Performing Best when it Matters Most: Evidence from professional tennis[1]

Julio González-Diaz, Olivier Gossner, and Brian W. Rogers[2]

[PRELIMINARY]

November 11, 2009

Abstract: Psychological skills impact performance. This suggests that when analyzing behavior, in addition to cognitive abilities and effort, one should also take into account, for instance, the pressure agents might face when making their decisions. Existing literature has provided important evidence suggesting that even highly qualified agents may perform worse when the stakes are very high or when they confront unusually high pressure. In this paper we study the degree of heterogeneity in agents' reactions to pressure. More precisely, we identify relative critical abilities with which professional tennis players' performance responds to the importance of a point. Adopting a natural definition of importance, the scoring rules of tennis generate significant variation in point importance. Using data from 12 years of the US Open tournament, we estimate the critical abilities for players in the tournaments controlling for serving and returning strengths. We then show that there is significant heterogeneity in critical abilities across players. Moreover, looking at professional rankings we find that this critical ability has a substantial impact on a player's career. These results suggest that when analyzing human behavior, beyond accounting for the fact that pressure may affect performance, it is also important to recognize that this effect may depend on the idiosyncrasies of each agent.

[1]We thank Sven Feldman for helpful comments, as well as audiences at Northwestern University and Malaga University.

[2]Gonzalez-Diáz: University of Vigo, Gossner: Paris School of Economics and London School of Economics, Rogers: MEDS, Kellogg School of management, Northwestern University.

1. Introduction

We are interested in dynamic environments in which agents interact, exert effort, and make decisions. A key feature is that at certain moments, the outcomes are more important for agents' welfare than at other moments. How do people respond to changes in importance, and to what extent do the responses vary across individuals? What are the economic implications of such responses?

To be clear from the outset, one of our goals is to distinguish between two different notions of skill. The first is a general ability to make good decisions, resulting in relatively good outcomes on average. The second is an ability to differentially improve one's performance in more important situations, without necessarily improving one's average performance. This ability will also generate better outcomes whenever it is not possible to achieve peak performance consistently. We focus primarily on measuring this second skill, which we call *critical ability.*

There are at least two reasons why it may not be possible to perform optimally in all situations, and thus why critical ability may be a significant determinant of outcomes. The first is that an agent may have limited resources to allocate towards a sequence of decisions. In this case, performing better in one situation comes at the opportunity cost of performing better in other situations. Critical ability then translates into skill in allocating resources optimally across decisions. The second reason is that important situations may involve feelings of high pressure which can affect performance [references], and in this case critical ability corresponds to an agents psychological ability to respond well to pressure. In either case, individuals are likely to react imperfectly to changes in importance. We study the ability of individuals to perform better in more important situations.

The suggestion that psychological factors are relevant for determining human decision making goes back to Hume (1739). Recently, there is increasing empirical evidence that psychology plays a role in performance across many settings. There is a great body of experimental work in both the psychology and economics litereatures demonstrating the importance of psychology in human decision making and performance.

Two recent papers argue convincingly that psychological responses to pressure affects performance. Apesteguia and Palacios-Huerta (2008) show that kicking first in professional soccer penalty shoot-outs results in a win rate exceeding 60%. Absent pyschological effects, kicking first would confer no advantage, and one would expect equiprobable outcomes. Given that the assignment of kicking order is random, this is strong evidence that psychology plays a role in determining outcomes. Dohmen (2008) demonstrates that penalty kicks during soccer matches exhibit decreased performance in the highest pressure situations. Absent psychological effects, performance should increase with the importance of the situation, so again the conclusion is that psychology affects performance.

Apesteguia and Palacios-Huerta (2008) point out that risk attitudes and the one-shot nature of penalty shoot-outs is crucial for their identification. This is because they are concerned specifically with the magnitude of psychological effects. A key departure of our work is that, as mentioned above, we seek to identify individuals' response to situations of varying importance, be it through psychological effects, resource allocation, or other means.

The fact that such effects exist and are important is clearly established. We focus instead on the effects of heterogeneity in individuals' abilities to respond to important situations. In our context, it is precisely the heterogeneity in critical abilities that is identifiable. As such, we are able to ask to what extent such heterogeneity exists, and how important critical ability is for determining individual-level outcomes.

Critical ability is likely to play a role in many contexts. Consider financial brokers who must make trading decisions quickly and repeatedly. Some decisions will have a steeper risk-reward tradeoff, and will thus be more important for long run success. How much of a trader's success is explained by an ability to adjust performance to such high pressure situations? One could also consider political campaigns, in which there are many decisions to make and performances to give, some of which will be of greater importance, such as a televised debate. Are our elected officials individuals who generally perform well, or are they simply skilled at delivering their best performance at the critical moment?

We explore these questions of general interest in a very specific context: professional mens tennis matches from the US Open, one of the four annual Grand Slam titles on the pro tour. While we are not interested in tennis matches per se, this setting offers a number of rather unique advantages. First, the stakes are very high. The agents are elite, well-trained, and highly motivated. For relatively low-ranked players, a good showing in a tournament can revolutionize their career. For top players, the amount of purse money and prestige on the line is substantial. Perhaps even more significantly, the outcomes will affect their rankings and ability to secure lucrative sponsorships.

Second, point-level tennis data allows for a particularly precise paired measurement of outcomes and importance. There is a clear winner of each and every point in the match. There is also a very natural notion of importance that can be attached to each point, which is discussed in detail below. Thus we have high quality information and an unambiguous means of measuring players' performance.

Third, in a single tournament, each participant plays many points, and the points typically vary significantly in their importance in determining the outcome of the match. Given the size of our data, it becomes possible to identify the critical abilities of players in a precise way. Our first main result is that these abilities indeed differ significantly across players. That is, the importance of a point in the match has a non-trivial effect on the outcome controlling, of course, for other relevant factors. In a match between players i and j, if i's critical ability is higher than j's, then the probability that i wins any given point is increasing in the importance of the point for determining the outcome of the match.

