match fixed effects, women are substantially more likely to make unforced errors at crucial junctions of the match. In the next section, we try to characterize further the determinants of player performance as the stakes increase, focusing in particular on heterogeneity in the response of players of different abilities.

5. Heterogeneous Effects

I now return to the set-level analysis, and I explore whether there are any differences in final set performance between players of different abilities. I add to the basic specification of Tables 2 and 3 the interactions between the final set dummy and own player rating, and between final set and opponent rating. The interactions are rescaled so that the main effect can be interpreted as the final set effect for the number 50 ranked player in the world playing the number 51 ranked player. I focus on just the percentage of unforced errors and the standardized performance index as the dependent variables.

The results are presented in Table 7. The representative player of both genders performs less well in the final set, especially in the specification that includes player-match fixed effects. Final set performance is strongly correlated with own player ranking (for men), and with both own and opponent ranking (for women). The slope of the performance function with respect to own rating is larger for women than for men, indicating that the differences in the ability to withstand pressure between high and low ranked players is more pronounced among women. The results imply that the top 11 men players and the top 15 women players are actually able to raise their performance in the decisive set of a match. Nevertheless, the gender differences in the coefficients are for the most part statistically insignificant.

I next use the point-by-point data and examine whether players of different abilities perform differently at important points in the match. To distinguish between the performance of the server and the receiver as a function of their abilities, I estimate a multinomial logit model for the *six* possible ways in which a point can end: a winner by the server, a winner by the receiver, an unforced error by the server, an unforced error by the receiver, a forced error by the server, and a forced error by the receiver (the base category). The importance variable is introduced linearly (as in specification A of Table 6), and interacted with the rescaled rating of the server and the receiver. To avoid cluttering the table with too many numbers, I present just the coefficients on the importance variables for the "unforced errors" equations (by both the server and the receiver).

The results are presented in Table 8. In the specification without fixed effects, we find that the representative woman commits more unforced errors (both when serving and when receiving) as the importance of the point grows. As in the set-level analysis, there is evidence that higher ability women are less likely to be negatively affected by the importance of the point: see the negative and significant coefficient on the importance×server rating interaction in the unforced errors by server equation, and the negative and significant coefficient on the importance×receiver rating in the unforced errors by receiver equation. For men, the evidence is more mixed, and most of the coefficients are not statistically significant. When one includes match fixed effects, all the coefficients in the women equation become insignificant, even though the pattern of signs is preserved.

Altogether, the results in this section point to higher ability players being less likely to make unforced errors at crucial stages of the match, with this tendency being more pronounced for women. This is not entirely surprising: one of the characteristics of top players is that they cope well with pressure and do not experience a drop in performance when the stakes are high.

7. Conclusion

In this paper I have used data from seven Grand Slam tournaments played between 2005 and 2006 to assess whether men and women respond differently to competitive pressure in a real-world setting with large monetary rewards. The aggregate set-level data reveals that the performance of both men and women deteriorates in the final and decisive set. Women's decline in performance is more pronounced than that of men, but the difference is not statistically significant. On the other hand, the analysis using detailed point-by-point data indicates that there are significant differences between men's and women's performances at crucial junctions of the match: the propensity of women to commit unforced errors increases significantly with the importance of the point, while men's propensity to commit unforced errors is unaffected by point importance.

To what extent then can we draw from this study more general lessons about gender differences in the labor market? An unforced error is by definition an error that cannot be attributed to any factor other than poor judgment by the player. Can we extrapolate from our findings that in general women's judgment becomes more clouded as the stakes become higher, and this may hinder their advancement to the upper echelons of management, science, and the professions? Clearly, the answer must be negative. The results are only relevant for the specific context, and it is questionable whether the conclusions can be even extended to athletes in other sports, let alone to managers, surgeons, or other professionals who must make quick and accurate decisions in high pressure situations.

