Hot heads, cool heads, and tacticians: Measuring the mental game in tennis (ID: 1464)

Stephanie Kovalchik
Tennis Australia
Victoria University
skovalchik@tennis.com.au

Martin Ingram
Stratagem Technologies
martin.ingram@gmail.com

It is often said that winning in tennis is as much a mental game as a physical one, yet there has been little quantitative study into the mental side of tennis. We present an approach to identify mentalities in tennis with dynamic response patterns that quantify how a player's probability of winning a point varies in response to the changing situations of a match. Using 3 million points played by professional male and female tennis players between 2011 and 2015, we found that, on average, players were affected by the state of the score and a variety of pressure situations: exhibiting hot hand effects when ahead, defeatist effects when down, and performing less effectively in clutch situations. Player-specific performance patterns suggested a diversity of player mentalities at the elite level, with subgroups of players responding more or less effectively to pressure, score history, and other match situations. One of the patterns found on the men's tour included four of the most decorated players in the current game, the 'Big Four', suggesting a champion's mentality that was characterized by cool-headedness on serve and adaptability on return. Accounting for player mentalities improved predicted outcomes of matches, substantiating the importance of the mental game for success in tennis.

1 Introduction

Mentality is an essential ingredient of all athletic performance. However, the influence a player's mental skills has on the outcomes of competition varies widely across sports. Tennis, the most popular individual sport in the world[1], is frequently said to be as much about the mind as the body. Indeed, former champion Jimmy Connors, holder of 109 career titles, has gone so far as to estimate that 95% of tennis is a matter of the mind[8]. Despite these claims, the mental side of tennis has received limited scientific study and quantitative measures of the mentalities of today's top tennis players are lacking.

Few studies in any sport have attempted to measure the mental aspect of elite athletic performance, which highlights the broad challenges of measuring the mental side of sport. Prior approaches have largely been qualitative in nature, relying on interviews[10] or surveys[9] of players and coaches to gain insight into player psychology. These studies presuppose that salient mental skills can be measured with a questionnaire or elicited from a conversation. A further drawback of these studies is that they have not taken advantage of the years of historical performance data that is available to researchers and what it can reveal about the mental game.

In this paper, we present a novel quantitative method to investigate the mentality of professional tennis players from observed match performance. Our approach is based on the premise that tennis

[1] http://www.biggestglobalsports.com

players reveal mental skills in the way they respond to changing situations in a match—the match dynamics. Using 3 million points played in professional singles matches for the men's and women's tours between 2011 and 2015, we quantify player response patterns to point dynamics, identify common mentalities, and evaluate the importance of mentality for match performance.

2 Methods

2.1 Data

Point-level data was obtained for singles matches played on the Association of Tennis Professionals (ATP) and Women's Tennis Association (WTA) tours during the 2011 to 2015 seasons. For the ATP tour, matches were restricted to those played at tournaments in the 250 series or above and included the 4 Grand Slams (Australian Open, French Open, Wimbledon, and US Open), where players have the opportunity to earn the most ranking points and prize money. The WTA data included matches from all International, Premier, and Grand Slam tournaments. To ensure an adequate sample size of points for each player, only players with 3 or more match appearances were eligible for inclusion. Point-level data was obtained from the Tennis Abstract[2] and accessed with functions from the R package `deuce`[3] .

The final datasets for the present paper included over 1.6 million points across 10,101 matches played by 434 ATP players and 1.4 million points in 9,668 matches of 424 WTA players (Table 1). Approximately 20% of the points in these datasets were played at the Grand Slams.

Table 1: Summary of Point-level Datasets of ATP and WTA Singles Matches, 2011-2015

Variable [Shorthand]	ATP	WTA
Points	1,610,439	1,373,095
Players	434	424
Matches	10,101	9,668
Grand Slam Matches	1,869	2,066
Candidate Predictors		
Tiebreak, %	3.4	1.9
Break point, %	8.4	11.7
Point away from break point, % [Break point -1]	17.6	21.3
Set or more up, % [Set+]	22.4	20.8
Set or more down, % [Set-]	23.8	21.5
Player won last point, % [Just won]	53.8	51.4
Serve game after missed break, % [Missed break, serve]	10.1	10.0
Return game after missed break,% [Missed break, return]	9.1	8.9
Importance, Mean (SD)	0.05 (0.05)	0.06 (0.05)
Last game's points, Mean (SD)	5.5 (2.9)	5.8 (3.2)
Point spread, Mean (SD)	-0.38 (4.5)	-0.19 (4.9)

The analytic datasets were divided into training and validation data for the purpose of model development and testing. The validation data were the points played at the 2015 Grand Slam tournaments: 105,717 points for the ATP and 71,341 for the WTA. For each validation tournament, the training data were all points played up to but not including the Grand Slam event.

[2]Github source: `github.com/JeffSackmann/tennis_pointbypoint`
[3]Github source: `github.com/skoval/deuce`

2.2 Player Dynamic Model

We introduce a player dynamic model (PDM) to quantify how players are uniquely affected by the situation of a point in a match. The player-specific dynamics estimated from the PDM will be the basis for characterizing player mentality. The dependent variable of the model was the win-loss outcome for a point, with respect to the server of that point. Let y_{ijk} be the point outcome (Win = 1, Loss = 0) for the ith player serving against the jth opponent on the kth service point.

The PDM describes the relationship between the situation of the point and the point outcome with the following linear mixed probability model,

$$E[y_{ijk}] = (\alpha_i + \beta_j + \theta)'\mathbf{X}_{ijk}. \tag{1}$$

The vector $\mathbf{X}_{ijk}$ contains an intercept, which defines the baseline serve and return ability of the player and opponent, and p dynamic features that affect performance on the point outcome (e.g. being a set down, facing a break point, etc.). The parameters θ are fixed dynamic effects that represent how players are affected by the point situation on average. The parameters α_i and β_j are player random effects for each feature for the ith player who is serving and jth player who is returning and each are drawn from a multivariate normal distribution with zero mean and general variance structure. The player random effects allow that some players could be more or less affected by the state of a point and that these effects could additionally depend on whether a player is serving or returning. The combination of the server and returner effects determine the overall win probability for the point.

Remarks. The PDM can be viewed as an extension of two established models for predicting point outcomes in tennis. When the feature matrix **X** is reduced to an intercept term, the PDM becomes a regression model for the opponent-adjustment proposed by Barnett and Clarke[2]. Klaassen and Magnus also proposed a dynamic model to test for deviations from the IID model, which says that points in a match are independent and identically distributed[3]. In contrast to the PDM, the model of Klaassen and Magnus considers a limited number of dynamic effects and does not incorporate player-specific effects, which are the parameters of primary interest in the present work.