Second, given individual level estimates of players' critical abilities, we ask how important these abilities are, in the sense of how strongly they contribute to determining economic outcomes for the players. Specifically, we are interested in the relationships between a player's critical ability and his professional tennis ranking and rating. To interpret these results, it is important to recall that a higher critical ability does not correspond to winning more points on average. Instead, it corresponds to being more likely to win more important points, and less likely to win less important points, all else equal. Thus this relationship is separate from the conventional notion of simply being a stronger player, in terms of being more likely to win any given point. We find that a one standard deviation increase in critical ability is equivalent to 40% of a standard deviation increase in the likelihood of winning a point, and thus that critical ability is a qualitatively important determinant of players' outcomes.

The remainder of the paper proceeds as follows. The next section formally defines the notion of the importance of points. Section 3 explains our econometric approach. The main results are described in Section 4. Section 5 concludes.

2. The data

Table 1: Descriptive statistics of the data

	Principle data	Reduced data	Secondary data
Tournaments	12	12	117
Grand Slams	12	12	36
Players	355	94	766
Matches	1,009	494	9,013
Points	223,139	110,033	2,137,238

We have point by point data from twelve U.S. Open tournaments from 1994-2006. The U.S. Open is one of the four major professional Grand Slam tournaments of the year. We focus on the men singles competition, which includes, collectively, 1009 matches. Table provides descriptive statistics on the data under the column"Principle data." For reasons discussed below, we use a subset of the data to perform our main econometric analysis, and we require an auxiliary data set in order to measure players' relative strengths; these data sets are described in the last two columns.

3. Importance of points

Heterogeneity in the importance of points within a match is the vehicle we use to identify critical abilities of the players. This heterogeneity arises under a very natural definition first proposed by Morris (1977), and employed in Klaasen and Magnus (2001) and Paserman (2008). The fundamental idea behind this definition is that a point's importance represents the extent to which the (stochastic) outcome of the match hinges on the outcome of the given point. This section formally develops this definition.

To understand the notion of importance, it is necessary to first describe the rules by which tennis is scored. The US Open is a single elimination tournament consisting of matches between pairs of players. Each player's objective is to win the match. In a match, the two players alternate service games, with a randomly chosen player electing who serves first. The winner of a game is the first player to reach four points, with a margin of victory of at least two points.[1] The first player to win six games in this fashion is awarded a set. Again, a margin of two games

[1] The game continues for as many points as necessary to declare a winner.

is necessary to win a set, but there is a tie-break played if the game score reaches 6-6 within a set. The winner of the match is the player to first win three sets.

The importance of a point is defined as follows. Assuming that, conditional on who is serving, the outcomes of points are *i.i.d.* random variables with known Bernoulli probabilities, one can compute, at any score, the probability that a player goes on to win the match. The (absolute) difference in this probability, as a function of the outcome of the current point, is the point's importance.

Arbitrarily name the players of a given match 1 and 2. Denote the score by $\{(p_i, g_i, s_i)\}_{i=1,2}$, with the interpretation that (p_i, g_i, s_i) denotes the number of points, games, and sets, respectively, that player i currently holds. Denote by θ the state of a match, which specifies the current score and the identity of the server. Let $W(\theta)$ denote the event that 1 wins the point at θ, and by $L(\theta)$ the complementary event that 1 loses the point at θ. Finally, denote by W^* the event that 1 wins the match. The importance of the point at θ is given by

$$I(\theta) = \Pr(W^* \,|\, W(\theta)) - \Pr(W^* \,|\, L(\theta)).$$

The importance of a point can be decomposed to the constituent probabilities of winning at the various hierarchical levels of the match. In particular, $I(\theta)$ is the product of (i) the importance of the point for determining the outcome of a game, (ii) the importance of the game in determining the outcome of the set, and (iii) the importance of the set in determining the outcome of the match.

We introduce notation to clarify this idea. First, let $I(\theta)$ be alternatively represented as $PiM(\theta)$. Denote the importance of the point for determining the outcome of the current game by $PiG(\theta)$. Similarly, define the importance of a game for determining the outcome of the current set by $GiS(\theta)$, and finally define the importance of the set by $SiM(\theta)$. We have the following

PROPOSITION **1** *For any point θ, we have*

$$PiM(\theta) = PiG(\theta) * GiS(\theta) * SiM(\theta).$$

Proof. See Appendix B. ■

This result expresses the independence of outcomes across levels of the scoring hierarchy. That is, conditional on the outcome of any particular game (set), the marginal importance of a point (game) beyond that game (set) is zero. While the idea is intuitive, it is very useful computationally, and previous papers adopting this notion of importance have not recognized it. Essential to our approach is to look at individual level point-by-point data. Estimation in this framework requires a large amount of data. A priori, computing the importance of a point involves tracing through all possible continuations of the game tree corresponding to the entire match. It is precisely Proposition 1 that allows our empirical approach to be practical.

A crucial aspect of adopting this definition is that importance varies in substantial and subtle ways across the points in a match. Given that our main objective is to measure players' ability to respond to changes in importance, much of the power of our approach relies on accurately representing the variation in points' importance during a match. For instance, as we demonstrate below, the obvious binary indicators

Before proceeding to the empirical sections of the paper, we make a few observations. First, closer matches have more points with high importance. This is because in unequal matches, the eventual outcome of the match is known with high probability early on, rendering the outcomes of individual points unimportant. Thus, as players' abilities diverge, almost all points converge to zero importance. Second, we can summarize which points tend to be important. As already mentioned, break points tend to be important whenever $q > 1/2$. When players are equally matched, the most important games are those when the current set is close, and the most important sets are those in which the match is close. Finally, when one player is significantly stronger than the other, the most important points are when the weaker player has the (surprising) opportunity to go ahead.

