INTERNATIONAL ECONOMIC REVIEW
Vol. 53, No. 3, August 2012

ON THE OPTIMALITY OF LINE CALL CHALLENGES IN PROFESSIONAL TENNIS*

BY RAN ABRAMITZKY, LIRAN EINAV, SHIMON KOLKOWITZ, AND ROY MILL[1]

Stanford University, U.S.A. and NBER, U.S.A; Stanford University, U.S.A. and NBER, U.S.A.; Harvard University, U.S.A.; Stanford University, U.S.A.

We study professional tennis players' decisions of whether to challenge umpires' calls using data on over 2,000 challenges in 35 tennis tournaments. The decision to challenge, which is simple to characterize, trades off reversing the umpire's call against losing subsequent challenge opportunities. Qualitatively, players are more likely to challenge when the stakes are greater and when the option value of challenging is lower, as theory predicts. Quantitatively, players' actual behavior is close to an optimal challenging strategy prescribed by a simple dynamic model. Our findings illustrate that professional decision makers develop decision rules that can approximate optimal behavior quite well.

1. INTRODUCTION

Decision theory is at the heart of almost any aspect of research in economics, and optimal decision making is often assumed in economic analyses. Nevertheless, it is challenging to test empirically this assumption outside of the laboratory, mainly because individuals' decisions and objective functions are typically complex, and it is often difficult to identify whether deviations from optimal behavior indeed reflect suboptimal decisions or simply misspecified objective functions.

In this article, we address this challenge by focusing on a setting in which the decision and objective function are simple. Specifically, we analyze professional tennis players' decisions of whether to challenge the umpire's call. A player's successful challenge reverses the umpire's decision, and an unsuccessful challenge leaves the score unchanged. Each player is only allowed a small number of challenges per set, thereby creating an opportunity cost of challenging. There are several features that make the challenge decision in professional tennis an attractive setting to test decision theory. First, a player's objective function is simple and well defined: to win the match. Second, as we show later, the trade off in deciding whether to challenge or not is easy to characterize. Third, professional tennis is an individual sport, so confounding issues such as group decision making or agency costs are not present. Fourth, beyond the intrinsic value of winning, the dollar stakes are reasonably large.[2] Finally, data on challenges and their outcomes are systematically recorded, rendering possible a study of a fairly large set of similar decision problems.

We assemble data on all challenges made in 35 men's tennis tournaments played in 2006–2008, the first three years in which the challenge technology was used in professional tennis. Overall, the data cover more than 2,000 challenges, and contain complete information about the players, the match, the point in the match at which the challenge was made, and the outcome of the challenge.

*Manuscript received November 2010; revised June 2011.

[1] We thank the ATP, the Sony Ericsson WTA Tour, and Hawk-Eye Innovations for providing the raw data used in this article, and Daniele Paserman for sharing his code. We are grateful to Hanming Fang (the Editor), two anonymous referees, Izi Sin, and Gui Woolston for many useful comments. Financial support from the National Science Foundation is gratefully acknowledged. Please address correspondence to: Liran Einav, Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305-6072. Phone: 650-723-3704; Fax: 650-725-5702. E-mail: *leinav@stanford.edu*.

[2] The direct incremental prize money from winning matches in our data ranges from several thousand to several hundred thousand dollars. The indirect value of winning a match is much greater due to the option it provides to keep advancing in the tournament and its effect on the player's ranking and endorsement deals.

Our analysis of the decision to challenge is complicated by the fact that we have information on points that were actually challenged only, rather than a more complete point-by-point data. We can nevertheless make inferences despite the selection problem by noting that, all else equal, a higher propensity to challenge implies challenging less conservatively, thereby leading to lower success rates in overturning the umpire's call. Thus, our inference relies on whether the challenge was successful and on the degree to which it was successful or not. From a conceptual perspective, our analysis thus demonstrates how a data selection problem can be overcome by using outcome data.