2.3 Predictors

We fit the PDM with eleven candidate predictors that are listed in Table 1. The candidates include the two dynamic factors previously examined by Klaassen and Magnus[3] and 9 additional variables that cover dynamic point, game, and set situations. The majority of the point conditions focus on various types of pressure, including indicators of whether a point occurs during a tiebreak, is a break point opportunity for the returner, or a point away from a break point opportunity (Break point -1). These predictors can be considered types of important points because they have a greater influence on the game or set outcomes than other points. We also include a probabilistic measure of point importance, defined by Morris[6], that is equal to the average change in match win probability when the current point is won compared to win it is lost. One final point condition is an indicator of whether the server won the previous point (Just won), which captures short-term correlation between points that could arise from a one-point hot hand effect, for example.

Three predictors contained information about game history. Two concerned performance on the service game (Missed break, serve) or next return game (Missed break, return) after missing a chance to break service, that is, failing to convert a break point opportunity. In the coding for these predictors, it is assumed that the psychological impact of a missed break is a within-set effect and does not carry over into the games of other sets or tiebreaks. We also examined how the number of points played in

Table 2: Average Effects of Dynamic Conditions on Point Performance, 2011-2015 ATP and WTA Tours

Dynamic	ATP			WTA		
	Estimate	95% CI	ΔAIC[a]	Estimate	95% CI	ΔAIC
Base rate	63.20	(62.72, 63.69)	–	55.79	(55.31, 56.27)	–
Tiebreak	-0.06	(-0.53, 0.41)	2.8	-0.61	(-1.27, 0.05)	-2.3
Break point	-0.76	(-1.08, -0.44)	0.9	-0.63	(-0.95, -0.31)	0.2
Break point -1	-0.39	(-0.61, -0.18)	8.9	-0.02	(-0.24, 0.21)	0.6
Set+	1.43	(1.24, 1.63)	5.2	1.71	(1.49, 1.92)	-0.4
Set-	-1.94	(-2.13, -1.75)	19.2	-2.01	(-2.22, -1.80)	0.4
Importance	-4.34	(-6.43, -2.26)	32.2	-5.56	(-7.77, -3.36)	0.6
Point spread	0.30	(0.28, 0.32)	13.3	0.31	(0.29, 0.33)	-2.5
Just won	0.63	(0.47, 0.78)	44.4	0.51	(0.33, 0.69)	0.6
Missed break, serve	0.21	(-0.09, 0.51)	-4.1	0.20	(-0.12, 0.52)	-2.6
Missed break, return	-0.22	(-0.49, 0.04)	-3.4	0.12	(-0.18, 0.41)	-3.3
Last game's points	-0.01	(-0.04, 0.02)	-5.3	-0.02	(-0.05, 0.01)	0.4

CI = Confidence interval, AIC = Akaike Information Criterion
a Change in AIC with inclusion of player-specific dynamic effects compared to a constant dynamic effect. Larger values indicate the player-specific model provided a better fit to observed point outcomes.

the previous game might influence play in the current game. A more closely contested game will have more points played and could serve as an indirect measure of player fatigue.

Three set conditions were also considered. Two factors indicated when the player serving was either up a set or more (Set+) or down a set or more (Set-) in the match. Finally, in order to capture longer term momentum effects than those due to the outcome of the previous point, we tracked the point spread across games within a set, subtracting the points won in the set by serving player from the returning player.

2.4 Model Estimation

The PDM was implemented with the R package `lme4` using a Gaussian family for the outcome distribution. The Gaussian model has a number of desirable features. Most importantly, the dynamic effects have a simple and meaningful interpretation, as they represent the absolute change in point win probability for a one unit increase in a feature. The model is also the most computationally efficient within the generalized linear family. However, the model is most appropriate for continuous outcomes, whose mean, unlike a binary outcome, does not have a constrained support.

Klaassen and Magnus have previously shown that the linear model works well in practice for modeling point win probabilities[3]. We also conducted our own investigation by obtaining marginal predictions for the dynamic effects using a logistic model and comparing these to their corresponding effects with the linear probability model. The effects differed by no more than one significant digit, indicating that the lack of constrained estimation had a negligible impact on the PDM estimates.

In addition to the PDM, we fit two alternative models for the purpose of comparison. One of these was a simpler version of the dynamic model that had average effects for the eleven dynamics and only an intercept term for the player effects. We will refer to this model as the average dynamic model (ADM). The second model had only an intercept term for the fixed and player effects, an IID model.

All models were fit in R and code for the estimation is available from the authors upon request.

2.5 Model Performance

To summarize the overall magnitude and significance of the dynamic effects, we fit the ADM with all of the available data and computed 95% confidence intervals for each effect. The added value of including player-specific effects for each dynamic feature was measured by the change in the Akaike Information Criterion (AIC). The AIC is a measure of overall model fit that allows comparisons between non-nested models and rewards more parsimonious models. For each dynamic predictor, we estimated the change in AIC with the inclusion of player-specific effects versus a constant effect for all players.

To assess the implications of mentality on points for predicting the outcomes of matches, we used a Monte Carlo simulation to test the predictive performance of the PDM. The simulator ran 5,000 trial matches for each pairing of the 2015 Grand Slams (a best-of-five format for the ATP and best-of-three for the WTA). Each simulated point on serve in a match adjusted the server's probability of winning the point according to the state of that point as described by the PDM, and the fraction of trial matches won gave an estimate of the probability that a player won the match. The player dynamic match predictions were compared to the alternative ADM and IID models. Our primary metric of performance was log loss, as it places a high penalty on overconfident predictions[11].

2.6 Identifying Mentality Types

The ways in which a player's performance is affected by the conditions of a match provide insight into the player's mentality. The set of player-specific dynamic effects from the PDM—which represent how much more or less a player's performance on a point shifts in response to the state of the match compared to the field—provide a mentality profile. To identify common mentalities on the tour, we applied a hierarchical clustering method to the dynamic profiles of the players who competed in the 2015 Grand Slams. Only the most salient features for distinguishing player types were included, which were defined as the features that showed an AIC improvement over the constant dynamic effects of the ADM. Because these features represent systematic variation about the baseline ability to win a point but do not include the baseline itself, we are able to separate a player's overall skill level from their mentality and allow for the possibility that players of different rank could share similar mental skills.

Prior to clustering, the dynamics effects of the mentality profiles were converted from their probability scale to a z-score so that each effect would have the same mean and variance. A distance measure was then applied to all possible pairs of the standardized profiles (e.g. Player A's standardized dynamics on tiebreaks, break points, etc. versus Player B), which results in a dissimilarity matrix. We then apply a linkage approach to identify clusters among players.