4. Empirical Strategy

Observe that we can hope to measure only relative, as opposed to absolute abilities of the players. To illustrate, if there is a match where player 1 wins most of his service points, one cannot identify whether this is because he is a strong server, or if his opponent is a weak returner. In the presence of additional information, such as service velocity, frequency of aces, etc., one could hope to identify an absolute level of performance, but we work without such information. Similarly, if it is true that players differ in their critical abilities, then there would be no difference in the outcomes of points if all players were to simultaneously improve their critical abilities. As such, the aim is to express abilities relative to the average ability in the population. In the process, we will also be able to determine cardinal rankings over the abilities.

One feature of the tournament structure of the U.S. Open is that it is very sensitive to missing data. Each tournament is a single elimination tournament beginning with a round of 128 players (64 matches). If for some reason, one of the matches from the tournament is excluded from the estimation, then not only is that data lost, but so also is the data from all matches below that match's node in the tournament graph. The reason is the following. Every match included in the estimation must be between two players each of whose abilities are being estimated. In the set of players whose abilities are being estimated, each player must be connected to each other player by some sequence of common opponents. We refer to this requirement as connectedness. Otherwise, there is no way to measure their abilities relative to each other.[2] Since every match has a loser, and the loser appears in no subsequent matches, estimating the abilities of a match's loser requires the presence of the matches in the branch above him in the tournament tree, or else such a chain of common opponents to the other players will not exist.

It is difficult to estimate precisely the abilities of some players due to having relatively few observations on them. This happens, for instance, when a player loses his first and, hence, only match, and thus plays relatively few total points in the tournament. While it is not problematic, per se, to have a few players for which reliable estimates cannot be obtained, the problem is that the noise in the estimates of these players' abilities affects how precisely we can estimate the average player's abilities, since the average is a function of the abilities of all players. This, in turn, affects our identification strategy for the entire sample. As a result we are forced to remove these players (more precisely, the matches containing observations from these players) from the data to obtain a set of sufficiently precise estimates on the remaining players.

There are also matches that are simply missing from the data we obtained from the USTA. We were not able to get an explanation for this fact. Most of these matches occur in the first two rounds of the tournament. While we can not rule out the possibility that this will cause a selection effect, the fact that the matches are from early rounds means that we do not lose too many additional matches from this source of omitted data.

[2] To illustrate, if the sample consists of two matches, the first between i and j, the second between distinct players k and l, then there is no way to measure the relative abilities of, say, i and k. However, if also there is a match between, say, j and l, then one can hope to measure all players relative to the others.

Thus, working on a tournament-by-tournament basis is not satisfactory. In order to retain a sufficiently large sample of players, as well as points per player, we proceed by pooling the matches from the twelve tournaments together.

The pooling has two advantages. First, many of the players appear in more than one tournament, and so we tend to have more data per player in the pooled data. Second, the aggregate data of matches no longer has the tree structure of a single tournament. This is very useful, because eliminating a particular match no longer necessarily implies the elimination of a number of other matches. That is, often times there are several chains of common opponents connecting a given pair of players,[3] and this helps to identify the relative abilities of two players who never compete directly in any of the tournaments.

Finally, there are still players for whom, even after pooling the data, have too few observations to estimate well. We proceed by eliminating players (and their matches) for whom we observe fewer than five matches.[4] After this is done, there are some remaining players for whom, even though we originally observed at least five matches, now have fewer than five matches left in the sample after the initial pruning. The pruning is iterated until all players remaining in the data have at least five matches in the data that remains. In the end, this leaves us with 94 players and roughly 110,000 points, and it is on this sample that we report results in the next section.

4.1. Consistent player abilities

There are many ways to handle the possibility that a given player's abilities may change over the course of their career, or even from one match to another. There is a tradeoff to be made here. On the one hand, allowing for greater flexibility in within-player differences can produce a very rich model, capable, in principle, of capturing many effects. On the other hand, the ability to identify the model decreases sharply with this flexibility.

The reason is the following. Because we want to estimate the ranking of players according to various abilities, we necessarily have to compare each player to every other player. However, there are many pairs of players in our data who never play directly against each other in a

[3]More precisely, the graph where players are vertices, and edges represent a match between two players, is connected, but not minimally connected, as in a single tournament.

[4]While somewhat arbitrary, five matches of data turns out to be sufficient to obtain estimates with tight enough standard errors for our purposes.

match, even after we pool the data across tournaments. In such cases, we must make an indirect inference about the relative abilities of these players, using the fact that they may have chains of common opponents. Say there is a particular player i who has played many matches against different opponents. Then the extent to which the presence of player i helps identify the relative abilities of other players depends on how much his abilities change across matches. On one extreme, if there is no constraint at all placed on the consistency of player i's abilities across his matches, then the fact that he played players j and k does not help at all to identify the relative abilities of j and k — indeed, it is as if j and k played against two different opponents altogether. On the other hand, if it is known that player i had exactly the same abilities in each of the two matches, then one can compare j and k readily by measuring their relative performances against i.

Since we are interested mainly in demonstrating that players differ meaningfully in their ability to perform well in critical situations, rather than capturing the subtle changes in abilities over time a player may undergo, we make the extreme assumption that each player can be described by a set of abilities that is constant across all the matches we observe. While this rules out potentially interesting patterns of career developments, it greatly simplifies the analysis and makes identification as easy as possible.

4.2. Service probabilities

Within a match, the importance of each point depends not just on the current score, but, crucially, on the estimated probabilities with which each player wins a given service point. Thus we begin by constructing, for each match, a measure of $q = (q_1, q_2)$.

One way to proceed would be to simply count the frequencies of service points won by the two players. However, this will create simultaneity problems with the estimation. The issue is that this procedure requires one to use the point-by-point data first to construct the measure of importance as a function of q, and then to regress players' abilities on the importances so obtained. Instead, we conduct the following exercise.

