Disequilibrium Play in Tennis*

Axel Anderson, *Georgetown University*†
Jeremy Rosen, *Georgetown University*‡
John Rust, *Georgetown University*§
Kin-Ping Wong, *Digonex*¶

February 20, 2024

Abstract

Do the world's best tennis pros play Nash equilibrium mixed strategies? We answer this question using data on serve direction choices (to the receiver's left, right or body) from the Match Charting Project. Using a new approach, we test and reject a key implication of a mixed strategy Nash equilibrium: that the probability of winning the service game is identical for all possible serve strategies. We calculate best-response serve strategies by dynamic programming and show that for most elite pro servers, the DP strategy significantly increases their win probability relative to the mixed strategies they actually use.

Keywords: tennis, games, Nash equilibrium, Minimax theorem, constant sum games, mixed strategies, dynamic directional games, binary Markov games, dynamic programming, structural estimation, muscle memory, magnification effect

JEL Codes: C61, C73, L21

*Acknowledgements:** We thank Georgetown University for research support, and Rust gratefully acknowledges financial support provided by the Gallagher Family Chair in Economics, as well as helpful feedback from the editor James J. Heckman; 3 referees; and Pierre-André Chiappori, Peter Cramton, Drew Fudenberg, Nicolas Legrand, David Levine, Ignacio Palacios-Huerta, David Romer, Ted Sichelman, Frank Vella, Mark Walker, and John Wooders.

†Department of Economics, Georgetown University, e-mail: aza@georgetown.edu

‡Department of Economics, Georgetown University, e-mail: jar361@georgetown.edu

§Department of Economics, Georgetown University, e-mail: jr1393@georgetown.edu

¶Digonex, e-mail: kjw65@georgetown.edu

1 Introduction

Walker and Wooders (2001) (WW) analyzed 40 tennis "point games" from Grand Slam tournaments, focusing on the server's choice of first serve direction: to the receiver's left or right. They analyzed first serves to the ad and deuce courts separately, with each treated as repeated *IID* one-shot simultaneous-move games between the server and receiver. They then concluded that serve location choices are consistent with a mixed strategy Nash equilibrium in their hypothesized static game. In particular, the server's chance of winning a point is the same whether the serve is to the left or the right. Equality of win rates across serve directions has been confirmed in several follow-up studies using additional data. In contrast, our tests typically reject the hypothesis of equal win probabilities across serve directions. We find that most elite professional players such as Roger Federer, Rafael Nadal, and Novak Djokovic could significantly increase their chances of winning if they were to systematically exploit these differences.

Our analysis differs from WW by considering three serve directions (left, right, and *body*) and modeling tennis as a *dynamic game*. We allow for body serves because tennis pros believe they are important, see e.g. Rive and Williams (2011). Dynamics are relevant because the server's strategy and probability of winning the service game depend on the score state as well as *muscle memory effects*. For instance, the server may be more likely to be successful when serving to the same location as the previous serve. Alternatively, the receiver may have more success receiving a serve hit to the same location as the previous serve. We capture muscle memory via the directions of the two previous first serves and show that it can explain serial correlation in serves even when play is in Nash equilibrium. Previous studies including WW have found serial correlation and interpreted it as evidence against Nash equilibrium.

Accounting for dynamics, a third serve direction, and state-dependent serve direction probabilities leads to more powerful tests of mixed strategy play. Our analysis is based on an online database called the Match Charting Project (MCP), run by Sackmann (2013), which crowdsources play-by-play data from professional tennis matches and records all three serve directions used in our analysis. Even after restricting our sample to matches played on hard courts,[1] we end

[1] We do this to eliminate a potential source of heterogeneity that could confound our results, since playing characteristics differ across surfaces. We extend our analysis to grass and clay in Section 5.3.

up with roughly ten times as many serves per server-receiver pair as WW.

However, the main reason why we reject the hypothesis of Nash equilibrium play is a new methodology that models tennis as a dynamic game in contrast to previous work which treated serves as choices in repeated static games. We model serves as decisions at each *subgame* of the overall *service game* between a server and receiver, which ends when one of the players has won at least four points and at least two more points than their opponent. The server chooses the location, speed, and spin of each serve, while the receiver allocates a fixed attention budget to the three serve locations. We test the null hypothesis that observed play in a tennis service game is realization of a Markov Perfect Equilibrium (MPE), in which the server and receiver's strategies depend only on the muscle memory and score state.[2]

We prove that a MPE exists and is unique in the sense that all subgame perfect equilibria result in the same win probability for the server. Serve strategies are also *completely mixed,* i.e. at every state of the service game, the server has a positive probability of choosing any of the three possible serve directions. We define *point outcome probabilities* (POPs) to be the equilibrium probabilities that a serve to a given direction is in, as well as the probability the server wins the rally given the serve is in. Both are conditional on muscle memory, score state and serve direction. The POPs are endogenous objects since they depend on unobserved choices by the server and receiver. However, in a completely mixed MPE, the POPs can be treated as fixed and invariant to temporary changes in serve strategy. The reason is that a receiver would not be able to detect any deviation in serve strategy from a small number of observations if serve directions are chosen randomly at every stage of the service game.

This fact allows us to estimate the POPs as "reduced-form objects" or "projections" that reflect the unobserved strategic decisions of both players in terms of observable outcomes (i.e. faults and points following rallies) conditional on the choice of serve direction which we do observe. This converts the dynamic game to a single-agent dynamic programming (DP) problem since the POPs constitute the payoff-relevant beliefs that the server needs to evaluate different serve strategies. According to the *one shot deviation principle* of game theory, a necessary condition for a serve strategy to be a MPE strategy is that there is no deviation at any stage of the

[2] While the underlying characteristics of the game do not directly depend on the current score, it is the case that with muscle memory effects, strategies generally depend on the score state.

dynamic game that strictly increases the server's expected win probability. In most games, this means that any deviation in serve strategy *strictly reduces* the probability of winning. However, in a completely mixed MPE, a much stronger restriction holds: *all temporary deviations in serve strategy have the same win probability.* This is an extremely strong implication of a completely mixed MPE that results in infinitely many testable restrictions which we exploit to develop powerful new tests of equilibrium play.

In particular, the service game win probability must be the same for all serve directions in all states of the service game. We test these strong implications of a mixed strategy equilibrium by estimating the POPs and the actual serve strategy used in the service game. Since our model has 324 muscle memory/score states and three serve directions, a fully unrestricted estimator of the serve strategy and the POPs would require 4536 parameters for each server-receiver pair — far too many to estimate given the size of our dataset. In Section 4, under a testable assumption that actual serve strategies and POPs are stationary and Markovian (but not necessarily MPE strategies), we estimate flexible reduced-form parametric models of serve strategies and the POPs that include the unrestricted specification as a special case. We use the Akaike Information Criterion (AIC) to select a preferred specification with 44 parameters (12 for the server's strategy and 32 for the POPs) that balances the desire for flexibility against the danger of overfitting.

Rather than separately testing for equal win probabilities across serve locations by aggregating data across individual points (treating first serves as independent, static games as WW did), we introduce a new more powerful *Omnibus Wald test* of the hypothesis of equal win probabilities that must hold *across all possible states of the service game, simultaneously.* We also derive Wald tests of the other key restriction of a completely mixed MPE: that win probabilities are the same for all possible deviation strategies. These tests strongly reject the equal win probabilities for all serve directions implied by completely mixed MPE play for the majority of the elite pros we analyzed, including the very top players, such as Federer, Nadal, and Djokovic. The tests are based on recursive calculations of the conditional probability of winning the entire service game for any given serve strategy in all game states and serve directions. Our tests allow for serial correlation in serve directions and the POPs due to muscle memory effects, thus allowing for state dependence in serve behavior and outcomes that cannot be captured in static approaches to testing for equal win probabilities. Muscle memory not only explains serial correlation in serve directions: accounting for it is the key to the strong rejections of equal win probabilities, even using tests similar to the static testing methodology that WW employed that assume servers maximize the probability of winning each point rather than the overall service game.

To quantify the potential deviation gains from systematically exploiting the unequal win probabilities that our tests reveal, we use DP to calculate best-response serve strategies for individual server-receiver pairs using the estimated POPs to provide outcome probabilities for each point given the choice of serve direction. For all the elite pros we analyze, the DP strategy significantly increases win probabilities relative to the mixed serve strategies implied by our reduced-form estimates of their serve behavior. Adopting the DP serve strategy would improve Nadal's probability of winning a service game against Djokovic from 71% (his current win rate) to 91.5%, and Djokovic's chance of winning against Nadal from 83% to 93.7%.[3]

Thus, the play of elite tennis pros does not constitute a MPE: our empirical analysis reveals many small advantageous one shot deviations (i.e. changes in serve direction at individual points), and the DP strategy takes maximal advantage of all of them, resulting in much more significant deviation gains at the level of the entire service game.[4] The reason why we find much larger deviation gains by modeling tennis as a dynamic game rather than as a sequence of repeated static games is an implication of the tennis scoring system we call the *magnification effect.*[5] For example, if we assume each point is an *IID* bernoulli draw with a 50% chance of a win for the server, the server will also win the service game with 50% probability since the rules of tennis imply that points evolve as a random walk with absorbing states of win and loss, respectively. However, if a change in serve strategy results in a small increase in winning each point, say an increase to 55% (a 10% increase), then tennis scores evolve as a random walk with drift. This causes the probability of winning the service game increases to 62.3%, a nearly 25% increase.

Though the majority of our analysis focuses on elite male pros playing on hard courts, we show that our findings extend to elite women and other less elite pros, as well as to play on clay

[3] Traditional game theory has little to say about "mental ability" since all players are equally rational and intelligent. In the context of our model, these increases in win rates result from a better mental approach to the game. This is because the estimates assume the receiver's strategy and other aspects of the server's play are unchanged under the DP serve strategy. Therefore, relative physical ability is held constant.

[4] We find significant improvements from the DP best response serve strategies for all 94 server-receiver-surface combinations for which we have sufficient data to precisely estimate our model. See Section 5.3 for details.

[5] See Online Appendix D for further discussion of this "magnification effect" that causes our Omnibus Wald test of equal win probabilities over the different possible serve directions to have far greater power than WW's tests.

and grass courts. In general, we find that the magnitude of deviation gains from adopting the DP best response serve strategy is a declining function of "relative ability," as proxied by the server's probability of winning the service game against specific opponents. We do not advise tennis pros to adopt our best-response serve strategies since they are pure strategies that the receiver would eventually learn and adapt to. However we also calculate "robust" mixed serve strategies that account for estimation error and uncertainty about the POPs and the strategy of the receiver. The robust strategies also significantly increase servers' win probabilities but are more difficult for a receiver to detect and adapt to.

To gain insight into the reasons for suboptimal serve choices, we estimate three structural models of the serve direction choices involving increasing degrees of farsightedness. These models allow for persistent shocks to server performance (muscle memory), as well as *IID* shocks that reflect unobserved transitory factors that affect servers' choices. In the fully-dynamic model, the server uses backward induction to maximize the probability of winning the entire *service game*, which is effectively an infinite horizon problem because service games must be won by at least two points. In the point-myopic model, the server solves a two period DP problem to maximize the probability of winning the current *point*. Here, the server accounts for the option value of a second serve but ignores the effect of current decisions on the future state of the service game. Finally, in the serve-myopic model, the server maximizes the probability of winning on each *serve*, a completely static problem that ignores even the option value of the second serve.

The serve-myopic model is typically rejected because of signficant differences in observed serve directions between first and second serves that result from the option value of the second serve. The fully-dynamic model is also nearly always rejected because it implies subjective POPs that are too "pessimistic" compared to our unrestricted estimate of the actual objective POPs. In most cases, the best fitting model is the point-myopic model. It implies mixed serve strategies that are close to the ones players actually use, while constituting a nearly optimal response to the "subjective POPs" in the sense that additional increases in win probability from adopting a full DP serve strategy are negligible. The suboptimality in serve behavior we identify is primarily driven by incorrect server beliefs, i.e. a lack of rational expectations of the server and receiver's strengths and weaknesses as captured by the POPs, rather than players' inability to optimize.

We address concerns that we only have and therefore only use *estimates* of the POPs rather than the *true* POPs which can result in spurious, upward-biased estimates of the deviation gains. To account for this, we derive an approximate probability distribution for the true POPs based on the observed data. We calculate win probabilities for the fully-dynamic, point-myopic, and serve-myopic serve strategies using a random sample of POPs drawn from the asymptotic distribution centered on the point estimates of the POPs. This robustness exercise confirms our core finding: the fully-dynamic and point-myopic strategies based on "rational POPs" have significantly higher win probabilities (in the sense of first-order stochastic dominance) than those implied by the mixed serve strategies the elite pros actually use.

The paper is organized as follows. In Section 2, we briefly review other previous work on testing for minimax play in tennis. Section 3 introduces our dynamic models of tennis serve behavior and the relevant implications from game theory for equilibrium play that we test empirically in this paper. In Section 4, we summarize the key findings from our reduced-form empirical analysis of the MCP database, including our key finding: the frequent rejection of the hypothesis of equal win probabilities for all serve directions. In Section 5, we present estimation results for the three structural models of tennis serve behavior discussed above, and we calculate the deviation gains from using unrestricted estimates of the "objective POPs" to compute optimal serve strategies. Section 6 concludes with further discussion/speculation as to why many elite tennis pros appear to fail to adopt optimal serve strategies given the strong incentives to do so.

2 Previous Literature

The first empirical analysis of tennis using statistical methods that we are aware of is by George (1973) who analyzed the decision of whether the serve should be strong (i.e. fast and more difficult to return, but higher probability of faulting) versus weak (i.e. slow and easier to return, but lower probability of faulting). The first analysis of the tennis *service game* using DP that we are aware of is by Norman (1985) who used it to determine "whether to serve fast or slow on either or both serves at each stage in a game, and a simple policy is found" (p. 1985).

We already noted the seminal work of Walker and Wooders (2001) who focused on first serves modeled as independent static games and were unable to reject the hypothesis of equal win probabilities for serving left or right. They also found negative serial correlation in serve directions across individual points in tennis, which they interpreted as evidence against equilibrium play. With a larger dataset, Hsu, Huang, and Tang (2007) confirmed WW's conclusions regarding equal win probabilities across observed serve directions, but they did not find serial correlation. Wiles (2006) showed that serial correlation may not necessarily be evidence of disequilibrium play due to the presence of a "timing variable," which is analogous to our muscle memory effects.

In addition, Walker, Wooders, and Amir (2011) showed that if a *monotonicity condition* holds, namely if it is always better to win the current point than lose it, then the strategy that maximizes the probability of winning each point also maximizes the probability of winning the service game. This could explain why the "point myopic" serve behavior we find is not necessarily suboptimal, as we discuss in Section 3.5. Also, Lasso de la Vega and Volij (2020) characterized and proved that MPEs exist in a class of recursive zero-sum games that include our model without muscle memory effects. Most recently, Gauriot, Page, and Wooders (2023), using data from 3000 matches and nearly 500,000 serves, confirmed WW's conclusions and noted that "the behavior in the field of more highly ranked (i.e. better) players conforms more closely to theory." But unlike Hsu et al. (2007), they "resoundingly reject the hypothesis that the direction of the serve is serially independent" (p. 1) due to having a large dataset that includes non-elite pros.

Other related papers include Klaassen and Magnus (2001) and (2009). Klaassen and Magnus (2001) tested whether successive points in tennis are independent and identically distributed (*IID*) binary random variables using 481 Wimbledon matches containing nearly 90,000 points. They rejected the *IID* hypothesis, but they found that "Deviations from iid are small, however, and hence the iid hypothesis will still provide a good approximation in many cases." Klaassen and Magnus (2009) abstracted from serve direction and focused on the tradeoff between making a serve hard to return and faulting on the serve, considering both the first and second serves of a point. They rejected the hypothesis that servers optimally solve this tradeoff, but found that "the estimated inefficiencies are not large." In the conclusion, we also discuss empirical evidence for disequilibrium play in other sports including soccer, football and baseball. The findings are mixed: there is strong support for minimax play in penalty kicks in soccer, but strong evidence against equilibrium play in first-down decisions in football and pitch locations in baseball.

3 Modeling Tennis as a Dynamic Game

Tennis is two-player game between a *server* and a *receiver* played in tournaments composed of *matches.* A match consists of a sequence of *sets.*[6] A set, in turn, is a sequence of *service games* in which one of the two players is the server. The server of the first game is chosen by a flip of a coin, and the identity of the server alternates in each game thereafter. Winning a set typically requires winning six games with a lead of at least two games.[7] Each service game consists of a sequence of subgames that are called *points.* A point consists of a first serve, plus an option for a second serve after a *faulted*, or missed, first serve.[8] First serves alternate between the right (deuce) and left (ad) side of the court. The service game ends when one of the players wins at least four points in total and at least two more points than their opponent.

3.1 Dynamic Theory of the Service Game

We use the scalar x to track both the cumulative points scored by each player in the current service game and whether or not the server is attempting a first or second serve. Figure 1 is a directed graph of all the transitions for the *point-state* variable x within a service game. The circular nodes indicate first serves, whereas the square nodes indicate second serves. The game starts in state $x = 1$, which corresponds to a first serve at the tennis score 0–0. If the server wins the point on that first serve, the point-state transits to $x = 3$, corresponding to a first serve at the score 15–0. If the server faults the first serve, the state transits to $x = 2$, which is a second serve at 0–0, and so forth. There are three possible transitions at every first-serve node, two possible transitions at every second-serve node, and two absorbing states (i.e. terminal nodes whose arrows only point in): the server wins ($x = 37$) or loses ($x = 38$) the game.

The arrows connecting most nodes are uni-directional, leading to higher states x. But state $x = 31$ (deuce) is connected by a bi-directional arrow to both $x = 33$ and $x = 35$. This follows from the fact that when the players are tied at 40–40 (i.e. deuce), one of the players must win by

[6] Depending on the tournament, a player must win two of three sets or three of five sets to win the match.

[7] Alternatively, if the score is tied at six-all, the set is decided by a *tiebreak*, in which the winner is the first to score seven points and be ahead by at least two points.

[8] If a serve touches the net and lands in the field of play (a “let”), then the serve (first or second) is redone. Since our data does not record lets, we do not include them in our model.

Figure 1: Score states and transitions in the service game

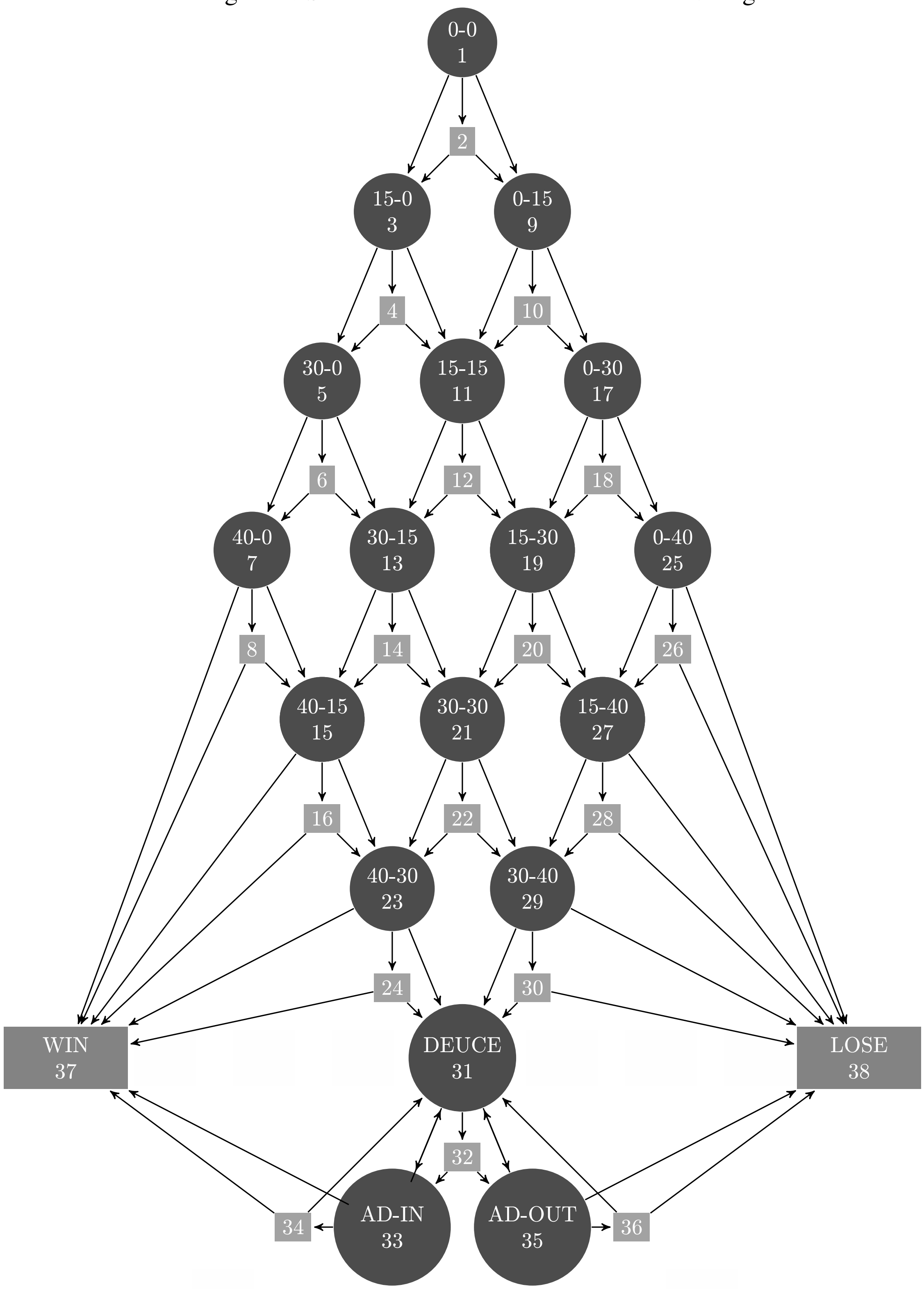
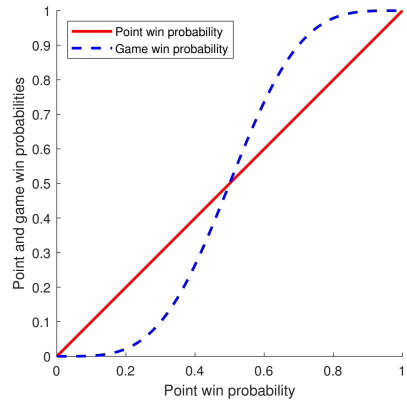


Figure 2: Probabilities of winning the point vs. service game



two points to win the game. Collectively, we refer to states $31-38$ as the *deuce endgame*.[9]

Given these scoring rules, the probability of winning an individual *point* is generally not the same as the probability of winning the *service game* as illustrated in Figure 2. It plots the service game win probability $g(p)$ as a function of the point win probability p under the assumption that each point of tennis is an *IID* Bernoulli draw with probability p of success. Although we relax the assumption that play at different points are independent draws in our model below, the *IID* Bernoulli assumption implies that the point state in tennis evolves as a random walk with drift, with absorbing states $x = 37$ (win for the server) and $x = 38$ (loss for the server), respectively. The game win probability $g(p)$ equals p at $p = 0.5$ as we noted in the introduction. However, any changes in serve strategy that increase the probability of winning each point have a *magnified effect* on the probability of winning the game. Near $p = 0.5$ the slope of the game win probability $g(p)$ is approximately 2.5, so each 10% increase in the point win probability increases the game win probability by 25%. The magnification effect shows how small, hard-to-detect deviation gains at each point of tennis cumulate into much bigger and easier to detect deviation gains in the overall service game, a feature we exploit to derive more powerful tests of Nash play.

At each tennis serve, the server chooses the serve *type* $t = (s,d)$, where $d \in \{l,r,b\}$ indicates the *direction:* to the receiver's left l or right r, or directly into the receiver's body b.[10] Moreover,

[9] Note that states 23 and 24 (29 and 30) are strategically equivalent to 33 and 34 (35 and 36): the transitions to future states in the game including winning or losing are identical.

[10] We follow the literature in assuming that servers "choose" a location, and feel that it is a reasonable fit for the players we analyze. After all, these location categories are broad and our servers are the "best of the best." Tea and Swartz (2022) group serves into our three categories based on "heat maps" of the directions of tens of thousands of

$s \in \mathcal{S} \subset \mathbb{R}^2$ indicates the *speed* and *spin* of the serve ($\mathcal{S}$ is non-empty, closed, and bounded). The receiver *anticipates* the direction choice of the server. Anticipation includes observable (e.g. where to stand) and unobservable choices. We model anticipation with an attention vector $(a^l, a^r, a^b) \geq 0$, where a^d denotes the attention the receiver devotes to serve location d. We normalize the attention budget $a^l + a^r + a^b = 1$. We assume throughout that the serve direction choice weakly follows the choice of a. This captures both the case in which a is a pure location choice, chosen strictly before the server chooses a direction, and the case in which a represents a simultaneous pure mental choice of anticipation.[11]

The probability ℓ that a serve *lands in* (i.e. is not a fault) depends on the court ($c \in \{0,1\}$)[12] and serve type t, while the probability ω that the server wins the subsequent rally (conditional on serving in) depends on the serve type t, court, and attention vector a. We also assume these probabilities can be affected by *muscle memory m*, which we encode as the directions of the previous two first serves. Thus, $m = (d_1, d_2)$ where d_1 is the direction of the previous first serve and d_2 is the direction chosen two first serves ago. We track the previous *two* first serves due to the alternation of serves between ad and deuce courts and the possibility that muscle memory may be more affected by the *last serve to the same court rather than by the last serve, which is to a different court.* We initialize muscle memory to null $m = (\emptyset, \emptyset)$ at the start of the service game, and after any first serve, we update muscle memory from $m = (d_1, d_2)$ to $m' = f(m, d) = (d, d_1)$, reflecting the direction of the current first serve. We assume muscle memory is only updated after first serve states. This still allows m to capture muscle memory effects of the faulted first serve on the subsequent second serve, as well as allowing first serve directions to depend on the direction of the previous first serve to the same court.

We assume the probabilities $\ell(m,d,c,s)$ and $\omega(m,d,c,s,a)$ are continuous in (s,a), satisfy $\ell\omega \in [\underline{w}, \overline{w}]$ for some $0 < \underline{w} < \overline{w} < 1$), and are stationary; namely:

Assumption 1 (Stationarity I)**.** *The functions ℓ and ω may vary across server-receiver pairs, but do not vary over time (independent of m and x) or across service games.*

We assume each player's objective is to win the service game, and so we normalize the win-

men and women's serves at the Grand Slam tournament Roland Garros in 2019 and 2020.

[11] All results extend to a model in which the receiver first chooses a subset of the unit triangle and then chooses a specific element of this subset, simultaneous to the server choosing t. This allows for the realistic case in which the physical location of the receiver on the court constrains, but does not fully determine, the attention vector.

[12] Recall that first serves alternate between the deuce ($= 0$) and ad ($= 1$) courts .

ning payoff to 1 and the losing payoff to 0. Since $\ell\omega$ is strictly interior, the game will almost surely end in a finite number of serves. But for completeness' sake, we assume each player earns payoff $1/2$ if the game never ends.