Nevertheless, there are at least two striking features in this study that still deserve attention. First, the women in our sample are among the very best in the world in their profession, and are without question extremely competitive. In this respect, they are probably quite distant from the typical woman in experimental studies, which underperforms in competitive settings and shies away from competition. Therefore, it is doubly surprising that even these highly competitive women exhibit a decline in performance in high pressure situations. Second, some experimental studies (e.g., Gneezy, Niederle, and Rustichini, 2003) found that women's tendency to underperform in competitive environments occurs only when they compete against men. By contrast, here we find that women's performance deteriorates as competitive pressure rises, even when the competition is clearly restricted to women alone. This may have implications for educational policies such as single-sex schooling, and deserves further investigation.

References

Gneezy, Uri; Niederle, Muriel, and Rustichini, Aldo. “Performance in Competitive Environments: Gender Differences.” *Quarterly Journal of Economics*, August 2003.

Gneezy, Uri and Rustichini, Aldo. “Gender and Competition at a Young Age”. *American Economic Review,* May 2004.

Lavy, Victor. “Do Gender Stereotypes Reduce Girls’ Human Capital Outcomes? Evidence from a Natural Experiment.” NBER Working Paper No. 10678, August 2004.

Klaassen, Franc J.G.M., and Magnus, Jan R. “Are Points in Tennis Independent and Identically Distributed? Evidence from a Dynamic Binary Panel Data Model.” *Journal of the American Statistical Association*, June 2001.

Morris,C. “The Most Important Points in Tennis,”in *Optimal Strategies in Sport*, eds. S.P.Ladany and R.E.Machol, 1977. Amsterdam: North-Holland Publishing Company, 131-140.

Niederle, Muriel, and Vesterlund, Lise. “Do Women Shy Away from Competition? Do Men Compete Too Much?” NBER Working Paper No. 11474, June 2005.

Sunde, Uwe. “Potential, Prizes and Performance: Testing Tournament Theory with Professional Tennis Data.” IZA Discussion Paper No. 947, December 2003.

Table 1: Summary Statistics
Set level data

	Men								Women							
	Total	Austr. Open 2005	French Open 2005	Wimb. 2005	US Open 2005	Austr. Open 2006	French Open 2006	Wimbl. 2006	**Total**	Austr. Open 2005	French Open 2005	Wimb. 2005	US Open 2005	Austr. Open 2006	French Open 2006	Wimb. 2006
Number of matches	827	126	124	125	81	127	125	119	829	126	127	127	78	126	127	118
Number of sets	6,018	906	902	888	616	944	898	864	3,816	586	592	594	348	570	580	546
Pct. unforced errors	16.42 (8.39)	16.62 (7.42)	22.45 (9.54)	11.53 (5.75)	16.50 (7.17)	17.31 (7.63)	19.50 (8.01)	10.67 (5.54)	20.43 (8.87)	21.59 (8.15)	25.01 (9.15)	15.61 (7.14)	20.45 (7.66)	23.17 (8.95)	22.42 (8.21)	14.46 (6.43)
Pct. winners	17.10 (6.67)	16.31 (7.04)	16.58 (6.46)	17.83 (6.29)	17.32 (7.19)	16.53 (6.53)	17.55 (6.56)	17.71 (6.55)	14.92 (7.23)	14.52 (7.29)	14.76 (7.14)	15.49 (6.69)	14.06 (7.75)	13.17 (6.74)	16.69 (7.41)	15.40 (7.25)
Pct. forced opponent errors	16.49 (6.98)	17.06 (5.94)	10.97 (6.32)	20.64 (6.12)	16.18 (5.38)	16.16 (5.64)	12.95 (6.09)	21.62 (6.24)	14.65 (7.34)	13.90 (6.05)	10.23 (6.90)	18.90 (6.66)	15.49 (5.90)	13.66 (5.88)	10.89 (6.00)	20.14 (6.86)
Pct. first serve	60.83 (10.93)	59.72 (10.54)	58.77 (11.00)	61.68 (10.63)	58.97 (10.75)	59.09 (11.34)	64.13 (10.79)	63.06 (10.09)	61.82 (11.34)	60.24 (11.12)	59.12 (11.31)	62.44 (11.31)	61.64 (10.52)	60.16 (11.49)	65.27 (11.27)	63.96 (10.74)
Pct. net approaches	14.66 (9.19)	14.06 (9.74)	13.24 (7.38)	16.73 (9.31)	14.77 (10.6)	15.62 (10.6)	13.35 (7.62)	14.89 (8.91)	9.51 (6.64)	8.94 (6.06)	8.60 (5.64)	10.87 (7.73)	8.93 (6.47)	9.51 (6.74)	9.03 (5.83)	10.50 (7.42)
Average player rank	84.97	84.58	83.33	88.05	81.73	85.01	84.39	86.57	80.02	81.22	85.05	85.62	68.20	79.05	76.86	79.79
Average player rating	2.19	2.21	2.19	2.12	2.42	2.11	2.20	2.17	2.24	2.17	2.17	2.15	2.64	2.25	2.25	2.21