The goal is to have the importance of a point depend only on events that happened before the point. So in the beginning of match, we require estimates of the q_i that do not depend at all on the match at hand. To accomplish this, we collect data on the players from a set of tournaments leading up to the U.S. Open. This data is summarized in the last column of Table . From this we construct a serving (s_i) and returning (r_i) ability for each player, and this allows us to define service point probabilities for a match between two players i and j by

$$q_i^0 = \frac{1}{1 + \exp(-(s_i - r_j))}.$$

The ideal measure of q will reflect players' beliefs about point outcomes at each stage of the match. At the beginning of the match, these beliefs should be well captured by the historical outcomes of points played between the two players. Thus, for each match, we take the initial value of q to be given by q^0. Next, due to idiosyncratic factors (court surface, time of day, etc), and to changes in the players' abilities that may have occurred since their last match together, it may be that the probability of winning a service point in the current match differs from q^0. When such a discrepancy occurs, players are likely to realize it, and revise their beliefs about the probability of winning future points on the basis of observed outcomes in the current match. We adopt a simple approach to this updating at the level of the set. Specifically, we assume that the value of q in set $s > 1$ of a match are the simple average of the initial values, q^0, and the realized service winning frequencies from the previous $s - 1$ sets. In this manner, players' beliefs, starting from the basis of historical averages, respond to data from the current match and, as the match progresses, converge to the realized service winning probabilities of the current match.

Given these values of q to be used for each point in the match, we can proceed to construct importances according to the definition of the previous section. At any particular point, the importance is defined relative to the current values of q, assuming that all remaining points are won and lost with the respective probabilities.

We emphasize, however, that the results below do not hinge on this particular construction of q, and the implied values of importance. The qualitative results are the same under various alternative specifications. In particular, we obtain the same results when q is taken to be the observed frequencies of service points won for each player, and also when q is taken to be, for every player and every match, the overall average probability of winning a service point in the data, which is roughly 0.63.

4.3. The model

The fundamental aspect of our approach is to model the outcome of each point as depending on the server's ability to serve, the returner's ability to return, and the interaction of the importance of the point with each player's critical abilities. Each point in our sample corresponds to a single observation in our primary regression.

Each match is a competition between two particular players, i and j. For each match, one of the players is arbitrarily labeled "Player 1" and his opponent is labeled as "Player 2." We introduce the following notation. For each player i, let δ_i^S (δ_i^R) denote i's serving (returning) dummy variable. It takes value zero for observations, i.e., points, in which i does not serve (return). It takes value $+1$ for observations in which i serves (returns) in the role of Player 1, and value -1 when i serves (returns) in the role of Player 2. Let δ_i^C denote player i's critical ability dummy, which takes value $+1$ (-1) for observations in which i has the role of Player 1 (Player 2), and value zero in observations not involving i. Notice that $\delta_i^C = \delta_i^S + \delta_i^R$, but δ_i^C will appear only when interacted with the point's importance, so that there is not a problem with multicolinearity. We set the dependent variable to be the probability with which (the arbitrarily designated, but fixed) Player 1 wins the point.

The basic form of our regression is then

$$\Pr(W(\theta)\,|\,\theta) = \Phi\left(\beta_0 + \sum_{i=1}^{n}\left(\beta_i^S\delta_i^S + \beta_i^R\delta_i^R + \beta_i^C\delta_i^C I(\theta)\right)\right) \tag{4.1}$$

for all points in the sample, where Φ is the logistic cummulative distribution.

The abilities of each player i are described by the triple of coefficients $(\beta_i^S, \beta_i^R, \beta_i^C)$. It is the estimation of the critical abilities that are presented in the next section.

5. Critical Abilities

We have two main goals in this section. First, we demonstrate that there is significant variation in critical abilities across professional tennis players. Second, we show that this variation has a significant effect on player's careers, as measured by their professional rankings and annual earnings.

Recall that we want our measure of critical ability to be independent from any ability to generally win more points on average. However, given our econometric specification (), the fact that all point importances are by definition (strictly) positive would mean that a player with high critical ability would tend to win more points on average than an otherwise identical player with a low critical ability. In order to circumvent this effect, we must demean the importance variable. Upon so doing, a higher critical ability has roughly no effect on the average number of points won, but instead implies that more important points are won with higher probability, and less important points are won with lower probability, all else equal.

Close matches tend to involve many important points, while one-sided matches involve many unimportant points. In particular, then, the average importance of a point varies across matches. Thus, the way in which we demean the importances matters. We demean match-by-match. Thus, a high critical ability means that a player is likely to win the relatively more important points in the match, independent of their absolute level of importance. It turns out that this specification produces more explanatory power than demeaning importance globally. Moreover, demeaning at the match level insures that each player is involved in points of the same average importance.

5.1. Heterogeneity

The first observation we make is that there is strong evidence that players differ significantly and substantially in their critical abilities. Summarizing from above, we work with the pooled data from all twelve years of the US Open in the sample, and under the restriction that players' abilities, serving, returning, and critical, are constant throughout this time period.

It is universally accepted that serving and returning abilities are crucial to tennis performance. Joint significance of each of these two sets of variables is very high in all of the regression specifications.

A priori, there are a number of definitions one can ascribe to capture importance points. Perhaps the first that comes to the mind of tennis fans is break points.[5] In order to test for heterogeneity in abilities of players to respond to break points, we regress point outcomes on the serving and returning abilities along with a dummy variable for breakpoints. Joint significance of critical abilities measured with breakpoints is not nearly significant, with a p-value of 0.79 (see Table).

[5] A break point is a point that, if won by the returning player, results in breaking the service game of the opponent.