Let (σ_S, σ_R) denote the server and receiver's strategies (perhaps mixed and arbitrarily history-dependent) in the service game. And let $\mathcal{W}(x,m)$ be the set of probabilities that the server wins the game starting in state (x,m) induced by some pair of (not necessarily Markovian) Subgame Perfect Equilibrium (SPE) strategies (σ_S^*, σ_R^*) for the server and receiver. Appendix A proves:

Theorem 1. *All subgames have a unique value (i.e.* $\mathcal{W}(x,m)$ *is a singleton), and there exists a Markov Perfect Equilibrium (MPE) in which strategies only depend on the current state* (x,m)*.*

Our empirical approach is valid in any MPE in which *all* strategies only depend on (x,m), and Theorem 1 guarantees that there exists an equilibrium with this property. But if there are multiple MPE, then one can construct non-stationary equilibria by making history dependent selections from the set of MPE. In fact, our empirical approach does not rely on Markovian choices of serve direction, meaning it is sufficient for the server's (speed, spin) strategy and the receiver's attention strategy to be Markovian in (x,m). The next assumption guarantees that this is true in any SPE; the proof of Theorem 2 is also in Appendix A.

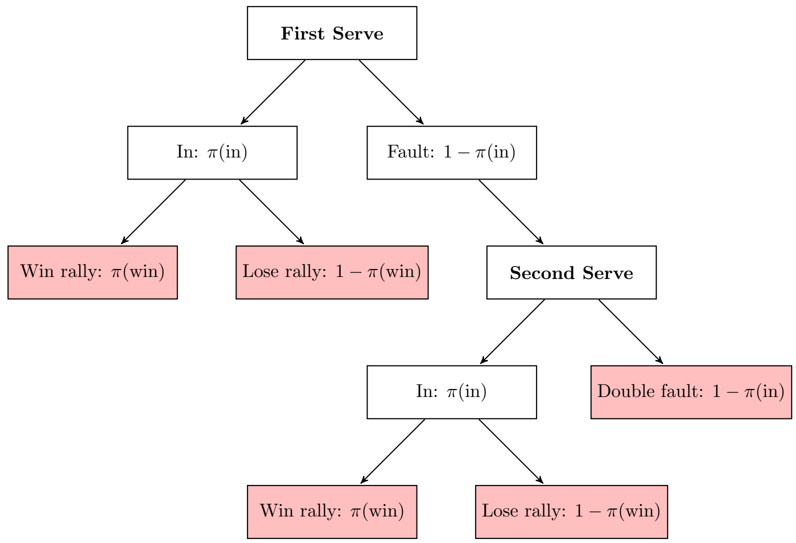
Assumption 2. *The win chance* ω *and chance of not faulting* ℓ *obey:* (i) ω *strictly convex in attention a,* (ii) $\ell\omega$ *strictly concave in (speed, spin) s, and* (iii) $\ell(\omega - 1)$ *concave in s.*

Theorem 2. *If Assumption 2 holds, then every SPE has the same attention strategy and the same (speed, spin) strategy, and each of these strategies is Markovian in the current state* (x,m)*.*

3.2 Serve Direction Strategies from the Induced Dynamic Program

Our empirical analysis uses Match Charting Project (MCP) data, which do not record the serve speed or spin, nor the location of the receiver. To overcome this shortcoming, we use Theorem 1 to project any MPE into the induced DP problem that the server faces when choosing serve directions to maximize the chances of winning the service game. To do so, let $\rho(s|x,m)$ denote a Markov mixed strategy over the speed and spin vector $s \in \mathcal{S}$ for the server, and let $\alpha(a|x,m)$ denote a Markov mixed strategy over attention for the receiver.

Figure 3: Details on the point subgame of tennis



Definition 1. *Given any MPE* (ρ^*, α^*)*, the* ***Point Outcome Probabilities (POPs)*** Π *are:*

$$\pi(in|x,m,d) \equiv \int \ell(m,d,c(x),s)d\rho^*(s|x,m)$$
$$\pi(win|x,m,d) \equiv \int\int \omega(m,d,c(x),s,a)d\rho^*(s|x,m)d\alpha^*(a|x,m).$$

Notice that the mixing probabilities (ρ^*, α^*) will generally depend on the state of the game (x,m), so the POPs will depend on (x,m) even if the underlying conditional probabilities ℓ and ω do not. Given any MPE strategies (ρ^*, α^*), the probabilities π define a single agent "game against nature," i.e. a dynamic optimization problem in which the server chooses a serve direction at each node in Figure 1 in order to maximize the probability of winning the service game. Figure 3 illustrates the extensive form of the *point subgame*, namely the subset of the larger directed graph starting at every odd point-state x. In the point subgame, the server chooses a serve direction for the first serve d_1, and in the event of a fault, the direction of a second serve d_2. The point subgame ends with the server winning or losing a point at each pink node.

Building on Norman (1985), we describe the server's DP problem given π. Let $W(x,m)$ denote the server's maximal *conditional win probabilities* in state (x,m). Let $W(x,m,d)$ be the conditional win probability for the server assuming he serves to direction d on the current serve and behaves optimally on all following serves. Finally, let $x^+(x)$ and $x^-(x)$ denote the successor state in the event that the server wins or loses the point on the current serve, respectively.

The optimal serve strategy can be calculated recursively with the Bellman equation given by:

$$W(x,m) = \max_{d \in \{l,b,r\}} W(x,m,d). \tag{1}$$

$$\begin{aligned} W(x,m,d) &= \pi(\text{in}|x,m,d)\left[\pi(\text{win}|x,m,d)W(x^+(x),m') + [1-\pi(\text{win}|x,m,d)]W(x^-(x),m')\right] \\ &+ [1-\pi(\text{in}|x,m,d)]W(x+1,m') \end{aligned} \tag{2}$$

when x is a first serve state (i.e. x is one of the odd-numbered circular nodes in Figure 1), and

$$\begin{aligned} W(x,m,d) &= \pi(\text{in}|x,m,d)\left[\pi(\text{win}|x,m,d)W(x^+(x),m) + [1-\pi(\text{win}|x,m,d)]W(x^-(x),m)\right] \\ &+ [1-\pi(\text{in}|x,m,d)]W(x^-(x),m) \end{aligned} \tag{3}$$

when x is a second serve state (i.e. x is one of the even-numbered square nodes in Figure 1). The optimal serve strategy $D^*(x,m)$ is the set of serve directions that maximize the win probability

$$D^*(x,m) = \underset{d \in \{l,b,r\}}{argmax}\ W(x,m,d). \tag{4}$$

A necessary condition for an MPE serve strategy to be a mixed strategy is that $D^*(x,m)$ contains more than one serve direction. A completely mixed MPE serve strategy requires equality of the three win probabilities $\{W(x,m,l),W(x,m,b),W(x,m,r)\}$ in all states (x,m). Since the win probability does not depend on the serve direction in any state (x,m) in a completely mixed MPE, this immediately implies the very strong *strategy independence result*, which states that in equilibrium, *any deviation* in serve strategy implies the same win probability $\mathcal{W}(x,m)$ in all states (x,m). We use this strong implication of completely mixed MPE play to construct powerful tests of equilibrium play in Section 4.

Tennis can be viewed as an example of a *directional dynamic game* (DDG) defined by Iskhakov, Rust, and Schjerning (2016), with the exception of the deuce endgame where directionality is not present. While most service games are reasonably short in practice (fewer than 10 points), there is no fixed upper bound on the duration of the *deuce endgame*, the subgame starting at $x = 31$.[13] As a result, tennis must be analyzed as an infinite-horizon dynamic game, starting with the deuce endgame, which is a fully recursive subgame where win probabilities are

[13] The longest deuce endgame that we are aware of was between Anthony Fawcett and Keith Glass in 1975. The score reverted back to deuce 37 times before Glass won the game. Fawcett, however, won the match.

determined by solving the Bellman equation simultaneously as the unique fixed point $W = \Gamma(W)$. After solving the deuce endgame, we use *state recursion* to solve the rest of the game by backward induction across the remaining directionally ordered states $x < 31$.[14]

3.3 Calculating Win Probabilities for Stationary Serve Strategies

It is *sufficient* for our empirical analysis that the unobserved elements of choice (speed, spin, and receiver attention) are Markovian in (x,m), but this condition is not *necessary*. Instead, we can make the following assumption directly on the induced probabilities:

Assumption 3 (Stationarity II)**.** *The **actual POPs** (those implied even if players are not using MPE strategies) are given by families of conditional probabilities $\{\pi(in|x,m,d), \pi(win|x,m,d)\}$ that do not vary over time (independent of (x,m)) or across service games.*

Assumption 1 and MPE (serve, speed, and attention) strategies are jointly sufficient, but not necessary, for Assumption 3. While Assumption 3 does not impose equilibrium, it does implicitly assume that the players are unaware if they are failing to play mutual best responses. Otherwise, they would have an incentive to alter their strategies, perhaps touching off a learning and adaptation process that would violate stationarity.

When stationarity holds and we have enough data, we can consistently estimate Π and use DP to calculate optimal serve strategies numerically. We then compare optimal win probabilities to win probabilities given actual serve strategies (which can also be consistently estimated given sufficient observations on serve directions). Specifically, let $P(d|x,m)$ be an arbitrary (potentially suboptimal) Markovian serve strategy, i.e. the probability that the server chooses direction d in state (x,m). Let $W_P(x,m)$ be the server's win probability starting in state (x,m) and using strategy P for all future serves, and let $W_P(x,m,d)$ be their win probability given choice of serve direction d in state (x,m) and using strategy P for all future serves. We then have the analog to (1):

$$W_P(x,m) = \sum_{d \in \{l,b,r\}} W_P(x,m,d) P(d|x,m). \tag{5}$$

where $W_P(x,m,d)$ is given by Equations (2) and (3) with W_P in place of W. These equations make it clear that W_P is an implicit function of the POPs Π and the serve strategy P.

[14] Norman (1985) recognized the directionality of tennis and grasped the essence of state recursion when he described how the optimal tennis serve strategy and corresponding win probabilities could be calculated by DP.

In fact, we can write an expression for W_P as the solution to a system of linear equations, as is well known in the DP literature on *policy evaluation*. Since there are 298 distinct states (x,m):[15]

$$W_P = w_P(P,\Pi) + M_P(P,\Pi)W_P. \tag{6}$$

Here, $w_P(P,\Pi)$ is a 298×1 vector providing the probability of directly winning the service game in each state (in most states this is zero), and $M_P(P,\Pi)$ is a 298×298 Markov sub-transition matrix (i.e. not all of its rows sum to 1)[16] representing the probability of transitioning between any two states induced by the serve strategy P and POPs Π. Since $M_P(P,\Pi)$ is a Markov sub-transition matrix, the linear system (6) has a unique solution W_P.[17] We can see from (6) that W_P is an implicit function of both (P,Π). We use this result later in the paper to rapidly calculate win probabilities, and via the Implicit Function Theorem, the gradients of the win and conditional win probabilities with respect to model parameters. This enables us to compute standard errors for win probabilities and conduct efficient Wald tests of the hypothesis of equal win probabilities.

With enough observations of service games between a given server and receiver, $W_P(x,m,d)$ can be consistently estimated as the fraction of service games won when the server chose direction d in state (x,m). If the POPs and serve strategies are stationary and Markovian, any win probability W_P must obey identity (6). With 298 states (x,m), non-parametric estimation of W_P involves $298 \times 3 = 894$ individual probabilities $W_P(x,m,d)$. Our analysis also requires an estimate of the actual Markovian serve strategy P and the observed POPs Π. Non-parametric estimation of P requires 298 probabilities and Π a total of $894 \times 2 = 1788$ probabilities. Together, non-parametric estimation of W_P, P, and Π involves a total of $894 + 298 + 1788 = 2980$ probabilities, which would require tens of thousands of service games to estimate with any accuracy.

However, in our dataset, we typically have only 100 to 200 service games per server-receiver pair. To overcome this data limitation, we introduce parametric reduced-form models for serve probabilities and the POPs in Section 4. We still refer to these models as *unrestricted* models of serve behavior because unlike the dynamic structural models we introduce next, we do not require serve direction choices to be best responses to the server's beliefs about the POPs.

[15] There is only one possible muscle memory state at the start of the service game $x = 1$, three possible muscle memory states for $x = 2,3$, and 9, and 9 possible muscle memory states for the remaining 32 score states. Thus, $1 \times 1 + 3 \times 3 + 32 \times 9 = 298$ states (x,m).

[16] The rows of M_P do not all sum to 1, $\ell\omega \in [\underline{w}, \overline{w}]$ implies $\pi(\text{in})\pi(\text{win}) \in [\underline{w}, \overline{w}]$ for any strategy choices.

[17] We prove this in a corollary of Lemma 1 in Appendix A.

3.4 Dynamic Discrete Choice Models of Serve Behavior

To get deeper insight into the behavior of elite servers, we introduce three different structural models of serve behavior that we use in our empirical analysis: 1) a *fully-dynamic model* that assumes the server chooses a strategy that maximizes the probability of winning the entire *service game;* 2) a *point-myopic model* that assumes the server chooses serve directions to maximize the probability of winning each *point;* and 3) a *serve-myopic model* that assumes the server chooses serve directions to maximize the probability of winning each *serve*, ignoring the option value of the second serve. For each of these models, we estimate the server's *subjective POPs* that rationalize observed serve behavior as a best response to the server's potentially subjective beliefs about their own performance and the performance of the receiver.

The other important aspect of these dynamic discrete choice models is the introduction of *unobserved shocks* affecting a server's choice of serve direction. These shocks can be interpreted as idiosyncratic factors that affect the server's choice, which unlike muscle memory are not persistent over states of the game. Technically, the introduction of these shocks implies that the server is using a *pure strategy* that only appears to be a mixed strategy due to the effect of the unobserved "serve shocks," though it is tempting to interpret the *conditional choice probabilities* $P(d|x,m)$ implied by these models as mixed strategies.[18]

We assume that these trembles or preference shocks are *IID* across successive serves and are observed only by the server but not by the receiver or econometrician. Let $\varepsilon(d)$ be the tremble associated with serving to direction d. We further assume that $\{\varepsilon(l),\varepsilon(b),\varepsilon(r)\}$ has a standardized Type 1 extreme value distribution with location parameter normalized so that $E\{\max_d \varepsilon(d)\} = 0$ scaled by $\lambda \geq 0$. If λ is large enough, the server's behavior can mimic a mixed serve strategy even when the win probabilities for different serve directions are unequal. However as $\lambda \downarrow 0$, the conditional choice probabilites converge to a mixed strategy only if the subjective POPs satisfy

[18] An alternative, game-theoretic interpretation is that these shocks represent *trembles*, or incomplete information on players' preferences that imply a Bayesian Nash equilibrium. McKelvey and Palfrey (1995) and McKelvey and Palfrey (1998) studied games of this type and referred to them as *quantal response equilibria.* However, the perspective we take is to model the server's direction choice as a single-agent dynamic discrete choice problem, taking the receiver's behavior as given and embodied in the POPs. Under this interpretation, following Rust (1987), the ε shocks are unobserved state variables, i.e. idiosyncratic *IID* private information or preference shocks known by the server (though not the receiver or econometrician), but which make the server's behavior appear random even though the actual serve strategy is a pure strategy (i.e. a deterministic function of the server's information).

the equal win probability restriction. Thus, dynamic discrete choice models are a natural way to model server behavior while converting the test for equal win probabilities into a simpler test of whether the estimated value of λ equals 0.

Let $\sigma_{FD}(x,m,\varepsilon)$ be the server's strategy under the fully-dynamic structural model as a function of the observed state (x,m) and the unobserved trembles $\varepsilon = (\varepsilon(l),\varepsilon(b),\varepsilon(r))$. The fully-dynamic model presumes that for each (x,m,ε), the server chooses the serve direction that maximizes the probability of winning the service game, given by:

$$\sigma_{FD}(x,m,\varepsilon) = \underset{d\in\{l,b,r\}}{argmax}[\lambda\varepsilon(d) + V_\lambda(x,m,d)], \tag{7}$$

where $V_\lambda(x,m,d)$ is a conditional value function, the analog of the conditional win probability $W(x,m,d)$ defined in Equations (2) and (3) of Section 3.[19] Here, the analog of the function $W(x,m)$ given by the Bellman equation (1) is replaced by $V_\lambda(x,m)$, which is given by:

$$V_\lambda(x,m) = \lambda \log\left(\sum_{d\in\{l,b,r\}} \exp\{V_\lambda(x,m,d)/\lambda\}\right). \tag{8}$$

The serve direction probability implied by the fully-dynamic model is denoted by $P_{FD}(d|x,m)$:

$$P_{FD}(d|x,m) = Pr\{d = \sigma_{FD}(x,m,\varepsilon)|x,m\} = \frac{\exp\{V_\lambda(x,m,d)/\lambda\}}{\sum_{d'\in\{l,b,r\}}\exp\{V_\lambda(x,m,d')/\lambda\}}. \tag{9}$$

$P_{FD}(d|x,m)$ gives the probability of choosing to serve to direction d in observed state (x,m) while accounting for the randomness of the unobserved trembles ε. Since the trembles are *IID* across serves, it would appear that this model should also imply conditional independence of serve directions across successive first and second serves. However, that will actually only be true if there is no muscle memory, i.e. the variable m does not enter $V_\lambda(x,m,d)$ (recall that m is a vector that stores the directions of the two most recent first serves). With muscle memory present, we can still have serial correlation of serves even though the trembles are *IID*.

[19] Generically the optimal strategy $\sigma_{FD}(x,m,\varepsilon)$ will be a *pure strategy* since the probability of more than one serve direction resulting in the same expected reward (including the shock $\varepsilon(d)$) is zero. Below we characterize a necessary condition for σ_{FD} to converge to a mixed strategy as $\lambda \downarrow 0$: this requires the limiting values of V_λ to be equal for all d. As a reviewer noted, using private information shocks to provide an alternative interpretation of what might otherwise appear to be mixed strategies dates back to Harsanyi (1973), and can be used "as a way of justifying the paper's structural model." Indeed, we show in Online Appendix E that Equations (7)–(36) hold in any Bayesian Nash Equilibrium in which the server, but not the receiver, learns ε before choosing his serve strategy.

By Theorem 3 of Iskhakov, Jørgensen, Rust, and Schjerning (2017) we have:

$$W(x,m,d) = \lim_{\lambda \downarrow 0} V_\lambda(x,m,d), \tag{10}$$

uniformly for all (x,m,d). This implies that the only way for $P_{FD}(d|x,m)$ to converge to a completely mixed serve strategy as $\lambda \downarrow 0$ is if the limiting conditional win probabilities $W(x,m,d)$ obey the equal win probability constraints, $W(x,m,l) = W(x,m,b) = W(x,m,r)$ for all (x,m).

The point-myopic and serve-myopic models have the same general structure as the fully-dynamic model, so the serve strategies, value functions, and choice (mixing) probabilities are given by the same equations: (7), (37), and (36). The difference is in the equations defining V_λ. In the serve-myopic model, we have:

$$V_\lambda(x,m,d) = \pi(\text{in}|d,x,m)\pi(\text{win}|d,x,m), \tag{11}$$

i.e. $V_\lambda(x,m,d)$ is the probability of winning the serve. A serve-myopic server maximizes the probability of winning each *serve* while the point-myopic server's objective is to win each *point.* Thus, a point-myopic server performs a two-period backward induction calculation. In any second-serve state, the value of the point myopic server $V_\lambda(x,m,d)$ coincides with the serve-myopic formula given in Equation (11) above. But in any non-terminal first serve state, V_λ is given by:

$$V_\lambda(x,m,d) = \pi(\text{in}|d,x,m)\pi(\text{win}|d,x,m) + [1-\pi(\text{in}|d,x,m)]V_\lambda(x+1,m'), \tag{12}$$

where $m' = (d^{-1},d)$, and $V_\lambda(x+1,m')$ is the maximum win probability over all the second serve directions given in Equation (11).

Note that all three structural models imply probabilistic serve strategies that are entirely determined by the POPs and the scale parameter λ for the trembles. In contrast, the reduced-form model of serve directions does not depend on the POPs since it is estimated separately with a flexible parameterization of serve directions. The structural models can be viewed as restricted special cases of the most flexible specification of the reduced-form model. This enables us to conduct likelihood-ratio specification tests for the three structural models relative to the unrestricted reduced-form specification.[20]

[20] A valid likelihood-ratio specification test would be based on a fully unrestricted version of the reduced-form model with a total of 624 parameters so that it has the flexibility to replicate any conditional probability $P(d|x,m)$.

3.5 The Monotonicity Condition and Myopic Optimality

Unlike chess, where a player's ability to look ahead and consider the consequences of different moves is critical to success, planning ahead may not be as critical to success in tennis. However, the ability to solve at least a two period DP is important, and in Section 4.1, we provide clear evidence that the option of a second serve affects the first serve strategy. In particular, first serves are significantly faster but have a higher chance of faulting than second serves. Second serves are also more likely to be body serves, which are less likely to miss wide in either direction.

But it is less clear whether there is a payoff to solving an infinite horizon DP to determine optimal serve strategies as we did in Section 3.2. Indeed, absent muscle memory effects, point myopic play *is* optimal. In particular, Walker et al. (2011) (WWA) show that tennis is a *binary Markov game*, which is a two player constant-sum game with only two possible outcomes, both for the overall game and all component subgames. WWA assume that the probability of winning a point is independent of the current score and all prior choices. They define a *minimax-stationary strategy* for the overall game as one where each player focuses only on winning the current point. They show that the minimax-stationary strategy coincides with the MPE of the overall game, provided a *Monotonicity Condition* (MC) holds; namely that the probability of winning the service game is always higher after winning any point than losing it. Thus, point-myopic play is optimal without muscle memory effects, since the MC holds in this special case of our model.

However, the MC is not sufficient for WWA's decomposition result to hold in our model with muscle memory effects.[21] Of course, WWA's decomposition result is sufficient, but not necessary, for a point-myopic serve strategy to be optimal. For example, a point-myopic serve strategy will be optimal whenever points are *IID* Bernoulli outcomes, since we showed in Section 3.1 that this implies the tennis score state is a random walk with drift. Therefore, a strategy that increases

Given the limited number of observations for specific server-receiver pairs, our specification for $P(d|x,m)$ depends on only 12 parameters, though it produces estimates that fit the data well. While our reduced-form specification does not strictly nest the structural models, it has sufficient flexibility to closely approximate the structural serve probabilities. We can also do tests using the non-nested specification test of Vuong (1989). However, we prefer the LR tests and also rely on the AIC model selection criterion to select our preferred structural specification, similar to the way we used it to select our preferred specification for the reduced-form model.

[21] In the presence of dynamic effects such as muscle memory, we can imagine there might be tradeoffs, such as serve directions that increase the chance of winning the current point but which compromise the ability to win subsequent points. For example, serving to the same direction as the previous serve may reduce the chance of faulting due to muscle memory effects, but doing so might improve the receiver's ability to return future serves hit to that direction, reducing the server's effectiveness in subsequent states of the game.

the probability of winning any point also increases the probability of winning the service game.

In fact, a point-myopic serve strategy will be trivially optimal in our data as long as we are observing a Nash equilibrium in which the server is using a completely mixed strategy, since *any* deviation serve strategy is optimal in that case. So in this section we assume that we are not observing a complete mixed NE, allowing for pure strategies in some states and/or disequilibrium play. The following testable *generalized monotonicity condition* (GMC) implies that point-myopic play is optimal in the presence of muscle memory:

Definition 2. ***Generalized Monotonicity Condition (GMC)*** *The probability of winning the game is always higher after winning a point than losing it:*

$$W(x^+(x),m) > W(x^-(x),m) \quad \textit{for all } (x,m). \tag{13}$$

Also, at any first serve state x and for any direction d_1 of the previous first serve, if the probability of winning the point is higher for serving to direction d than d', then:

$$\begin{array}{rcl} W(x^+(x),(d,d_1)) & \geq & W(x^+(x),(d',d_1)) \\ W(x^-(x),(d,d_1)) & \geq & W(x^-(x),(d',d_1)). \end{array} \tag{14}$$

The first inequality in (13) is the same Montonicity Condition that WWA showed is sufficient to establish that an optimal point-myopic strategy also maximizes the probability of winning the service game when there is no "state dependence" other than through the score state x. When there are dynamic effects such as muscle memory, the MC alone will no longer be sufficient to establish this result. The new condition (14) imposes the additional restriction that if a particular serve direction d results in a higher probability of winning the current point than some other serve direction d', that the choice of d will not lower the server's probability of winning the service game relative to d' in the subsequent states $x^+(x)$ and $x^-(x)$.

A stronger sufficient condition that implies condition (14) is to require that at any first serve state x, the service game win probability does not depend on the direction of the previous first serve (though it can depend on the serve direction d_2 two first serves ago when the server was serving to the same court, unless $d_2 = \emptyset$ for the first serves in the game to the deuce and ad courts). When there is muscle memory, the choice of first serve direction has future consequences because it affects the evolution of muscle memory, which in turn affects the server effectiveness in future states of the game. However if muscle memory only operates across successive serves to the same court, then condition (14) will hold, and the server will not have to consider the current serve direction's effects on the probabilities of winning in subsequent game states.

Theorem 3. *If GMC holds, then the optimal point myopic and fully dynamic serve strategies coincide and result in the same service game win probabilities for the server.*

The proof of Theorem 3 is in Appendix A.4. In Section 4, we show that the GMC is testable, and there are server-receiver pairs for which the GMC fails for our empirically estimated W. In these cases, point-myopic strategies are suboptimal, although we show that typically the cost of suboptimality in terms of reduced service game win probability is small.

Consider the implications for serial independence of serve directions. If there are no muscle memory effects, then WWA's monotonicity condition is equivalent to Inequality 13 and implies that any strategy that maximizes the probability of winning each point also maximizes the probability of winning the service game. This leads WW to the wrong conclusion that, "In addition to equality of players' winning probabilities, equilibrium play also requires that each player's choices be independent draws from a random process" (p. 1522). Independence in serve directions is a *consequence of their assumption* that serve strategies do not depend on previous choices and outcomes. When there is history dependence such as muscle memory, equilibrium strategies will generally depend on both the score state x and muscle memory m. This history dependence implies that a point-myopic serve strategy will generally be suboptimal in terms of maximizing the probability of winning the service game. However, even when the stronger form of the monotonicity condition — our GMC Assumption 14 holds, and a point-myopic serve strategy is optimal, serial correlation in serve directions will still be a general property of a MPE, as we show in Online Appendix F. Thus, serial independence of serve directions is generally *not* an implication (i.e. necessary condition) of a mixed strategy MPE in the presence of muscle memory, though independence does hold in the absence of dynamic effects such as muscle memory.

We conclude by summarizing the testable implications of the theory we have presented:

1. **Nash equilibrium:** There should not be any alternative serve strategy that increases the server's probability of winning.

2. **Mixed-strategy equilibrium:** The probability of winning in state (x,m) should be equal for all serve directions chosen with positive probability in state (x,m).

We also test the following behavioral implications of the GMC:

3. **Optimality of point-myopic serve strategies:** When the GMC holds, it is optimal for the server to adopt a point-myopic strategy that focuses only on the goal of maximizing the probability of winning each point.

4. **Serial independence:** If GMC holds and there are no muscle memory effects, the direction of a first serve should not depend on the direction of any previous first serve.