Note: Data refers to all completed matches for which detailed statistics are available. Standard deviations in parentheses.

Table 2: The Effect of the Decisive Set on Performance
Unit of Observation: Set

	Individual Controls			Individual controls and match fixed effects		
Dependent variable:	Men	Women	Difference	Men	Women	Difference
Pct. unforced errors	1.2562 [1.56]	2.5759 [3.03]	1.3196 [1.13]	1.6323 [2.11]	2.5339 [3.09]	0.9016 [0.80]
Pct. winners	-1.1151 [-1.66]	-1.3767 [-2.09]	-0.2615 [-0.28]	-1.3101 [-2.01]	-1.1998 [-1.85]	0.1103 [0.12]
Pct. first serve	-0.3828 [-0.57]	-0.2654 [-0.38]	0.1176 [0.12]	0.2744 [0.42]	-0.7212 [-1.04]	-0.9956 [-1.05]
Match fixed effects	No	No	No	Yes	Yes	Yes
Number of observations (sets)	3009	1908	4917	3009	1908	4917

Note: Entries in the table represent the coefficient on the "decisive set" dummy, t-statistics in parentheses. All regressions control for a full set of tournament dummies interacted with: a constant, the high-ranked player's rating, the low-ranked player's rating, the cumulative number of points played up to the beginning of the set, and the tournament round. Standard errors are corrected for clustering at the match level.

Table 3: The Effect of the Decisive Set on Performance
Unit of Observation: Player-Set

	Individual Controls			Individual controls and match fixed effects		
Dependent variable:	Men	Women	Difference	Men	Women	Difference
Pct. unforced Errors	0.6281 [1.40]	1.3197 [2.70]	0.6916 [1.04]	0.8162 [1.91]	1.2670 [2.68]	0.4508 [0.71]
Pct. Winners	-0.5516 [-1.45]	-0.7112 [-1.78]	-0.1596 [-0.29]	-0.6550 [-1.83]	-0.5999 [-1.62]	0.0552 [0.11]
Pct. first Serve	-0.3933 [-0.57]	-0.4234 [-0.58]	-0.0301 [-0.03]	0.3121 [0.47]	-0.8307 [-1.18]	-1.1428 [-1.19]
Standardized Performance Index	-1.0536 [-1.68]	-1.8357 [-3.26]	-0.7821 [-0.93]	-1.1715 [-1.89]	-1.7061 [-3.15]	-0.5346 [-0.65]
Match fixed effects	No	No	No	Yes	Yes	Yes
Number of observations (player-sets)	6018	3816	9834	6018	3816	9834

Note: Entries in the table represent the coefficient on the "decisive set" dummy, t-statistics in brackets. All regressions control for a full set of tournament dummies interacted with: a constant, the own player's rating, the opposite player's rating, the cumulative number of points played up to the beginning of the set, and the tournament round. Standard errors are corrected for clustering at the player-match level.