There are a number of options available for both the distance metric and linkage approaches. Three common measures of distance are the Euclidean (L_2-norm), the Manhattan (L_1-norm), and absolute maximum (or Chebyshev's distance). The methods primarily differ in their response to outliers with the maximum distance being most sensitive to extremes. The linkage method is the technique applied to the pairwise dissimilarities to determine a measure of the cluster distance. *Single linkage* is a nearest neighbor method that assigns clusters based on the minimum distance. Average linkage chooses the cluster that minimizes the average between-group distances. Complete linkage is the opposite extreme of single linkage in that it assigns clusters by maximizing the difference between clusters. Each of the hierarchical techniques begin with all units assigned to a single cluster, i.e. a 'bottom up' approach.

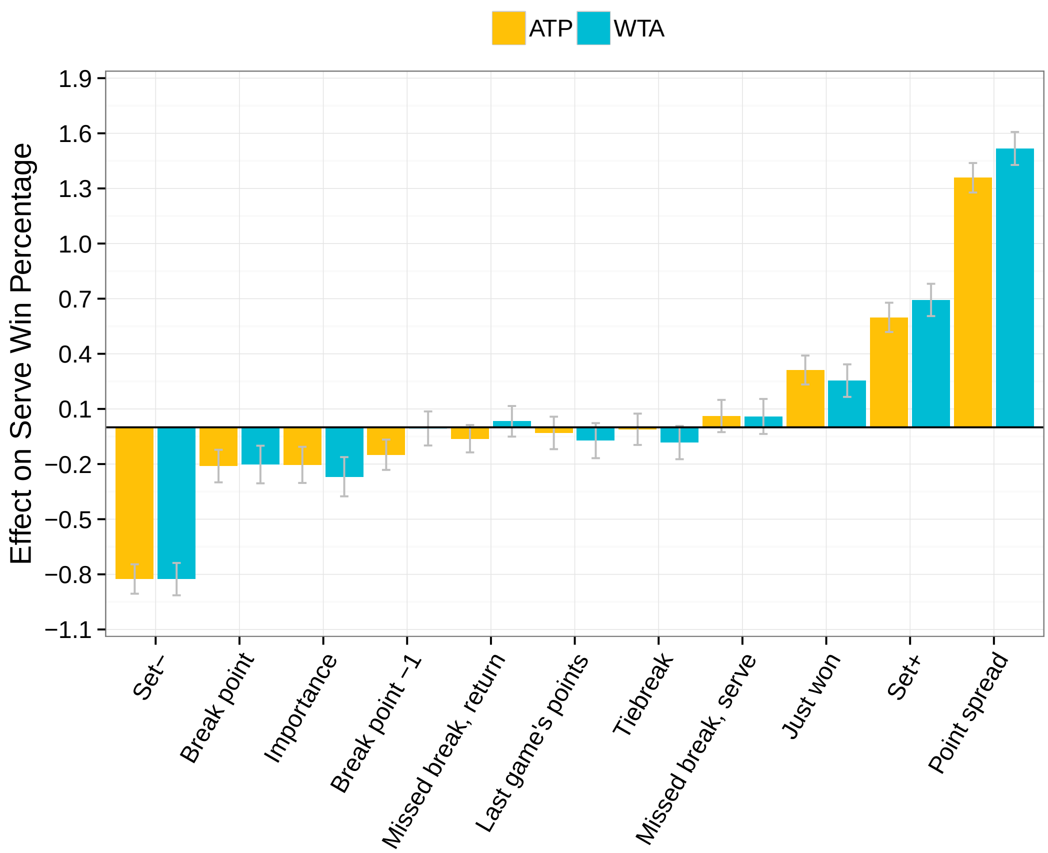

Figure 1: Average dynamic effects on the probability of winning a point on serve in 2011-2015 ATP and WTA singles matches. The y-axis shows the estimated change point win probability (in percentage points) for a one standard deviation change in the dynamic predictor. Error bars denote the 95% confidence interval.

There is no universally best method among these; the performance, instead, depends on the properties of the data used [4]. For this reason, each combination of distance measure and linkage measure were examined. The approach selected was one that showed the largest number of patterns containing two or more players per cluster. The choice of the number of clusters was selected by beginning with a large number ($K = 12$) and visually inspecting the patterns of each cluster with parallel coordinate plots. When two or more clusters could not be easily distinguished the number of clusters was reduced by one. This process was repeated until all patterns could be uniquely described.

2.7 Player Unpredictability

It is possible that some players will not easily fit into any of the identified mentality types. One way this could arise is if a player's shifts in performance are essentially random, i.e. ups and downs that are unrelated to the specific situation of the point. To examine which players exhibited more or less of this kind of volatility, we computed the mean Brier score (the variance in observed minus predicted outcomes) of the PDM predictions for each player on serve and return when applied to the Grand Slam validation data. Outliers were flagged as players with scores that were 2 or more standard deviations from the mean.

3 Results

3.1 Characteristics of Dynamic Predictors

Among the eight categorical dynamics, a server's win on the previous point was the most common, occurring slightly more than half of the time for each tour (Table 1). For one of every five points on serve, at least one competitor would be expected to be a set up or down in a match based on recent years of matchplay. A similar percentage of points would be one point from a break point opportunity. Break points and points played in games following a missed break opportunity were some of the least common events among the predictors, happening approximately 10% of the time during a match. The rarest dynamic event was a tiebreak, which occurred for 3% of points on the men's tour and 2% of points on the women's tour.

The average importance of a point in a tennis match was 5 probability points, corresponding to an expected increase in match win probability of 5 percentage points when the point was won versus lost (Table 1). In recent years, we also found that the average number of points played in a game was 6 for both tours. The average point spread over the games in a set was approximately zero but it was not unusual to observe differences as large as 10 points.

3.2 Average Dynamic Effects

The average dynamic effects on point outcomes identified factors that significantly increased the advantage of the server and other effects that increased the advantage of the returner. For both tours, a server's win opportunity was negatively affected when playing a set down, when facing a break point, or when facing more important points overall (Table 2). By contrast, when a server was ahead a set (or more) in a match, had won the previous point, or otherwise had a lead in the point spread, we found the server generally had a significantly greater probability of winning a point. Because the point spread can take negative values when the server is behind in the set, the impact of spread would, in this case, have a comparable negative effect on the server's win probability. In data not shown, we assessed the

Table 3: Summary of Match Prediction Performance for the 2015 Grand Slams

Tournament	Accuracy			Log Loss		
	IID	ADM	PDM	IID	ADM	PDM
ATP						
Australian Open	73.3	74.1	75.9	0.516	0.505	0.491
French Open	70.6	71.4	73.9	0.552	0.541	0.532
US Open	75.5	75.5	76.4	0.507	0.496	0.493
Wimbledon	73.3	72.5	72.5	0.540	0.523	0.517
Overall	73.1	73.3	74.6	0.529	0.517	0.509
WTA						
Australian Open	74.0	73.2	73.2	0.595	0.573	0.576
French Open	72.7	71.1	71.9	0.568	0.561	0.563
US Open	65.8	65.0	64.2	0.663	0.632	0.634
Wimbledon	70.4	70.4	71.2	0.568	0.555	0.560
Overall	70.8	69.9	70.1	0.598	0.580	0.583

IID = Independent identically distributed, ADM = Average dynamic model, PDM = Player dynamic model

linear assumption for point spread and found that the relationship was best described by a line with approximately equal slope for positive and negative spreads with respect to the player serving. Thus, momentum is found to be a two-sided coin.