4 Reduced-Form Analysis of Serve Strategies

In this section, we start with a model-free descriptive analysis of our data. Then we introduce a flexible *reduced-form* model of tennis that we use to test several of the key implications of game theory summarized in Section 3, particularly the implication that conditional win probabilities are the same for all serve directions.[22] Most of our analysis focuses on a set of elite professional tennis players, who have all been ranked number one in the world and won multiple Grand Slams. These players are Roger Federer, Rafael Nadal, Novak Djokovic, Andy Murray, Pete Sampras, and Andre Agassi.[23] We focus on these players for two reasons: first, we have the most observations for them, and second, if we can show that they serve suboptimally, that means even the best of the best are susceptible to strategic errors.

4.1 Analysis of Play of Specific Server-Receiver Pairs

We have sufficient observations to analyze serve decisions of specific server-receiver pairs. Table 1 summarizes some of the key statistics for 10 selected elite server-receiver pairs, revealing a great deal of player-specific heterogeneity that would be masked in pooled statistics. The table presents the total number of service games and serves we observe for each pair. A typical service game ends after seven to nine serves. The third column breaks down the total number of serves we observe into first and second serves. We can see that the "crude fault rate" (fraction of total

[22] Our analysis is not assumption-free, as we maintain Assumption 2 for validity of our statistical tests.
[23] We provide results for women and additional men in Section 5.3.

Table 1: Win probabilities and mixed serve strategies for selected elite server-receiver pairs

Server $\rightarrow$ receiver	Games, serves Serves/game	1st serves 2nd serves	Serve directions L	B	R	Win prob (std) P-value: $P_1 = P_2$
Roger Federer $\rightarrow$	523, 4732	3208	.4402	.1007	.4592	.7686 (.0184)
Rafael Nadal	8.36	1164	.2174	.2698	.5129	5.1×10^{-60}
Rafael Nadal $\rightarrow$	519, 4081	3227	.6616	.2048	.1336	.8092 (.0172)
Roger Federer	7.86	854	.5937	.3208	.0855	6.3×10^{-12}
Roger Federer $\rightarrow$	411, 3501	2524	.4521	.0939	.4540	.8200 (.0190)
Novak Djokovic	8.52	977	.4084	.3408	.2508	6.7×10^{-68}
Novak Djokovic $\rightarrow$	407, 3653	2696	.4640	.1565	.3795	.8010 (.0198)
Roger Federer	8.98	957	.4389	.3365	.2247	1.0×10^{-33}
Rafael Nadal $\rightarrow$	346, 2937	2230	.3964	.2825	.3211	.7197 (.0241)
Novak Djokovic	8.49	707	.4073	.5403	.0523	2.4×10^{-64}
Novak Djokovic $\rightarrow$	356, 2877	2149	.4067	.1619	.4314	.7528 (.0222)
Rafael Nadal	8.08	728	.1484	.2940	.5577	1.2×10^{-40}
Novak Djokovic $\rightarrow$	230, 1958	1447	.4651	.1244	.4105	.7696 (.0278)
Andy Murray	8.51	511	.2192	.4618	.3190	3.0×10^{-53}
Andy Murray $\rightarrow$	230, 2141	1522	.3863	.0841	.5296	.7435 (.0288)
Novak Djokovic	9.31	619	.4233	.4782	.0985	5.8×10^{-122}
Pete Sampras $\rightarrow$	140, 1275	884	.4434	.0724	.4842	.9000 (.0254)
Andre Agassi	9.11	391	.4680	.1765	.3555	5.3×10^{-8}
Andre Agassi $\rightarrow$	135, 1125	825	.5127	.1115	.3758	.8666 (.0293)
Pete Sampras	8.33	300	.5766	.2700	.1533	7.2×10^{-16}

serves that are second serves) differs across servers, ranging from a low of 21% for Nadal serving to Federer to a high of 30% for Sampras serving to Agassi.

The three columns labelled L, B, and R list the fraction of first and second serves to the receiver's left, body, and right for each server-receiver pair. We see that in general, servers use mixed strategies, but the mixing probabilities for second serves differ significantly from those for first serves. The last column of the table includes the P-value of a likelihood-ratio test of the null hypothesis that the mixing probabilities for the first and second serves are equal. We see that for all servers, we can decisively reject this hypothesis. In general, the fraction of second serves to the body is about twice as large as for first serves.

We also see that servers adjust their serve strategy for different receivers. For example, from Table 1, we can see that Nadal uses a different serve strategy when serving to Federer than when serving to Djokovic. The final column of Table 1 shows the empirical service game win probability for the server and its estimated standard error (i.e. the fraction of games the server won). We see quite a bit of variation in service game win probabilities across different server-receiver pairs, ranging from a low of 72% for Nadal serving to Djokovic, to a high of 90% for Sampras serving to Agassi. Even controlling for the same server, we see a fairly big variation in win probabilities depending on the receiver: for example, Nadal has an 81% service game win probability when serving to Federer, as Federer is a weaker receiver than Djokovic. Given the relatively small standard deviations in estimated win probabilities, we can strongly reject the null hypothesis that the variation in estimated win probabilities is due to sampling error.

4.2 A Flexible, Agnostic Reduced-Form Probability Model of Tennis

In order to test the key necessary condition for a mixed strategy equilibrium — equality of win probabilities for all serve directions — a deeper econometric analysis is required. As discussed in Section 3, we do not have enough data to estimate a non-parametric model. Instead, we estimate a flexibly parameterized reduced-form specification for serve strategies $P(d|x,m)$ and POPs $(\pi(in|x,m,d),\pi(win|x,m,d))$. Following standard terminology in the dynamic discrete choice literature, we refer to the serve probabilities below as Conditional Choice Probabilities (CCPs). Let $f(x,m,d)$ be a $1 \times K_P$ vector of indicators for various subsets of the state/action space. We will describe specific choices for f below. In general, f will partition the state space into subsets where serve direction probabilities are similar. Let θ_P be a conformable $K_P \times 1$ vector of coefficients to be estimated. We use the following flexible logit model for the CCPs:

$$P(d|x,m,\theta_P) = \frac{\exp\{f(x,m,d)'\theta_P\}}{\sum_{\delta\in\{l,b,r\}}\exp\{f(x,m,\delta)'\theta_P\}}. \tag{15}$$

Similarly, let $g_{in}(x,m,d)$ and $g_{win}(x,m,d)$ be $1 \times K_{in}$ and $1 \times K_{win}$ vectors of indicators used to define the following binary logit models for $\pi(in|x,m,\theta_{in})$ and $\pi(win|x,m,\theta_{win})$ that depend on parameter vectors $(\theta_{in},\theta_{win})$:

$$\pi(in|x,m,d,\theta_{in}) = \frac{\exp\{g_{in}(x,m,d)'\theta_{in}\}}{1+\exp\{g_{in}(x,m,d)'\theta_{in}\}} \tag{16}$$

$$\pi(win|x,m,d,\theta_{win}) = \frac{\exp\{g_{win}(x,m,d)'\theta_{win}\}}{1+\exp\{g_{win}(x,m,d)'\theta_{win}\}} \tag{17}$$

We estimate the parameter vector $\theta = (\theta_P, \theta_{in}, \theta_{win})$ by maximum likeihood using the log-likelihood function $L(\theta)$ given by:

$$L(\theta) = \sum_{n=1}^{N} \sum_{s=1}^{S_n} \left[\log\left(P(d_{s,n}|x_{s,n}, m_{s,n}, \theta_P)\right) + \log\left(h(o_{s,n}|x_{s,n}, m_{s,n}, d_{s,n}, \theta_{in}, \theta_{win})\right)\right], \qquad (18)$$

where N is the total number of service games observed for a particular server-receiver pair, S_n is the number of serves in game n, and $(d_{s,n}, x_{s,n}, m_{s,n})$ are the observed serve direction, game state, and muscle memory state at serve s in game n. The variable $o_{s,n}$ is the outcome of serve s of game n and takes one of three possible values: $o_{s,n} = 1$ if the serve is in (i.e. not faulted) and the server wins the subsequent rally, $o_{s,n} = 2$ if the serve is in and the server loses the subsequent rally, or $o_{s,n} = 3$ if the serve is faulted. In all first serve states (i.e. odd values of x), the service game state transits to a second serve in the event that $o = 3$, but in any second serve state the server loses the point when $o = 3$ (i.e. the server "double faults"). The conditional probability $h(o|x, m, d, \theta_{in}, \theta_{win})$ is defined in terms of the POPs as follows:

$$h(o|x,m,d,\theta_{in},\theta_{win}) = \begin{cases} \pi(in|x,m,d,\theta_{in})\pi(win|x,m,d,\theta_{win}) & \text{if } o = 1 \\ \pi(in|x,m,d,\theta_{in})[1-\pi(win|x,m,d,\theta_{win})] & \text{if } o = 2 \\ 1-\pi(in|x,m,d,\theta_{in}) & \text{if } o = 3. \end{cases} \qquad (19)$$

Different specifications correspond to different partitions of the state/action space. The finest partition, in which every pair (x, m) is a partition element, yields the full non-parametric specification for (P, Π). Since we do not have sufficient observations to reliably estimate a fully non-parametric model, we face a classic bias/variance tradeoff between estimating a flexible model with many parameters, versus a more parsimonius model with sufficiently many observations per parameter to guard against overfitting plus outliers that could distort our estimates of the POPs.

We manage this tradeoff using model selection techniques, particularly the Akaike Information Criterion (AIC), which penalizes model complexity. Specifically, $\text{AIC} = 2[K - L(\hat{\theta})]$, where K is the total number of parameters estimated in a given model, $L(\hat{\theta})$ is the maximized value of the log-likelihood function, and $\hat{\theta}$ is the maximum likelihood estimate of the parameters of the particular model. We evaluated several different specifications (i.e. choices for f, g_{in} and g_{win} with different numbers of parameters and different partitions of the state space) and chose as our preferred specification (detailed in Appendix B) the model with the lowest AIC.[24]

[24] We also used the Bayesian Information Criterion $BIC = K\log(n) - 2L(\hat{\theta})$, which has a stronger penalty for

Our preferred specification still involves a large number of parameters per server-receiver pair. In particular, the serve direction probabilities $P(d|\cdot)$ are governed by 12 parameters. Eight are for the full set of interactions between direction $d \in \{l, r\}$[25] and court (deuce vs. ad) and serve (first vs. second) dummy variables. The other four are muscle memory parameters that interact the court and serve dummies with a dummy indicating whether the direction of the current serve equals the direction of the previous first serve to the same court. Each of the POPs $(\theta_{in}, \theta_{win})$ is determined by 16 parameters, which correspond to the same indicators as just described for serve probabilities, except that the current serve direction must include all three directions, since the probabilities $\pi(\text{in})$ and $\pi(\text{win})$ need not sum to 1 across serve directions.[26]

We do not have the space to present all these parameter estimates and the associated standard errors for each of the server-receiver pairs we analyzed, though we provide them for Federer vs. Djokovic in Appendix B and can provide the rest on request. As we will describe further in the next sections, our preferred specification balances the tradeoff described above: it provides an accurate probability model of the entire service game for individual server-receiver pairs while avoiding the dangers of overfitting. In the remainder of this section, we will use this model to test several of our key assumptions, including the key hypothesis of Nash equilibrium play in tennis.

4.3 Testing for Stationarity Across Matches

We now test Assumption 3, i.e. stationarity of the POPs $(\pi(in|x, m, d, \theta_{in}), \pi(win|x, m, d, \theta_{win}))$ over time and across service games. Suppose the CCPs are also stationary in this same sense. Then the stochastic processes of serves and serve outcomes in any given service game between a given server and receiver on a given type of court are Markovian, and the realizations of these Markov processes are *IID* across successive service games. While the presence of muscle memory and the scoring rules of tennis imply that the sequences of serve directions and serve outcomes

---

model complexity. But we found that the higher complexity penalty caused the BIC to select models with fewer parameters. In cases where one model specification was nested within another encompassing specification, the BIC would choose the more parsimonious restricted specification even though likelihood-ratio tests would lead us to reject the parsimonious restricted specification relative to the less restricted encompassing model.

[25] Since $P(d|x, m, \theta_P)$ sums to 1 across directions, we need only include two dummies for current serve direction.

[26] We also estimated our models with a reduced-form specification that adds a binary partition of the score state capturing how far ahead (or behind) the server is in the current service game to the reduced-form specification of the POPs and CCPs. All of our qualitative results are robust to this alternative specification.

will be serially correlated *within a service game,* there will be no dependence across successive games because we assume muscle memory is reset at the start of each service game, and there are no other effects that lead to dependence across successive games.

It is easy to think of reasons why Assumption 3 may not hold. For example, if a server injures his shoulder, this can adversely affect the POPs. Or there might be psychological effects, such as confidence or a "hot hand," that could lead to serial correlation across successive service games served by the same player. Finally, if a player is learning and adapting, his strategy may slowly evolve as he learns more about his opponent's weaknesses and adjusts to exploit them.

On the other hand, we need to pool across service games to have any hope of efficiently estimating the parameters determining the CCPs and POPs. From the previous section, our preferred reduced-form model has a total of 32 parameters (24 if we exclude muscle memory effects). Given that a typical service game lasts for about seven to nine serves, we need at least 100 games of data to estimate these 32 (or even 24 parameters) with sufficient accuracy. We are particularly concerned about *overfitting*, along with the possibility that the model's predictions of conditional win probabilities will be incredibly high or low due to the lack of sufficient observations.

The stationarity assumption is testable, and we present results from a simple way of testing for stationarity in Tables 2 and 3 below. For the same set of 10 server-receiver pairs in Table 1, we estimate separate CCPs and POPs for different subsets of service games based on year groupings of our data.[27] For example, for Agassi and Sampras, we divide the data into two subperiods, one from 1995-1999 where we have 67 service games and another from 2000-2002 where we have 60 games. For Federer and Nadal, we have sufficient data to create three subperiods: 2004-2007, 2008-2012, and 2013-2017 with 67, 81, and 91 service games, respectively. We estimate a pooled, or "restricted," model using all games in all years and imposing stationarity. Next, we estimate an "unrestricted" model that allows the CCPs and POPs to be different in each subperiod.

We calculate a likelihood-ratio (LR) test statistic of the stationarity hypothesis by taking two times the difference between the log-likelihood for the unrestricted model (i.e. summing the individual subperiod log-likelihoods) and the log-likelihood for the restricted model. The unrestricted model with two subperiods has a total of $2 \times 32 = 64$ (with muscle memory) or $2 \times 24 = 48$ (without muscle memory) parameters that are estimated separately without placing any equal-

[27] Appendix C tests stationarity with an alternative data partition.

Table 2: Tests for stationarity of POPs: $\{\pi(in|x,m,d,\theta_{in}),\pi(win|x,m,d,\theta_{win})\}$

Server → receiver	Muscle Memory			No Muscle Memory		
	Restricted	Unrestricted	LR test (df)	Restricted	Unrestricted	LR test (df)
	LL, AIC	LL, AIC	P-value	LL, AIC	LL, AIC	P-value
Roger Federer →	-1934.3	-1901.9	64.9 (64)	-1940.1	-1916.9	46.5 (48)
Rafael Nadal	**3932.6**	3995.7	.447	**3928.2**	3977.7	.533
Rafael Nadal →	-1880.9	-1843.1	75.7 (64)	-1883.2	-1853.9	58.6 (48)
Roger Federer	**3825.9**	3878.2	.150	**3814.5**	3851.9	.140
Roger Federer →	-2280.7	-2242.0	77.4 (64)	-2284.7	-2256.9	55.8 (48)
Novak Djokovic	**4625.4**	4676.0	.122	**4617.5**	4657.7	.206
Novak Djokovic →	-2403.9	-2363.9	80.1 (64)	-2411.7	-2383.3	56.7 (48)
Roger Federer	**4871.9**	4919.7	.084	**4871.3**	4910.7	.183
Rafael Nadal →	-1414.2	-1402.0	24.3 (32)	-1415.8	-1408.5	14.6 (24)
Novak Djokovic	**2892.4**	2932.1	.832	**2879.6**	2913.0	.932
Novak Djokovic →	-1302.1	-1280.7	42.7 (32)	-1304.5	-1285.9	37.1 (24)
Rafael Nadal	**2668.1**	2689.4	.098	**2656.9**	2667.9	.043*
Novak Djokovic →	-1183.2	-1165.6	35.0 (32)	-1188.7	-1175.4	26.7 (24)
Andy Murray	**2430.3**	2459.3	.326	**2425.5**	2446.8	.317
Andy Murray →	-1280.1	-1258.5	43.0 (32)	-1287.9	-1273.0	30.0 (24)
Novak Djokovic	**2624.1**	2645.1	.092	**2623.9**	2641.9	.186
Pete Sampras →	-1117.9	-1097.4	40.9 (32)	-1124.1	-1107.3	33.7 (24)
Andre Agassi	**2299.7**	2322.9	.135	**2296.2**	2310.5	.091
Andre Agassi →	-1031.1	-1009.1	44.0 (32)	-1032.6	-1012.3	40.6 (24)
Pete Sampras	**2126.2**	2146.2	.077	**2113.2**	2120.6	.019*

ity restrictions across the two sample subsets. Thus, the LR test has 32 degrees of freedom for the specification with muscle memory and 24 degrees of freedom for the specification without muscle memory. For the player pairs where we have enough data to divide the sample into three subperiods, the test has 64 and 48 degrees of freedom, respectively.

Table 2 shows that we are unable to reject our stationarity Assumption 3 at the 5% level for any of the 10 player pairs we analyzed under the muscle memory specification. For the specification without muscle memory (which is the preferred one for all 10 pairs under the AIC criterion), we only reject stationarity for Agassi serving to Sampras. We conclude that Assumption 3 is a reasonable approximation to the data, which justifies pooling across service games to get the most reliable possible estimates of the CCPs and POPs.

Table 3: Tests for stationarity of CCPs: $\{P(d|x,m)\}$

Server → receiver	Muscle Memory			No Muscle Memory		
	Restricted	Unrestricted	LR test (df)	Restricted	Unrestricted	LR test (df)
	LL, AIC	LL, AIC	P-value	LL, AIC	LL, AIC	P-value
Roger Federer →	-1844.8	-1812.0	65.6 (24)	-1874.4	-1844.0	60.6 (16)
Rafael Nadal	3713.5	**3696.0**	.000*	3764.7	**3736.1**	.000*
Rafael Nadal →	-1688.2	-1636.1	104.2 (24)	-1690.9	-1643.3	95.2 (16)
Roger Federer	3400.3	**3344.1**	.000*	3397.8	**3334.5**	.000*
Roger Federer →	-2265.1	-2233.0	64.1 (24)	-2293.9	-2264.6	58.4 (16)
Novak Djokovic	4554.1	**4538.1**	.000*	4603.7	**4577.3**	.000*
Novak Djokovic →	-2423.8	-2392.7	62.2 (24)	-2454.8	-2425.7	58.1 (16)
Roger Federer	4871.6	**4857.4**	.000*	4925.5	**4899.4**	.000*
Rafael Nadal →	-1432.6	-1419.5	26.2 (12)	-1437.5	-1426.8	21.4 (8)
Novak Djokovic	2889.3	**2887.0**	.010*	2891.0	**2885.6**	.006*
Novak Djokovic →	-1347.4	-1343.2	8.6 (12)	-1364.2	-1360.3	7.8 (8)
Rafael Nadal	**2718.9**	2734.3	.740	**2744.4**	2752.6	.450
Novak Djokovic →	-1201.6	-1173.0	57.2 (12)	-1221.0	-1191.7	58.6 (8)
Andy Murray	2427.1	**2393.9**	.000*	2458.0	**2415.4**	.000*
Andy Murray →	-1250.0	-1182.7	134.6 (12)	-1254.9	-1189.5	130.9 (8)
Novak Djokovic	2524.0	**2413.4**	.000*	2525.9	**2411.0**	.000*
Pete Sampras →	-1085.4	-1070.3	30.3 (12)	-1096.4	-1083.9	25.0 (8)
Andre Agassi	2194.9	**2188.6**	.003*	2208.8	**2199.8**	.002*
Andre Agassi →	-931.8	-919.7	24.2 (12)	-945.2	-934.2	22.1 (8)
Pete Sampras	1887.6	**1887.4**	.019*	1906.5	**1900.4**	.005*

Table 3 displays the results of LR tests of stationarity of the CCPs. The AIC is lowest for the muscle memory specification for 7 of the 10 player pairs. We reject stationarity of the CCPs for 8 of the 10 and 9 of the 10 pairs under the muscle memory and no muscle memory specifications, respectively. Under the assumptions in Section 3, if the POPs are stationary, players use MPE strategies, and if the MPE is unique, then the CCPs must be stationary as well. Thus, we conclude that the rejections in Table 3 indicate either a) there are mutiple MPE, and the players “select” different MPEs in different time periods, or b) players are not playing MPE strategies, and the variation in CCPs reflects the effect of some sort of learning or experimentation with different serve strategies over time.

4.4 Testing Equality of Win Probabilities Over Directions and Strategies

We now present tests of the key implication of a completely mixed MPE that point and service game win probabilities are independent of serve direction, plus the stronger implication that all deviation serve strategies imply the same win probability. We strongly reject these implications in models with muscle memory. As we show below, the data support the presence of muscle memory for almost all player pairs due to strong evidence of serial dependence in serve directions. Accounting for this dependence is key to our ability to detect violations of equal win probabilities.

Table 4 compares the recursively calculated game win probabilities from Equation (6) to non-parametric estimates (i.e. simply the fraction of games won) of these probabilities at the first serve of each service game. We restrict attention to the first serve of the game beause it provides the most observations to reliably estimate the game win probability non-parametrically. The final column shows the P-value of a Durbin-Hausman-Wu (DHW) test of our preferred reduced-form specification. Recall that the DHW test compares a consistent but inefficient non-parametric estimator of the game win probability to a relatively efficient estimate of it from Equation (6).[28]

We see that the calculated win probabilities are close to the non-parametric estimates and are almost always within a standard deviation of each other. The high P-values of the DHW specification tests in the final column of the table show that for all servers except Federer serving to Nadal, we are unable to reject the reduced-form specification and its implied win probability. In the case of Federer serving to Nadal, the RF estimate of the win probability is .796, slightly more than one standard deviation away from the non-parametric estimate of the win probability, .829. The middle columns compare the non-parametric estimates of the conditional win probabilities with the corresponding estimates implied by the reduced-form model, $W_P(1,1,d)$ for $d \in \{l, b.r\}$. The estimates are generally close to each other, though there are some cases where there are large differences due to small numbers of observations resulting in noisy non-parametric estimates.[29]

[28] The DHW specification test compares two estimators of a given quantity or parameter: an inefficient but $\sqrt{N}$-consistent estimator that is consistent under both the null and alternative hypotheses, and an efficient estimator that is also $\sqrt{N}$-consistent for the true parameter under the null hypothesis but may be inconsistent under the alternative hypothesis (N denotes the sample size). In our case, the relevant null hypothesis is that our reduced-form specification for (P,Π) is correct, and the non-parametric estimates of the win probabilities in Table 4 are inefficient but consistent even if the null hypothesis is false (i.e. our reduced-form model is misspecified). Under the null, the DHW test statistic is equal to the square of the two estimates of the win probability divided by the differences in the asymptotic variances, and it converges to a χ^2 random variable with one degree of freedom.

[29] For example, in the case of Sampras serving to Agassi, due to the low probability that Sampras serves to the

Table 4: Estimated win and conditional win probabilities at first serve of service game

Server → receiver	Est.	Win prob 1st serve	Conditional win probability, 1st serve **L**	**B**	**R**	DHW test P-value
Roger Federer →	NP	.796 (.026)	.816 (.025)	.650 (.030)	.803 (.025)	.004
Rafael Nadal	RF	.829 (.023)	.828 (.024)	.819 (.027)	.833 (.022)	
Rafael Nadal →	NP	.786 (.026)	.748 (.028)	.896 (.020)	.762 (.028)	.107
Roger Federer	RF	.807 (.023)	.808 (.023)	.807 (.025)	.803 (.025)	
Roger Federer →	NP	.810 (.024)	.844 (.022)	.867 (.021)	.767 (.025)	.504
Novak Djokovic	RF	.818 (.020)	.826 (.020)	.812 (.023)	.813 (.021)	
Novak Djokovic →	NP	.782 (.025)	.769 (.026)	.710 (.028)	.815 (.024)	.910
Roger Federer	RF	.781 (.022)	.792 (.022)	.769 (.026)	.774 (.024)	
Rafael Nadal →	NP	.712 (.035)	.685 (.036)	.726 (.035)	.750 (.034)	.992
Novak Djokovic	RF	.712 (.034)	.712 (.034)	.701 (.035)	.718 (.034)	
Novak Djokovic →	NP	.829 (.029)	.868 (.026)	.735 (.034)	.833 (.029)	.278
Rafael Nadal	RF	.848 (.023)	.854 (.023)	.830 (.027)	.849 (.023)	
Novak Djokovic →	NP	.794 (.034)	.759 (.036)	.750 (.036)	.841 (.031)	.871
Andy Murray	RF	.791 (.029)	.796 (.031)	.758 (.034)	.799 (.029)	
Andy Murray →	NP	.721 (.038)	.816 (.033)	.500 (.042)	.701 (.038)	.675
Novak Djokovic	RF	.717 (.036)	.735 (.036)	.712 (.039)	.703 (.037)	
Pete Sampras →	NP	.885 (.028)	.894 (.027)	1.00 (.000)	.859 (.030)	.150
Andre Agassi	RF	.866 (.024)	.866 (.025)	.839 (.029)	.872 (.024)	
Andre Agassi →	NP	.874 (.029)	.907 (.026)	.867 (.030)	.852 (.032)	.362
Pete Sampras	RF	.859 (.024)	.861 (.026)	.853 (.026)	.859 (.024)	

In contrast, the middle columns of Table 4 reveal big differences in game win probabilities for different serve directions. The largest gap is a 31 percentage point difference between the win probability for left (81.6%) vs. body (50%) for Murray serving to Djokovic, roughly 10 times higher than their estimated standard errors. The average value of the maximum deviation in game win probabilities over all states and serve directions in Table 4 is 19 percentage points, nearly four times as large as the estimated standard errors of these maximum deviations.

Although Table 4 reassures us that our recursive calculation of game win probabilities results in accurate and efficient estimates, some readers may be skeptical that the evidence against equal

body (approximately 7%, see Table 1) and the relatively low number of games in which we observe him serving (140), the non-parametric estimate of the conditional win probability of serving to the body equals 1. Of course, this non-parametric estimate is probably not a reasonable estimate: instead it is likely to be a lucky outcome for Sampras who happened to win every one of the 8 games where he served to Agassi's body on the very first serve of the game.

game win probabilities is as convincing as the tests of equal *point* win probabilities that WW and most of the subsequent literature have focused on. In Table 5 we present Omnibus Wald tests of equality of the *point win probabilities* at all states (x,m) of tennis simultaneously. Recall that under the point-myopic theory of play, the server does not consider the future consequences of different serve directions and instead maximizes the probability of winning each point, which is a two-period DP problem. Starting at the second serve, the restriction that point win probabilities are the same for all serve directions holds if the serve win probability $V(x,m,d)$ given by:

$$V(x,m,d) = \pi(\text{in}|x,m,d)\pi(\text{win}|x,m,d), \tag{20}$$

is the same for all threee serve directions in all second serve states (x,m). In any first serve state, the point win probability $V(x,m,d)$ is given by:

$$\begin{aligned} V(x,m,d) &= \pi(\text{in}|x,m,d)\pi(\text{win}|x,m,d) + \\ & [1-\pi(\text{in}|x,m,d)][\sum_{d' \in \{l,b,r\}} P(d'|x,m')\pi(\text{in}|x+1,m',d')\pi(\text{win}|x+1,m',d')], \end{aligned} \tag{21}$$

and it should also be the same for all d where $m' = f(m,d')$ is the new muscle memory state implied by serve direction d', which is updated only in first serve states.