Table 2: Joint significance for ability identification specifications

	Regression Specifications			
Serve	0 (510.2)	0 (550.7)	0 (551.5)	0 (617.0)
Return	0 (949.2)	0 (1010.0)	0 (1009.3)	0 (903.7)
Critical (PiM)			.062 (114.8)	.01 (127.6)
Critical (BP)	.79 (81.8)			
Critical ($p = .63$)		.38 (96.6)		
Endurance				.0002 (149.1)

Notes: Joint significance tests are based on a χ^2 distribution with d.f. 93, except for Return, which has d.f. 94, due to how we normalized the regressions. The χ^2 statistics are reported in parentheses. For joint significance values below 10^{-10} we report zero.

As described above, our notion of importance is much more subtle than any indicator variable can be. One of the suggestions of Figure 1 is that PiG is quite sensitive to q. Of course, the relationship between serving probabilities and importance becomes even more complex when considering the whole match, i.e., PiM depends in complicated ways on (q_1, q_2). As such, using an accurate measure of q is essential to our identification.

Pooling over all of our point observations, the global probability of winning a service point is $q = .63$. Consider defining importance under the assumption that this value of q describes every point in the data set, and denote the importances so obtained by $PiM(.63)$. When we regress point outcomes on the serving and returning abilities along with $PiM(.63)$ we find that critical abilities are not jointly significant (p-value .38, see Table).

This finding is not surprising, given that we have already demonstrated that players differ significantly in terms of their serving and returning abilities. As a direct result, there is a great deal of variation across matches (and across the two players within each match) in terms of the likelihood of winning any given service point. This variation cannot be ignored when estimating players' critical abilities. Thus we adopt the approach described in Section 3.4 for the remainder of the empirical analysis, whereby serving probabilities are estimated exogenously and updated during the match based on observed point outcomes thus far (and denote the corresponding importances simply by PiM).

We now present what may be considered the first of two main results, that players' critical abilities differ significantly, and are important in explaining point outcomes. Point outcomes are regressed on the serving and returning variables along with *PiM*. As to be expected, serving and returning retain high levels of significance. More importantly critical abilities are now significant as well at the 6% level.

In order to increase confidence in this finding, we explore a number of robustness checks. First, we add to the regressors the other importance measures from above, PiG, GiS, and SiM. There are a number of reasons to believe these variables may add additional explanatory power to the regression. For example, PiG may capture a player's ability to adjust their performance in a more myopic way, respecting the kind of effects illustrated in Figure 1. GiS is to some extent considered important in the conventional wisdom, including statements to the effect of the seventh or eight games in set being particularly important. While our definition agrees with this assessment only some of the time, we believe it is reasonable that a player may have in mind such effects. When we run this regression each collection of importance measures is jointly significant. This suggests that, controlling for players' abilities to react to the overall importance of the point, they also differ significantly in their performance responses to more "local" notions of the point's importance. Table summarizes these findings.

The second robustness check we perform is to add a variable that corresponds to the point number in the match. This allows us to estimate an endurance effect, whereby a player may perform better or worse as the match goes on. While interesting in its own right, it is also important to include for the reason that it may be correlated with *PiM*. If later points tend to be more important, then we could be attributing effects on point outcomes due to differential endurance incorrectly to critical abilities. When point outcomes are regressed on the serving and returning variables, *PiM*, and *endurance*, critical abilities are jointly significant at the 1% level, and *endurance* is highly significant, indicating that endurance is an important determinant of point outcomes, and players differ significantly on that dimension.

5.2. Player estimates

Given the evidence for significance of critical abilities, we proceed to examine the estimates at the individual level. We begin by assessing our estimates of serving and returning ability.

Lists of the top twenty five servers and returners are included in Table 5. There is probably a strong level agreement that these lists fairly well describe the tennis players who are known to be particularly good servers and returners, at least through the top ten positions where outstanding abilities are most salient. Table 3 shows the correlation between serving and returning abilities among the 94 players estimated is 0.57. We interpret this correlation as an expression of the fact that winning points while serving and returning involves substantial overlap in skill sets (for instance, quality of ground strokes and volleys) but at the same time each involves a significant component of specialized skill.

Examining next the list of top players according to critical ability we find many of the same top players, but there is a significant presence of somewhat lesser know players as well. The correlations between serving and critical ability and between returning and critical ability are lower: 0.34 and 0.20, respectively. We turn now to assessing the impact of critical ability on career success.

Table 3: Correlations between abilities

	log(rating)	Serving	Returning	Critical
log(rating)	1	-	-	-
Serving	0.57	1	-	-
Returning	0.39	0.26	1	-
Critical	0.38	0.34	0.20	1

5.3. Economic Impact

We have demonstrated significant variation not only in players serving and returning performance, but also in their critical abilities. The first two abilities clearly translate into making a player stronger, in the sense that he will win more points, all else equal, the higher is his serving or returning ability. Critical ability also translates into making a player strong but in the sense that, given a baseline probability of winning points, he is more likely to win a point the more important is the point in determining the outcome of a match. This generally leads to higher probability of winning a match, all else equal, without changing the frequency of points won. We now seek to demonstrate that critical ability has a substantial impact on the career of players relative to their serving and returning abilities.

To this end, we require a proxy for the overall success of a player. The main proxy we use is the (log of the) player's rating from the ATP points system. Points are accumulated based on tournament outcomes according to a set points scheme, and reflect a player's results from the previous twelve months of play. As a robustness check we also use a player's (log) ranking, which is simply the ordinal counterpart of the points rating.

The main question we want to address is: To what extent are players' serving, returning, and critical abilities able to explain ATP ratings? Table 3 already hints at the potential of critical ability to be a key factor in success, as its correlation with the ATP rating is 0.38, approximately as strong as returning ability's correlation with rating. To provide a formal answer to this question, we take the estimates from our previous results in Section 4.1, and use them as regressors in explaining ATP ratings.

There are several issues to address in order to operationalize the procedure. First, the estimates come from data that cover a thirteen year time span. Since the ATP rating reflects only a twelve month window, the ratings can vary substantially over this time span. It is also the case that some players have periods of inactivity, during which their ratings drop but which do not necessarily reflect drops in ability. In order to address this issue, we proceed by computing, for each player in our sample, the mean ATP rating over whatever periods the player was active from 1994-2006, the time span of our data. This generates a single value for each player, which we call simply rating, and take as a proxy of their professional success.