Points of greater importance were associated with decreased win probability, suggesting that players are typically less effective when the pressure is on. Interestingly, even after accounting for the importance of the point, break points had an additional negative effect on performance, indicating a tendency for players to exaggerate the importance of break points.

While the majority of the effects were nearly identical for both the men's and women's tours, there were a few interesting exceptions. Being a point away from a break point opportunity had a modest negative effect on the server's advantage among male players but not female players (Table 2). For the men's tour, missed break opportunities had a modest boost to the next return game of the returner who failed to convert but no evidence of an effect for the women's tour. However, we found some evidence of a decrease in serve advantage for the women's tour during tiebreak points, which was not observed for the men's game.

The estimates in Table 2 represent the estimated change in serve probability (in percentage points units) associated with a one-unit increase in a dynamic factor. Because a one-unit increase might not be meaningful for all of the predictors (e.g. point importance), we compare the average effects on a standardized scale where each bar corresponds to the change in serve win probability for one standard deviation increase in the corresponding dynamic factor (Figure 1). This plot reveals that the strongest predictor for both tours was point spread, where a server with a one standard deviation lead in the point score was estimated to have a 1.4-1.5 percentage point increase in point win probability. Being a set up or set down were runners up with a roughly 1 percentage point effect size, a server being a set up adding to the server's advantage and server being a set down adding to the returner's advantage. Other game and point conditions had more moderate effects, and there was some indication that the negative effects of important points and tiebreak points were greater for female players compared to male players.

3.3 Player Dynamic Effects

The estimates shown in Table 2 represent the effect of each dynamic feature if all players were equally influenced by the state of the match. However, because not all players have the same mentality on court, we would expect some players to respond differently to point conditions than others. The player dynamic model allows for player-to-player differences in their response to the state of the match by estimating a separate dynamic effect for each player.

To determine when the player-specific dynamics better explained observed performance, we calculated the improvement in AIC with the PDM. The changes in AIC shown in Table 2 reveal the presence of important player-to-player differences in the dynamic effect, positive changes reflecting a better fit with the PDM. For the men's game, all but three of the predictors demonstrated important player dynamics. The exceptions were both indicators of missed breaks of service and the total points played in the last game. While dynamic factors with a significant average effect were generally found to also have important player-to-player variation, tiebreak points were found to have important player-to-player variation in performance despite a weak average effect for the men's tour.

For the WTA, improvements with the PDM were fewer and smaller in magnitude than for the ATP. Six of the 11 factors—break points, being one point from a break point, being down a set, point importance, the outcome of the previous point, and the total points played in the previous game—showed important player-to-player variation (Table 2). Thus, male player responses to tiebreak points, being a set up, and having a lead in point were more variable than for female players, whereas female player responses to the total points played in a game were more variable.

3.4 Model Performance

On the ATP, the importance of mental effects was confirmed by the greater accuracy and lower log loss for the predictions of the dynamic models compared to the IID model (Table 3). The consistent superior performance of the PDM over the ADM substantiates the importance of player differences in response to the changing situations of matchplay on the men's tour.

While the dynamic models also improved match predictions for the women's tour, the differences were smaller than for the men's and the performance of the average and player-specific dynamic models were statistically equivalent.

3.5 Mentality Profiles

3.5.1 Men's Tour

The eight dynamic factors that improved the predictive performance of point outcomes for the men's tour revealed eight unique mentalities among the male players who had competed in one or more of the 2015 Grand Slams. The players with each mentality type are displayed in Figure 2 as a dendrogram in which mentalities that are more similar are closer together in their order from top to bottom. The underlying feature profiles for each group are displayed in Figure 3. Here, each line is a specific player's set of dynamic effects on serve and return, with effects scaled to have an equal standard deviation of one. A smoothed regression line is plotted over the observed profiles in each panel to highlight the key differences from the status quo ('The Field') shown in gray.

The Field. We begin with a description of the mentality suggested by the cluster with the largest number of players and, consequently, the most common profile among top male players. This group exhibits a drop in performance when pressure is on the serve, as indicated by the negative effects when