Table 5 provides the test statistics, P-values, and degrees of freedom for the Omnibus Wald test of equality of conditional win probabilities for all serve directions, i.e. the restrictions that $V(x,m,l) = V(x,m,b) = V(x,m,r)$ for all 298 states (x,m), where $V(x,m,d)$ is given in Equations (20) and (22) above. We see that there are strong rejections of the hypothesis of equal win probabilities for all player pairs except for Federer serving to Djokovic. Overall, we also see big differences in point win probabilities across different serve directions: the average maximum deviation over all 10 player pairs is .275 with a standard deviation of .083.

Why are we able to reject the hypothesis of equal point win probabilities so strongly when previous studies were unable to do so? We believe that accounting for muscle memory is a large part of the story. If we repeat the Wald tests in Table 5 under the no muscle memory specification, we find smaller maximum deviations in win probabilities over serve directions over the reduced state space, and we only reject the equal win probability hypothesis for 2 of the 10 player pairs above.[30] Previous tests such as those by WW focused only on first serves

[30] These pairs are Djokovic serving to Murray and Djokovic serving to Nadal (the Wald statistics are 18.4 and 23.0

Table 5: Wald tests of equal point win probabilities, MM specification

Server → Receiver	Wald statistic	Degrees of freedom	P-value
Roger Federer → Rafael Nadal	405.4	29	5.9×10^{-68}
Rafael Nadal → Roger Federer	243.2	30	2.9×10^{-35}
Roger Federer → Novak Djokovic	23.6	30	.75
Novak Djokovic → Roger Federer	274.5	27	8.9×10^{-43}
Rafael Nadal → Novak Djokovic	83.5	29	3.5×10^{-7}
Novak Djokovic → Rafael Nadal	69.6	28	2.1×10^{-5}
Novak Djokovic → Andy Murray	52.3	30	0.007
Andy Murray → Novak Djokovic	212.0	30	2.7×10^{-29}
Pete Sampras → Andre Agassi	146.4	30	2.9×10^{-17}
Andre Agassi → Pete Sampras	198.6	30	8.9×10^{-27}

and pooled all first serve observations into just two groups: the deuce and ad courts. Pooling the data in this way masks big differences in win probabilities for different serve directions that appear once we control for serial correlation in serves by conditioning on previous serve history via the muscle memory state. The importance of controlling for muscle memory is confirmed in Online Appendix D, where we present the results of Wald tests of equal win probabilities under the no muscle memory specification, essentially replicating WW's approach but using our data and including second serves. Like WW, these tests usually fail to reject equal win probabilities.

We now turn to testing for *equal service game win probabilities* using the fully dynamic version of the model with recursively calculated game win probabilities W_P from Equation (5). We also use Equations (2), (3), and (5) to calculate the direction-specific win probabilities $W_P(x,m,d)$ entering the recursive formula for $W_P(x,m)$. The Omnibus Wald test of equal win probabilities for all serve directions involves testing in all 298 states (x,m) the equality restrictions:

$$W_P(x,m,l) = W_P(x,m,b) = W_P(x,m,r). \tag{22}$$

In our preferred specification with muscle memory, this test amounts to a test of 596 equality restrictions of the form given in Equation (22).[31]

with P-values of .018 and .003 under 8 degrees of freedom, respectively). Also, the average value of the maximum difference in point win probabilities over all directions and states for these 10 pairs is .20 (standard deviation = .08).

[31] The specification with muscle memory is our preferred specification, since the no muscle memory specification is strongly rejected for all but one of the player-pairs (see Table 7 in Section 4.5 below).

Table 6: Wald tests of equal service game win probabilities, MM specification

Server → Receiver	4 fixed serve strategies at 3 states, 9 df	RF serve strategy at 4 states, 12 df
Roger Federer → Rafael Nadal	1.4×10^{-11}	.605
Rafael Nadal → Roger Federer	.873	.018
Roger Federer → Novak Djokovic	6.5×10^{-30}	1.6×10^{-68}
Novak Djokovic → Roger Federer	.0009	.220
Rafael Nadal → Novak Djokovic	2.2×10^{-254}	.526
Novak Djokovic → Rafael Nadal	4.0×10^{-91}	4.5×10^{-44}
Novak Djokovic → Andy Murray	1.4×10^{-69}	.018
Andy Murray → Novak Djokovic	9.1×10^{-91}	.00001
Pete Sampras → Andre Agassi	.787	.003
Andre Agassi → Pete Sampras	.764	.667

Since the conditional win probabilities are implicit functions of (P,Π), and (P,Π) are functions of the 44-dimensional parameter vector $\hat{\theta} = (\hat{\theta}_P, \hat{\theta}_{in}, \hat{\theta}_{win})$, we use the delta method to construct the Omnibus Wald test statistic. This is a quadratic form in the 596×1 vector of differences in conditional win probabilities between serve directions over all states, using the Moore-Penrose inverse of the 596×596 covariance matrix of win probability differences, expressed as a sandwich formula in terms of the 44×44 variance covariance matrix for the reduced-form parameter vector $\hat{\theta}$. We need to use the Moore-Penrose inverse rather than the standard matrix inverse because the rank of the covariance matrix (which equals the degrees of freedom of the χ^2 distribution of the Omnibus test statistic under the null hypothesis) is at most 44.[32]

We find that the Omnibus Wald test results in even stronger rejections of equal service game win probabilities than we obtained when testing for the equality of point win probabilities in Table 5, with P-values of nearly 0 for all player pairs (see Table 6). However, there are reasons to distrust such strong rejections due to small sample numerical issues with the Moore-Penrose inverse, which is not a continuous function of its matrix argument. The discontinuity can invalidate the standard χ^2 asymptotic distribution of the Wald test statistic under the null hypothesis. Andrews (1987) provides a sufficient condition for "generalized Wald tests" (which rely on the Moore-Penrose inverse) to have the usual asymptotic χ^2 distribution: namely, the rank of the

[32] The rank of the covariance matrix is generally even lower than 44 (the number of parameters) because the rank of the 596×44 gradient matrix of the win probability differences is often less than 44.

finite sample covariance matrix of the restrictions must converge with probability 1 to the rank of the limiting covariance matrix. Matrix rank is not a continuous function either, but it is semi-continuous, so the rank condition of Andrews (1987) should hold generically.[33]

Nevertheless, we have observed a tendency for Wald test statistics to grow rapidly with the total number of retrictions being tested, so we have opted to adopt a more conservative approach to testing for equal win probabilities using *small subsets of the total number of restrictions.* Since matrix inversion is continuous, our conservative approach reduces the problem of spurious rejections, though it does lead to power/size tradeoffs in the choice of how many resrictions to test. It also requires additional choices over which subset of restrictions we choose to test.

The last column of Table 6 presents the P-values for our more conservative test of equal win probabilities at a subset of six points in the state space: 1) 0–0, 2) 15–0, 3) 0–15, 4) 40–15, 5) 15–40, and 6) deuce. This test has 12 restrictions, and since the covariance matrix for these restrictions is invertible, the test has 12 degrees of freedom. It rejects the hypothesis of equal win probabilities at the 5% level for 6 of the 10 player pairs in the table.[34]

The second column of Table 6 reports P-values for a Wald test of the invariance of win probabilities with respect to strategy deviations that must hold when MPE serve strategies are completely mixed. We compute the win probabilities of four different fixed strategies at three points in the state space, resulting in a test with 9 restrictions and degrees of freedom (since the covariance matrix for this reduced set of restrictions is invertible). The four fixed serve strategies are: 1) always serve left, 2) always serve to the body, 3) always serve right, and 4) serve to each direction with probability 1/3 (i.e. a uniform distribution across the serve directions). The three particular score states used in these tests are 1) 40–15, 2) 15–40, and 3) deuce. This test strongly rejects the hypothesis of equal win probabilities for 7 of the 10 player pairs. Thus, our new approach to testing for equal win probabilities, allowing for serial correlation in serve directions via muscle memory effects, and our inclusion of body serves and more observations, explain why we reject mixed strategy Nash play for the majority of the elite player pairs in our data set.

[33] A sufficient condition for the rank condition in Andrews (1987) to hold is that the limit covariance matrix of the restrictions must be *regular*. That is, the rank must be the same for all covariance matrices in a neighborhood of the limiting value. In addition, Proposition 4 of A. D. Lewis (2009) establishes that the set of regular matrices is an open and dense set of the space of all matrices.

[34] A test of equal point win probabilities over the same score states rejects for 3 of the 10 pairs at the 5% level.

4.5 Testing for "Muscle Memory" Effects

We conclude this section by presenting evidence of serial dependence in the CCPs and, to a lesser extent, the POPs. We have already shown in Section 4.1 that there are significant differences between the mixture probabilities for first vs. second serves, so it should not be surprising that we also find significant serial dependence between first and second serves. However, this serial dependence is not necessarily inconsistent with equilibrium play, since the server considers the option value of the second serve when choosing the speed, spin, and direction of the first serve.

The more important question is whether there is serial correlation *across successive first serves hit to the same court*. Our preferred specification for our reduced-form model of serve directions conditions on the deuce vs. ad court, so the server's strategy can alternate across courts. However, this effect does not induce serial correlation in serve directions across successive first serves to the same court. We capture serial correlation in such serve directions via muscle memory. The muscle memory specification also induces serial correlation between first and second serves, since the CCPs for second serves depend on the direction of the faulted first serve. The specification without muscle memory allows for serial dependence as play alternates between courts, but it implies zero serial correlation across successive first serves to the same court.

We use likelihood-ratio tests for serial correlation by comparing the likelihood that includes muscle memory with restricted likelihood that excludes the muscle memory variable m since as we showed in Section 3, serve directions become serially independent under this specification. Table 7 presents the results of LR tests of the hypothesis of "no muscle memory effects." The last column of this table shows that except for Nadal serving to Federer, we can reject the hypothesis of no muscle memory in the CCPs at the 5% level. However, when it comes to the POPs, we have far weaker evidence of serial correlation. For most of the server-receiver pairs in Table 7, we are unable to reject the hypothesis of no muscle memory effects in the POPs.

Why is this the case? We think it may have to do with the receiver's behavior. Specifically, if muscle memory effects are real, and the receiver shifts his position accordingly, then the receiver can effectively cancel out any effect that muscle memory would impart on the POPs. As a result, we observe serial correlation in the server's directional choices but not in the POPs.[35] This results

[35] We solved for the MPE in a two-direction version of our model and observe that muscle memory effects induce much larger changes in the server and receiver's equilibrium mixed strategies than in the POPs.

Table 7: Tests for muscle memory effects in (P,Π)

Server → receiver	Model	No muscle memory		Muscle memory		LR test
		AIC	**LL**	**AIC**	**LL**	P-value
Roger Federer →	Serves	3764.7	-1874.4	3713.5*	-1844.8	4.3×10^{-12}
Rafael Nadal	POPs	3928.3*	-1940.1	3932.6	-1934.3	.170
Rafael Nadal →	Serves	3397.8*	-1690.9	3400.3	-1688.2	.249
Roger Federer	POPs	3814.5*	-1883.3	3824.9	-1880.9	.779
Roger Federer →	Serves	4603.7	-2293.9	4554.1*	-2265.1	9.3×10^{-12}
Novak Djokovic	POPs	4617.4*	-2284.8	4625.4	-2280.7	.414
Novak Djokovic →	Serves	4925.5	-2454.8	4871.6*	-2423.8	1.1×10^{-12}
Roger Federer	POPs	4871.3*	-2411.7	4871.9	-2403.9	.048
Rafael Nadal →	Serves	2891.0	-1437.5	2889.3*	-1432.6	.044
Novak Djokovic	POPs	2879.6*	-1415.8	2892.4	-1414.2	.921
Novak Djokovic →	Serves	2744.8	-1364.2	2718.9*	-1347.4	9.0×10^{-7}
Rafael Nadal	POPs	2656.9*	-1304.5	2668.1	-1302.1	.779
Novak Djokovic →	Serves	2458.0	-1221.0	2427.1*	-1201.6	7.7×10^{-8}
Andy Murray	POPs	2425.5*	-1188.7	2430.3	-1183.2	.202
Andy Murray →	Serves	2525.9	-1254.9	2524.0*	-1250.0	.044
Novak Djokovic	POPs	2623.9	-1287.9	2524.1*	-1280.1	.049
Pete Sampras →	Serves	2208.8	-1096.4	2194.9*	-1085.4	2×10^{-4}
Andre Agassi	POPs	2296.2*	-1124.1	2299.7	-1117.9	.134
Andre Agassi →	Serves	1906.5	-945.3	1887.6*	-931.8	2×10^{-5}
Pete Sampras	POPs	2113.2*	-1032.6	2126.2	-1031.1	.934

can be consistent with Nash equilibrium play, as we demonstrate in Online Appendix F.

5 Dynamic Structural Analysis of Serve Strategies

In the previous section, we estimated an unrestricted reduced-form model of serve directions and POPs and showed that this flexible, agnostic model of tennis rejects the key implication of a mixed strategy Nash equilibrium: namely that the POPs satisfy the restriction that the server's win probability is the same for all serve directions in every state of the service game (and thus, all possible serve strategies have equal win probabilities). These tests did not require us to make any assumptions about server behavior beyond Assumption 3 (Stationarity). This section provides more insight into server behavior by presenting estimation results for the three structural models we introduced in Section 3.4. We estimate their parameters by maximum likelihood using the full panel likelihood function (18) on data from hard courts for the 10 elite server-receiver pairs listed in Table 8.[36] Based on our findings in Section 4.5, which provide strong evidence of serial correlation in serve directions across successive points, we focus on the specification with muscle memory. For comparability, we use the same specification of the POPs as in our reduced-form model presented in Section 4. Therefore, our structural models have a total of 33 parameters: the 32×1 vector of POP parameters $(\theta_{in}, \theta_{win})$, plus the extreme-value scaling parameter λ.

The structural estimates of the POPs can be regarded as estimates of the server's *subjective beliefs* that may or may not correspond to rational *objective beliefs* about the true POPs, which we estimate via our unrestricted POP estimates. As we discussed in Section 3.4, the structural model implies mixed strategy Nash equilibrium play if two key restrictions are satisfied: 1) $\lambda = 0$ (i.e. players use mixed strategies, which can only hold if the POPs obey the equal win probability restrictions), and 2) the subjective POPs equal the objective POPs.

Unlike the reduced-form specification, the assumption of optimal play implicit in the structural models imposes "cross equation restrictions" on the serve probabilities: they are implicit functions of the POP parameters as well as the scale parameter λ for the extreme value distributed trembles. This implies that the likelihood function is no longer block-diagonal between the POP parameters $(\theta_{in}, \theta_{win})$ and λ, unlike the unrestricted reduced-form model where we have block-diagonality between the 12×1 parameter θ_P determining serve direction probabilities and the POPs $(\theta_{in}, \theta_{win})$. Thus, in the structural model, there is a tension between maximizing the likelihood for the POPs versus the likelihood for the serve directions. As we will see, maximum likelihood resolves this tension by distorting the estimates of the POPs while also driving the estimate of λ close to zero. As we noted in Equation (10) of Section 3.4, the only way the model can explain mixed strategy play as $\lambda \downarrow 0$ is to force the POPs to obey the equal win probability restrictions. Maximum likelihood results in distorted POPs that satisfy equal win probability restrictions because it enables the model to match observed serve direction probabilities.

Table 8 summarizes the structural estimation results for the same 10 elite server-receiver pairs that we analyzed in Section 4.[37] For comparison, we show the optimized log-likelihood function

[36] We extend our analysis to grass and clay courts and a much larger set of server-receiver pairs in Section 5.3.

[37] Due to limited space, we do not provide the 32 parameter estimates of $(\theta_{in}, \theta_{win})$ and their standard errors for all 10 servers and all three structural models. We are happy to provide these results to interested readers on request.

Table 8: Summary of structural estimation results for selected elite server-receiver pairs

Player pair	**Reduced-Form**	**Serve-Myopic**	**Point-Myopic**	**Fully-Dynamic**
Server$\rightarrow$ receiver	LL, N AIC	LL, $\hat{\lambda}$ AIC, LR P-value	LL, $\hat{\lambda}$ AIC, LR P-value	LL, $\hat{\lambda}$ AIC, LR P-value
Roger Federer $\rightarrow$ Rafael Nadal	-3779.1, 2011 7646.1	-3788.2, 6.1×10^{-4} 7642.7, .074	-3783.8, 5.5×10^{-3} **7633.7, .571**	-3817.3, 2.9×10^{-5} 7700.7, 7.1×10^{-12}
Rafael Nadal $\rightarrow$ Roger Federer	-3569.1, 1882 7226.2	-3571.3, 6.1×10^{-3} 7208.6, .957	-3570.6, 2.7×10^{-3} **7207.3, .990**	-3632.1, 8.4×10^{-5} 7330.3, 1.1×10^{-21}
Roger Federer $\rightarrow$ Novak Djokovic	-4545.8, 2333 9179.5	-4551.2, .010 **9168.4, .457**	-4552.2, 4.4×10^{-3} 9170.4, .300	-4576.0, 9.0×10^{-4} 9128.0, 7.5×10^{-9}
Novak Djokovic $\rightarrow$ Roger Federer	-4827.7, 2372 **9743.5**	-4840.0, .011 9746.0, .010	-4842.0, 1.8×10^{-3} 9750.0, 2.7×10^{-3}	-4844.8, 2.4×10^{-4} 9755.6, 3.5×10^{-4}
Rafael Nadal $\rightarrow$ Novak Djokovic	-2846.8, 1405 5781.7	-2853.8, 5.8×10^{-6} 5773.7, .232	-2853.2, 6.5×10^{-6} **5772.4, .310**	-2864.5, 6.4×10^{-7} 5795.0, 2.2×10^{-4}
Novak Djokovic $\rightarrow$ Rafael Nadal	-2649.5, 1344 5387.0	-2659.9, .070 5385.9, .035	-2656.1,.097 5378.2, .285	-2654.7, 5.8×10^{-6} **5375.3, .505**
Novak Djokovic $\rightarrow$ Andy Murray	-2384.7, 1201 4857.5	-2396.2, 9.6×10^{-3} **4858.3, .018**	-2396.9, .044 4859.8, .011	-2413.0, 2.2×10^{-4} 4892.0, 4.1×10^{-8}
Andy Murray $\rightarrow$ Novak Djokovic	-2530.1, 1328 5148.1	-2536.4, .014 **5138.9, .310**	-2539.8, 6.8×10^{-3} 5145.7, .052	-2556.3, $1.1\times1-^{-5}$ 5178.6, 2.2×10^{-7}
Pete Sampras $\rightarrow$ Andre Agassi	-2203.3, 1181 **4494.6**	-2219.6, .031 4505.3, 5.9×10^{-4}	-2217.7, .037 4501.4 2.5×10^{-3}	-2240.2, 3.1×10^{-4} 4546.3, 2.4×10^{-11}
Andre Agassi $\rightarrow$ Pete Sampras	-1962.9, 1050 4013.8	-1973.0, 6.7×10^{-4} 4011.9, .043	-1970.9, 3.2×10^{-6} **4007.8, .140**	-2005.8, 5.1×10^{-6} 4077.6, 1.1×10^{-13}

for the reduced-form model and the number of serve observations used to estimate the parameters, along with the point estimates of λ for each of the structural models. The second row of numbers for each server-receiver pair reports the AIC value along with the P-value of a "likelihood-ratio test" of each structural model relative to the reduced-form model. As per our discussion above, these models are not strictly nested within each other, though the reduced-form model is the more flexible specification with a total of 44 parameters.

In light of this, we follow our approach in Section 4 and select our preferred model as the one with the smallest value of the AIC, which we label in bold font. Notice that the best-fitting model per the AIC is also the model with the highest P-value for a quasi-likelihood-ratio test of each the structural models relative to the reduced-form model. Thus, the model with the lowest AIC is generally also the model for which there is the least evidence (from the quasi-likelihood-ratio test) against it relative to the reduced-form model. In two cases, Djokovic serving to Federer and Sampras serving to Agassi, the AIC selects the reduced-form model, and the likelihood-ratio test strongly rejects all three structural models.

For the other eight servers, the AIC selects the fully-dynamic model in only one case, Djokovic serving to Nadal. It selects the point-myopic model for four other servers and the serve-myopic model for the remaining three. We would expect the serve-myopic model to be resoundingly rejected because it does not allow the server to look even just one serve ahead to take advantage of the option value of a second serve when hitting a first serve. However, the serve-myopic model does implicitly reflect adjustments in serve strategy via the POPs that may reflect a server's ability to look ahead. For example, the estimated POPs for the second serve in the serve-myopic specification show a lower proability of faulting (presumably because the server reduces the speed of the second serve), but a lower probability of winning the rally given that the second serve is in (presumably due to the receiver's improved ability to return a slower serve). Therefore, the serve-myopic model is able to reflect state-dependence in tennis serves via its effect on the POPs, which is why it is not so surprising that this model performs as well as it does.

Notice that the estimated scale parameters $\hat{\lambda}$ for all specifications are uniformly small, so we find a limited role for "trembles" to explain the observed mixed serve strategies of these players. Instead, the maximum-likelihood estimates of the POPs $(\hat{\theta}_{in}, \hat{\theta}_{win})$ are distorted in a manner that results in conditional win probabilities much closer to equal than the ones implied by the reduced-form estimates of the POPs. Note that the λ estimates decline for the structural models that require increasingly "farsighted" calculations by the server. When λ is sufficiently small, the conditional value functions $V_{\lambda}(x,m,d)$ are extremely close to the conditional win probabilities, per the limiting result in Equation (10). But when λ is larger, the trembles play a more important role in the mixed serve strategies, allowing more freedom for the conditional value functions (and the conditional win probabilities) to differ across serve directions.

Table 9 provides the estimated service game win probabilities and P-values of Hausman-Wu-Durbin tests of the different model specifications. Recall that this test is based on comparing the implied win probabilities calculated via Equation (6) to the non-parametric estimate of those probabilities; the latter is simply the fraction of service games between a given server-receiver pair that the server won. The first column of Table 9 presents the non-parametric estimates of the

Table 9: Service game win probabilities and Hausman tests

Player pair	**Nonparametric win probability**	**Reduced-Form**	**Serve-Myopic**	**Point-Myopic**	**Fully-Dynamic**
Server$\rightarrow$ receiver	$\hat{W}(1,1)$	$W(1,1)$ P-value	$W(1,1)$ P-value	$W(1,1)$ P-value	$W(1,1)$ P-value
Roger Federer $\rightarrow$ Rafael Nadal	.796 (.026)	.829 (.023) .004	.825 (.021) .050	.830 (.021) .025	.749 (.021) 1.3×10^{-3}
Rafael Nadal $\rightarrow$ Roger Federer	.786 (.026)	.807 (.023) .107	.806 (.022) .147	.807 (.022) .132	.641 (.022) 1.0×10^{-22}
Roger Federer $\rightarrow$ Novak Djokovic	.810 (.024)	.818 (.020) .504	.818 (.020) .506	.818 (.020) .509	.759 (.020) 1.0×10^{-4}
Novak Djokovic $\rightarrow$ Roger Federer	.781 (.025)	.781 (.023) .910	.779 (.022) .838	.778 (.022) .756	.746 (.021) .011
Rafael Nadal $\rightarrow$ Novak Djokovic	.712 (.035)	.712 (.034) .992	.703 (.033) .485	.706 (.033) .675	.650 (.032) 4.1×10^{-5}
Novak Djokovic $\rightarrow$ Rafael Nadal	.829 (.029)	.848 (.023) .278	.846 (.023) .318	.850 (.023) .231	.797 (.024) .052
Novak Djokovic $\rightarrow$ Andy Murray	.794 (.034)	.792 (.029) .871	.791 (.030) .840	.791 (.030) .844	.750 (.029) .012
Andy Murray $\rightarrow$ Novak Djokovic	.721 (.038)	.717 (.036) .675	.717 (.034) .792	.718 (.035) .825	.584 (.032) 1.6×10^{-11}
Pete Sampras $\rightarrow$ Andre Agassi	.885 (.028)	.866 (.024) .150	.863 (.024) .130	.866 (.024) .187	.757 (.028) 0
Andre Agassi $\rightarrow$ Pete Sampras	.874 (.029)	.859 (.024) .362	.854 (.026) .183	.853 (.026) .150	.715 (.028) 1.4×10^{-58}

win probabilities and their standard errors, and the remaining columns present the estimated win probabilities implied by Equation (6) with standard errors calculated via the delta method.[38]

We see that the specification tests strongly reject the fully-dynamic model, with the exception of Djokovic serving to Nadal. Recall from Table 8 that the AIC criterion selects the fully-dynamic model as the preferred specification for Djokovic serving to Nadal, so it is reassuring to know that it is not rejected by the specification tests. But for the other servers, we note that the fully-dynamic model typically significantly underestimates the true service game win probability. This is caused by the need to distort the POPs to rationalize serve behavior as a best response to the estimated

[38] Note that the model estimates are relatively efficient estimates of the win probabilities (as reflected by their smaller standard errors), but they are consistent only if the model specification is correct. The less efficient non-parametric estimator of the win probabilities is consistent regardless.

POPs in the fully-dynamic model. As we will show in the next subsection, the serve strategy for the fully-dynamic model is close to the "true" serve strategy captured by the reduced-form model, but the estimated POPs from the fully-dynamic model imply far less favorable performance for the server than the POPs estimated from the reduced-form model. Indeed, the fully-dynamic POPs generally imply both a higher probability of faults and a lower probability of winning the rally given a serve is in compared to the reduced-form POPs. In contrast, the specification tests are generally unable to reject the point-myopic and serve-myopic models. This is consistent with the results we reported in Table 8, where we showed that these models were the ones most frequently selected as having the lowest AIC values.

Note that when λ is sufficiently small, the structural models predict that the effect of trembles are negligible, and servers will choose to serve to the direction with the highest win probability. In this situation, in order to fit the observed mixed serve strategies, the model is forced to equate conditional win probabilities. We see this most clearly in the inability of the Omnibus Wald test to reject the hypothesis of equal conditional win probabilities for the fully-dynamic model (not reported due to space considerations). For the point-myopic and serve-myopic models, we showed that the estimated λ values are larger, so trembles play a greater role in explaining serve strategies. This allows more freedom for these models to rationalize the observed mixed strategies without having to equate conditional win probabilities, which is reflected in turn by more rejections of equal win probabilities for these models, especially the serve-myopic model. The reduced-form model places no constraint on the estimation of the POPs since it estimates separate parameters and likelihoods for the CCPs and POPs. This flexibility results in nearly unbiased estimates of the POPs and their implied service game win probabilities.