We begin by regressing rating on serving and returning abilities. Tests for joint significance yield very low p-values, showing that the individual-level estimates we recovered from the first set of regressions are useful in explaining professional success of tennis players.

Our second main result comes from adding critical abilities to this regression. Serving and returning abilities each remain jointly significant in this regression. Most importantly from our perspective is that critical abilities are also significant (p-value 0.035). This means that the ability to perform better in more important points is a significant factor in a player's overall success. The adjusted R^2 from this regression is 0.40, and so even with this limited set of variables one is able to capture a lot of the variance in ratings across players. For comparison, the regression omitting critical abilities has a lower adjusted R^2 of 0.38.

The previous result is based on a simple OLS regression. However, there are some specification issues that need to be addressed. Most importantly, when performing the regression of ratings on abilities, it is clear that the regressors are themselves the estimates of a previous regression. Since they should be viewed as measured with noise, using them in a secondary

Table 4: Impact of Abilities on Career

Ind. Var.	Rating	Rating	Rating	Rating	Rank	Rank	Rank	Rank
Intercept	7.89(.21)	7.86(.21)	7.32(.14)	6.14(20.3)	1.73(.34)	1.77(.33)	2.69(.23)	23.5(33.2)
Serve	1.68(.28)	1.48(.29)	1.22(.29)	1.23(.33)	-2.75(.45)	-2.46(.47)	-2.04(.47)	-1.98(.54)
Return	.84(.28)	.78(.27)	.77(.29)	.87(.33)	-1.43(.44)	-1.33(.44)	-1.34(.47)	-1.35(.54)
Critical		.036(.017)**	.031(.015)**	.034(.016)**		-.054(.027)**	-.047(.025)*	-.053(-.026)**
Endurance			66.9(52.6)	53.7(56.4)			-95.3(85.4)	-93.5(92.2)
Controls	No	No	No	Yes	No	No	No	Yes
Adj. R^2	.38		.40		.39	.40		

Notes: Each specification uses as regressors abilities estimated from a first stage in which only those abilities are used to explain point outcomes. The ** denotes significance at the 5% level, and * at the 10% level. We report these only for Critical Ability. Serve and Return are highly significant in each specification. The controls are place of birth, height, left/right-handed, GDP of home country, if the person is American, and if they attended the Bollettieri school.

regression presents some issues.

First, there is the fact that the noise inherent in our estimates may produce attenuation bias. Although we do not have a good way to address the issue, it is clear that if the results suggest significant effects of players' abilities on ratings, then correcting for attenuation bias would only strengthen the significance of the results.

Second, we want to use optimally the information from our set of first stage estimates. In particular, we have 94*3 relevant estimates of player abilities, and the complete variance-covariance matrix. In order to use this information efficiently, there are several approaches one could take. One option is to run a GLS model, which is reported in Table along with the other regression specifications.

A second approach is to bootstrap the estimation procedure. Bootstrapping will correct for the patterns of noise from our estimates. We proceed as follows. Starting with the first stage of equation (), we re-sample, with replacement, n point observations from the data, where n is the roughly 110,000 points. Then, on the re-sample, we run (), producing estimates for each player. Given these estimates, we proceed to the second regression, adopting GLS using these re-sampled estimates as the regressors. In this second stage, we again bootstrap the regression to achieve a distribution of parameter estimates. We run many second-stage regression for each first stage regression because the second stage regressions are computationally much cheaper. This entire process is then repeated to generate parameter distributions.

We report results from this procedure using 1000 full iterations, where for each iteration, the second stage regression is run 1000 times. This gives us a sample distribution of size one million for estimating the coefficients on the players' abilities. The bootstrapped estimates are also reported in Table . Based on the bootstrapping procedure, both serving and returning abilities are significant at the 1% level, while critical ability is significant at the 8% level. Our conclusion is that there is strong evidence that players differential critical abilities plays a role in their long term career success.

One way to express the impact of critical ability on career is to look at standardized coefficients. Relative to the average player, the impact on ATP rating of increasing one's critical ability by one standard deviation in the population is worth 41% (76%) of the impact of increasing one's serving (returning) ability by one standard deviation.

6. Conclusion

The main issue motivating this study is the potential impact caused by reactions to critical situations. In particular, if it is the case that individuals respond to critical situations heterogeneously, then one should expect to find observable effects on individuals' outcomes. We chose a very specific context in which to assess the empirical content of this argument. In the case of professional tennis players, there is very strong evidence that the importance of a point has a first order effect on the point's outcome. In fact, we identify which players are best at winning the most important points. While it is suggestive that such a skill will translate into better outcomes for a player, an important second step is to quantify that effect. In this regard we find that the skill of critical ability explains a substantial part of a player's professional success, controlling for the other main skills involved in winning tennis points.

Only relatively recently have psychological effects been given primary roles in economic analyses. This paper contributes to the body of research showing that such effects can be important determinants of individual success. The setting of tennis permitted particularly clean data with which to estimate critical abilities. An important goal of future work should be to understand the impact of psychological responses in other competitive environments and market settings such as job search.

Table 5: Best performance according to our estimates and average ATP ratings.