We start in Section 2 by providing a short background about tennis challenges, and we briefly describe the data in Section 3. Section 4 develops a dynamic model of tennis that provides a simple characterization of the trade-off involved in the decision whether or not to challenge. There are only two underlying factors that enter this decision: the value of reversing the umpire's call and the opportunity cost of the challenge. The former essentially captures the relative importance of the point (in a way that the model makes precise), whereas the latter captures the option value associated with retaining one additional challenge. In Section 5, we report a set of regressions that show that tennis players behave in a way that is qualitatively consistent with what optimal decision making would prescribe. They are *more* likely to challenge (as measured by being *less* successful conditional on challenge) more important points and when the option value associated with an additional challenge is lower.

In Section 6, we use the model to quantify how close challenge behavior is to optimal. We find that although tennis players seem to challenge too little relative to what optimal decision making would prescribe, they seem to capture much of the benefits associated with the challenge technology. In our baseline specification, we find that the observed challenge behavior increases the player's probability of winning the match by 1.55 percentage points, which is 97% of what could be attained by optimal behavior. We estimate that a more naive challenging technology would capture only 75% of this value. These general results—that players challenge too little but yet capture almost all the value—are fairly robust across several alternative specifications that we explore.[3]

Two other papers in economics exploit—like us—the relative simplicity of the game of tennis for empirical work, although they are motivated by different aspects of the game. Walker and Wooders (2001) test predictions of game theory by studying tennis serves, and Paserman (2010) uses point-by-point data from tennis to investigate gender differences.

Our finding that observed decisions are close to optimal contrasts with the suboptimal behavior found by Levitt (2006) and Romer (2006) in the context of an owner–operator bagels distribution firm and professional football, respectively. Levitt (2006) points to the lack of feedback about bad decisions as a possible source for suboptimal behavior. In our context, tennis players are routinely exposed to feedback through coaches, fans, television, post-game analysis, and in-play situations, which may explain why the results differ. Romer (2006) raises two reasons for his negative findings: either genuinely suboptimal behavior of football coaches or agency costs that give rise to objectives that go beyond winning a football game. Because, unlike football coaches, tennis players' careers only depend on their own performance (rather than on NFL owners' perceptions), agency costs are not important in our context. Thus, our findings suggest that Romer's more negative results may well be driven by such career concerns. We develop this point and further discuss the potential relevance of our findings to other settings in the concluding section.

2. TENNIS CHALLENGES: A BRIEF BACKGROUND

Tennis matches are played between two opponents,[4] typically within a tournament in which the winner advances to the next round and the loser is eliminated. Scoring is based on points,

[3] Avid tennis fans may note that we abstract from secondary elements that are sometimes mentioned in the context of challenges, such as gaining rest or changing momentum. As we discuss later in the article, the data suggest that these factors are unlikely to be driving the vast majority of the points in our data.

[4] We focus our analysis on singles. Tennis can also be played in doubles, where two players are on each side.

which add up to games, which in turn add up to sets. The first player who wins two sets is the winner of the entire match.[5] To win a set one has to win six games (or seven, if the opponent won more than four), and to win a game one has to win at least four points (and two more than the opponent). If a set reaches a score of 6–6 in games, a longer decisive game is played in which a player must win at least seven points (and two more than the opponent). This decisive game is called a tiebreak.

The focus of our analysis is a point. A tennis point is a sequence of alternating shots among the two players. One way in which a player wins a point is if his opponent hits the ball and it does not pass the net. When the ball passes above the net after being hit, it has to land within a designated area of the court marked by lines. If the ball lands outside this area, the ball is called "out" and the player who hit it loses the point. If the ball lands inside this area the ball is called "in," and the player receiving the ball must hit it back before it lands on the ground for a second time, or else his opponent wins the point. To be ruled "in," the ball must land within the designated area or at least some portion of the ball must touch the designated area's boundary line. The umpire is the one responsible for calling the ball in or out.[6]

Professional tennis players hit the ball hard and accurately. It is therefore not uncommon for the ball to land very close to the line, making the umpire's in/out call potentially controversial. As in several other sports, improved technology allowed the governing body of professional tennis to introduce automated machines that can simulate the ball's movement and produce reasonably precise reads of the landing location of the ball. In order to avoid altering traditional tennis too much and in an attempt to avoid frequent disturbances to the regular flow of the game, these automated machines do not replace the umpire. Instead, they are used when the player who lost the point challenges the umpire's call, which can occur only a limited number of times.