We also observe significant *dynamic attenuation* in the restricted structural estimates of the POPs. That is, as we noted in the previous section, the reduced-form estimation results reveal much stronger evidence of serial correlation in serve directions compared to the POPs. In the fully-dynamic model, the degree of serial correlation in both serve directions and the POPs is attenuated (i.e. closer to zero), and it is thus less likely to be statistically significant. In fact, for most servers, the fully-dynamic model does not exhibit any statistically detectable serial correlation in the structural estimates of the POPs, though it does predict serial correlation in serve directions. What explains this paradox? The explanation is that when λ is close to zero, *serve strategies are* *very sensitive to small changes in the POPs*, since trembles play a negligible role, and the server chooses to serve to the direction with the highest win probability. Thus, it is possible to produce significant muscle memory effects in serve strategies (i.e. the current serve direction depending on the direction of the previous serve to the same court) via *very tiny oscillations in the POPs that are hard to detect statistically*.

Now we return to the key question of this paper: do these distorted/attenuated estimates of the POPs enable the structural models to rationalize observed serve behavior as mixed strategies consistent with Nash equilibrium? We have shown that at best, the structural models are able to rationalize observed serve behavior as a best response, but only relative to the server's *subjective perceptions* of their environment and the receiver, as captured by the structural estimates of the POPs. These subjective beliefs are distorted estimates of the true POPs, which are consistently estimated by the unrestricted reduced-form model. A Nash equilibrium entails a key assumption of *rationality*, i.e. the players' subjective beliefs about each other coincide with the truth. In the next section, we will use dynamic programming to directly calculate best response strategies for our estimates of the true POPs and compare how well these strategies perform relative to the mixed serve strategies the players actually use.

5.1 Calculating Best-Response Serve Strategies

We now provide a more powerful direct test of Nash equilibrium play in tennis: we construct alternative *deviation* serve strategies that significantly increase a server's chance of winning the service game compared to the mixed strategy they are actually using. If the hypothesis of Nash equilibrium is correct, it should be impossible to construct any such deviation strategies. *We construct optimal deviation strategies via DP, setting* $\lambda = 0$ *and using the reduced-form estimates of the POPs.* The DP solution results in pure serve strategies that exploit the unequal win probabilities reflected in the reduced-form estimates of the POPs. At each stage of the game, the DP serve strategy chooses the serve direction that has the maximum conditional win probability (see Equation (4) of Section 3), where the optimal conditional win probability $W(x,m,d)$ is calculated via the Bellman equations given in Equations (1), (2), and (3) of Section 3.

Table 10 presents the optimal DP service game win probability, as well as game win probabilities implied by three other potentially suboptimal serve strategies. For convenience, we repeat the first three columns of Table 9, which show the 10 player pairs, the non-parametric win probabilities, and the reduced-form estimates of the win probabilities. The latter are calculated from the estimated POPs and mixed serve strategies of the reduced-form model using Equation 5 in Section 3.3. As we noted, the reduced-form estimates generally closely match the non-parametric estimates, and thus they constitute our best estimates of each server's win probability implied by the mixed serve strategy they actually use. The next three columns of Table 10 show counterfactual game win probabilities for the serve-myopic, point-myopic and fully-dynamic serve strategies, also using Equation (5). In all three cases, we calculate game win probabilities with the reduced-form estimates of the POPs, not the distorted structural estimates of the POPs in Table 9. We also fix $\lambda = 0$, so we do not allow for any "trembles" in our calculated serve strategies.

By construction, the fully-dynamic serve strategy maximizes the game win probability, which we see in Table 10. However, if the generalized monotonicity condition (GMC) holds, the optimal point-myopic serve strategy coincides with the fully-dynamic strategy and implies the same game win probability. Therefore, failures in the GMC are revealed by cases where the fully-dynamic game win probability is strictly higher than the win probability implied by the optimal point-myopic serve strategy. We do observe some violations of the GMC in Table 10, but in all cases, the incremental gain from using DP to compute an optimal dynamic serve strategy is small.

The last column of Table 10 presents the P-value of a Wald test for Nash equilibrium. The test is constructed by appealing to the *one shot deviation principle*, which states that there is *no deviation at any stage of a dynamic game that can increase the server's chance of winning, given the strategy of the receiver and the service game continuation values*. We find that there are profitable one-shot deviations at many stages of the service game, and while each such deviation yields a modest improvement in the game win probability, the cumulative effect of all profitable deviations is often a large improvement in the game win probability. Of course, if a server were to switch to the DP best response, the receiver would eventually detect the change and adjust his own strategy, which would change the POPs and thus offset some of the gains we predict.

Recall that σ_S was used in Section 3 to denote the optimal serve strategy, which is an implicit function of the POPs Π, that we now make explicit by writing $\sigma_S(\Pi)$. Let P^* and Π^* denote the true equilibrium mixed serve strategy and POPs, respectively, in a Markov Perfect Equilibrium.

Table 10: Improvements in service game win probabilities

Player pair	**Nonparametric win probability**	**Reduced-Form**	**Serve-Myopic**	**Point-Myopic**	**Fully-Dynamic**	**Wald test P-value**
Roger Federer → Rafael Nadal	.796 (.026)	.829 (.023)	.856	.894	.894	.002
Rafael Nadal → Roger Federer	.785 (.026)	.807 (.023)	.840	.884	.884	.014
Roger Federer → Novak Djokovic	.810 (.024)	.818 (.020)	.823	.870	.877	.024
Novak Djokovic → Roger Federer	.782 (.025)	.781 (.023)	.850	.863	.869	.002
Rafael Nadal → Novak Djokovic	.712 (.035)	.712 (.034)	.855	.916	.916	.0004
Novak Djokovic → Rafael Nadal	.829 (.029)	.848 (.023)	.937	.927	.937	.0001
Novak Djokovic → Andy Murray	.794 (.034)	.792 (.029)	.901	.905	.905	.0002
Andy Murray → Novak Djokovic	.721 (.038)	.717 (.036)	.845	.860	.869	.001
Pete Sampras → Andre Agassi	.885 (.028)	.866 (.024)	.942	.949	.949	.003
Andre Agassi → Pete Sampras	.874 (.029)	.859 (.024)	.912	.935	.936	.040

By assumption, the players have common knowledge of these POPs. While we do not directly observe P^* and Π^*, we can consistently estimate them with sufficient data. In particular, the hypothesis of Nash equilibrium implies that for any alternative serve strategy σ, we have:

$$W(P^*,\Pi^*) \geq W(\sigma,\Pi^*). \tag{23}$$

Let $\sigma_S(\Pi^*)$ be the optimal dynamic serve strategy (generally a pure strategy) calculated by DP for the true Nash equilibrium POPs Π^*. Then by definition of optimality, we have:

$$W(\sigma_S(\Pi^*),\Pi^*) \geq W(P^*,\Pi^*) \geq W(\sigma,\Pi^*) \tag{24}$$

for all stationary Markovian serve strategies σ. Together, Inequalities (23) and (24) imply the key equality:

$$W(P^*,\Pi^*) = W(\sigma_S(\Pi^*),\Pi^*), \tag{25}$$

which serves as the basis for our direct test of a mixed strategy Nash equilibrium in tennis: the optimal DP serve strategy should not result in a higher win probability compared to the mixed serve strategy P^* that the server actually uses.

Using consistent estimators of the game win probabilities on the left and right hand sides of Equation (25), we can construct a test statistic based on the squared standardized difference of these win probabilities, which has a χ^2 distribution with 1 degree of freedom if the null hypothesis is true. The last column of Table 10 presents the P-values for this test, and it shows that we strongly reject the best response property for a mixed strategy equilibrium (see Equation (25)) for all 10 player pairs.

5.2 Evaluating the Robustness of Deviation Gains

Our tests of the hypothesis of Nash equilibrium are of course based on *estimates* of the POPs rather than the *true* POPs. In small samples, estimation error could result in spurious, upward-biased estimates of service game win probabilities from noisy estimates of the POPs (instead of the unobservable true POPs) in best response serve strategies calculated by DP. To address this possibility, we use stochastic simulations to demonstrate the robustness of our conclusions by comparing game win probabilities of the estimated mixed strategies the players use vs. our calculated DP best response strategies over a large number of randomly drawn POPs.

We draw the random POPs from the asymptotic distribution of the maximum likelihood estimator centered on the point estimates of the reduced-form POP parameters $(\hat{\theta}_{in}, \hat{\theta}_{win})$. We then calculate POPs implied by these simulated parameter values to generate a set of POPs that are randomly distributed about the true POPs. For each realization of the POPs, we calculate the game win probability when fixing the mixed serve strategy at its estimated value $\hat{P}$ and fixing our estimated DP best response serve strategy at the value calculated using the reduced-form point estimate of the POPs, $\sigma(\hat{\Pi})$. This results in a distribution of simulated win probabilities for the two fixed serve strategies, allowing us to determine if the DP serve strategy outperforms the estimated mixed serve strategy in a range of environments near the true POPs. This eliminates any advantage the DP strategy obtains from assuming the estimated POPs are the same as the true POPs. Thus, we force the DP strategy to confront POPs it is not "expecting."

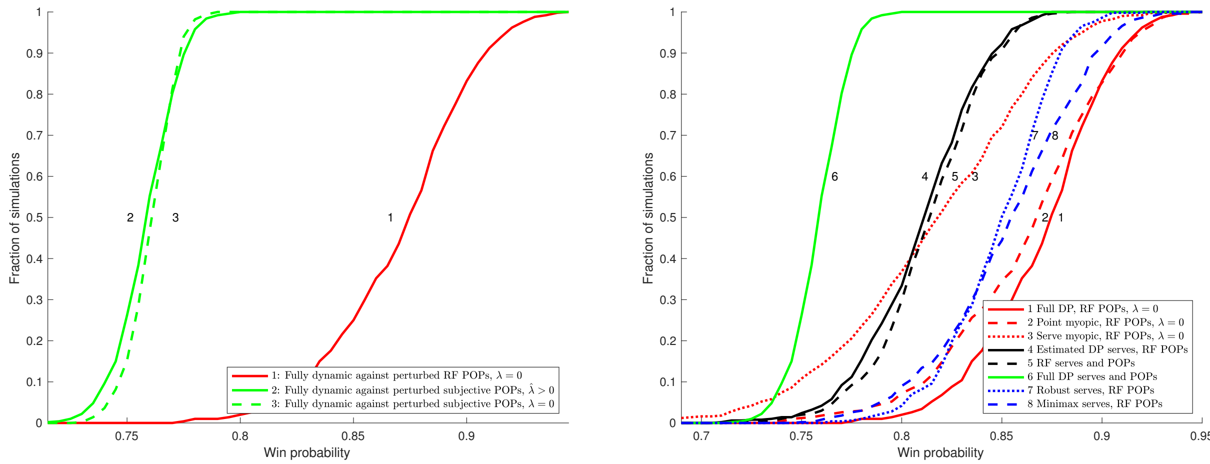
We also calculate similar distributions of win probabilities, but using simulated draws from the *structural estimates of the POPs.* We call these random draws the *perturbed POPs*, and the left panel of Figure 4 shows the CDFs of simulated game win probabilities for three different cases of Federer serving to Djokovic: 1) **Red line:** the fully-dynamic serve strategy $\sigma(\hat{\Pi})$ (calculated with $\lambda = 0$) against 500 random perturbations of the reduced-form estimates of the POPs $\hat{\Pi}$, 2) **Solid green line:** the fully-dynamic strategy (calculated with $\hat{\lambda} > 0$) against 500 perturbations of the structural estimates of the POPs, and 3) **Dashed green line:** the fully-dynamic strategy (calculated with $\lambda = 0$) against 500 perturbations of the structural estimates of the POPs. Note that the strategy used to construct the solid green CDF is a *mixed strategy*, whereas the red and dashed green CDFs are calculated by DP with $\lambda = 0$ and thus are *pure strategies.*

We see that CDFs 2 and 3 (solid and dashed green lines) are nearly identical, which is an illustration of the "no deviation gains" condition in Equation (25) when we assume the server is using an MPE strategy. Even though equal win probabilities do not hold for the 500 perturbations of the structural estimates of the POPs, they are close enough to holding that the win probabilities implied by the pure strategy (i.e. the dashed green CDF) do not systematically outperform those implied by the mixed strategy (i.e. the solid green CDF). In contrast, CDF 1 (red line) is the win probability CDF implied by the DP serve strategy (with $\lambda = 0$) against perturbations of the reduced-form estimates of the POPs, and it clearly stochastically dominates the other two CDFs.

This result illustrates the large increase in game win probabilities resulting from an optimal serve strategy based on an unbiased estimate of the POPs. As we already noted, the structural estimates of the POPs are distorted to rationalize the observed mixed serve strategy as a best response to the POPs. Even though the reduced-form estimates of the POPs may reflect some small sample noise, they indicate sufficiently large departures from the equal win probabilities restriction to result in the large deviation gains illustrated by the red CDF in Figure 4. Essentially, while the structural model can "rationalize" mixed serve strategies, it can only do so via irrational, distorted estimates of the subjective POPs.

The right panel of Figure 4 plots distributions of CDFs for other serve strategies. CDF 4 (solid black line) is the CDF of game win probabilities implied by the estimated mixed strategy from the reduced-form model. Whereas CDF 5 (dashed black line) is the CDF of game win probabilities implied by the estimated strategy from the fully-dynamic structural model. Both

Figure 4: Distributions of win probabilities, Federer serving to Djokovic



CDFs are calculated using random perturbations of the reduced-form estimates of the POPs. We see that CDF 4 lies nearly on top of CDF 5, indicating that the estimated strategy from the fully-dynamic model is virtually the same as the mixed strategy estimated by the reduced-form model. This result illustrates how the fully-dynamic structural model of serve behavior succeeds in "rationalizing" observed mixed serve strategies.

Note that CDF 6 (solid green line) and CDF 1 (solid red line) are the same as those in the left panel of Figure 4 and are included for reference (except that CDF 6 in the right panel is labelled CDF 2 in the left panel, but both are solid green lines). Recall that CDF 6 (solid green line) plots the game win probabilities implied by the estimated fully-dynamic mixed strategy (i.e. with $\hat{\lambda} > 0$) *against perturbations of the subjective POPs*, whereas CDF 1 (solid red line) is the CDF of game win probabilities implied by running the fully-dynamic model (i.e. with $\lambda = 0$) on *pertubations of the unrestricted reduced-form estimate of the POPs.* Thus, the improvement from CDF 6 (solid green line) to CDF 5 (dashed black line) can be thought of as the increase in win probabilities from replacing the distorted subjective POPs with the unrestricted reduced-form, or "rational," POPs (but in both cases fixing the estimated mixed strategy from the fully-dynamic model). Meanwhile, the additional gain in win probabilities by moving from CDF 4 or 5 (black lines) to CDF 1 (solid red line) comes from systematically exploiting unequal win probabilities at all points in the game tree. The fact that the CDF 1 clearly stochastically dominates CDFs 4 and 5 illustrates that Federer's statistically significant 5.9 percentage point increase in game win probabilities in Table 10 not only holds at the point estimate of the POPs but also is robust to significant unexpected deviations of the POPs.

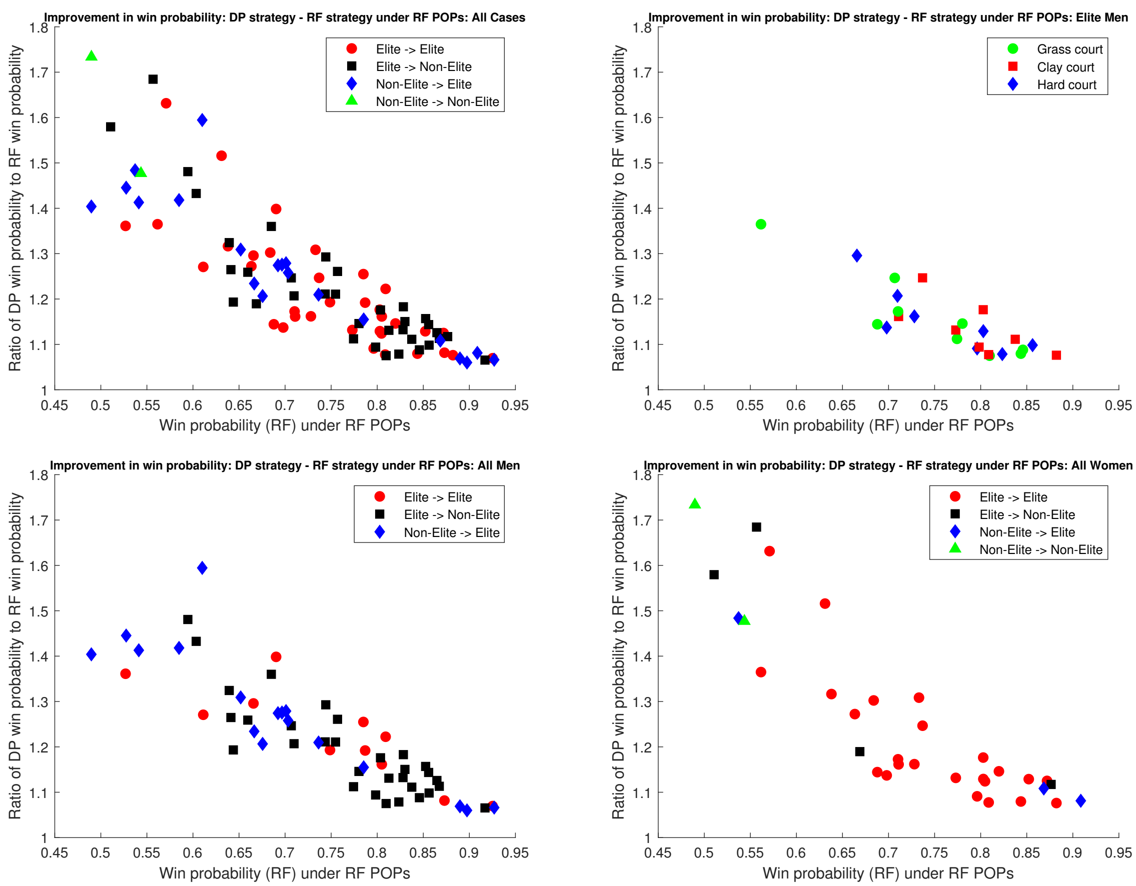
The other two dashed and dotted red CDFs in Figure 4 (also labelled as CDFs 2 and 3) are the distributions of win probabilities implied by the serve-myopic strategy (dotted red CDF 3) and point-myopic strategy (dashed red CDF 2) both computed with $\lambda = 0$ against 500 perturbations of the reduced-form estimates of the POPs. The fact that the three red CDFs (i.e. CDFs 1, 2 and 3) are also ordered by stochastic dominance shows that the point-myopic strategy outperforms the serve-myopic strategy, and the fully-dynamic strategy outperforms both. This figure illustrates a case where the GMC does not hold, though most of the gain comes from using a point-myopic strategy over a serve-myopic one. However, Federer would still benefit from the small but nonetheless significant additional gain from adopting the fully-dynamic strategy.

We also use an informal "robust control" approach to calculate two additional serve strategies illustrated by CDFs 7 and 8 (dotted and dashed blue lines). CDF 7 is computed using a "robust strategy" that constitutes a simple average of the optimal fully-dynamic strategies for each randomly drawn set of POPs. The robust strategy is a mixed strategy, which is a desirable property if receivers have more difficulty finding best responses to mixed than to pure strategies. In addition, we calculate a minimax strategy (CDF 8) by computing the fully-dynamic strategy for the worst-case draw of POPs, i.e. the fully-dynamic best response for the set of POPs that results in the lowest game win probability over all the randomly drawn sets of POPs. Note that to construct these CDFs, we independently draw *another* 500 random perturbations of the POPs and run the robust and minimax strategies on them. Both these CDFs stochastically dominate CDF 4 (solid black line), where the latter is computed using our reduced form estimate of Federer's actual mixed serve strategy. Neither of the robust strategies does as well as the fully-dynamic strategy, however (CDF 1, solid red line).

Overall, it appears that the optimal DP serve strategy, which is a pure strategy, performs surprisingly well in environments it is not "expecting." This could be because it is a pure strategy, and pure strategies may be fairly robust to perturbations in the POPs because they are frequently "corner solutions" that will not change in response to sufficiently small changes in the POPs. In any event, we leave further exploration of this topic, and a deeper assessment of the value of more sophisticated verions of robust control, to future work. Finally, we note that the optimal pure strategies that we calculate by DP are intuitive and relatively simple to describe verbally. For example, in the case of Djokovic serving to Nadal, the fully-dynamic serve strategy generally entails serving to Nadal's right (i.e. backhand since Nadal is a lefty) on first serves, whereas on second serves, the optimal direction depends on the whether Djokovic is serving to the deuce or ad court. To the deuce court, he should serve to Nadal's backhand, whereas to the ad court, he should serve to Nadal's forehand. That is, Djokovic should hit his second serve wide.

5.3 Results for Additional Server-Receiver Pairs and Surfaces

Figure 5: Relation between player ability and deviation gains for different player groups



We summarize our core findings for every server-receiver-surface combination for which we have sufficient data to estimate our model. In total, we analyzed 76 distinct server-receiver pairs and 94 distinct server-receiver-surface combinations. We decisively reject the hypothesis that the estimated mixed serve strategies are consistent with equilibrium play for all 94 combinations by using the Wald test of the absence of deviation gains, i.e. tests that Equation (25) holds.

Figure 5 summarizes the deviation gains from switching to the best-response strategies we calculate by DP. This figure has four panels that plot the gain in win probability for different groups of players. In each panel, the vertical axis is the ratio of the mean win probabilities from the fully-dynamic best-response serve strategy to the win probability implied by our reduced form estimates of each server's actual serve strategy. And the horizontal axis is the mean win probability under the actual serve strategy. Thus, the ratio of mean win probabilities shows the relative improvement in the win probability from adopting the fully-dynamic serve strategy.

We calculate the probabilities in Figure 5 using the same procedure as in Section 5.2; namely, by calculating win probabilities that are robust to estimation error in the POPs. Specifically, we estimate our structural model for each server-receiver pair. We then calculate win probabilities for the observed server strategies and the fully dynamic best response for 500 different POPs that are *IID* draws from the estimated asymptotic distribution about the point estimates of the POPs for each server in each server-receiver pair.

The top-left panel shows a scatter plot of the improvements for all 76 top-ranked server-receiver pairs for which we had sufficient data from the Match Charting Project to reliably estimate the POPs. The points are color-coded so that the red circles are our calculated mean deviation gains for "elite" servers playing "elite" receivers (where we classify a player as "elite" if they were ever ranked first or second worldwide in their career), the black squares are elite servers playing non-elite receivers, the blue diamonds are non-elite servers playing elite receivers, and the green triangles are non-elite servers playing non-elite receivers. The most striking finding in this graph is the obvious downward sloping pattern in the scatter plot: we predict that servers with lower win probabilities experience the biggest relative deviation gains from switching to the DP serve strategy. Of course, since win probabilities cannot be higher than 1, the relative gain is constrained to decline as the win probability under the server's existing serve strategy approaches 1. Nevertheless, the results indicate a clear correlation between "ability" as measured by the server's existing win probability and the extent of their suboptimality: we predict that the less-effective servers have the most to gain from using DP to optimize their serve strategies.

The top-right panel of Figure 5 plots the deviation gains across surfaces (green circles for grass, red squares for clay, and blue diamonds for hard courts) for the 10 "most elite" server-receiver pairs that we have focused our analysis on throughout this paper (i.e. the 10 pairs in Table 10). We see that these players have higher service game win probabilities and also lower deviation gains compared to the set of all players plotted in the left-hand panel. On average for all players, the mean win probability under the existing serve strategy is 73%, and they can expect a 24% increase in win probability on average from adopting the fully-dynamic serve strategy. However, for the most elite players, the average win probability when serving is 76%, and they can expect a 15% increase in win probability on average from deviating to the fully-dynamic strategy. There does not appear to be any clear relation between court type, server ability, and deviation gain for these most elite players.

Finally, the bottom two panels of Figure 5 plot the results for men (left panel) and women (right panel). The same negative correlation between deviation gains and ability as measured by win probability under their existing serve strategy is apparent for both men and women servers. The relationship between average win probability and the calculated gain to switching to the fully-dynamic serve strategy is also robust across the two sexes. In particular, the male servers in our analysis have a higher average probability of winning under their current serve strategy (74% for men vs 70% for women), and a lower average deviation gain from switching to the fully-dynamic serve strategy (22% for men vs 28% for women).

6 Conclusion

There is substantial evidence against Nash equilibrium and minimax play in laboratory experiments: see for example Brown and Rosenthal (1990) and Camerer (2003). However, a standard critique is that laboratory subjects are not sufficiently trained and incentivized to behave sufficiently closely to the predictions of game theory. The influential study by Walker and Wooders (2001) concluded that "the theory has performed far better in explaining the play of top professional tennis players in our data set" (p. 1535). Similar results have been found in other sports such as soccer (see e.g. Chiappori, Levitt, and Groseclose (2002) who studied the direction of penalty kicks). The general conclusion is encapsulated in the title of the study by Palacios-Huerta (2003), "Professionals Play Minimax" (see also Palacios-Huerta 2014).

In contrast, we show that the serve strategies of elite tennis pros are inconsistent with the minimax prediction. Though they use mixed strategies, win probabilities are not the same for all serve directions at all stages of the game — the key restriction of the Nash equilibrium/minimax solution. There has also been considerable work on testing for serial independence in serve directions as an additional implication of mixed strategy equilibrium. We argue that serial dependence, which has been found in many previous studies including Walker and Wooders (2001), is not necessarily inconsistent with equilibrium play when we account for *muscle memory effects* that reflect natural improvements from repeating recently-performed actions. We also show that such muscle memory effects can induce both positive and negative serial correlation in serve directions and and that it is important to account for it to explain observed serve behavior.

Our empirical analysis exploits a new source of data, the Match Charting Project, that allows us to analyze a large number of professional tennis matches at the level of individual server-receiver pairs. We also include body serves — a feature of the MCP data — along with the left and right serves in the previous literature. Tennis players and coaches consider body serves to be an important component of an optimal serve strategy. Our analysis supports this view, since body serves are used frequently in the data and in the calculated optimal serve strategies.

However, the inclusion of body serves and access to more observations are not the main reason why we reject the hypothesis of Nash equilibrium play. Our main innovation is to provide new, more powerful tests of the *dynamic implications* of Nash equilibrium. Specifically, we introduce new tests of the *one shot deviation principle* and an Omnibus Wald test of the equal win probability restriction for all serve directions in all states that must hold in a completely mixed strategy MPE. The latter test strongly rejects the hypothesis of equal win probabilities for the majority of the 10 elite professional server-receiver pairs we analyze, as well as the majority of an additional 66 male and female top ranked professional pairs. We also introduce a new test of the one shot deviation principle, i.e. the restriction that in an MPE there is no deviation strategy that strictly improves the payoff of the players. Using numerical dynamic programming and our econometric estimates of the point outcome probabilities (POPs) that capture the probabilistic outcomes of serves to each possible direction, we reject the hypothesis that the observed mixed strategies of these elite pro servers constitute best responses.