Table 4: The importance of points – summary statistics

	Men			Women			ALL
	French Open 2006	Wimbledon 2006	All	French Open 2006	Wimbledon 2006	All	
Mean	0.0220	0.0243	0.0231	0.0316	0.0237	0.0288	**0.0250**
Standard deviation	0.0299	0.0329	0.0314	0.0411	0.0343	0.0390	**0.0342**
25th percentile	0.0023	0.0030	0.0027	0.0052	0.0013	0.0033	**0.0029**
50th percentile	0.0116	0.0129	0.0122	0.0170	0.0092	0.0144	**0.0128**
75th percentile	0.0305	0.0322	0.0314	0.0440	0.0314	0.0403	**0.0340**
Number of points	7,776	7,076	14,852	4,728	2,587	7,315	**22,167**

Note: The importance of a point is defined as the probability that player 1 wins the entire match conditional on him/her winning the current point, minus the probability that player 1 wins the entire match conditional on him/her winning the current point. See text for details.

Table 5: Typology of points, by importance quartiles

		Men				Women			
		Number of points	Pct. winners	Pct. unforced	Pct. forced errors	Number of points	Pct. winners	Pct. unforced	Pct. forced errors
Importance quartile 1 (least important)		**3857**	**33.42**	**31.45**	**35.13**	**1677**	**33.04**	**33.21**	**33.75**
	Server		22.71	17.68	8.92		19.08	19.38	10.67
	Receiver		10.71	13.77	26.21		13.95	13.83	23.08
Importance quartile 2		**3756**	**34.56**	**30.01**	**35.44**	**1780**	**32.02**	**37.87**	**30.11**
	Server		24.23	17.17	9.45		18.54	21.85	10.45
	Receiver		10.33	12.83	25.99		13.48	16.01	19.66
Importance quartile 3		**3839**	**33.21**	**31.39**	**35.40**	**1693**	**29.65**	**39.87**	**30.48**
	Server		23.29	17.90	9.61		18.02	22.33	9.39
	Receiver		9.92	13.49	25.79		11.64	17.54	21.09
Importance quartile 4 (most important)		**3378**	**33.87**	**29.96**	**36.18**	**2155**	**31.97**	**40.93**	**27.10**
	Server		23.09	17.85	9.59		19.63	22.55	8.35
	Receiver		10.78	12.11	26.58		12.34	18.38	18.75

Note: The importance quartiles are based on the overall distribution of the importance variable.

Table 6: The Effect of Importance on Performance
Multinomial Logistic Regression

	Individual Controls			Individual controls and match fixed effects		
	Men	Women	Difference	Men	Women	Difference
Winners						
A: Linear						
Importance	-0.4026	0.0355	0.4381	-0.6946	-0.5104	0.1843
	[-0.58]	[0.04]	[0.38]	[-0.89]	[-0.47]	[0.14]
B: Quadratic						
Importance	-0.0426	2.5016	2.5442	-0.4499	2.9418	3.3917
	[-0.03]	[1.67]	[1.31]	[-0.31]	[1.45]	[1.36]
Importance	-2.4418	-12.377	-9.9354	-1.5103	-14.175	-12.664
squared	[-0.36]	[-1.95]	[-1.07]	[-0.20]	[-1.89]	[-1.20]
C: Piecewise constant						
Importance	0.0465	-0.0576	-0.1042	-0.0321	0.0602	0.0923
quartile 2	[0.83]	[-0.66]	[-1.00]	[-0.46]	[0.52]	[0.69]
Importance	0.0165	-0.1362	-0.1527	-0.0700	-0.1038	-0.0338
quartile 3	[0.29]	[-1.51]	[-1.43]	[-0.89]	[-0.76]	[-0.21]
Importance	-0.0034	0.0508	0.0541	-0.0716	0.0601	0.1317
quartile 4	[-0.06]	[0.56]	[0.49]	[-0.86]	[0.40]	[0.77]
Unforced Errors						
A: Linear						
Importance	-0.7732	0.6394	1.4126	-0.9246	0.3244	1.2490
	[-1.05]	[0.74]	[1.24]	[-1.10]	[0.31]	[0.92]
B: Quadratic						
Importance	-0.7290	4.2257	4.9547	-1.4549	4.0979	5.5528
	[-0.58]	[2.83]	[2.53]	[-0.96]	[2.07]	[2.22]
Importance	-0.2555	-18.578	-18.323	3.2290	-15.448	-18.677
squared	[-0.04]	[-2.78]	[-1.88]	[0.43]	[-2.13]	[-1.78]
C: Piecewise constant						
Importance	-0.0277	0.0488	0.0765	-0.1528	0.0214	0.1742
quartile 2	[-0.48]	[0.56]	[0.73]	[-2.06]	[0.19]	[1.29]
Importance	0.0147	0.0761	0.0614	-0.1346	-0.0253	0.1094
quartile 3	[0.25]	[0.86]	[0.58]	[-1.64]	[-0.19]	[0.70]
Importance	-0.0831	0.2077	0.2908	-0.2139	0.1477	0.3616
quartile 4	[-1.32]	[2.31]	[2.65]	[-2.42]	[1.00]	[2.11]
Match fixed effects	No	No	No	Yes	Yes	Yes
Number of observations (points)	14,830	7,305	22,135	14,830	7,305	22,135