Starting in 2006, most important tournaments installed and employed this new technology, thus allowing players to challenge a call.[7] The technology, named "Hawk-Eye," is based on quick processing of visual images (taken by several high-speed video cameras located around the court) to locate precisely where the ball landed. When a challenge is initiated, it takes 20–30 seconds for the computerized path and final landing location of the ball to be calculated and shown to the umpire, players, and the crowd on a large screen. According to the system's proponents, this leads to an additional layer of excitement. It is important to recognize that the Hawk-Eye verdict is a statistical prediction, and lab testing indicates a mean prediction error of 3.6 mm (leading to occasional complaints about the system). However, because the Hawk-Eye call is all that is relevant for the decision of whether to challenge or not, we treat the Hawk-Eye prediction as the "truth" throughout the article.

In most tournaments covered by our data, players were endowed with two opportunities to challenge per set.[8] If a player challenged and the umpire's decision was not overturned, the player lost one opportunity to challenge. If, however, the challenge was successful and the umpire's decision was overturned, the player won the point and did not lose the challenge opportunity. If a player's number of unsuccessful challenges in the set reached the maximum

[5] The precise scoring rules vary somewhat across tournaments. For example, in the most important tournaments of the year one has to win three sets (rather than two) to win the match. Throughout this section we describe the rules that govern the tournaments for which we have data.

[6] There are several (eight) linemen who are spread around the court in order to help the chief umpire make the right call. However, it is the umpire's ultimate responsibility to make the call, and it is not uncommon for the umpire (who sits on a high chair in the middle of the court) to overrule a lineman's call if he (or she) feels that he had a better view. Throughout the article, we simply treat the entire officiating crew as a single unit and call it the umpire.

[7] Tennis can be played on different surfaces (grass, clay, and indoor and outdoor hard courts). The technology is not used in tournaments played on clay, where a ball leaves a mark on the court, meaning its landing location can be verified by the umpire with no additional assistance.

[8] From the initial adoption of the technology (in January 2006) until April 2008, tournaments were allowed discretion on the precise rules, especially in terms of the number of challenges per set each player is allowed. Two tournaments in our data allowed three challenge opportunities for each player per set. In April 2008, the governing bodies of professional tennis unified these rules across tournaments.

number of allowable challenge opportunities, the player could not challenge any more until a new endowment of challenge opportunities was given to him.[9]

3. DATA

We obtained data from the Association of Tennis Professionals (ATP, the governing body of men's professional tennis) on all challenges that were called in 35 tennis tournaments around the world over the years 2006–2008. The data set covers all tournaments in 2006–2008 for which the ATP had challenge information, mostly covering tournaments that are part of the "ATP World Tour Masters 1000." These tournaments are more important than regular tournaments—in terms of their prize money and effect on players' ranking—but not as important as the four "grand slam" events. Although some of these tournaments include doubles matches (in addition to singles) and one tournament includes women's matches (in addition to men's), we restrict attention throughout the article to men's singles matches, which constitute the majority of the challenges observed in the data (2,784 out of 3,343, or 83%).