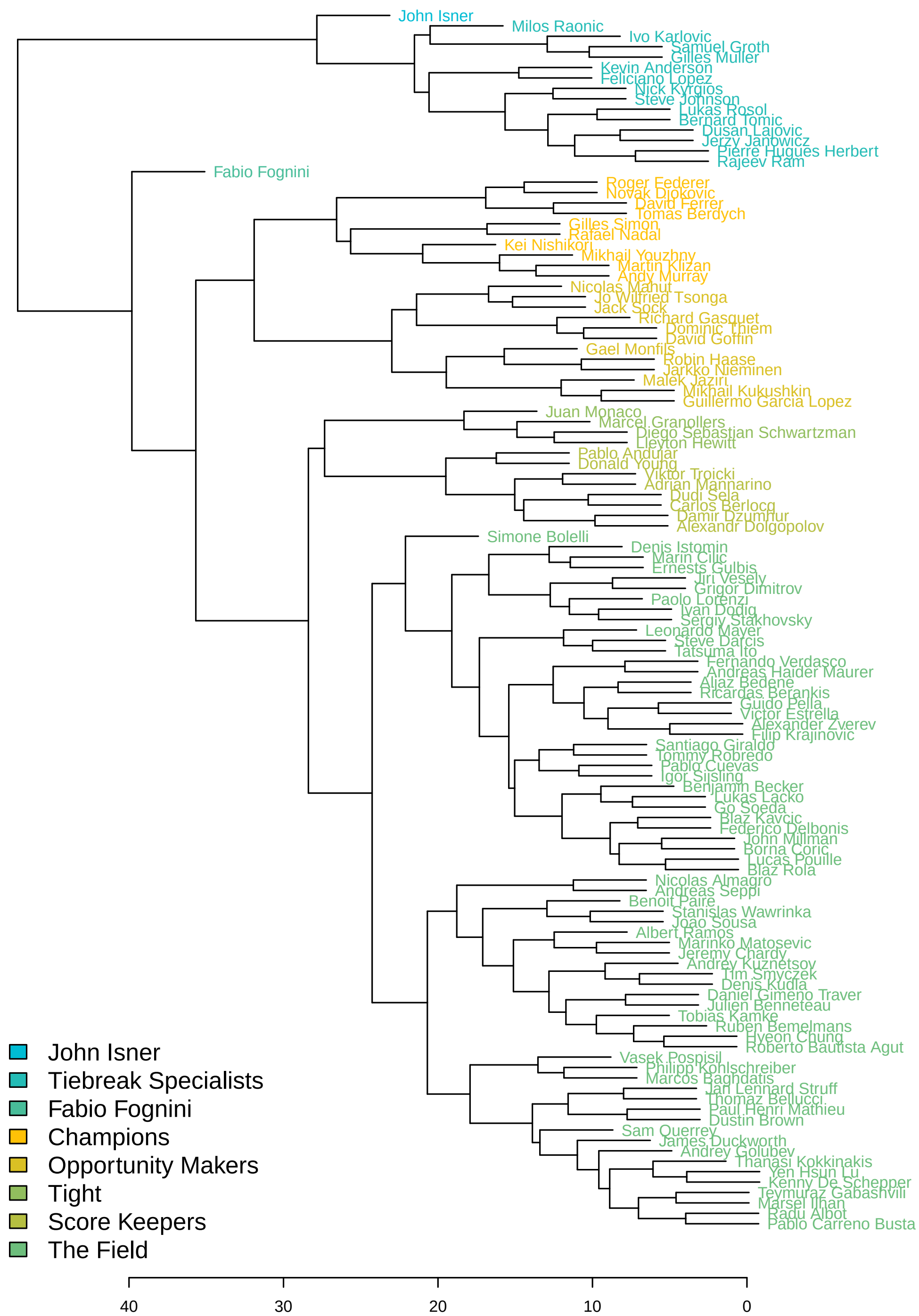

Figure 2: Dendrogram of ATP mentality profiles for players competing in the 2015 Grand Slams. Profiles consisted of 8 dynamic predictors on serve and return. Dissimilarity was measured with a Manhattan distance and players were clustered using complete linkage.

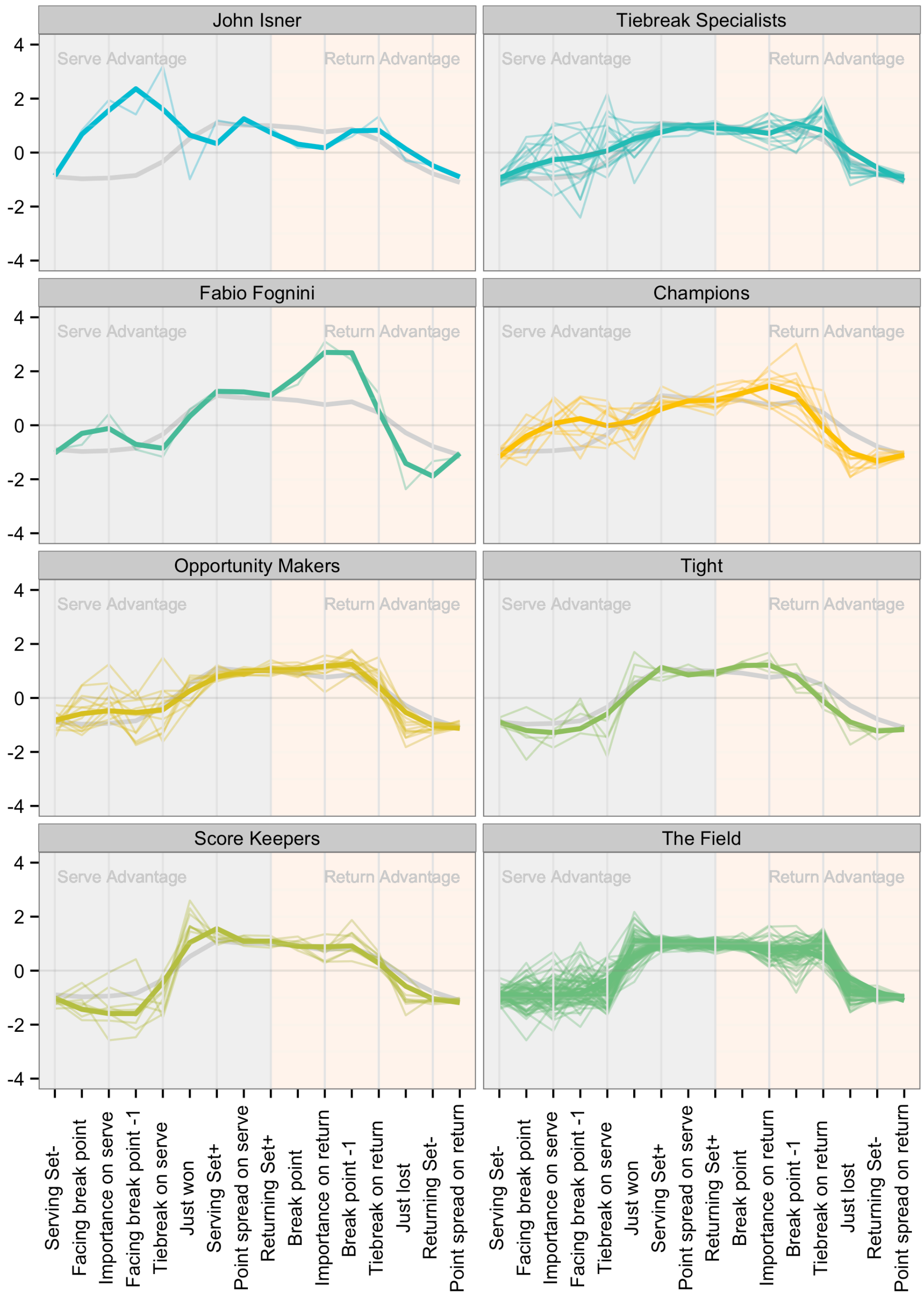

Figure 3: Parallel coordinates plot of mentality types for the ATP players in Figure 2. Effects were scaled to have a common standard deviation of one but were not centered.

a set down or facing important points, such as break points and tiebreaks (Figure 3). These players also exhibit sensitivity to the state of the point score, as is indicated by the positive effects on winning the previous point, having an edge in point spread, or being a set up. These 'hot hand' effects induce a corresponding loser's curse on the return game, in which players who fall behind are even less likely to win a point than when even or ahead in the score.

John Isner. One of two players with a unique profile was big server John Isner. The large positive effects on serve indicate greater overall mental toughness when serving than any other player evaluated. On the defense game, Isner's pattern may indicate a lack of confidence on break point and other important points. His performance on the return game was otherwise similar to the field, with the exception of tiebreaks where he showed strong performance whether returning or serving.

Tiebreak Specialists. Like Isner, these players shine on tiebreak points, raising their performance when serving or returning. On other point types, they also exhibit a similar disparity between the service and return games, with an implied greater overall confidence on serve, but to a less extreme degree than Isner.