Top 25	Serving	Point Estimate	Returning	Point Estimate	Critical	Point Estimate	ATP Rating	log(rating)
1	A.RODDICK	0.268	L.HEWITT	-0.457	T.ROBREDO	6.026	P.SAMPRAS	8.109
2	P.SAMPRAS	0.203	R.FEDERER	-0.485	A.CORRETJA	4.491	R.FEDERER	8.077
3	R.KRAJICEK	0.145	K.KUCERA	-0.5	J.FERRERO	2.892	M.STICH	7.856
4	R.FEDERER	0.065	A.AGASSI	-0.519	A.COSTA	2.631	L.HEWITT	7.813
5	M.MIRNYI	0.036	J.BJORKMAN	-0.534	M.ROSSET	1.434	A.AGASSI	7.78
6	M.STICH	0.017	M.YOUZHNY	-0.549	M.ZABALETA	1.43	A.RODDICK	7.779
7	A.AGASSI	0.	N.ESCUDE	-0.558	G.POZZI	1.324	Y.KAFELNIKOV	7.766
8	P.RAFTER	-0.002	Y.KAFELNIKOV	-0.562	R.SCHUETTLER	1.257	G.KUERTEN	7.742
9	G.RUSEDSKI	-0.005	D.NALBANDIAN	-0.574	A.RODDICK	1.043	T.MUSTER	7.702
10	N.ESCUDE	-0.024	G.CORIA	-0.577	G.IVANISEVIC	0.872	J.FERRERO	7.583
11	G.KUERTEN	-0.04	J.BLAKE	-0.582	B.BLACK	0.841	P.RAFTER	7.573
12	L.HEWITT	-0.047	P.KORDA	-0.587	M.WOODFORDE	0.799	R.NADAL	7.523
13	M.LARSSON	-0.047	A.RODDICK	-0.603	B.KARBACHER	0.762	P.KORDA	7.431
14	M.SAFIN	-0.081	G.CANAS	-0.617	P.SAMPRAS	0.655	T.HENMAN	7.407
15	B.BECKER	-0.087	D.HRBATY	-0.634	L.HEWITT	0.637	C.MOYA	7.377
16	T.MARTIN	-0.087	P.RAFTER	-0.638	N.ESCUDE	0.569	D.NALBANDIAN	7.37
17	J.BLAKE	-0.092	V.SPADEA	-0.651	A.MEDVEDEV	0.261	A.CORRETJA	7.293
18	X.MALISSE	-0.101	H.LEE	-0.655	P.RAFTER	0.151	M.SAFIN	7.281
19	W.ARTHURS	-0.104	S.SARGSIAN	-0.664	S.DOSEDEL	0.098	B.BECKER	7.264
20	M.ZABALETA	-0.112	J.COURIER	-0.674	M.SAFIN	0.096	R.KRAJICEK	7.236
21	M.DAMM	-0.121	M.ZABALETA	-0.678	W.ARTHURS	0.009	G.CORIA	7.22
22	J.COURIER	-0.134	T.ENQVIST	-0.68	A.AGASSI	0.	J.COURIER	7.192
23	D.NALBANDIAN	-0.135	A.CLEMENT	-0.683	R.FEDERER	-0.196	C.PIOLINE	7.188
24	R.GINEPRI	-0.135	T.HAAS	-0.685	C.PIOLINE	-0.236	T.ROBREDO	7.168
25	J.FERRERO	-0.139	M.SAFIN	-0.689	T.MARTIN	-0.238	A.MEDVEDEV	7.16

References

Apesteguia and Palacios-Huerta 2008

Ariely, Gneezy, and Mazar (2009)

Dohmen (2008)

Hume 1739

Klaasen and Magnus (2001)

Morris 1977

Passerman (2008)

Hume 1739

8. Appendices

8.1. Appendix A: Qualitative features of point importance

Unless the players are very unequally matched, there is substantial variation in the importance of points in a typical match.[6] It is this natural variability that will allow us to identify critical abilities.

As explained, one component of a point's importance is its importance in determining the outcome of its game (this is the first term in the decomposition of Proposition 1). We begin by describing how this game importance (PiG) responds to the relative strength of the server at various scores in the game.

[6] When one player is much stronger than the other, his probability of winning the match is close to one for all scores except those at which he is very close to already losing the match. Since in this case such scores have very low probability of being realized, most points have importances close to zero in this case.

Suppose that player 1 is serving and that, given his ability to serve and his opponent's ability to return serve, player 1 has a probability q of winning any given service point, on average. Serving is generally an advantage, and so typically $q > 1/2$. What are the importances of points in the game at the different possible scores, and how do they depend on q? We start with the tied score, $p_1 = p_2 \geq 2$, known as "deuce."[7] Recall that W_G denotes the event that player 1 wins the game, and let D denote a deuce score with player 1 serving. We have $\Pr(W_G \,|\, D) = q^2 + 2q(1-q)\Pr(W_G \,|\, D)$, so that

$$\Pr(W_G \,|\, D) = \frac{q^2}{1 - 2q(1-q)}.$$

The probability of player 1 winning the game at all other points can be computed recursively, using the relationship

$$\Pr(W_G \,|\, p_1, p_2) = q\Pr(W_G \,|\, p_1 + 1, p_2) + (1-q)\Pr(W_G \,|\, p_1, p_2 + 1).$$

Then, importances (PiG) are computed by taking the relevant differences in outcome probabilities as before.

Table 1 shows the PiG importance of each score at the average value of q in our data, which is approximately 0.63.

$p_2 \backslash p_1$	0	1	2	3
0	.24	.18	.10	.03
1	.34	.31	.22	.09
2	.39	.45	.44	.25
3	.30	.47	.75	.44

Notice that the most game-important point is $(p_1, p_2) = (2, 3)$, a score known as break point. This coincides with the conventional wisdom that break points represent important situations in the match. This arises under our definition because typically $q > 1/2$, and so the server is expected to win service games. The most pivotal situations are those in which this outcome has one chance to be reversed, which is the score $(2, 3)$.

[7] The score of $p_1 = p_2 = 2$ is strategically equivalent, but not known as deuce.