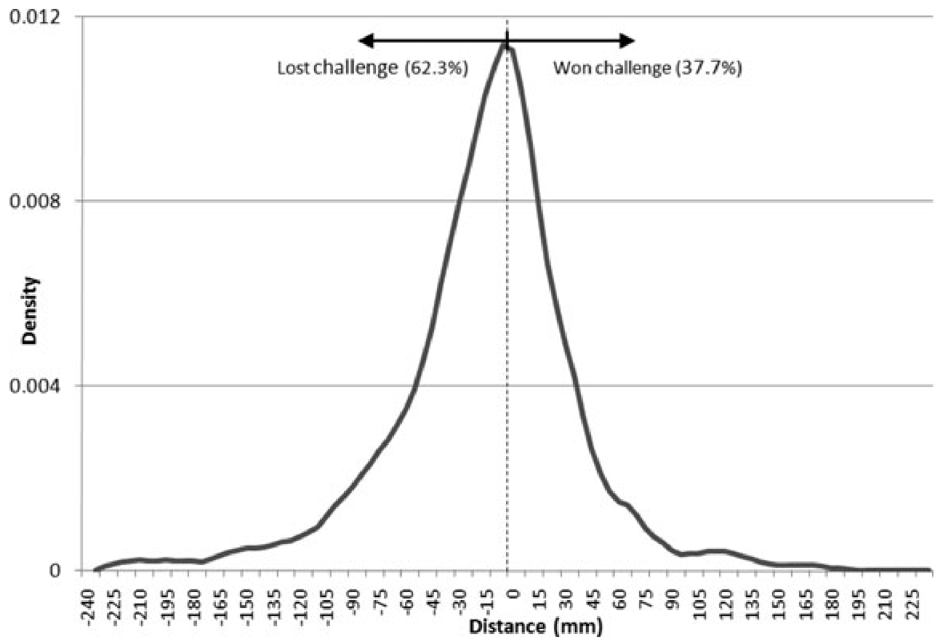
Because challenge statistics are not computerized or uniformly organized, to our knowledge the data we have collected are the largest to date.[10] The raw data come in the form of umpire sheets. Umpires have one sheet per match, on which they note the details of the match (the players, the tournament round, and the final score), including an entry for each challenge that was called during the match. For each challenge, the sheet contains information about the identity of the challenging player, the original umpire call ("in" or "out"), the Hawk-Eye verdict ("in" or "out"), the Hawk-Eye prediction regarding the distance (in mm) between the ball's landing location and the line (and which court boundary line it is). We merged these data with ancillary data (available online) on the match statistics, the tournament characteristics, and players' characteristics (as of June–July 2008).

The final data set contains information about 2,784 challenges made by 179 distinct players in 741 different tennis matches in which at least one challenge was made (out of approximately 1,050 matches played in the tournaments covered by the data). In the 741 matches we have, a total of 106,166 points were played. Of these, 2,784 points (2.6%) were challenged. Our data therefore contain full information on these selected points that were actually challenged. We do not have point-by-point data on points that were not challenged (and, to the best of our knowledge, such data do not exist for the tournaments covered by our data set).

In our analysis below, we restrict attention to our baseline sample, which consists of a subset of 2,008 challenges (associated with 158 distinct players and 479 distinct matches, which cover 71,555 total points). About 20% of the lost observations are due to umpires improperly completing the challenge forms (either obvious errors or, more commonly, simply not completing all fields). The other 80% of the lost observations are due to our inability to infer the identity of the server at that point of the game. The identity of the server is an important input into the model described in the next section, but this information is not part of the umpire sheet. Although we were able to infer the identity of the server in most cases by relying on related information (e.g., the court line associated with a challenge or with other challenges in the same match), we lose several hundred observations that become unusable in the primary part of the analysis. We are not particularly concerned about sample selection issues because a priori it seems unlikely that the above described reason for lost observations are systematically

[9] If the set enters a tiebreak, each player is endowed with an additional challenge opportunity, which is added to the number of challenge opportunities he already possesses. At the beginning of each set, the number of challenges is reset and does not roll over from set to set.

[10] Other papers that use data on challenges in tennis are Mather (2008) and Whitney et al. (2008). See also some mentions of these in the popular press (Chang, 2008; Schwartz, 2009).



NOTES: The figure presents the estimated kernel density of the distribution of the distance of challenged points in the baseline sample (2,008 challenges). This is the distance between the actual location of the ball (as measured by the Hawk-Eye technology) and the nearest line. Here and throughout the article, we use positive distances to correspond to successful challenges and negative ones to correspond to unsuccessful ones. The mean distance is −16.4, with a standard deviation of 59.7. Some percentiles from the distance distribution are as follows: −110 (5th), −80 (10th), −39 (25th), −10 (median), +12 (75th), +38 (90th), +62 (95th).