Fabio Fognini. The second player found to have a unique mentality was Fabio Fognini, Italian No. 1 at the time of this writing. Fognini's distinctive profile backs up the mercurial label he has often been given by the media[4]. While being unusually mentally strong on more important points (especially on the return game) and on making break point opportunities, the large negative effects when a point or set down on return indicate that he is one of the players most susceptible to collapse.

Champions. The players who currently hold the most Grand Slam titles Novak Djokovic, Roger Federer, Andy Murray and Rafael Nadal (colloquially referred to as 'The Big Four') were all found in the same mentality cluster, suggesting a 'Champion's mentality'. The players in this group exhibit similar strength on serve as the big servers among the tiebreak specialists, being less affected by the state of the point than the average player on tour. On the return game, these players set themselves apart with the mental toughness they show in clutch situations: important points and creating break point opportunities. While the majority of these players also showed a greater ability to convert break points than other competitors, Roger Federer was notably the most negatively affected on break points in this group.

Opportunity Makers. These players had many of the tendencies of the champions group but to a lesser degree. The most consistent positive trait observed compared to the field was the tendency to raise their game to create opportunities to break serve, shown by the positive effect on the point away from break point on the return game. Several players considered to be the most exciting in today's game—Jo Wilfried Tsonga and Gael Monfils—are included in this group.

Tight. This mentality was the only one that was noteworthy for being weaker on certain points than the average top player. Specifically, in clutch situations on serve and when down a point or a set on the return game, these players showed a greater drop in win probability than any others. Finding former World No. 1 Lleyton Hewitt in this group was unexpected but could be explained by the point-level data only covering the final years of his career.

Score Keepers. In addition to appearing generally less confident on serve, the final group of players were unique in their response to the outcome of the previous point, showing a hot hand response when winning a point on serve and a corresponding 'cool hand' after losing a point on return. Thus, the performance of these players are unusually sensitive to the short-term state of the score.

[4]Medlock, W. (June 14, 2015) 'Ranking the Most Unpredictable Tennis Players Today'. Retrieved from: `http://bleacherreport.com/articles/2495031-ranking-the-most-unpredictable-tennis-players-today/`

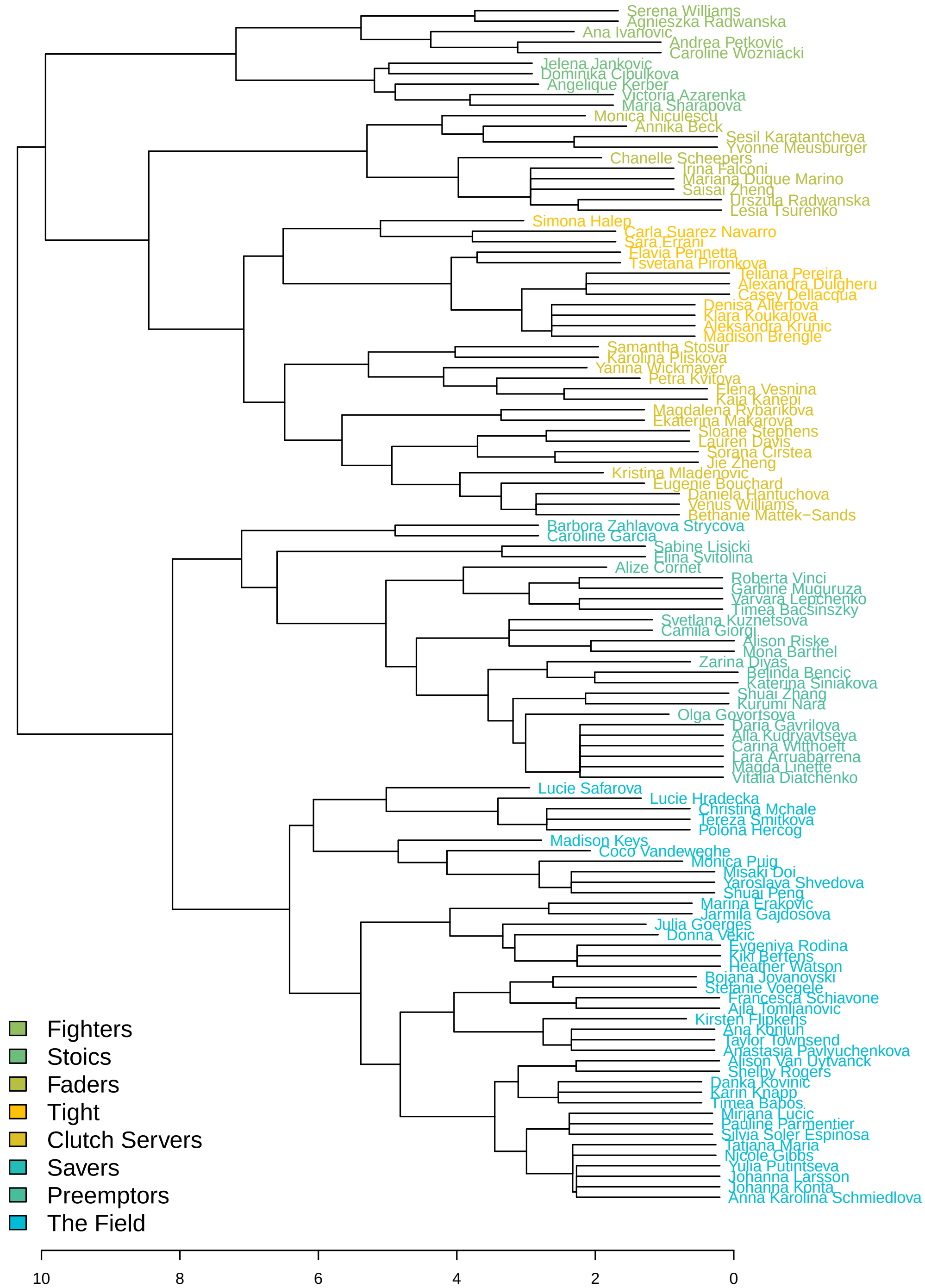

Figure 4: Dendrogram of WTA mentality profiles for players competing in the 2015 Grand Slams. Profiles consisted of 6 dynamic predictors on serve and return. Dissimilarity was measured with a Manhattan distance and players were clustered using complete linkage.