Previous approaches to testing for equal win probabilities over serve directions focused on the *probability of winning individual points*, whereas we recursively calculate how the choice of serve direction affects the *probability of winning the entire service game*. Tests based on the former have low power to detect evidence of disequilibrium play because (as we show in Online Appendix D), deviation gains for individual points are smaller and statistically more difficult to detect. By focusing on the conditional win probabilities for the entire service game, we develop much more powerful tests of the key implications of Nash equilibrium play that exploit the *magnification effect* — that small deviation gains at each individual points cumulate into much more substantial and easier to detect deviation gains in the service game as a whole. Using DP to construct best-response serve strategies, we show that they significantly increase the probability of winning the overall service game. Then using stochastic simulations, we show that our calculated deviation gains are *robust* in the sense that they result in significantly higher win probabilities even when the true POPs differ from the estimated POPs that the strategies are "expecting."

Similar to WW, our conclusion is based on a key *stationarity assumption* that all learning and strategy experimentation has already taken place, and that strategies do not change across games. However, our stationarity assumption is substantially weaker than WW's: like them, we assume stationary play across service games, but unlike them, we relax the assumption of stationary play *over different states within a service game.* We show that serve strategies and win probabilities vary significantly across states within an individual service game in tennis. We also show that the stationarity assumption is testable, and that we cannot reject stationarity of the POPs across service games, though we do reject stationarity of the serve strategies (CCPs) both within and across games. We interpret the latter rejection as further evidence against minimax play, since if the POPs are stationary across games and serve strategies correspond to a unique MPE of the service game, then serve strategies should be stationary across games as well.

A reviewer of this article pointed out that our structural model fails to account for persistent private information of the players, such as information related to their health or stamina during a match. As the reviewer points out, if players have private information about the POPs, this can result "in the econometrician observing POPs that are very different from the POPs observed by the player. Therefore, the econometrician will be using the wrong statistics to test the 'equal win probabilities' hypothesis, rendering the test invalid."

We acknowledge that our structural models do not allow for many potential sources of persistent changes within and across matches, such as changes in health or declining stamina which could affect serve strategies, and that allowing for private information would be theoretically interesting. On the other hand, introducing an additional state variable would reduce the precision of our empirical estimates. Ultimately, the question is whether a private information model of this sort is likely to be of sufficient empirical relevance in our applications to justify the reduced precision. Our opinion is that it is not *given the data we have to analyze*. The main support for this opinion is the results of our stationarity tests. In particular, we conjecture that the POPs would be non-stationary in a model with persistent private information. For example, if the server has private information about some aspect of her current ability, then presumably the receiver will update his beliefs about the server's ability throughout the match and adjust his strategy, which, intuitively, will change the POPs. However, we cannot reject stationarity of the POPs across calendar years or between first sets and later sets in a match. We discuss private information in significantly more detail in Online Appendix E.

Our finding that many elite tennis pros fail to play serve strategies that are best responses to their opponents may also seem surprising given the stakes involved in top level tennis matches, and it is clearly contrary to the consensus in the literature noted above. We believe that we have convincing evidence of suboptimal serve strategies, but the ultimate test would be to run field experiments to verify whether our DP serve strategies really do deliver the increased win probabilities that we predict. Our predicted gains may dissipate rapidly in the field as the receiver recognizes and adapts to a change in the server's strategy. Ultimately, the issues raised by the possibility of learning and adaptation to changes in strategy are fascinating topics for further exploration but are beyond the scope of this analysis.

Though we introduce behavioral models that can explain disequilibrium play as a result of "distorted subjective beliefs," we have not explained why elite players seem to have less than fully rational expectations about their own strengths and weaknesses and those of their opponents. The monetary rewards to increasing the probability of winning by the magnitudes we estimate are very high. The usual presumption in economics and much of the previous literature on tennis is that when there are high rewards, we can expect that some incompletely defined learning process will lead to behavior consistent with Nash equilibrium. At the very least we should not see *large* gains left unexploited. Thus, our findings are only partially consistent with Simon (1956)'s principle of *satisficing*: "However adaptive the behavior of organisms in learning and choice situations, this adaptiveness falls far short of the ideal of ‘maximizing’ postulated in economic theory. Evidently, organisms adapt well enough to ‘satisfice’; they do not, in general, ‘optimize.’ ”

Our conclusion that even elite pro tennis players may have inadequate statistical knowledge or an inaccurate “mental model” of the POPs is corroborated by the nascent industry of *sports analytics*, which provides statistical analysis and advice to improve athletes’ play in tennis and other sports. It is unlikely that the growth in tennis analytics would be as large as it is if most of the elite pros already have “rational beliefs” and are already playing best response strategies on their own. Over time, if more elite pros use increasingly powerful analytics to help improve their play, the long run outcome of this process of learning and experimentation could well be something that more closely approximates Nash equilibrium play.

We are not the first study to have provided evidence that suggests highly compensated and motivated sports professionals may not be behaving optimally. There is the famous book *Moneyball* by M. Lewis (2003) that showed how analytics can improve the performance of entire baseball teams. Also, focusing on individual baseball players, Bhattacharya and Howard (2022) found that while pitchers use mixed strategies over different pitches (fastball, curve, etc.), “payoffs differ significantly across pitch types” (p. 350). In football, Romer (2006) used dynamic programming to demonstrate that teams were making suboptimal decisions regarding when to go for it on fourth down, punt, or kick a field goal. Tennis may be another sport where econometrics, dynamic programming, and analytics can affect thinking, change behavior, and help guide players to play in a way that more closely corresponds to the predictions of Nash equilibrium.

Appendices

A Existence and Uniqueness

A.1 A Value Contraction for Markov Chains with Terminal Rewards

Let Z be a finite set of states. Let $q(z',z)$ be the transition chance from state z to z', and let $q^n(z',z)$ be the chance that the stochastic process is in state z' after exactly n steps when starting in state z. Define $q(Z',z) \equiv \sum_{z'\in Z'} q(z',z)$ and $q^n(Z',z) \equiv \sum_{z'\in Z'} q^n(z',z)$ for all $Z' \subseteq Z$. States $Z_A \subseteq Z$ are terminating states, i.e. $q(z,z) = 1$ for all $z \in Z_A$. Assume terminal reward function $g : Z_A \to [0,1]$, i.e. $g(z)$ is the reward when the process first enters terminal state z. Let $\mathcal{W}$ be the space of functions mapping from Z to $[0,1]$, with $W(z) = g(z)$ for all $z \in Z_A$ and define the value mapping:

$$TW(z) = \sum_{z'} q(z',z)W(z') \quad \forall z \in Z\setminus Z_A \text{ and } W \in \mathcal{W} \tag{26}$$

The iterated map is defined recursively: $T^2W = T(TW)$ and $T^nW = T(T^{n-1}W)$.

Lemma 1. *Assume there exists an integer $k < \infty$ and a constant $q_0 > 0$, s.t. $q^k(Z_A,z) \geq q_0$ for all $z \in Z\setminus Z_A$, then T^k is a contraction of modulus $1-q_0$ in the sup-norm.*

Proof. First, we show via induction on k that:

$$T^kW(z) = \sum_{z'\in Z_A} q^k(z',z)g(z') + \sum_{z'\in Z\setminus Z_A} q^k(z',z)W(z') \quad \forall z \in Z\setminus Z_A \text{ and } W \in \mathcal{W} \tag{27}$$

To verify (27) for $k = 1$, simply rewrite (26), as follows:

$$TW(z) = \sum_{z'\in Z_A} q(z',z)g(z') + \sum_{z'\in Z\setminus Z_A} q(z',z)W(z') \quad \forall z \in Z\setminus Z_A \text{ and } W \in \mathcal{W}$$

For the inductive step assume (27) is satisfied for $k-1$, then for $z \in Z\setminus Z_A$ and $W \in \mathcal{W}$:

$$\begin{aligned}
T^kW(z) &= \sum_{z'\in Z_A} q(z',z)g(z') + \sum_{z'\in Z\setminus Z_A} q(z',z)T^{k-1}W(z') \\
&= \sum_{z'\in Z_A} q(z',z)g(z') + \sum_{z'\in Z\setminus Z_A} q(z',z)\left[\sum_{z''\in Z_A} q^{k-1}(z'',z')g(z'') + \sum_{z''\in Z\setminus Z_A} q^{k-1}(z'',z')W(z'')\right] \\
&= \sum_{z'\in Z_A} q(z',z)g(z') + \sum_{z''\in Z\setminus Z_A} q^k(z'',z)g(z'') + \sum_{z''\in Z\setminus Z_A} q^k(z'',z)W(z'') \\
&= \sum_{z'\in Z_A} q^k(z',z)g(z') + \sum_{z'\in Z\setminus Z_A} q^k(z',z)W(z')
\end{aligned}$$

Having verified (27), the fact that T^k is a contraction of modulus $1-q_0$ whenever $q^k(Z_A,z) > q_0$ for all $z \in Z \setminus Z_A$ follows easily from standard reasoning. □

A.2 Value and MPE Existence: Proof of Theorem 1

Proof. STEP 1: A UNIQUE EQUILIBRIUM VALUE. Let $Z = \{1,2,\ldots,38\} \times \{l,r,b\}^2$ be the set of states, and $\Gamma(z)$ be the subgame starting in state z. For any pure strategies (d,s,a), define the transition function $q(z',z|d,s,a)$. Easily, q inherits continuity in (s,a) from ℓ and ω. For any function $v : \{1,2,\ldots,38\} \times \{l,r,b\}^2 \mapsto [0,1]$ with $v(38,m) = 0$ and $v(37,m) = 1$, define the static zero-sum game with server payoff:

$$u(z,d,s,a|v) \equiv \sum_{z' \in Z} q(z',z|d,s,a)v(z')$$

Let $\mathcal{B}$ be the set of probability distributions over $\{l,r,b\} \times \mathcal{S}$ and let $\mathcal{A}$ be the set of probability distributions over receiver attention vectors. Since q is continuous in (s,a), u is continuous in (s,a) for any fixed v; the Minimax theorem in Ville (1938) applies:

$$\min_{\alpha \in \mathcal{A}} \max_{\beta \in \mathcal{B}} \int u(z,d,s,a|v) d\beta(d,s) d\alpha(a) = \max_{\beta \in \mathcal{B}} \min_{\alpha \in \mathcal{A}} \int u(z,d,s,a|v) d\beta(d,s) d\alpha(a)$$

Altogether, the recursive sub-game $\Gamma(z)$ meets the premise of Theorem 6 in Everett (1957); and thus, there exists a unique value $v^* : \{1,2,\ldots,38\} \times \{l,r,b\}^2 \mapsto [0,1]$ in each $\Gamma(z)$.

STEP 2: PAYOFFS ARE WELL-DEFINED FOR ALL STATIONARY STRATEGIES. Let $Z_A = \{37,38\} \times \{l,r,b\}^2$, $g(37,m) = 1$, and $g(38,m) = 0$ for all muscle memory states m. Stationary pure strategies are functions $B : Z \setminus Z_A \mapsto \{l,r,b\} \times \mathcal{S}$ and $A : Z \setminus Z_A \mapsto \mathcal{A}$. Any pair of pure strategies induces transition chance $q(z',z|B(z),A(z))$. Since $\omega\ell \geq \underline{w} > 0$, there is an integer $k < \infty$ and a uniform lower bound $q_0 > 0$ on the chance that the server wins ($x = 37$) within k serves starting from any non-terminal state. Thus, the transition function $q^n(z',z|B,A)$ obeys the premise of Lemma 1. Consequently, the chance that the server wins the service game starting in any state z given strategies B,A is the unique fixed point of the contraction mapping (27), which we write as:

$$T^k W(z|B,A) = \sum_{z' \in Z_A} q^k(z',z|B,A)g(z') + \sum_{z' \in Z \setminus Z_A} q^k(z',z|B,A)W(z'|B,A) \tag{28}$$

to emphasize the dependence on strategies (B,A). Let $U(z|B,A)$ be this unique fixed point.

STEP 3: MPE EXISTENCE. We now consider variation in strategies; and thus, reinterpret (28) as a mapping on the space of functions that map from $Z \setminus Z_A \times \{l,b,r\} \times S \times A$ to $[0,1]$ and are continuous in (s,a). Since, $q(\cdot|d,s,a)$ is continuous in (s,a) by Step 1 and T^k is a contraction, the fixed point $U(z|B,A)$ is continuous in (B,A). Furthermore, the strategy sets S and A are both compact. Thus, by Glicksberg (1952) there exists a Nash Equilibrium $(\sigma_R^*(z), \sigma_S^*(z))$ of the static game with payoffs $U(z|B,A)$. By construction this is an MPE of the dynamic game. □

A.3 Unique Attention, Speed, and Spin: Proof of Theorem 2

STEP 0: WINNING A SERVE IS STRICTLY PREFERRED TO LOSING. Easily, if we fix the serve direction, the server strictly prefers: winning on the current serve to faulting, faulting a first serve to losing the point on the first serve, and winning a point to losing i.e.:

$$\begin{aligned} W(x^+(x),m) &> W(x+1,m) \quad \text{and} \quad W(x+1,m) > W(x^-(x),m) \quad \forall x \text{ odd} \\ W(x^+(x),m) &> W(x^-(x),m) \quad \forall x \end{aligned}$$

This follows easily from our assumption that $\ell\omega$ is bounded away from zero and 1 for all strategy pairs. In particular, for any arbitrarily (perhaps history contingent) equilibrium strategy for the receiver, the server could always adopt the (perhaps history contingent) strategy following a point win (or faulted first serve) that he would have adopted in the alternate universe that he lost the point. Easily, whatever this strategy is, the chance of eventually winning the game is strictly higher following a point win (or fault) than following a lost point.

STEP 1: THE SET OF MPE IS CONVEX. Since the equilibrium value is unique, we can write the static zero sum payoff for all first serves (x odd) as:

$$\begin{aligned} u(x,m,d,s,a) \equiv\ & W(x+1,m') + \ell(m,d,c(x),s)\omega(m,d,c(x),s,a)\left[W(x^+(x),m') - W(x+1,m')\right] \\ & + \ell(m,d,c(x),s)(\omega(m,d,c(x),s,a)-1)\left[W(x+1,m') - W(x^-(x),m')\right] \end{aligned} \quad (29)$$

where $m' = (d, m_1)$, and for all second serves (x even):

$$u(x,m,d,s,a) \equiv W(x^-(x),m) + \ell(m,d,c(x),s)\omega(m,d,c(x),s,a)\left[W(x^+(x),m) - W(x^-(x),m)\right]$$

Note that W is not a function of a or s, and all terms in square brackets are strictly positive by Step 0; and thus, by Assumption 2, u is strictly convex in a and strictly concave in s for both odd and even x. Furthermore, the server has a finite choice of serve directions d. Altogether, by Proposition 3.1(i) in Hwang and Rey-Bellet (2020) the set of MPE is convex.

STEP 2: UNIQUE a AND s FOR ALL (x,m). We prove that there is a unique NE attention in each state (x,m). The proof of uniqueness of the server's NE (speed, spin) choice in each state follows parallel steps. First, given any arbitrary mixed server strategy $\beta \in \mathcal{B}$ define the payoff:

$$\nu(x,m,\beta,a) \equiv \int u(x,m,d,s,a)d\beta(d,s)$$

Trivially, ν is linear in β, and inherits strict convexity in a from u. Since the receiver's best response attention minimizes ν, which is strictly convex in a, the receiver uses a pure strategy at every (x,m) in any NE.

Now, toward a contradiction assume that $(\hat{\beta},\hat{a})$ and (β',a') are both Nash Equilibria in state (x,m) with $\hat{a} \neq a'$. Define $(\beta'',a'') \equiv \lambda(\hat{\beta},\hat{a}) + (1-\lambda)(\beta',a')$ for some $\lambda \in (0,1)$. Then use $\nu^* = \nu(x,m,\hat{\beta},\hat{a}) = \nu(x,m,\beta',a')$, followed by the fact that $\hat{\beta}(\beta')$ maximizes ν given $\hat{a}$ (a'), to get:

$$\begin{aligned}
\nu^* &= \nu(x,m,\hat{\beta},\hat{a}) && \geq \nu(x,m,\beta'',\hat{a}) \\
\nu^* &= \nu(x,m,\beta',a') && \geq \nu(x,m,\beta'',a')
\end{aligned}$$

Now take a $(\lambda,1-\lambda)$ weighted average of the above two inequalities, and then use ν strictly convex in a and $\hat{a} \neq a'$ to discover:

$$\nu^* \geq \lambda\nu(x,m,\beta'',\hat{a}) + (1-\lambda)\nu(x,m,\beta'',a') > \nu(x,m,\beta'',a'')$$

And thus, (β'',a'') cannot be a NE at (x,m), since all NE must have the same value ν^*. But this contradicts Step 1, i.e. the set of NE convex for all (x,m). □

A.4 Optimality of Point-Myopic Play: Proof of Theorem 3

Call $\Pi(x,m,d)$ the chance the server wins the point when choosing direction d on a first serve in state (x,m), assuming an optimal serve strategy on a second serve (if necessary). Let $D^{PM}(x,m)$ and $D^{FD}(x,m)$ be the set of point myopically optimal and dynamically optimal serve strategies in score state x and muscle memory state m. We must show $D^{PM}(x,m) = D^{FD}(x,m)$ for all (x,m).

STEP 1: $d \in D^{PM}(x,m) \;\Rightarrow\; d \in D^{FD}(x,m)$. By assumption, both (13) and (14) hold for any $d \in D^{PM}(x,m)$. Now for any arbitrary d' use in sequence for each implication below: condition (14) and $\Pi \in [0,1]$, rearrangement of terms, then $\Pi(x,m,d) \geq \Pi(x,m,d')$ along with inequality (13), and finally the definition of W:

$$\Pi(x,m,d')\,[W(x^+,(d,d_1)) - W(x^+,(d',d_1))] + (1-\Pi(x,m,d'))\,[W(x^-,(d,d_1)) - W(x^-,(d',d_1))] \geq 0$$

$$\Rightarrow \Pi(x,m,d')W(x^+,(d,d_1)) + (1-\Pi(x,m,d'))W(x^-,(d,d_1))$$
$$\geq \Pi(x,m,d')W(x^+,(d',d_1)) + (1-\Pi(x,m,d'))W(x^-,(d',d_1))$$

$$\Rightarrow \Pi(x,m,d)W(x^+,(d,d_1)) + (1-\Pi(x,m,d))W(x^-,(d,d_1))$$
$$\geq \Pi(x,m,d')W(x^+,(d',d_1)) + (1-\Pi(x,m,d'))W(x^-,(d',d_1))$$

$$\Leftrightarrow W(x,m,d) \geq W(x,m,d')$$

And since the sequence above is true for all d' we may conclude that $d \in D^{FD}(x,m)$.

STEP 2: $d' \notin D^{PM}(x,m) \;\Rightarrow\; d' \notin D^{FD}(x,m)$. If $d' \notin D^{PM}(x,m)$, then there exists d such that $\Pi(x,m,d) > \Pi(x,m,d')$. Now we simply mimic the same steps as above to get:

$$\Pi(x,m,d')\,[W(x^+,(d,d_1)) - W(x^+,(d',d_1))] + (1-\Pi(x,m,d'))\,[W(x^-,(d,d_1)) - W(x^-,(d',d_1))] \geq 0$$

$$\Rightarrow \Pi(x,m,d')W(x^+,(d,d_1)) + (1-\Pi(x,m,d'))W(x^-,(d,d_1))$$
$$\geq \Pi(x,m,d')W(x^+,(d',d_1)) + (1-\Pi(x,m,d'))W(x^-,(d',d_1))$$

$$\Rightarrow \Pi(x,m,d)W(x^+,(d,d_1)) + (1-\Pi(x,m,d))W(x^-,(d,d_1))$$
$$> \Pi(x,m,d')W(x^+,(d',d_1)) + (1-\Pi(x,m,d'))W(x^-,(d',d_1))$$

$$\Leftrightarrow W(x,m,d) > W(x,m,d') \quad \Rightarrow \quad d' \notin D^{FD}(x,m)$$

where the *strict* inequality in the penultimate line follows from $\Pi(x,m,d) > \Pi(x,m,d')$ and (13).

□

B Preferred Specification Estimates: Federer and Djokovic

Tables 11 and 12 present the maximum-likelihood estimates of the parameters $\hat{\theta} = (\hat{\theta}_P, \hat{\theta}_{in}, \hat{\theta}_{win})$ of our preferred specification of the reduced-form model of serves and POPs, where we restrict the sample to hard courts only. Our preferred specification has the smallest AIC over the four alternative specifications we consider, including a restricted specification that rules out muscle memory effects by excluding the variable m from the set of indicator functions

$\{f(x,m,d), g_{in}(x,m,d), g_{win}(x,m,d)\}$ defining each specification. Our preferred specification has 12 parameters for $P(d|x,m,\theta_P)$ and 32 parameters for $(\pi(in|x,m,d), \pi(win|x,m,d))$ (16 parameters each). We do not have the space to present the estimates for all the players we analyze in this paper. However, we do show the parameters for two specific server-receiver pairs, Djokovic serving to Federer and Federer serving to Djokovic.

In Table 11, we present the parameter estimates $\hat{\theta}_P$ determining the reduced-form serve probabilities $P(d|x,m,\hat{\theta}_P)$. Recall that we code muscle memory as $m = (d_{-2}, d_{-1})$, where d_{-2} is the direction of the serve two first serves ago (i.e. the direction of the previous first serve to the same court) and d_{-1} is the direction of the previous first serve. Parameters 3 and 9 are significantly positive for both Djokovic and Federer, which indicates that their first serves exhibit *positive* serial correlation: there is an increased likelihood of serving to direction $d = d_{-2}$, where the latter is the direction of the most recent first serve to the same court. However, the coefficients on Parameters 6 and 12 are not significantly different from zero, which is consistent with a lack of significant serial correlation between the directions of the faulted first and subsequent second serves.

Table 11: ML parameter estimates, reduced-form model of serve directions, $P(d|x,m,\theta_P)$

Parameter		**Djokovic** $\to$ Federer		**Federer** $\to$ Djokovic	
Name		Estimate	Standard Error	Estimate	Standard Error
1	1st serve, ad court, $d = l$	.861	(.095)	1.367	(.133)
2	1st serve, ad court, $d = r$	.601	(.100)	1.319	(.132)
3	1st serve, ad court, $d = d_{-2}$	.294	(.079)	.396	(.086)
4	2nd serve, ad court, $d = l$	.462	(.146)	.551	(.152)
5	2nd serve, ad court, $d = r$	−.386	(.183)	−.215	(.171)
6	2nd serve, ad court, $d = d_{-1}$	−.051	(.136)	−.143	(.142)
7	1st serve, deuce court, $d = l$	.836	(.102)	1.234	(.145)
8	1st serve, deuce court, $d = r$	.795	(.106)	1.402	(.143)
9	1st serve, deuce court, $d = d_{-2}$	.566	(.077)	.483	(.103)
10	2nd serve, deuce court, $d = l$	−.208	(.145)	−.293	(.132)
11	2nd serve, deuce court, $d = r$	−.650	(.151)	−.531	(.159)
12	2nd serve, deuce court, $d = d_{-1}$	−.002	(.168)	.137	(.144)
Observations, log-likelihood		2372, -2324.8		2333, -2265.06	
AIC, BIC		4871.6, 4940.8		4554.1, 4623.2	

The other parameter estimates in Table 11 determine the directions of the first and second serves to the ad and deuce courts separately. Our specification normalizes $f(x,m,d) = 0$ when $d = b$ (i.e. body serves), so the coefficient estimates for the other serve directions $d \in \{l, r\}$ are enough to determine all three probabilities $P(d|x,m)$ for any given (x,m) value. The implicit restriction in our specification is that serve directions only depend on whether the current serve is a first serve or second serve, and whether it is to the ad or deuce court. The large positive values of Parameters 1, 2, 7, and 8 indicate that on first serves, Djokovic and Federer typically serve more often to the left or right than to the body (approximately 18% to the body for Djokovic, which is still significantly greater than 11% for Federer). However, the negative estimates of Parameters 10 and 11 indicate that on second serves to the deuce court, both players increase their probability of hitting a body serve significantly (to about 43% for both servers).

Table 12 shows the parameter estimates of our preferred specification for the POPs $(\hat{\theta}_{in}, \hat{\theta}_{win})$. None of the parameters for indicators that depend on the muscle memory state variable m are significant for $\pi(in|x,m,\hat{\theta}_{in})$ (see Parameters 4, 8, 12 and 16), which indicates that serving to the same direction as the previous serve does not reduce their probability of faulting. However, Parameter 4 of $\hat{\theta}_{win}$ is positive and significant for Djokovic serving to Federer, so here, serving to the same direction as the previous first serve to the ad court increases Djokovic's probability of winning given that the serve is in.

Comparing the parameters (1,2,3,9,10,11) of the vector $\hat{\theta}_{win}$, we see that they are uniformly (i.e. across all three directions) significantly larger for Federer than for Djokovic. This implies that conditional on a first serve going in, Federer has a higher chance of winning the subsequent rally when serving to Djokovic than Djokovic does when serving to Federer. However, when we do a similar comparison of the corresponding coefficient estimates of $\hat{\theta}_{in}$, the inequality is almost uniformly reversed, indicating that Djokovic has a higher probability of making his first serve than Federer.[39] Thus, our estimates reflect an intuitive trade-off: a faster serve or one aimed closer to the lines has a higher chance of missing, but conditional on it going in, the receiver has a lower chance of returning it successfully or winning the subsequent rally.

[39] Though for second serves to the deuce court, Federer has a uniformly lower chance of double faulting.