Note: Entries in the table are the coefficients in a multinomial logit model for the typology of points (winners/ unforced errors/ forced errors) on the importance quartile dummies. Additional control variables: set dummies, server’s rating, receiver’s rating, and a Wimbledon 2006 dummy interacted with all of the above. The base category is “forced errors.” T-statistics in parentheses.

Table 7: Differential Effects of the Decisive Set by Player Quality

Set-level analysis

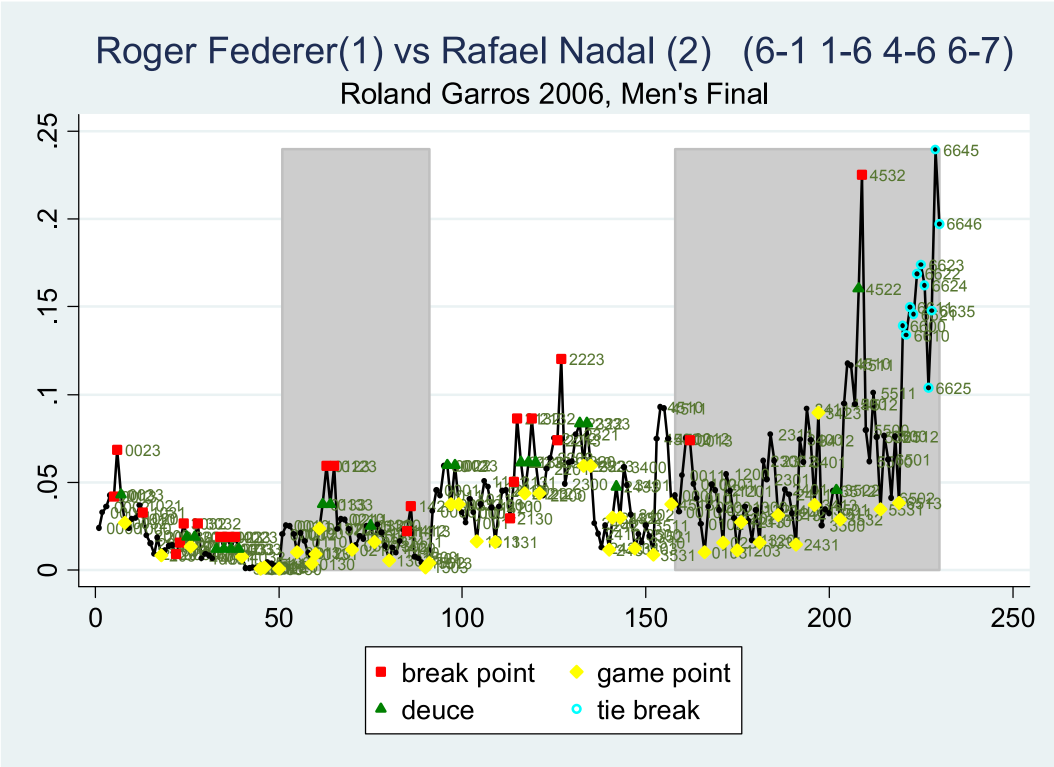
	Individual Controls			Individual controls and match fixed effects		
	Men	Women	Difference	Men	Women	Difference
Unforced errors						
Decisive set	0.7154 [1.48]	1.3569 [2.68]	0.6415 [0.92]	0.9601 [2.10]	1.2705 [2.63]	0.3104 [0.47]
Decisive set × (own rating – $rating_{50}$)	0.1129 [0.45]	-0.2669 [-1.36]	-0.3798 [-1.19]	-0.4845 [-1.94]	-0.7150 [-3.73]	-0.2305 [-0.73]
Decisive set × (opponent rating – $rating_{51}$)	-0.3214 [-1.36]	0.1729 [0.82]	0.4943 [1.56]	0.0890 [0.40]	0.6666 [3.31]	0.5776 [1.94]
Standardized performance index						
Decisive set	-1.2207 [-1.84]	-1.9886 [-3.40]	-0.7678 [-0.87]	-1.4041 [-2.17]	-1.7867 [-3.21]	-0.3826 [-0.45]
Decisive set × (own rating – $rating_{50}$)	-0.4688 [-1.48]	-0.1344 [-0.58]	0.3344 [0.85]	0.6736 [2.09]	1.0360 [4.68]	0.3625 [0.93]
Decisive set × (opponent rating – $rating_{51}$)	0.8501 [2.63]	0.4475 [1.97]	-0.4026 [-1.02]	-0.0422 [-0.13]	-0.8127 [-3.70]	-0.7705 [-1.97]
Match fixed effects	No	No	No	Yes	Yes	Yes
Number of observations	6018	3816	9834	6018	3816	9834