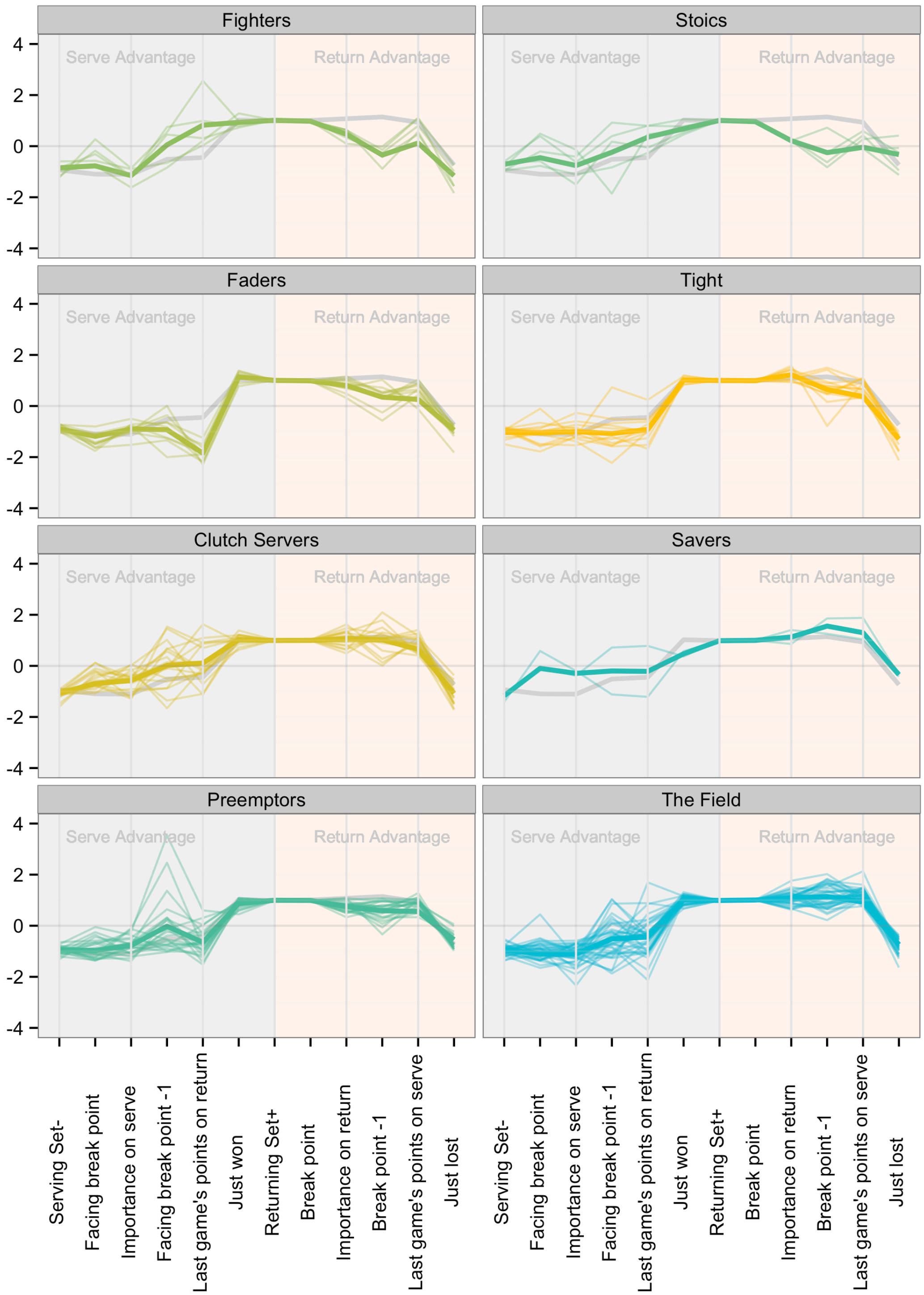

Figure 5: Parallel coordinates plot of mentality types for the WTA players in Figure 4. Effects were scaled to have a common standard deviation of one but were not centered.

3.5.2 Women's Tour

Considering the six player-specific dynamics for the women's tour, eight unique mentalities were also found among players competing in the 2015 Grand Slams (Figure 4).

The Field. Like male players, the majority of the service game of top female players is negatively affected under pressure but benefits from a recent point win—a mini hot hand effect.

Fighters. Several of the players were found to raise the level of their play after tightly contested games (Figure 5). These players had large positive effects associated with more points played in the last game whether serving or returning but especially on serve. These players also show signs of greater cool-headedness in clutch situations on the return game, but their most unique characteristic is the fighter's mentality suggested by their improved performance after long points. It is worth noting that World No. 1 Serena Williams, known for her mastery of the comeback, had the largest positive effect when serving after a long game.

Stoics. Another group of players showed even greater cool-headedness on the return game than the 'Fighters'. The defense performance for these players was virtual unaffected by the state of the score or the importance of points other than break points. On the service game, these players were also the least negatively affected by pressure and the least phased by being a set down. Two players often praised for their mental toughness—Maria Sharapova and Victoria Azarenka—were found in this group.

Faders. In sharp contrast to the 'Fighters' described above, another set of players had a notable negative effect in their service game after a closely contested game, suggestive of mental or physical fatigue.

Tight. While nearly all players show some decline in serve performance in pressure situations, only one group of players had strong and nearly equal negative effects when facing a break point, a break point opportunity, or other important points. Although less pronounced on the return game, the greater negative effect on points away from break point suggest these players are generally more vulnerable in clutch situations.

Clutch Servers. We also observed a group of players that were generally unaffected by pressure on serve, having little or no effect on break points and other important points. There was also some evidence of improved performance on serve after long points like that observed for the 'Fighters' group. Several players found in this group, like Sam Stosur and Petra Kvitova, are known for inconsistent displays of excellent play.

Savers. Two players, Barbora Strycova and Caroline Garcia, stood out from the rest of their cohort for being unusually unmoved when facing a break point.

Preemptors. One of the larger group of players were noteworthy for their mentality on serve when a point from facing a break point. Unlike the field, these players tended to increase their win probability to avoid a possible break of service. Several rising stars of the WTA tour, including Garbine Muguruza and Belinda Bencic, were members of this group.

3.6 Unpredictable Performance

When we measured the prediction error of the PDM for each player (a metric of a player's unpredictability), we found more outliers who were unusually predictable than outliers who were unusually unpredictable. For both tours, a small but roughly equal number of players were highly predictable on serve and return.

Figure 6 highlights the ten players on each tour who were the most extremely predictable. On the men's side, the group clustered in the lower left quadrant are players who had very little variation

Figure 6: Unexplained variance on serve and return according to the PDM Brier scores for each player when applied to the Grand Slam validation data. Brier scores are shown as z-scores and outliers with magnitude of 2 or more are highlighted in blue. The ten most extreme outliers for each tour are labeled.

on the serve or return game after accounting for the dynamic effects of the PDM. Notably, 3 of the strongest servers on tour (John Isner, Milos Raonic, and Ivo Karlovic) were among this group. The lower right quadrant consisted of players who were predictable on serve but much less so on return. It was surprising to observe 3 of the greatest players of the current era (Roger Federer, Novak Djokovic, and Rafael Nadal) in this group.