FIGURE 1

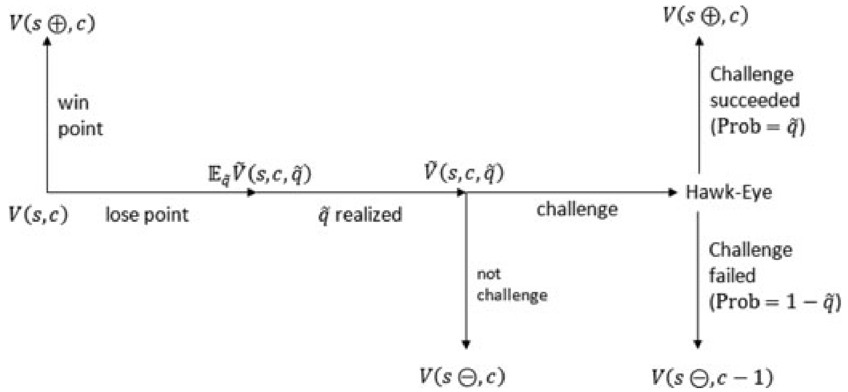
THE DISTRIBUTION OF DISTANCE IN CHALLENGED POINTS

associated with challenge incentives. Indeed, whenever we could, we computed results for the larger sample and confirmed that all the main summary statistics remain virtually the same.[11,12]

The overall challenge success rate in our sample is 38%. That is, 38% of the challenges resulted in overturning the umpire's call. In the baseline sample, 361 of the challenges (18%) were made when the player who challenged the call had a single challenge remaining, and in 219 of these cases (61%, or 11% of the entire baseline sample) the challenge was unsuccessful, leaving the player who challenged without any opportunity to challenge for the rest of the set. Figure 1 presents the distribution of the distance between the ball and the nearest line across all the challenges, with positive numbers representing distances of successful challenges—that is, distances that imply reversal of the original umpire call—and negative numbers representing distances of failed challenges that maintain the original call and result in the player losing a challenge. This convention will be adopted throughout the analysis. Not surprisingly, the vast majority of challenges are over borderline calls: in 92% of the challenges the reported distance is no more than 100 mm (roughly 4 inches) away from the line. The figure also conveys the fact that additional aspects that may be associated with challenges, such as a way to obtain rest or

[11] In particular, Figure 1 and Table 1, which are described below, do not require information about the identity of the server, so we replicate them for the entire sample in Figure A.1 and Table A.1, respectively.

[12] An interesting pattern in our data is that challenges of "out" calls by the umpire have higher success rate than challenges of "in" calls. We find it encouraging that this finding is consistent with the casual observation quoted in Chang (2008), although—unlike the interpretation provided in the press—we think that this finding likely reflects inherent differences in these two scenarios: an "in" call is associated with the player being close to the ball but probably running to get to it, whereas an "out" call is associated with the player who is farther away from the ball, who perhaps has more time to watch the ball closely. None of the results we report are sensitive to the inclusion of a control for whether the umpire call was in or out.



NOTES: The figure presents the timing of the model of whether or not to challenge. See text (Section 4) for details.

FIGURE 2

TIMING AND NOTATION

to change momentum, are unlikely to be a common driver of observed challenges; if they were, one would have expected to see a less symmetric distance distribution and a much fatter left tail.

4. A MODEL

We now provide a simple dynamic model through which we can think about the game of tennis, the objective function of the players, and the decision of whether or not to challenge.

4.1. *Notation and the Challenge Trade-off.* Let us first introduce simple notation. We denote by s the state of the game from player i's perspective. The state s takes the form $\{sets_i - sets_j; games_i - games_j; points_i - points_j; serve_i\}$, which indicates the current score and whether player i is to serve the next point. For example, a tennis match begins with $s = \{0 - 0; 0 - 0; 0 - 0; 0\}$ or $s = \{0 - 0; 0 - 0; 0 - 0; 1\}$, and could end with, say, $s = \{2 - 0; 0 - 0; 0 - 0; 0\}$. We use the operators $s\oplus$ and $s\ominus$ to map a score s into the score that would result if player i won or lost the subsequent point, respectively. For example, if $s = \{1 - 1; 2 - 0; 3 - 1; 1\}$ then $s\oplus$ would be $\{1 - 1; 3 - 0; 0 - 0; 0\}$ and $s\ominus$ would be $\{1 - 1; 2 - 0; 3 - 2; 1\}$.[13]