Table 12: ML parameter estimates of POPs $(\pi(in|x,m,d,\theta_{in}),\pi(win|x,m,d,\theta_{win}))$

Parameter		**Djokovic** → Federer		**Federer** → Djokovic	
θ_{in}		Estimate	Standard Error	Estimate	Standard Error
1	1st serve, ad court, $d=l$	.465	(.147)	.486	(.140)
2	1st serve, ad court, $d=b$	.945	(.213)	.864	(.230)
3	1st serve, ad court, $d=r$	.744	(.134)	.614	(.115)
4	1st serve, ad court, $d=d_{-2}$	−.093	(.156)	−.113	(.145)
5	2nd serve, ad court, $d=l$	3.468	(.572)	2.277	(.403)
6	2nd serve, ad court, $d=b$	2.137	(.288)	3.249	(.459)
7	2nd serve, ad court, $d=r$	2.150	(.445)	1.898	(.360)
8	2nd serve, ad court, $d=d_{-2}$	.277	(.513)	.118	(.440)
9	1st serve, deuce court, $d=l$	.604	(.159)	.122	(.148)
10	1st serve, deuce court, $d=b$	.928	(.173)	.530	(.232)
11	1st serve, deuce court, $d=r$	.813	(.142)	.422	(.128)
12	1st serve, deuce court, $d=d_{-2}$	−.138	(.165)	.164	(.154)
13	2nd serve, deuce court, $d=l$	2.214	(.406)	2.652	(.362)
14	2nd serve, deuce court, $d=b$	1.917	(.352)	3.820	(.719)
15	2nd serve, deuce court, $d=r$	1.430	(.374)	2.033	(.387)
16	2nd serve, deuce court, $d=d_{-2}$	.294	(.452)	.329	(.499)
θ_{win}		Estimate	Standard Error	Estimate	Standard Error
1	1st serve, ad court, $d=l$	.641	(.182)	1.092	(.208)
2	1st serve, ad court, $d=b$	.470	(.196)	.760	(.250)
3	1st serve, ad court, $d=r$	.439	(.143)	.795	(.154)
4	1st serve, ad court, $d=d_{-2}$	.456	(.188)	.314	(.210)
5	2nd serve, ad court, $d=l$	.975	(.235)	.148	(.247)
6	2nd serve, ad court, $d=b$	.593	(.195)	.073	(.182)
7	2nd serve, ad court, $d=r$	.650	(.290)	.063	(.285)
8	2nd serve, ad court, $d=d_{-2}$	−.537	(.273)	.038	(.270)
9	1st serve, deuce court, $d=l$	.878	(.185)	1.397	(.232)
10	1st serve, deuce court, $d=b$	.614	(.223)	1.055	(.305)
11	1st serve, deuce court, $d=r$	.728	(.171)	.933	(.167)
12	1st serve, deuce court, $d=d_{-2}$	−.182	(.194)	−.392	(.216)
13	2nd serve, deuce court, $d=l$	−.090	(.221)	.292	(.221)
14	2nd serve, deuce court, $d=b$	−.282	(.222)	−.022	(.222)
15	2nd serve, deuce court, $d=r$	.530	(.295)	.142	(.263)
16	2nd serve, deuce court, $d=d_{-2}$	.489	(.281)	.206	(.259)
Observations, log-likelihood		2333, -2403.9		2372, -2324.8	
AIC, BIC		4871.9, 5056.6		4625.4, 4809.6	

Table 13: Within-Match Tests for stationarity of POPs: $\{\pi(in|x,m,d,\theta_{in}),\pi(win|x,m,d,\theta_{win})\}$

Server → receiver	Muscle Memory			No Muscle Memory		
	Restricted	Unrestricted	LR test (df)	Restricted	Unrestricted	LR test (df)
	LL, AIC	LL, AIC	P-value	LL, AIC	LL, AIC	P-value
Roger Federer →	-1934.3	-1919.8	29.0 (32)	-1940.1	-1928.0	24.3 (24)
Rafael Nadal	**3932.6**	3967.6	.620	**3928.2**	3952.0	.445
Rafael Nadal →	-1880.9	-1860.3	41.3 (32)	-1883.2	-1867.5	31.4 (24)
Roger Federer	**3825.9**	3848.6	.125	**3814.5**	3831.1	.142
Roger Federer →	-2280.7	-2265.7	30.0 (32)	-2284.7	-2272.6	24.2 (24)
Novak Djokovic	**4625.4**	4659.4	.570	**4617.5**	4641.3	.448
Novak Djokovic →	-2403.9	-2389.0	29.9 (32)	-2411.7	-2399.0	25.2 (24)
Roger Federer	**4871.9**	4906.0	.572	**4871.3**	4894.1	.393
Rafael Nadal →	-1414.2	-1397.0	34.4 (32)	-1415.8	-1404.7	22.2 (24)
Novak Djokovic	**2892.4**	2922.0	.355	**2879.6**	2905.4	.565
Novak Djokovic →	-1302.1	-1289.7	24.7 (32)	-1304.5	-1294.3	20.3 (24)
Rafael Nadal	**2668.1**	2707.5	.819	**2656.9**	2684.6	.681
Novak Djokovic →	-1183.2	-1166.0	34.4 (32)	-1188.7	-1179.3	19.0 (24)
Andy Murray	**2430.3**	2459.9	.355	**2425.5**	2454.5	.753
Andy Murray →	-1280.1	-1263.7	32.7 (32)	-1287.9	-1277.1	21.6 (24)
Novak Djokovic	**2624.1**	2655.4	.431	**2623.9**	2650.2	.601
Pete Sampras →	-1117.9	-1103.5	28.6 (32)	-1124.1	-1111.9	24.4 (24)
Andre Agassi	**2299.7**	2335.1	.638	**2296.2**	2319.7	.437
Andre Agassi →	-1031.1	-1005.3	51.6 (32)	-1032.6	-1017.4	30.3 (24)
Pete Sampras	**2126.2**	2183.6	.015*	**2113.2**	2130.9	.174

C Alternative Subsample Stationarity Tests

We revisit the stationarity tests in Section 4.3, except that here we divide all serves on hard courts for each of our ten elite server-receiver pairs into serves from the first set in the match (no matter when that match was played) and serves from the later sets of a match. We then estimate POPs and CCPs separately for these two subsamples and perform stationarity tests as in Section 4.3.

Table 13 presents results for the POPs, which are strikingly similar to those for the calendar year groupings in Section 4.3. We get one rejection at the 5% level for the muscle memory specification, and no rejections (even at the 10%) level for the no muscle memory specification. As in Section 4.3, stationarity is rejected more frequently for the CCPs (Table 14).

Table 14: Within-match tests for stationarity of CCPs: $\{P(d|x,m)\}$

Server → receiver	Muscle Memory			No Muscle Memory		
	Restricted	Unrestricted	LR test (df)	Restricted	Unrestricted	LR test (df)
	LL, AIC	LL, AIC	P-value	LL, AIC	LL, AIC	P-value
Roger Federer →	-1844.8	-1837.9	13.7 (12)	-1874.4	-1870.6	7.5 (8)
Rafael Nadal	**3713.5**	3723.8	.321	**3764.7**	3773.2	.479
Rafael Nadal →	-1688.2	-1680.3	15.7 (12)	-1690.9	-1687.9	5.9 (8)
Roger Federer	**3400.3**	3408.6	.205	**3397.8**	3407.9	.659
Roger Federer →	-2265.1	-2256.5	17.1 (12)	-2293.9	-2290.4	6.9 (8)
Novak Djokovic	**4554.1**	4561.0	.145	**4603.7**	4612.8	.545
Novak Djokovic →	-2423.8	-2411.7	24.2 (12)	-2454.8	-2441.2	27.2 (8)
Roger Federer	4871.6	**4871.3**	.019*	4925.5	**4914.3**	.001*
Rafael Nadal →	-1432.6	-1420.2	24.8 (12)	-1437.5	-1431.9	11.1 (8)
Novak Djokovic	2889.3	**2888.5**	.016*	**2891.0**	2895.9	.198
Novak Djokovic →	-1347.4	-1340.5	13.8 (12)	-1364.2	-1360.7	7.0 (8)
Rafael Nadal	**2718.9**	2729.1	.315	**2744.4**	2753.4	.539
Novak Djokovic →	-1201.6	-1191.7	19.8 (12)	-1221.0	-1217.4	7.2 (8)
Andy Murray	**2427.1**	2431.3	.071	**2458.0**	2466.8	.514
Andy Murray →	-1250.0	-1240.7	18.5 (12)	-1254.9	-1248.2	13.4 (8)
Novak Djokovic	**2524.0**	2529.5	.100	**2525.9**	2528.5	.098
Pete Sampras →	-1085.4	-1077.0	16.9 (12)	-1096.4	-1092.5	7.7 (8)
Andre Agassi	**2194.9**	2202.0	.154	**2208.8**	2217.1	.466
Andre Agassi →	-931.8	-923.6	16.4 (12)	-945.2	-939.5	11.5 (8)
Pete Sampras	**1887.6**	1895.2	.176	**1906.5**	1911.0	.176

Online Appendices

D Replicating Walker and Wooders' (2001) Analysis

We now replicate the static approach to testing for equal win probabilities used by Walker and Wooders (2001) (WW), but with our much larger dataset from the Match Charting Project. As we noted in Section 3, the WW approach is valid if 1) the Generalized Monotonicity Condition (GMC) holds (i.e. the Markov Perfect Equilbrium (MPE) for the overall service game decomposes into independent MPEs for the individual "point subgames" of the service game), and 2) their stationarity assumption holds (i.e. serves to the ad or deuce courts by a particular server-receiver pair on a given court surface can be treated as repeated static games for which serve directions and the "stage game" outcomes are *IID* random variables across points within a single service game and across service games on the same court surface for any server-receiver pair).

Our approach to testing the hypothesis that win probabilities are the same for all serve directions via the "Omnibus Wald test" introduced in Section 4 is robust to relaxations of both of these assumptions. In addition, our test is significantly more powerful due to the"magnification effect;" namely, that the variation in service game win chances are typically much larger than the variation in the probability of winning any particular point in the service game. Thus, the variability in win probabilities of individual points across serve directions is far harder to detect compared to the variability in conditional win probabilities for the entire service game.

However, it is useful for comparison purposes to see if WW's findings can be replicated using our larger data set using their approach to testing for equal win probabilities. It is also useful to see if their findings are robust to several other changes to how we model the "point subgame" that they analyzed, as well as changes to the statistical method used to test the null hypothesis of equal win probabilities. We follow their assumptions and approach, except that in our "replication" below we:

1. Incorporate a third serve direction, body serves, not just serves to the left and right,
2. Include second serves after a faulted first serve, not just first serves,
3. Use both Wald and likelihood-ratio tests of the null hypothesis of equal win probabilities.

We continue to impose WW's stationarity assumption. That is, all serve directions and point-subgame outcomes to the ad and deuce courts, respectively, are *IID* random variables reflecting repeated play of the same point-subgame MPE at all states of the service game. However, unlike WW (who only considered first serves), we allow for the option of a second serve by treating the point subgame as a two-person, two-stage, constant-sum game between the server and receiver, as illustrated in Figure 3 of Section 3.2. As we noted in Table 1, there is clear evidence that servers use different mixing probabilities for first vs. second serves, which we account below in our tests of equal win probabilities for all serve directions for first and second serves.

Our tests for equal win probabilities are based on a conditional likelihood function for four possible outcomes of the point subgame conditional on the serve directions d_1 and d_2, which are, respectively, the first and second serves of the subgame. Since second serves are only attempted in response to a faulted first serve, we let $d_2 = \emptyset$ denote a null outcome when the first serve was not faulted, thereby obviating the need for a second serve. Let o denote the outcome of a generic point subgame. It has four possible values $o \in \{1,2,3,4\}$ defined as follows:

$o = 1$ First serve is in, and the server wins the subsequent rally,

$o = 2$ First serve is in, and the server loses the subsequent rally,

$o = 3$ First serve is out (faulted first serve), and the server wins with the second serve,

$o = 4$ First serve is out (faulted first serve), and the server loses with the second serve.

Let $\pi(\text{in}_1|d_1)$ denote the probability that the first serve is in (i.e. not faulted) conditional on its direction being $d_1 \in \{l,b,r\}$. Let $\pi(\text{win}_1|d_1)$ denote the probability that the server wins the rally on the first serve conditional on the first serve being in and its direction being d_1. Similarly, let $\pi(\text{lose}_1|d_1)$ be the probability that the server loses the rally following the first serve, conditional on the first serve being in and its direction being d_1. Then we have:

$$1 = \pi(\text{in}_1|d_1)\pi(\text{win}_1|d_1) + \pi(\text{in}_1|d_1)\pi(\text{lose}_1|d_1) + [1 - \pi(\text{in}_1|d_1)] \tag{30}$$

reflecting the three possible outcomes of the first serve: 1) first serve is in and the server wins the rally, 2) first serve is in and the server loses the rally, or 3) the server faults the first serve.

Let $\pi(\text{win}_2|d_2)$ denote the conditional probability of winning with the second serve given that the first serve is faulted (thus resulting in a second serve) and the second serve direction is d_2. Since there are only two possible outcomes for the second serve (i.e. the server wins or loses with it), it follows that $\pi(\text{lose}_2|d_2) = 1 - \pi(\text{win}_2|d_2)$. And similarly for first serves, $\pi(\text{lose}_1|d_1) = 1 - \pi(\text{win}_1|d_1)$.

Let $f(o|d_1,d_2)$ denote the probability of the outcome o of a point subgame conditional on the serve directions for the first and second serves (d_1,d_2), where $d_2 = \emptyset$ in cases where the first serve is not faulted. We can express this conditional probability in terms of the other conditional probabilities π defined above as:

$$\begin{aligned} f(1|d_1,d_2) &= \pi(\text{in}_1|d_1)\pi(\text{win}_1|d_1) \\ f(2|d_1,d_2) &= \pi(\text{in}_1|d_1)\pi(\text{lose}_1|d_1) \\ f(3|d_1,d_2) &= [1-\pi(in_1|d_1)]\pi(\text{win}_2|d_2) \\ f(4|d_1,d_2) &= [1-\pi(in_1|d_1)][1-\pi(\text{win}_2|d_2)]. \end{aligned} \tag{31}$$

It is easy to see that $\sum_{o=1}^{4} f(o|d_1,d_2) = 1$, so $f(o|d_1,d_2)$ is a valid conditional probability. We use f as the basis for our estimation and tests of equality of win probabilities for all serve directions. Note that this distribution is fully described by nine parameters:

1. $\pi(\text{in}_1|d_1)$, $d_1 \in \{l,b,r\}$
2. $\pi(\text{win}_1|d_1)$, $d_1 \in \{l,b,r\}$
3. $\pi(\text{win}_2|d_2)$, $d_2 \in \{l,b,r\}$.

This nine-parameter model for $f(o|d_1,d_2)$ forms the basis for what we refer to as the *unrestricted log-likelihood* that does not impose any constraint that win probabilities for the first or second serve are equal across all serve directions. Note that we can define the *ex ante* win probability for the overall point subgame, accounting for the option of a second serve, as follows:

$$\text{Prob}\{\text{win point subgame}|d_1,d_2\} = \pi(\text{in}_1|d_1)\pi(\text{win}_1|d_1) + [1-\pi(\text{in}_1|d_1)]\pi(\text{win}_2|d_2). \tag{32}$$

Thus, the *hypothesis of equal win probabilities for all serve directions* (for both first and second serves) amounts to the following four restrictions on the probabilities (i.e. parameters) of the unrestricted model:

$$\begin{aligned}\pi(\text{win}_2|l) &= \pi(\text{win}_2|r) \equiv \Pi(\text{win}_2) \\ \pi(\text{win}_2|b) &= \pi(\text{win}_2|r) \equiv \Pi(\text{win}_2) \\ \pi(\text{in}_1|l)\pi(\text{win}_1|l) + [1-\pi(\text{in}_1|l)]\Pi(\text{win}_2) &= \pi(\text{in}_1|r)\pi(\text{win}_1|r) + [1-\pi(\text{in}_1|r)]\Pi(\text{win}_2) \\ \pi(\text{in}_1|b)\pi(\text{win}_1|b) + [1-\pi(\text{in}_1|b)]\Pi(\text{win}_2) &= \pi(\text{iin}_1|r)\pi(\text{win}_1|r) + [1-\pi(\text{in}_1|r)]\Pi(\text{win}_2).\end{aligned} \tag{33}$$

We use both Wald and likelihood-ratio (LR) tests to test the hypothesis of equal win probabilities above. Under the Wald test, we estimate the nine parameters of the unrestricted model by maximum likelihood and then use a standard Wald test to see if the four nonlinear restrictions on the parameters given in Equation (33) hold. For the LR test, we estimate a restricted likelihood that imposes the equal win probability restrictions (33). Call the log-likelihood of the restricted model L_r and the log-likelihood of the unrestricted model L_u. Then the LR test statistic is $\text{LR} = 2(L_u - L_r)$, and it is asymptotically distributed as a χ^2 random variable with 4 degrees of freedom if the null hypothesis of equal win probabilities for all serve directions is true. Of course, the Wald test statistic is also asymptotically distributed as a χ^2 random variable with 4 degrees of freedom under the null hypothesis.

Tables 15 and 16 present the results of these tests for the subsets of first serves to the ad and deuce courts, respectively. We also present maximum-likelihood estimates of the conditional win probabilities for the first and second serves, as defined above, and we show the total number of observations of first and second serves on which these estimates are based. The numbers in parentheses below the estimated conditional win probabilities are the estimated standard errors of the parameters. These are computed from the inverse of the Hessian matrix of the unrestricted likelihood L_u at the maximum-likelihood estimates. The final column of each table shows the P-values of the LR and Wald tests.

In Table 15, we reject the null hypothesis of equal win probabilities at the 5% significance level for only one pair, Djokovic serving to Nadal. In Table 16, we reject the null hypothesis for 3 of the 10 pairs shown: 1) Djokovic serving to Murray, 2) Murray serving to Djokovic, and 3) Sampras serving to Agassi. In general, both the Wald and LR tests result in similar test statistics and P-values, and both tests agree on the server pairs for which we reject equal win probabilities.

We can get more insight into why these rejections occur by looking at the estimated con-

Table 15: Tests of equal point win probabilities, Ad court, selected elite server-receiver pairs

Server → receiver	Serves 1st, 2nd	1st serve win probs **L**	**B**	**R**	2nd serve win probs **L**	**B**	**R**	P-values Wald, LR
Roger Federer →	692	.650	.456	.665	.493	.609	.508	.275
Rafael Nadal	274	(.025)	(.066)	(.028)	(.057)	(.059)	(.044)	.271
Rafael Nadal →	698	.647	.636	.741	.593	.527	.400	.351
Roger Federer	188	(.022)	(.043)	(.041)	(.044)	(.067)	(.155)	.331
Roger Federer →	805	.649	.622	.636	.563	.490	.493	.581
Novak Djokovic	329	(.024)	(.053)	(.026)	(.039)	(.051)	(.058)	.580
Novak Djokovic →	848	.598	.614	.596	.496	.393	.533	.463
Roger Federer	288	(.025)	(.039)	(.029)	(.042)	(.052)	(.064)	.474
Rafael Nadal →	512	.575	.567	.579	.463	.467	.750	.783
Novak Djokovic	148	(.034)	(.041)	(.039)	(.068)	(.053)	(.217)	.849
Novak Djokovic →	476	.641	.587	.716	.710	.606	.453	.0006
Rafael Nadal	161	.035)	(.050)	(.033)	(.082)	(.060)	(.062)	.0005
Novak Djokovic →	414	.641	.492	.657	.548	.591	.531	.134
Andy Murray	157	(.033)	(.062)	(.040)	(.077)	(.060)	(.071)	.122
Andy Murray →	448	.543	.534	.623	.469	.427	.440	.662
Novak Djokovic	203	(.040)	(.065)	(.031)	(.051)	(.055)	(.100)	.661
Pete Sampras →	392	.642	.650	.666	.475	.600	.436	.717
Andre Agassi	176	(.038)	(.107)	(.033)	(.050)	(.110)	(.068)	.724
Andre Agassi →	369	.674	.680	.666	.640	.555	.312	.123
Pete Sampras	128	(.030)	(.093)	(.046)	(.047)	(.165)	(.116)	.164

ditional win probabilties. For Djokovic serving to Nadal in the ad court, Table 15 shows big differences in conditional win probabilities across different serve directions, ranging from a low of 45.2% for second serves to the right to a high of 71.0% for second serves to the left. This difference of 25.8% is more than two standard errors in magnitude (when using the standard error for either win probability). The spread in conditional win probabilities for first serves is also more than two standard errors (though the standard errors for first serves are about half as large as those for second serves due to 466 first serve vs. 157 second serve observations).

Even in the cases where we are unable to reject equal win probabilities, we see pretty big differences in win probabilities across serve directions. However, since these differences are typically within two standard errors of each other, the Wald and LR tests do not reject. However, this suggests to us that the failure of these tests to reject in so many of the player pairs may

Table 16: Tests of equal point win probabilities, Deuce court, selected elite server-receiver pairs

Server → receiver	Serves 1st, 2nd	1st serve win probs			2nd serve win probs			P-values Wald, LR
		L	**B**	**R**	**L**	**B**	**R**	
Roger Federer →	769	.637	.645	.680	.467	.459	.529	.772
Rafael Nadal	276	(.027)	(.061)	(.023)	(.074)	(.064)	(.038)	.767
Rafael Nadal →	758	.613	.635	.578	.492	.616	.607	.316
Roger Federer	238	(.022)	(.036)	(.047)	(.045)	(.052)	(.092)	.320
Roger Federer →	875	.652	.624	.645	.495	.500	.458	.672
Novak Djokovic	324	(.025)	(.050)	(.023)	(.049)	(.043)	(.055)	.672
Novak Djokovic →	914	.661	.612	.637	.652	.558	.556	.270
Roger Federer	322	(.024)	(.039)	(.024)	(.045)	(.042)	(.059)	.282
Rafael Nadal →	551	.588	.578	.641	.574	.504	.615	.187
Novak Djokovic	194	(.035)	(.040)	(.034)	(.060)	(.047)	(.135)	.205
Novak Djokovic →	527	.665	.610	.680	.591	.511	.602	.493
Rafael Nadal	180	(.032)	(.056)	(.030)	(.105)	(.075)	(.036)	.473
Novak Djokovic →	465	.626	.515	.679	.667	.424	.560	.017
Andy Murray	165	(.035)	(.060)	(.032)	(.086)	(.054)	(.070)	.019
Andy Murray →	488	.639	.600	.567	.461	.436	.632	.040
Novak Djokovic	189	(.034)	(.077)	(.031)	(.057)	(.051)	(.111)	.047
Pete Sampras →	427	.675	.590	.730	.422	.628	.583	.021
Andre Agassi	186	(.033)	(.078)	(.032)	(.059)	(.074)	(.058)	.018
Andre Agassi →	405	.658	.625	.683	.481	.594	.704	.356
Pete Sampras	148	(.038)	(.061)	(.034)	(.069)	(.059)	(.088)	.384

be due to their limited power when there are relatively few observations. The lack of sufficient observations results in high estimated standard errors in conditional win probabilities, which makes it difficult to determine if win probabilities really are different across serve directions, or if these differences are purely a reflection of random sampling error.

To try to get a handle on the question of the power of these tests, we present a summary of a limited Monte Carlo study in Tables 17 and 18, which display results for serves to the ad and deuce courts, respectively. Each table summarizes the results of a Monte Carlo experiment where we simulate the play of 46 elite server-receiver pairs (which include the 10 pairs analyzed above and in the paper) across 2000 tennis points. We use our point estimates from the Match Charting Project data of the CCPs and POPs for each pair as the true data generating process. Since the point estimates generally entail unequal win probabilties across serve directions, we expect that a

Table 17: Summary of equal point win probability tests, Ad court, 46 elite player pairs

Item	**Data**	**Unrestricted Simulation**	**Restricted Simulation**
Average number of points	227	2000	2000
Rejections at 5% Wald test	1 out of 46, or 2.1% of cases	34 out of 46, or 73.9% of cases	1 out of 46, or 2.1% of cases
Average P-value Wald rejections	.0006	.008	.0005
Rejections at 5% LR test	1 out of 46, or 2.1% of cases	34 out of 46, or 73.9% of cases	3 out of 46, or 6.5% of cases
Average P-value LR rejections	.0005	.008	.0005
Average number of points for rejected pairs	476	2000	2000

sufficiently powerful test will reject the null hypothesis in the majority of the 46 cases considered.

This is indeed the case, as we can see in the "Unrestricted Simulation" column. Here, the Wald test rejects equal win probabilities at the 5% significance level for 39 of the 46 pairs when serving to the ad court, and for 34 of the 46 pairs when serving to the deuce court. In contrast, when we perform the tests on the actual Match Charting Project data, the Wald test rejects the null hypothesis for only three and one of the 46 player pairs, respectively. We conclude that power may be a concern, and having a dataset that is approximately ten times larger than the Match Charting Project (which in turn is about ten times larger than WW's dataset) is necessary to have sufficient power to detect even fairly substantial violations of the equal win probabilities for all serve directions.

There is, of course, a related concern about the possibility of "over-rejecting" the null hypothesis in large sample sizes. This danger is measured by the Type 1 error rate, which is controlled to be 5% in our example. The last column of Tables 17 and 18 titled "Restricted Simulation" presents the results of Wald and likelihood-ratio tests for another set of simulated samples of size 2000 for the ad and deuce courts, repectively. But in this case, we use as the true data-generating process maximum-likelihood estimates of the conditional win probabilities under the constraint of equal win probabilities for all three serve directions. Although we perform only a single Monte Carlo simulation per server-receiver pair, when considering all 46 pairs, we see that we have re-

Table 18: Summary of equal point win probability tests, Deuce court, 46 elite player pairs

Item	**Data**	**Unrestricted Simulation**	**Restricted Simulation**
Average number of points	224	2000	2000
Rejections at 5% Wald test	3 out of 46, or 6.5% of cases	40 out of 46, or 87.0% of cases	2 out of 46, or 4.3% of cases
Average P-value Wald rejections	.026	.009	.028
Rejections at 5% LR test	3 out of 46, or 6.5% of cases	40 out of 46, or 87.0% of cases	2 out of 46, or 4.3% of cases
Average P-value LR rejections	.028	.008	.028
Average number of points for rejected pairs	460	2000	2000

jections of equal win probabilities ranging from 2.2% to 8.7% of the pairs. This is the range of rejections we would expect to see from a test of size 5% when the null hypothesis is really true.

Thus, we have attempted to replicate WW's main findings using a testing approach similar to the one they used using our larger Match Charting Project dataset. We reach a different conclusion than they do: we do not conclude that the inability to reject the null hypothesis of equal win probabilities justifies a conclusion that elite tennis pros are playing minimax serve strategies. Instead we interpret the tests of equal win probabilities as frequently failing to reject as a result of low power due to limited number of observations of first and second serves. Our Monte Carlo experiments suggest that if we had approximately 10 times as many point game observations as we currently have (even in our larger data set, which already has about 10 times more point game observations than the Wimbledon tennis match data that WW analyzed), there would be sufficient power to reject the null hypothesis of equal win probabilities for the most of the server/receiver pairs we analyzed.

E Private Information Discussion

Private information could refer to information that is common knowledge to the players, but hidden from the econometrician, or it could refer to information that is private to one of the players. An example of the former might be the weather (wind, sun, temperature, etc.). An example of the latter might be the server knowing that she has been struggling in practice to serve close to the line without faulting when serving to the left in the ad court. In particular, *private information* need not imply *asymmetric information between the players*. We revisit our results in the context of three alternative types of private information models below.

E.1 Private Information Observed by Both Players

Augment the current (score, muscle memory) state by an additional finite set Θ. The state could be persistent within a service game or match, or change more frequently. Assume that the server (she) and receiver (he) both learn the realized value $\theta \in \Theta$ prior to choosing their actions, but that this realization is unobserved to the econometrician. Assume further that our data is generated by a Markov Perfect Equilibrium, in which the server serves in direction $d \in \{l, b, r\}$ with chance $\beta_d^\theta(x,m)$ in state (x,m,θ). Assume further that β is completely mixed in each state, which seems natural when the receiver observes θ before choosing his attention.

Let $V^\theta(x,m,d)$ be the chance of eventually winning the service game when choosing direction d for the current serve in state (x,m,θ). Since we are observing a MPE and β^θ is completely mixed, the server's conditional win chances are the same across directions, conditional on θ:

$$V^\theta(x,m,l) = V^\theta(x,m,r) = V^\theta(x,m,b) \equiv V^\theta(x,m) \quad \forall\ (x,m,\theta) \tag{34}$$

But it is not necessarily true that $V^\theta = V^{\theta'}$ for $\theta \neq \theta'$.

Assume that the data contains $n^\theta(x,m)$ serves in state (x,m,θ). Then since θ is unobservable to the econometrician, we must pool observations across $\theta \in \Theta$; and thus, equal win rates in state (x,m) across serve directions l and r (for example) requires:

$$\frac{\sum_\theta \beta_l^\theta(x,m) n^\theta(x,m) V^\theta(x,m)}{\sum_\theta \beta_l^\theta(x,m) n^\theta(x,m)} = \frac{\sum_\theta \beta_r^\theta(x,m) n^\theta(x,m) V^\theta(x,m)}{\sum_\theta \beta_r^\theta(x,m) n^\theta(x,m)} \tag{35}$$

Trivially, Equality (34) implies (35) if either: (i) $V^\theta(x,m)$ is constant across θ, or (ii) $\beta_l^\theta(x,m)$ and $\beta_r^\theta(x,m)$ are both constant across θ.[40]

[40] But if (34) fails to hold, then aggregation across θ can reduce the difference in conditional win rates; and thus, reduce the power of our tests to reject the null hypothesis of equal win rates.