Note: The unit of observation is the player-set. Entries in the table represent the coefficients on the "decisive set" dummy and its interactions with the player's rating and the opponent's rating. T-statistics in brackets. All regressions control for a full set of tournament dummies interacted with: a constant, the own player's rating, the opposite player's rating, the cumulative number of points played up to the beginning of the set, and the tournament round. Standard errors are corrected for clustering at the player-match level.

Table 8: Differential Effects of Importance by Player Quality
Point-by-point

	Individual Controls			Individual controls and match fixed effects		
	Men	Women	Difference	Men	Women	Difference
Unforced errors by server						
Importance	-1.7000	3.7172	5.4172	-0.5518	1.1870	1.7389
	[-1.28]	[1.85]	[2.25]	[-0.38]	[0.49]	[0.61]
Importance × (server rating – $rating_{50}$)	-0.0198	-1.1236	-1.1038	-0.8435	-0.9809	-0.1374
	[-0.04]	[-2.02]	[-1.55]	[-1.81]	[-1.57]	[-0.18]
Importance × (receiver rating – $rating_{51}$)	1.1792	-0.3698	-1.5490	1.6508	1.0049	-0.6459
	[2.63]	[-0.62]	[-2.08]	[3.54]	[1.52]	[-0.80]
Unforced errors by receiver						
Importance	-3.2313	5.8062	9.0375	-2.9146	1.4598	4.3744
	[-2.10]	[2.70]	[3.41]	[-1.71]	[0.56]	[1.41]
Importance × (server rating – $rating_{50}$)	0.4137	-1.1516	-1.5653	-0.0013	-0.9731	-0.9718
	[0.78]	[-1.87]	[-1.93]	[-0.00]	[-1.39]	[-1.10]
Importance × (receiver rating – $rating_{51}$)	0.7713	-1.5835	-2.3549	0.8010	0.2782	-0.5228
	[1.48]	[-2.43]	[-2.82]	[1.50]	[0.38]	[-0.58]
Match fixed effects	No	No	No	Yes	Yes	Yes
Number of observations	14,830	7305	22,135	14,830	7305	22,135

Note: Entries in the table are the coefficients in a multinomial logit model for the six possible ways in which a point can end (winner by server / winner by receiver / unforced error by server / unforced error by receiver / forced error by server / forced error by receiver) on the linear importance measure, and its interactions with the server rating and the receiver's rating. Additional control variables: set dummies, server's rating, receiver's rating, and a Wimbledon 2006 dummy interacted with all of the above. The base category is "forced errors by receiver." T-statistics in brackets.