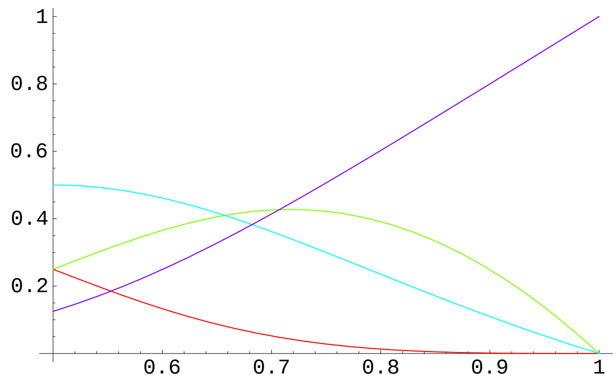

Figure 1: Red: 2-0, Turquoise: deuce, Green: 0-2, Purple: 0-3

Table 1 shows the importance of all possible situations within a game, but only for a particular value of q. Figure 1, on the other hand, plots the importance of several prominent scores as a function of q, in order to show how relative importances change depending on the relative strength of the server. Notice that the relative game importance of points depends very much on q. In other words, the points that are most important in determining the outcome of a game depend on the distribution of point outcomes and, moreover, their rankings vary non-trivially in the range of q observed in our data.

For instance, of the points shown in Figure 1, we see that $(2, 2)$ is the most important point in the game whenever q is below about 0.66. However, for slightly larger values of q, $(0, 2)$ is the more pivotal point, and then for even larger values of q $(0, 3)$ is the most important point. As the server becomes relatively stronger and stronger, the only points with non-trivial importance are break points: those at which the point would award the game to the returner. All other point importances tend to zero as q grows, since then the server will win the game with arbitrarily high probablity, indpendent of the outcome of the current point.

Thus, while considering only the importance of a point in the game, there is already a complex relationship between importance and the characteristics of the players involved. However, one must consider as well the importance of the game in the set, and also the set in the match. Notice that when considering games and sets, the importance will depend on the probability of winning a service point for both players. So define q_i as the probability that player i wins a typical service point, $i = 1, 2$. The following two tables present these importances (GiS and SiM), respectively, when $q_1 = .75$ and $q_2 = .63$.

$g_2 \backslash g_1$	0	1	2	3	4	5
0	.34	.13	.12	.02	.01	.00
1	.39	.40	.14	.13	.01	.01
2	.33	.46	.47	.16	.13	.01
3	.30	.31	.55	.57	.17	.14
4	.08	.25	.26	.67	.69	.17
5	.03	.03	.16	.17	.83	.69

$s_2 \backslash s_1$	0	1	2
0	.12	.08	.03
1	.36	.29	.17
2	.69	.83	1.0

In the example given, the most important game is at the game score $(4, 5)$, the outcome of which affects the outcome of the set with importance 0.83. Notice that this contrasts with the conventional wisdom that the seventh game in the set is the most important.[8] Of course, these importances will vary with the particular values of q_1 and q_2 for any given match.

The final table shows, for instance, that the final set determines with probability one the outcome of the match whenever it is reached. Also notice that the most important sets are when player 2, the weaker player, is winning. When player 1 is winning, the sets are less important because even if player 1 loses the set, he will still win the match with high probability.

To illustrate the heterogeneity in importance, consider the following scenario from the 1995 U.S. Open final match between Agassi and Sampras. The difference in purse winnings between the winner and loser of the match was $287, 500. In that match, Agassi won 65% of his service points, and Samprass won 72% of his service points, so we have $q_A = .65$ and $q_S = .72$. At one

[8]The origin of this "wisdom" is disputed. See, for instance, Magnus and Klaassen (1996).

point late in the match the score was such that the point importance was 0.0008, implying that \$380 was riding on the point in expected terms. At an earlier point when Agassi, the weaker player in the match, had a break point, the importance was .13, so that \$64,000 was riding on the point in expected terms. Such variation is typical in later rounds of the tournaments.

8.2. Appendix B: Proof of Proposition 1

The proof can be illustrated by showing that the importance of a point θ in determining the outcome of the current set (denoted $PiS(\theta)$, is equal to $PiG(\theta) * GiS(\theta)$. Define the following notation: $W_G(\theta)$ ($L_G(\theta)$) is the event that player 1 wins (loses) the current game at point θ, while $W_S(\theta)$ ($L_S(\theta)$) is the event that player 1 wins (loses) the current set at point θ. We have the following, suppressing the dependence on θ.

$$\begin{aligned}
PiS &= P(W_S \,|\, W) - P(W_S \,|\, L) \\
&= P(W_S \,|\, W_G)P(W_G \,|\, W) + P(W_S \,|\, L_G)P(L_G \,|\, W) - \\
&\quad [P(W_S \,|\, W_G)P(W_G \,|\, L) + P(W_S \,|\, L_G)P(L_G \,|\, L)] \\
&= P(W_S \,|\, W_G)[P(W_G \,|\, W) - P(W_G \,|\, L)] + P(W_S \,|\, L_G)[P(L_G \,|\, W) - P(L_G \,|\, L)] \\
&= P(W_S \,|\, W_G) * PiG - P(W_S \,|\, L_G) * PiG \\
&= PiG * GiS
\end{aligned}$$

The first equality holds by definition, the second expresses Bayes Law, the third equality rearranges terms, and the remainder applies our definitions. The argument to complete the proof proceeds in an analogous fashion.

8.3. Appendix C: Robustness of main results

We report below the results of GLS and bootstrap versions to asses the impact of critical abilities on career outcomes.

Table 6: Alternative Specifications for Impact of Abilities on Career

Regression Type	Independent Variable	Intercept	Dependent Variables Serve	Return	Critical	Fit Adj. R^2
GLS	Rating	7.92(.21)	1.52(.27)	.83(.27)	.038(.017)**	
Boot Strap	Rating	7.74(.25)	1.26(.32)	.71(.31)	.023(.017)*	
GLS	Rank	1.65(.34)	-2.54(.45)	-1.42(.43)	-.058(.028)**	
Boot Strap	Rank	1.98(.41)	-2.09(.54)	-1.20(.50)	-.037(.026)*	