These operators can be used recursively, so that, for example, $s\oplus\oplus$ implies the resulting score after starting at score s and winning two points, and so on. We let $c \in \{0, 1, 2, 3\}$ denote the number of challenges left in the set for player i, and denote by $V(s, c)$ player i 's value when the score is s and he has c challenges remaining, prior to playing the next point. We will define $V(s, c)$ more precisely later, but for now it is sufficient to naturally assume that $V(s, c)$ is increasing in both s and c.[14]

Figure 2 illustrates the sequence of events and decisions from player i's perspective. At the beginning of a point, player i's value is given by $V(s, c)$. He can either win or lose the point. If he wins the point, he moves on to playing the next point in the match, and his value becomes $V(s\oplus, c)$. If he loses the point, he forms his subjective assessment, which we denote by $\tilde{q}$, of

[13] The tennis convention is to count points within a game as 15, 30, and 40 for the first, second, and third points in the game, respectively. For those who are familiar with tennis score conventions, the examples above translate into $s = \{1 - 1; 2 - 0; 40 - 15; 1\}$, $s\oplus = \{1 - 1; 3 - 0; 0 - 0; 0\}$, and $s\ominus = \{1 - 1; 2 - 0; 40 - 30; 1\}$.

[14] We say that s is greater than s' if either $s = s'\oplus^n$ for some $n > 0$, or $s' = s\ominus^m$ for some $m > 0$. Of course, this only defines a partial order over scores.

the probability that the Hawk-Eye technology would overturn the umpire's call.[15] We denote the interim value at this decision point by $\widetilde{V}(s, c, \widetilde{q})$. At this point, player i makes a decision of whether to challenge the point or not. If he does not, he loses the point and moves on to playing the next point in the match, and his value becomes $V(s\ominus, c)$. If he decides to challenge, with probability $\widetilde{q}$ the challenge is successful, leading to the next point with a value of $V(s\oplus, c)$. With probability $1 - \widetilde{q}$ the challenge fails, in which case player i loses the point and also loses a challenge opportunity, leading to the next point with a value of $V(s\ominus, c - 1)$.

Consider now the decision of whether to challenge or not. As long as $c > 0$, player i is faced with a binary decision. The trade-off is simple. The value from not challenging is given by $V(s\ominus, c)$, and the value from challenging is given by $\widetilde{q}V(s\oplus, c) + (1 - \widetilde{q})V(s\ominus, c - 1)$.[16] It is then optimal to challenge if and only if[17]

$$\widetilde{q}(V(s\oplus, c) - V(s\ominus, c)) > (1 - \widetilde{q})(V(s\ominus, c) - V(s\ominus, c - 1)), \tag{1}$$

or if

$$\widetilde{q} > q^*(s, c) \equiv \frac{V(s\ominus, c) - V(s\ominus, c - 1)}{(V(s\oplus, c) - V(s\ominus, c)) + (V(s\ominus, c) - V(s\ominus, c - 1))}. \tag{2}$$

Thus, the decision whether to challenge or not can be simply characterized by a cutoff strategy $q^*(s, c)$. As long as $\widetilde{q} > q^*(s, c)$, it is optimal for the player to challenge. The cutoff depends on two relatively simple objects. The first is $V(s\oplus, c) - V(s\ominus, c)$. This is the value of reversing the call, which is associated with the relative importance of the particular point within the match. The second is $V(s\ominus, c) - V(s\ominus, c - 1)$. This is the value of keeping an additional challenge, which is associated with the option value of remaining challenges. This option value decreases with the number of challenges left and increases with the number of points that are expected to be subsequently played within the set. Our model below is geared toward quantifying these two objects.

4.2. *A Statistical Model of Tennis.* Before we describe the full model, it is instructive to start with a simpler framework, which allows us to introduce some of the key ideas. This simpler framework is also used later in the qualitative empirical analysis (Section 5).