Our stationarity results can be viewed as tests of (i) and (ii) for particular models of θ. In particular, if Θ is the partition of our service games across the 2–3 calendar year bins we analyze for each server-receiver pair, then our failure to reject stationarity in the POPs (in Table 2), supports the hypothesis that $V^{\theta}(x,m)$ constant across θ. Similarly, our stationarity tests based on dividing the data between first sets and non first sets can be viewed as a test of (i) and (ii) if θ changes from the first set to later sets, but is otherwise fixed. Again, our failure to reject stationarity in the POPs for this early vs. late division of the data (in Table 13) supports the hypothesis that $V^{\theta}(x,m)$ is constant across θ.

E.2 Transitory Asymmetric Information.

We now consider two models in which the server has transitory private information. In order to streamline the notation, we assume that serves always land in (thus, no second serves), that the game is played in only one court, and that the server does not choose spin or speed.[41] Altogether, $w(m,d,a)$ is the chance the server wins the current serve (and point) in muscle memory state m serving in direction d when the receiver chooses attention vector a. Since we have assumed the game is played in only one court and that there are no second serves, we assume that m simply encodes the last serve direction.

We now define the *taste shock model* as in Section 3.4 of the paper, but with the receiver's behavior explicitly modeled. The taste shock $\varepsilon(d)$ is an *iid* draw across serves and across directions $d \in \{L,B,R\}$ from a standardized Type 1 extreme value distribution with location parameter normalized so that $E\{\max_d \varepsilon(d)\} = 0$. The server observes the score and muscle memory states and the realization $\varepsilon = (\varepsilon(L), \varepsilon(B), \varepsilon(R)) \in \mathbb{R}^3$ prior to choosing her mixture over serve locations, where X is the set of score states. Thus, the server's strategy is a mapping from $X \times \{L,B,R\} \times \mathbb{R}^3$ to mixtures over serve directions. Since the receiver does not observe the realization ε, his strategy α is a mapping from $X \times \{L,B,R\}$ to mixtures over attention vectors $(a^l, a^b, a^r) \geq 0$ with $a^l + a^b + a^r = 1$.

Let $V_\lambda(x,m|\alpha,\beta)$ be the server's expected payoff starting in score and muscle memory state x given that the receiver and server follow strategies α and β for all serves. Let $V_\lambda(x,m,d,a|\alpha,\beta)$, be

[41] However, it should be clear that the analysis generalizes to the more general case in the paper.

the server's expected payoff assuming *current* serve direction d and attention a, with the receiver and server employing strategies α and β for all *future* serves. These value functions obey:

$$\begin{aligned} V_\lambda(x,m,d,a|\alpha,\beta) &= w(m,d,a)V_\lambda(x^+(x),d|\alpha,\beta)+(1-w(m,d,a))V_\lambda(x^-(x),d|\alpha,\beta) \\ V_\lambda(x,m|\alpha,\beta) &= \int\int\sum_d V_\lambda(x,m,d,a|\alpha,\beta)\beta_d(x,m,\varepsilon)f(\varepsilon)d\varepsilon\, d\alpha(a|x,m), \end{aligned}$$

where f is the pdf over the vector of taste shocks $\varepsilon = (\varepsilon(L),\varepsilon(B),\varepsilon(R))$. We again overuse V_λ for both the expected payoff conditional on serving in direction d and unconditional on serve direction. Similarly, we define the expected payoffs for the receiver:

$$\begin{aligned} V_\lambda(x,m,d|\alpha,\beta) &\equiv \int V_\lambda(x,m,d,a|\alpha,\beta)d\alpha(a|x,m) \\ U_\lambda(x,m,a|\alpha,\beta) &\equiv 1-\int\sum_d V_\lambda(x,m,d,a|\alpha,\beta)\beta_d(x,m,\varepsilon)f(\varepsilon)d\varepsilon. \end{aligned}$$

Then using the distribution over taste shocks, the server's best response β^* obeys:

$$\beta_d^*(x,m|\alpha) = \frac{\exp\{V_\lambda(x,m,d|\alpha,\beta^*)/\lambda\}}{\sum_{d'\in\{l,b,r\}}\exp\{V_\lambda(x,m,d'|\alpha,\beta^*)/\lambda\}}, \tag{36}$$

which implies the following optimal expected payoff for the server:

$$V_\lambda(x,m|\alpha) = \lambda\log\left(\sum_{d\in\{l,b,r\}}\exp\{V_\lambda(x,m,d|\alpha,\beta^*)/\lambda\}\right). \tag{37}$$

Meanwhile, α^* is a best response to β as long as the support of $\alpha^*(x,m)$ only contains a solving:

$$\max_a U_\lambda(x,m,a|\alpha,\beta).$$

Then the expected payoff will be $V_\lambda(x,m) = V_\lambda(x,m|\alpha^*,\beta^*)$ for any Bayesian Nash Equilibrium (BNE) (α^*,β^*), which corresponds to the expected payoff in Section 3.4. Furthermore, since any BNE converges to a MPE of the model without taste shocks in Section 3.1, the taste shock model *purifies* the equilibria of our model without taste shocks.

In the taste shock model, the server considers the shocks when choosing serve directions, but the shocks do not directly impact her payoff. In the closely related *payoff shock model*, the shocks $\varepsilon(d)$ directly enter the payoff functions. That is, given payoff shock $\varepsilon(d) = e$, the server's expected payoff when choosing direction d is:

$$V_\lambda(x,m,d,a,e|\alpha,\beta) = \lambda e + w(m,d,a)V_\lambda(x^+(x),d|\alpha,\beta)+(1-w(m,d,a))V_\lambda(x^-(x),d|\alpha,\beta).$$

The server's best response remains (36), while the expected payoff is adjusted by a constant:

$$V_\lambda(x,m|\alpha) = \lambda \log\left(\sum_{d\in\{l,b,r\}} \exp\{V_\lambda(x,m,d|\alpha,\beta^*)/\lambda\}\right) - \lambda\log(3).$$

This payoff shock model has the following *theoretical* shortcomings:[42]

1. The additive shocks have full support over the entire real line, and thus the realized win rate need not be in $[0,1]$.

2. The location parameter is set so that $E[\max_d \varepsilon(d)] = 0$, and thus $E[\varepsilon(d)] < 0$.

3. The shocks are additive to the endogenous values, rather than perturbing the underlying chances of serving in and chances of winning conditional on serving in.

However, given that the estimated values of lambda are so small, we do not think any of these issues is a substantive *empirical concern in our application*.

First, consider Issue 1. We can calculate the chance that the realized service game win rate falls outside of the interval $[0,1]$ given the estimated values of λ and V_λ for each of the three structural specifications (serve-myopic, point-myopic, and fully-dynamic). For the fully-dynamic model, this chance rounds to zero at many decimal places for all sever-receiver pairs in all states (x,m). While the λ shocks are larger in the serve-myopic and point-myopic models, the chance that the point or game win rate falls outside of the interval $[0,1]$ is at most 0.07 across *all* server-receiver pairs and all three model specifications. Given that the shocks are additive and the game is constant sum, we think our specification is a very close approximation to a model that constrains the realized win chances to $[0,1]$ (for example, by using a truncated Gumbel distribution).

A similar point applies to Issue 2: normalizing shocks so that $E[\varepsilon(d)] < 0$. In particular, our normalization $E[\max_d \varepsilon(d)] = 0$ implies that $E[\varepsilon(d)] = -\lambda\log(3)$. It seems quite unlikely that changing the location parameter by less than 0.001 (the maximum change across all server-receiver pairs in our fully-dynamic model) would have a significant impact on our results.

Next consider Issue 3: we have shocked *dynamic* values instead of the stage game POPs. This issue is not a problem in the serve-myopic model, but additive shocks to values become

[42] We thank an anonymous referee for this list.

more problematic as we make the optimization more far-sighted. In particular, if we adjust the fully-dynamic model so the shocks are to the point win chances, the *POP shock model* assumes:

$$V_\lambda(x,m,d,a,e|\alpha,\beta) = (w(m,d,a)+\lambda e)V_\lambda(x^+(x),d|\alpha,\beta)+(1-w(m,d,a)-\lambda e)V_\lambda(x^-(x),d|\alpha,\beta).$$

If we then define:

$$\Delta_\lambda(x,m,d) = V_\lambda(x^+(x),d|\alpha,\beta) - V_\lambda(x^-(x),d|\alpha,\beta),$$

we can think of this as a model with value shocks by writing (suppressing α and β):

$$V_\lambda(x,m,d,a,e|\alpha,\beta) = \lambda\Delta_\lambda(x,m,d)e + w(m,d,a)V_\lambda(x^+(x),d) + (1-w(m,d,a))V_\lambda(x^-(x),d).$$

The added complication is that now the scale parameter for the value shocks is $\lambda\Delta_\lambda(x,m,d)$, i.e. the scale parameter depends on the current score, muscle memory state, and serve direction. This last dependence is problematic, since it implies that the maximized value of the shock no longer follows a Gumbel distribution. But in practice, we think this is unlikely to be a serious empirical concern. In particular, in all of our estimates, muscle memory has a small impact on service game win rates, such that $\Delta_\lambda(x,m,d)$ is nearly a constant across (m,d). Thus, we think that replacing $\Delta_\lambda(x,m,d)$ by the average value across all muscle memory states and serve directions $\Delta(x)$ is unlikely to bias the results. Altogether, we could estimate the POP shock model by setting $\lambda(x) = \lambda\Delta(x)$ with (suppressing α and β on the RHS):

$$\beta^*_d(x,m) = \frac{\exp\{V_\lambda(x,m,d)/\lambda(x)\}}{\sum_{d'\in\{l,b,r\}}\exp\{V_\lambda(x,m,d')/\lambda(x)\}}.$$

We agree that this is a more satisfactory model of payoff shocks. But would allowing the scale parameter λ to depend on the current score significantly alter the results? Given that the estimated values of λ in the current model are so low, we do not think that allowing the scale of our shocks to be score specific would make any difference to our bottom line conclusions.

Altogether, we do not think a more sophisticated model of transitory *payoff* shocks would substantively change our empirical results.

E.3 Persistent Private Information.

To be concrete, assume the underlying stage game payoffs are determined by $\theta \in \Theta$ as in Section E.1 above, but that the server is informed of the realized value of θ prior to the start of a match, and that this value is fixed throughout the match. Let $\rho = (\rho_1, \rho_2, \ldots, \rho_N)$ denote the receiver's beliefs about θ, i.e. ρ_i is the probability that the state is θ_i from the receiver's point of view given the history of actions up through the current serve. We could then consider a Perfect Bayesian Equilibrium, in which the receiver's mixture over attention α depends on the score and muscle memory state (x,m) as well as the current belief vector ρ, while the server's mixture β depends on (x,m,ρ) and θ.

We feel that this would be an interesting theoretical generalization of our model, and analyzing it would be a novel theoretical contribution. On the other hand, introducing an additional state variable would reduce the precision of our empirical estimates. Ultimately, the question is whether a private information model of this sort is likely to be of sufficient empirical relevance in our applications to justify the reduced precision. Our opinion is that it is not *given the data we have to analyze*.

Our main justification the results of our stationarity tests for the first set vs. non-first set split of the data. In particular, private information will only change the empirical results if the server conditions on θ, in which case the receiver's belief ρ will drift toward the truth throughout the match. Consequently, the receiver's mixture should then change throughout the match as his beliefs drift toward the truth, which in turn implies that the POPs will change throughout the match. But we find that the POPs are statistically unchanged from early (first set) to later (subsequent sets) in a match (Table 13) for no pairs in the model with no muscle memory and only one of ten pairs in the model with muscle memory.

Of course, it could be the case that the first set vs. non-first set split is unable to detect the changes in the POPs induced by changes in ρ, i.e. that a finer partition is required. More generally, we conjecture that an identification problem exists that makes testing for equilibrium problematic: there may always be an alternative model of private information that is consistent with both PBE and the data we observe. Exploring this conjecture is beyond the scope of this paper. Ideally, future work with richer data (e.g. including receiver location, serve speed, rally length, etc.) may help to mitigate the identification problems and provide more illuminating analyses of the effect of persistent private information.

F Muscle Memory and Serial Correlation in Serve Locations

We now explore serial correlation in serve location choices in MPE. We do this in a simple version of the model, but the core insights remain valid in the general model.

F.1 Three Sources of Serial Correlation in Serve Locations

Assume that attention to serve location d, only directly impacts win rates if the server chooses location d, i.e. $\omega(m,d,s,a) = \omega(m,d,a^d)$, and that $\omega(m,d,a^d)$ is differentiable, strictly decreasing, and weakly convex in a^d for all m. Let speed and spin choices in some MPE be given by $s^*(x,m)$ and define the induced conditional probabilities:

$$\ell^d(x,m) \equiv \ell(m,d,s^*(x,m)) \quad \text{and} \quad \omega^d(x,m,a^d) \equiv \omega(m,d,s^*(x,m),a^d),$$

Recall that $W(x,m)$ is the server's chance of wining the service game in state (x,m) and let $w^d(x,m,a^d)$ be the conditional chance that the server wins the service game when choosing location d given that the receiver chooses attention a^d at location d. Then for first serves (x odd):

$$\begin{aligned} W^d(x,m,a^d) &= \ell^d(x,m)\left(\omega^d(x,m,a^d)W(x^+(x),(m_2,d)) + (1-\omega^d(x,m,a^d))W(x^-(x),(m_2,d))\right) \\ &\quad + (1-\ell^d(x,m))W(x+1,(m_2,d)) \end{aligned}$$

To further simplify, assume an MPE in which the server strictly mixes over l,r with respective chances $\sigma_S(x,m), 1-\sigma_S(x,m)$ on first serves.[43] Since the receiver has no direct effect on the muscle memory state, the receiver will best respond by setting $a^b = 0$; and thus, we have $a^r = 1-a^l$. Altogether, the chance that the server wins the service game in state (x,m) for first serves given receiver attention a^l is:

$$W(x,m,a^l) \equiv \sigma_S(x,m)W^l(x,m,a^l) + (1-\sigma_S(x,m))W^r(x,m,a^r)$$

Further assuming $a^l \in (0,1)$,[44] it must be the case that the receiver cannot lower this probability by adjusting a^l up *or* down; and thus, the receiver's MPE attention $a^l(x,m)$ must obey

[43] A sufficient condition for an MPE with no body first serves is that serving left or right on first serves gives the server a better chance of winning the current point *and* enhances his chances of winning future points: $\ell^b(x,m)\omega^b(x,m,a^b) < \ell^d(x,m)\omega^d(x,m,a^d)$ for all $m,a,d \in \{l,r\}$, and x odd *and* $\ell^d(x,(d'',b))\omega^d(x,(d'',b),a^d) \leq \ell^d(x,(d'',d'))\omega^d(x,(d'',d'),a^d)$ for all x,d'',a^d and $d,d' \in \{l,r\}$.

[44] An assumption that *implies* $a^l \in (0,1)$ is that the ratio $\omega^l_{a^l}(x,m,a^l)/\omega^r_{a^r}(x,m,1-a^l)$ converges to ∞ as $a^l \to 0$ and converging to 0 as $a^l \to 1$.

$W_{a^l}(x,m,a^l(x,m)) = 0$, i.e.:

$$\frac{\sigma_S(x,m)}{1-\sigma_S(x,m)} = \frac{W^r_{a^r}(x,m,1-a^l(x,m))}{W^l_{a^l}(x,m,a^l(x,m))} = \frac{\ell^r(x,m)\omega^r_{a^r}(x,m,1-a^l(x,m))\Delta^r(x,m_2)}{\ell^l(x,m)\omega^l_{a^l}(x,m,a^l(x,m))\Delta^l(x,m_2)} \tag{38}$$

where $\Delta^d(x,m_2) = W(x^+(x),(m_2,d)) - W(x^-(x),(m_2,d))$ is the increase in the service game win chance from winning vs. losing the current point on the first serve.

Equation (38) affords a way to formalize equilibrium serial correlation in first serve strategies. Specifically, a sufficient condition for negatively serial correlation is that the server is less likely to serve left following a left serve, i.e. when $\sigma(x,(d,l)) < \sigma(x,(d,r))$, for all odd x and first serve locations d chosen two first serves prior, which by (38) is equivalent to:

$$\frac{W^r_{a^r}(x,(d,l),1-a^l(x,(d,l)))}{W^l_{a^l}(x,(d,l),a^l(x,(d,l)))} < \frac{W^r_{a^r}(x,(d,r),1-a^l(x,(d,r)))}{W^l_{a^l}(x,(d,r),a^l(x,(d,r)))} \tag{39}$$

Since $W^d_{a^d} = \ell^d\omega^d_{a^d}\Delta$, inequality (39) compares the product of three separate ratios, and thus, there are three logically separate ways to generate negative serial correlation with muscle memory. One is when muscle memory affects the server's chance of landing a serve in, as follows:

$$\frac{\ell^r(x,(d,l))}{\ell^l(x,(d,l))} < \frac{\ell^r(x,(d,r))}{\ell^l(x,(d,r))} \tag{40}$$

This comparison of likelihood ratios states that the server's relative chance of landing a right serve in is higher following a right serve. While this makes intuitive sense, it is an empirical question whether such short term muscle memory effects exist for elite pro serves.

Muscle memory can also generate (39) if the following inequality holds:

$$\frac{\omega^r_{a^r}(x,(d,l),1-a^l(x,(d,l)))}{\omega^l_{a^l}(x,(d,l),a^l(x,(d,l)))} < \frac{\omega^r_{a^r}(x,(d,r),1-a^l(x,(d,r)))}{\omega^l_{a^l}(x,(d,r),a^l(x,(d,r)))} \tag{41}$$

For an interpretation of this condition, notice that $\omega^r_{a^r}/\omega^l_{a^l}$ is the marginal rate of technical substitution (MRTS) between attention at location r and attention at location l. Inequality (41) states that this MRTS is larger following a serve to the right than it is following a serve to the left. This could be the result of a *direct effect* of muscle memory, the MRTS larger following a serve to the right holding the receiver's attention constant, or an *indirect effect*, the receiver's attention changes following a serve to the right inducing a larger MRTS.

Notice that inequalities (40) and (41) are about the impact of *past* serve locations on the *current* point game ratios ℓ^r/ℓ^d and $\omega^r_{a^r}/\omega^l_{a^l}$. These effects may be present even if the server and receiver behave *myopically*, maximizing their chances of winning the current point and ignoring the future. When the players are forward looking, there is third potential source of negative serial correlation; namely:

$$\frac{\Delta^r(x,l)}{\Delta^l(x,l)} < \frac{\Delta^r(x,r)}{\Delta^l(x,r)}$$

which after substituting in for all Δ^d becomes:

$$\frac{W(x^+(x),(l,r)) - W(x^-(x),(l,r))}{W(x^+(x),(l,l)) - W(x^-(x),(l,l))} < \frac{W(x^+(x),(r,r)) - W(x^-(x),(r,r))}{W(x^+(x),(r,l)) - W(x^-(x),(r,l))} \tag{42}$$

For an interpretation, recall that muscle memory $m = (m_1, m_2)$, where m_1 is the location of the previous first serve and m_2 is the location of the second to last first serve. Now, arbitrarily order serve locations $l > r$ (or vice versa), then inequality (42) states that the increase in the probability of winning the service game from winning vs. losing the *current* point $W(x^+(x),m) - W(x^-(x),m)$ is strictly log-supermodular in (m_1,m_2). This necessarily requires muscle memory to depend on the two previous serve locations.

F.2 Example: Serial Correlation in A Linear Model

We now simplify the model further in order to sign the equilibrium serial correlation in serve locations and show that this serial correlation can be strictly negative (or strictly positive), even if the POPs in Definition 1 are independent of muscle memory.

The *two location linear model* removes spin, speed and body serves as choice variables and assumes that ℓ and ω only depend on the prior first serve location, m_1, and that the conditional win chance ω is linear in attention, i.e. $\omega^d(m_1) = \bar{\omega} - \eta^d(m_1)a^d$ with $\bar{\omega} \in (0,1)$ and $\eta^d(m_1) \in (0,\bar{\omega})$. Thus, this model is fully determined by the nine scalars: $\bar{\omega}$ and $\ell^d(m_1), \eta^d(m_1)$ for $(d,m_1) \in \{l,r\}^2$. The two location linear model is *log-supermodular* when $\ell^r(l)\ell^l(r)\eta^r(l)\eta^l(r) < \ell^r(r)\ell^l(l)\eta^r(r)\eta^l(l)$ and *log-submodular* when the opposite inequality obtains. The *symmetric two location linear model* further restricts: $\ell^d(m_1) = \bar{\ell}$, $\eta^r(l) = \eta^l(r) = \eta$, and $\eta^l(l) = \eta^r(r) = \hat{\eta}$.

Theorem 4. *Serve locations are negatively (positively) serially correlated in the log-supermodular (log-submodular) two location linear model in any MPE in which the server strictly mixes over first serve locations. The POPs are independent of muscle memory in the symmetric two location linear model.*

STEP 1: SERIAL CORRELATION IN SERVE LOCATIONS. Direct substitution establishes that inequality (39) is equivalent to $\ell^r(l)\ell^l(r)\eta^r(l)\eta^l(r) < \ell^r(r)\ell^l(l)\eta^r(r)\eta^l(l)$ (i.e. log-supermodularity) in the two location linear model; and thus, negative serial correlation obtains. Similarly, under log-submodularity, inequality (39) flips, implying positive serial correlation.

STEP 2: SERIALLY INDEPENDENT POPS. The chance of serving in ℓ is a constant by assumption. Routine algebra establishes that the following strategies constitute a MPE:

$$\frac{a^l(r)}{1-a^l(r)} = \frac{1-a^l(l)}{a^l(l)} = \frac{\sigma_S(r)}{1-\sigma_S(r)} = \frac{1-\sigma_S(l)}{\sigma_S(l)} = \frac{\hat{\eta}}{\eta} \tag{43}$$

Given these strategies, the win chance is $\omega^d(m_1) = \bar{\omega} - \frac{\eta\hat{\eta}}{\eta+\hat{\eta}}$, independent of muscle memory. This implies the continuation value function W is independent of muscle memory; and thus, the generalized monotonicity condition 2 holds. Altogether, the service game can be decomposed into a sequence of identical static games. It is straightforward to verify that strategies (43) constitute the unique equilibrium in these static games. □

The symmetric model is log-supermodular when $\eta < \hat{\eta}$ and log-submodular when $\eta > \hat{\eta}$; and thus, serve locations are generically either negatively serially correlated or positively serially correlated in the symmetric two location linear model, despite the fact that the POPs are independent of muscle memory.

References

Andrews, D. W. K. (1987). Asymptotic results for generalized wald tests. *Econometric Theory*, *3-3*, 348-358.

Bhattacharya, V., & Howard, G. (2022). Rational inattention in the infield. *American Economic Journal: Microeconomics*, *14-4*, 348-393.

Brown, J. N., & Rosenthal, R. (1990). Testing the minimax hypothesis: A re-examination of O'Neill's game experiment. *Econometrica*, *58*(5), 1065-1081.

Camerer, C. F. (2003). *Behavioral game theory: Experiments on strategic interaction*. Princeton University Press.

Chiappori, P., Levitt, S., & Groseclose, T. (2002). Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review*, *92*(4), 1139–1151.

Everett, H. (1957). Recursive games. In M. Dresher, A. Tucker, & P. Wolfe (Eds.), *Contributions to the theory of games, in ann. of math. stud., vol. iii* (pp. 47–78). Princeton, NJ: Princeton University Press.

Gauriot, R., Page, L., & Wooders, J. (2023). Expertise, gender, and equilibrium play. *Quantitative Economics (forthcoming)*.

George, S. L. (1973). Optimal strategy in tennis: a simple probabilistic model. *The Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *22*(1), 97–104.

Glicksberg, I. (1952). A further generalization of the kakutani fixed point theorem, with application to nash equilibrium points. *Proceedings of the American Mathematical Society*, *3*, 170–174.

Harsanyi, J. D. (1973). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory*, *2-1*, 1-23.

Hsu, S., Huang, C., & Tang, C. (2007). Minimax play at wimbledon: comment. *American Economic Review*, *97*(3), 517–523.

Hwang, S.-H., & Rey-Bellet, L. (2020). Strategic decompositions of normal form games: Zero-sum games and potential games. *Games and Economic Behavior*, *122*, 370–390.

Iskhakov, F., Jørgensen, T. H., Rust, J., & Schjerning, B. (2017). The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks. *Quantitative Economics*, *8*, 317–365.

Iskhakov, F., Rust, J., & Schjerning, B. (2016). Recursive lexicographical search: Finding all markov perfect equilibria of finite state directional dynamic games. *Review of Economic Studies*, *83*(2), 658–702.

Klaassen, F., & Magnus, J. (2001). 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.

Klaassen, F., & Magnus, J. (2009). The efficiency of top agents: An analysis through service strategy in tennis. *Journal of Econometrics*, *148*(1), 72–85.

Lasso de la Vega, C., & Volij, O. (2020). The value of a draw. *Economic Theory*, *70-4*, 1023-1044.

Lewis, A. D. (2009). *Semicontinuity of rank and nullity and some consequences.*

Lewis, M. (2003). *Moneyball: The art of winning an unfair game*. W.W. Norton & Company.

McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, *10*, 6-38.

McKelvey, R. D., & Palfrey, T. R. (1998). Quantal response equilibria for extensive form games. *Experimental Economics*, *1*, 9-41.

Norman, J. M. (1985). Dynamic programming in tennis-when to use a fast serve. *The Journal of the Operational Research Society*, *36*(1), 75–77.

Palacios-Huerta, I. (2003). Professionals play minimax. *Review of Economic Studies*, *70*, 395–415.

Palacios-Huerta, I. (2014). *Beautiful game theory: How soccer can help economics*. Princeton University Press.

Rive, J., & Williams, S. C. (2011). *Tennis skills & drills*. Human Kinetics, Inc.

Romer, D. (2006). Do firms maximize? evidence from professional football. *Journal of Political Economy*, *114*(2), 340–365.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, *55*(5), 993–1033.

Sackmann, J. (2013). *The match charting project.* Retrieved from `TennisAbstract.com`

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138.

Tea, P., & Swartz, T. B. (2022). The analysis of serve decisions in tennis using bayesian hierarchical models. *Annals of Operations Research*.

Ville, J. (1938). Note sur la théorie générale des jeux ou intervient l'habilité des joueurs. In *Traité du calcul des probabilités et de ses applications.*

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, *57-2*, 307–333.

Walker, M., & Wooders, J. (2001). Minimax play at wimbledon. *American Economic Review*, *91*(5), 1521–1538.

Walker, M., Wooders, J., & Amir, R. (2011). Equilibrium play in matches: Binary markov games. *Games and Economic Behavior*, *71*, 487–502.

Wiles, J. (2006). *Mixed strategy equilibrium in tennis serves* (Tech. Rep.).