While a similar pattern in predictability was found for the women's tour, fewer of the outlying players were as highly ranked the male outliers. The exception was for the lower right quadrant where we, as with the men, we found several of the tour's greatest champions: Serena Williams and Maria Sharapova. The similarity of this result for the men's and women's tours makes the intriguing suggestion that mental steadiness on serve combined with variety on return could be defining characteristics of a champion at the professional level.

4 Discussion

We have presented a novel method to quantify and describe the mental side of tennis. This approach measures the way player performance varies on service and return points in response to the dynamics of a match. Our analysis of millions of points of recent performance data found that the typical elite player is influenced by pressure situations and score history, and accounting for these changes in performance improved match predictions. Thus, our study rejects the conclusion of previous work that has questioned the practical significance of point-to-point variation in performance[3] and provides comprehensive quantitative evidence for the importance of player mentality for success in tennis.

The dynamic factors of a match that were found to have the most influence on player performance fall into two broad categories: score history and pressure. The larger debate over the existence of 'hot hand' and 'back to the wall' effects in sport points to a general concern with how an athlete's future performance is influenced by past successes and failures within a competition[5]. Not only did we find evidence that the difference in past sets won and points won each influence subsequent performance of elite tennis players—the yin and yang of momentum influencing players positively when ahead and negatively when behind—, we also found that the state of the score had the strongest effects on performance of all the dynamic factors considered.

Situations that increased the pressure of points were also important and were generally detrimental to performance. Although this result was expected, we were surprised to find that players reacted more negatively to some critical point situations, like break points, than the importance of those points would warrant. In the same way that individuals use heuristics to deal with uncertainty in daily life, this finding suggests that, when there is uncertainty about how the outcome of a point could influence the outcome a match, players may use heuristics that overstate the importance of certain points.

While there are common trends in how players respond to the circumstances of play, not all players at the top level share a common mentality. We found a variety of unique patterns of responses to pressure points, score history, and other dynamics. In fact, two players on the men's game (John Isner and Fabio Fognini), showed response patterns that were unlike any other player on tour. Such individualized profiles have direct implications for coaching as they can highlight areas of ineffectiveness and suggest strategies for improvement. The 'Score Keepers' profile on the men's side, as an example, suggests that tactics for playing more 'in the flow'[1] of points would be expected to improve effectiveness on the return game. On the women's side, having a profile of a 'Fader' would suggest training to focus on player psychological and physical recovery from tightly contested points. Having a description of the mentality patterns of opponents could also be useful information for game strategizing.

The fact that some of the greatest rivalries in tennis history—McEnroe-Borg, Agassi-Sampras, and Evert-Navratilova, to name a few—have also been a clash of personalities implies that tennis may not have a single formula for the mentality of a champion. While personalities may differ, our study found that when it comes to dealing with clutch situations, momentum, and other dynamics of a match, today's male champions have very similar patterns. All of the 'Big Four' (Novak Djokovic, Roger Federer, Andy Murray, and Rafael Nadal) shared a common dynamic profile that was characterized by clutch performance on the return and imperviousness to conditions on serve. It was noteworthy that 3 of this group (Roger Federer, Novak Djokovic, and Rafael Nadal) were also some of the most predictable players on serve and the most unpredictable on return. Taken together, these findings challenge the idea that players should 'play every point as it comes'[3] and argue instead for players who want to play with the mind of a champion to be mentally steady on serve and adaptable on the return.

While the average dynamic effects for the men's and women's tours were remarkably similar, deviations from the status quo mentality were less extensive on the women's tour compared to the men's tour. Fewer of the dynamic variables showed significant player-to-player variation for female players, and the variation that was observed was largely restricted to the service game. This suggests that the variety of mentalities on the women's tour might be less numerous than the men's or that other dynamic factors not considered in this paper are needed to explain variation in female point performance.

When we examined the unexplained variation among female players, we found that two of the most decorated competitors, Serena Williams and Maria Sharapova, had properties that were similar to the male champions: extreme steadiness on serve but greater unpredictability on return. The observation of this pattern on both tours adds strength to the conclusion that champions share common mental skills on court and these skills might be independent of gender.

We have focused on the role of mental skills when interpreting point-to-point fluctuations in performance, but it is important to acknowledge other possible explanatory factors. Variation in a player's within-match performance could also be due to conscious tactics or systematic adjustments made by opponents. Even with some uncertainty about cause, the ability to measure player-specific changes in performance at the point level and summarize these changes in a visually appealing way are significant advances for tennis analytics. Moreover, the tools presented in this paper are not only of academic interest but have practical significance for tennis sports psychologists, coaches, and broadcasters.

With the growth in camera tracking of ball and player position and crowd-sourcing efforts to collect shot-by-shot outcomes in matches[7], we can look forward to a future with richer features to characterize events in tennis. As new features become available, the framework we have presented here will continue to be a valuable resource for analysts evaluating player mentality and could help to quantify the characteristics of player performance with increasingly greater detail.

References

[1] Cian Aherne, Aidan P Moran, Chris Lonsdale, et al. The effect of mindfulness training on athletes' flow: An initial investigation. *Sport Psychologist*, 25(2):177–189, 2011.

[2] Tristan Barnett and Stephen R Clarke. Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*, 16(2):113–120, 2005.

[3] Franc Klaassen and Jan R Magnus. Are points in tennis independent and identically distributed? evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, 96(454):500–509, 2001.

[4] Vijay Kumar, Jitender Kumar Chhabra, and Dinesh Kumar. Performance evaluation of distance metrics in the clustering algorithms. *INFOCOMP Journal of Computer Science*, 13(1):38–52, 2014.

[5] Clare MacMahon, Jörn Köppen, and Markus Raab. The hot hand belief and framing effects. *Research Quarterly for Exercise and Sport*, 85(3):341–350, 2014.

[6] Carl Morris. The most important points in tennis. *Optimal Strategies in Sports*, 5:131–140, 1977.

[7] Jeff Sackmann. First service: The advent of actionable tennis analytics. Sloan Sports Analytics Conference, 2015. `http://www.sloansportsconference.com/?p=15758`.

[8] Dietmar Samulski. Tennis is a mental game: Part One. *ITF Coaching and Sport Science Review*, 40: 14, 2006.

[9] Jim Taylor. Predicting athletic performance with self-confidence and somatic and cognitive anxiety as a function of motor and physiological requirements in six sports. *Journal of Personality*, 55(1):139–153, 1987.

[10] Janet A Young and Alan J Pearce. Understanding, developing and maintaining mental toughness in tennis. *Medicine and Science in Tennis*, 16(2):18–22, 2011.

[11] Lo-Hua Yuan, Anthony Liu, Alec Yeh, Aaron Kaufman, Andrew Reece, Peter Bull, Alex Franks, Sherrie Wang, Dmitri Illushin, and Luke Bornn. A mixture-of-modelers approach to forecasting ncaa tournament outcomes. *Journal of Quantitative Analysis in Sports*, 11(1):13–27, 2015.