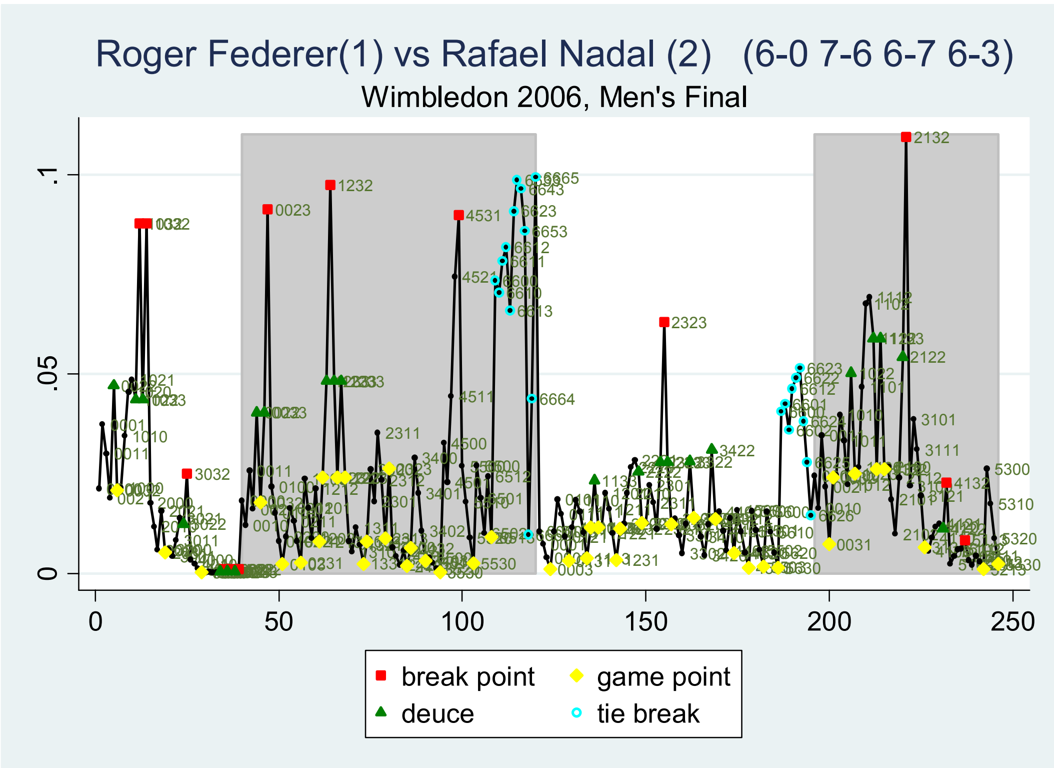
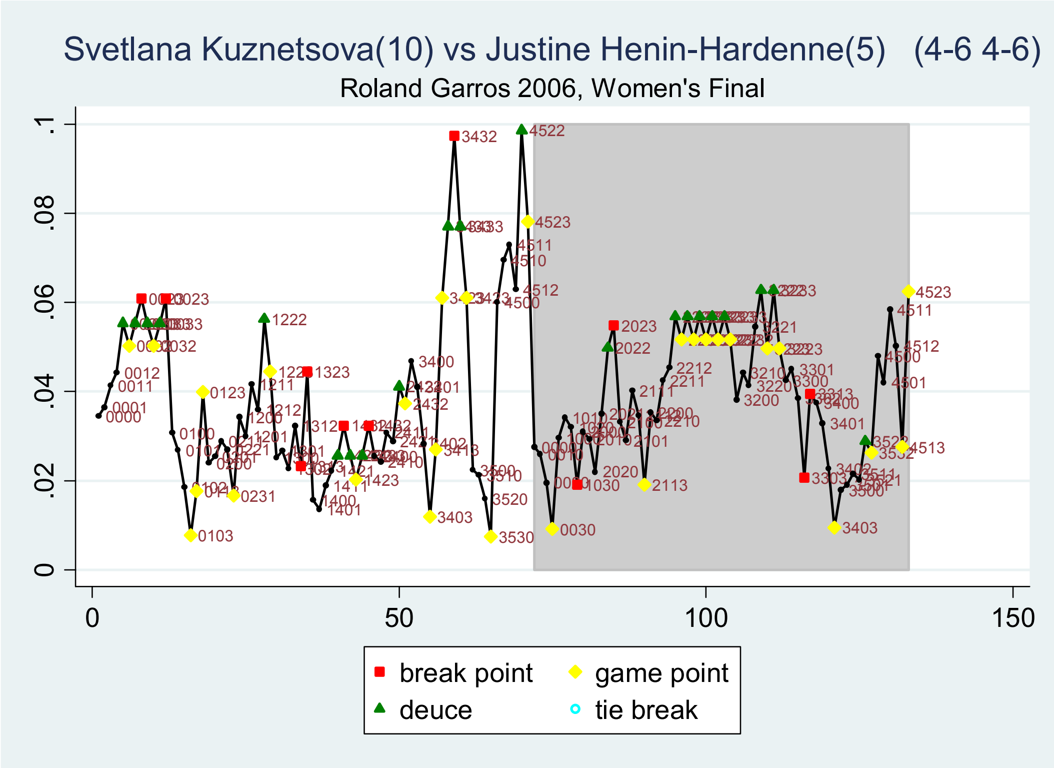