To be able to quantify $V(s, c)$, we start by making the natural assumption that players' only objective is to win the match. That is, players do not care whether they win or lose decisively, how long or how many sets they play, and so on. We then normalize the value they obtain from winning or losing the match to 1 and 0, respectively. Thus, at any intermediate point in the match, the value functions can be viewed as the probability of winning the match at that point. Therefore, the key objects of interest that would be key in our subsequent analysis, $V(s\ominus, c) - V(s\ominus, c - 1)$ and $V(s\oplus, c) - V(s\ominus, c)$, simply measure the difference in probabilities of winning the match.

To be able to compute $V(s, c)$ we draw heavily on Paserman (2010), who presents a statistical model of tennis, which abstracts from challenge opportunities. The model views the game of tennis as a sequence of statistically independent realizations of points. Indeed, Klaassen and

[15] There are cases in which this assessment is clearly zero, for example, when player i's shot did not cross the net, or when the umpire called the point in player i's favor but his opponent successfully challenged the call. Our notation will capture these cases by $\widetilde{q} = 0$.

[16] This expression assumes that player i wins the point if the umpire's call is overturned. In a small fraction of challenges, this is not the case. If player i challenges a first serve ace, a successful challenge implies that the point is replayed with the opponent serving a second serve. In other cases, player i may challenge an "out" call that caused the opponent to stop playing. Winning the challenge would make the players repeat the point. Because we have incomplete data on these special cases, throughout the rest of the article we abstract from these possibilities and assume that a successful challenge results in the player winning the point.

[17] Although this is inconsequential, we break ties by assuming that players, when indifferent, prefer to not challenge (for example, due to fan base or sportsmanship considerations).

Magnus (2001) use point-by-point data from many tennis matches and show that independence is a good approximation for actual play. We thus assume that the probability of a given player winning a point remains constant within a given match, conditional on whether he serves or receives. In tennis, players alternate their serves from game to game, and it is well known that serves generate a great advantage. It is therefore important to account for this advantage, and define p^s and p^r as the probability of winning a point when serving and when receiving, respectively. It follows that the realization of each point is drawn from a Bernoulli distribution with parameter p^s or p^r, depending on the server's identity at this point of the match.

When abstracting from challenge opportunities, this is a statistical model in which players do not have any strategic role or any room for decision making. It is essentially assumed that players always play their best possible tennis to maximize their probability of winning each point. Conditional on both players playing their best, players distinguish themselves from each other by their probabilities of winning each point, which are assumed to be common knowledge.

Consider now a game between player i and player j. Given our assumptions above, the two probabilities, p_i^s and p_i^r (the corresponding probabilities of player j are then given by $p_j^s = 1 - p_i^r$ and $p_j^r = 1 - p_i^s$) are sufficient to describe a match,[18] and the value function for player i can be written as $V(s, c; p^s, p^r)$. Abstracting from challenges (that is, contrary to the fact, we assume $c = 0$ for now), it is easy to see that $V(s; p^s, p^r)$ can be defined and computed recursively. One can start with the set of terminal scores, in which the match is already won by one of the players, and then compute recursively $V(s; p^s, p^r)$, going backwards.

If $V(s\oplus; p^s, p^r)$ and $V(s\ominus; p^s, p^r)$ are already known, then it is easy to see that

$$V(s; p^s, p^r) = pV(s\oplus; p^s, p^r) + (1-p)V(s\ominus; p^s, p^r), \tag{3}$$

where p is equal to p^s or to p^r, depending on the server's identity at s. A slight complication arises because tennis is not a finite game (to win a game, one needs to win by two points, so at least in principle a game can go on forever), but here the independence assumption is useful.[19]

4.3. *Optimal Challenge Strategy.* We now incorporate into the above statistical model the possibility of challenges. Again, we start from terminal scores and solve the model recursively. To incorporate the challenge opportunity, consider the simple case where $V(s\oplus, c; p^s, p^r)$ and $V(s\ominus, c; p^s, p^r)$ are already known (for all $c \in \{0, 1, 2, 3\}$), as we try to evaluate $V(s, c; p^s, p^r)$. To do this, we return to Equation (2), which defines the optimal decision of whether to challenge or not. Given the recursive nature of the exercise, all the elements of the right hand side of Equation (2) are assumed to be known, implying that $q^*(s, c; p^s, p^r)$, the success probability above which one should challenge, is also known, thus defining the optimal challenge strategy.[20]