Figure 1: The evolution of importance over the course of selected matches – men



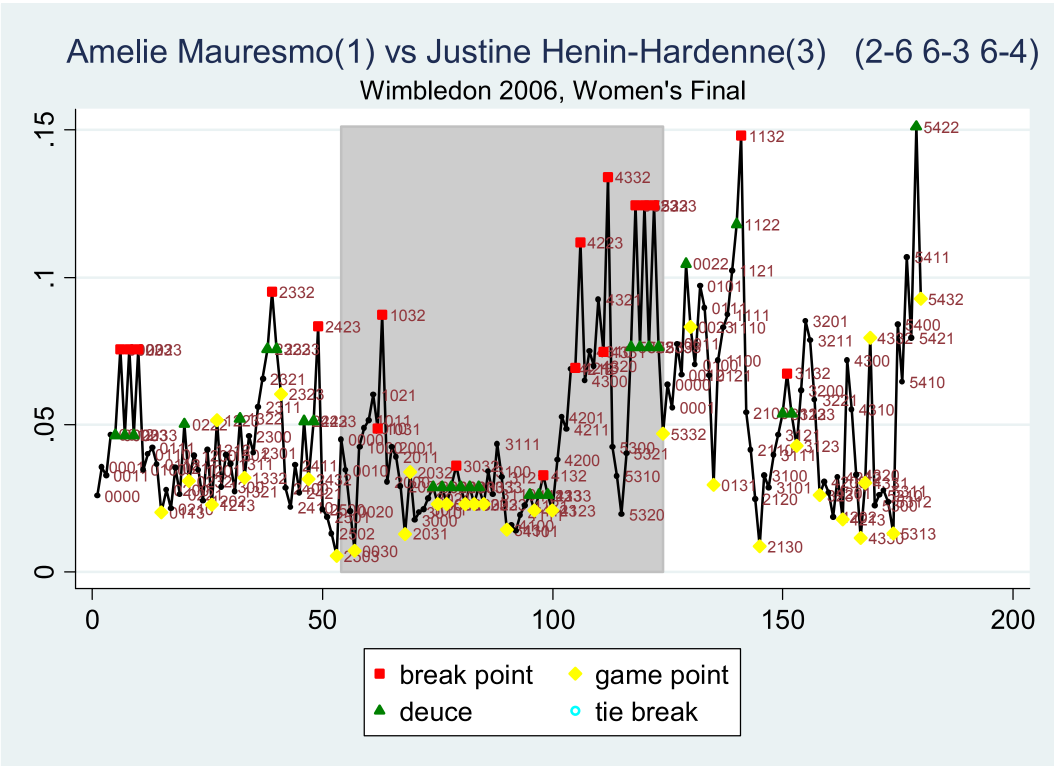


Figure 2: The evolution of importance over the course of selected matches – women