To continue with the recursive exercise, we need to obtain the value function $V(s, c; p^s, p^r)$. For the rest of this section, we omit the conditioning on p^s and p^r to simplify notation. We first solve for the interim value function $\widetilde{V}(s, c, \widetilde{q})$ (see Figure 2), which reflects player i's value conditional on the umpire call going against player i and conditional on player i's subjective

[18] We note that the identity of the player who serves first in the match (determined by a coin toss) or the set is inconsequential for our analysis. This is a result of our assumption about the i.i.d. realization of points and the fact that winning a set or a tiebreak requires winning by two games or points, respectively.

[19] In such situations, one can observe that $V(s\oplus\ominus; p^s, p^r) = V(s\ominus\oplus; p^s, p^r) = V(s; p^s, p^r)$, so instead of solving recursively point by point, we can iterate two points at a time. For example, if the game score is $3-3$ and $V(s\oplus\oplus; p^s, p^r)$ and $V(s\ominus\ominus; p^s, p^r)$ are known, then $V(s; p^s, p^r) = (p^2/(p^2 + (1-p)^2))V(s\oplus\oplus; p^s, p^r) + ((1-p)^2/(p^2 + (1-p)^2))V(s\ominus\ominus; p^s, p^r)$, where, again, p is equal to p^s or to p^r, depending on the server's identity in that game. To see this, denote $V(s; p^s, p^r) = x$ and observe that $x = p^2 V(s\oplus\oplus; p^s, p^r) + (1-p)^2 V(s\ominus\ominus; p^s, p^r) + p(1-p)x + (1-p)px$. This expression is similar but slightly different in a tiebreak, when serves alternate every two points within the game.

[20] Again, as in the case without challenges (see footnote 19), a technical complication arises because tennis is not a finite game. We address it numerically in the following way. In each such case we guess a strategy, compute the resulting values, and then calculate the optimal strategy implied by those values. We iterate until convergence. The full algorithm is available online, on our web pages.

probability of overturning the call. Given the optimal policy (and the known objects), this interim value function is given by[21]

$$\widetilde{V}(s,c,\widetilde{q}) = \begin{cases} V(s\ominus,c) & \text{if} \quad \widetilde{q} \leq q^*(s,c) \\ \widetilde{q}V(s\oplus,c) + (1-\widetilde{q})V(s\ominus,c-1) & \text{if} \quad \widetilde{q} > q^*(s,c) \end{cases}. \tag{4}$$

That is, the first case corresponds to the case where $\widetilde{q}$ is low, so player i does not challenge, and the second case corresponds to the case where player i challenges and either succeeds (in which case the call is overturned and player i does not lose a challenge) or fails (in which case he loses the point and also loses one challenge opportunity).

Finally, we can now compute the value function

$$V(s,c) = pV(s\oplus,c) + (1-p)\mathbb{E}_{\widetilde{q}}\widetilde{V}(s,c,\widetilde{q}), \tag{5}$$

where p, as before, is the probability to win the point (equal to p^s or to p^r, depending on the server's identity), and

$$\mathbb{E}_{\widetilde{q}}\widetilde{V}(s,c,\widetilde{q}) = \int_0^1 \widetilde{V}(s,c,\widetilde{q})\,dF(\widetilde{q}), \tag{6}$$

where $F(\widetilde{q})$ is the cumulative distribution function of $\widetilde{q} \in [0,1]$. In words, $F(\widetilde{q})$ can be thought of as the perceptual uncertainty and it defines the distribution of the (unconditional) call-reversal probabilities that each player expects to face. It is natural to assume that $F(\widetilde{q})$ has a large mass point at zero, its lower bound, corresponding to points in which the ball hits the net or was clearly out. We therefore assume that $F(\widetilde{q})$ is a mixture of a continuous distribution (with continuous density) $\widetilde{F}(\widetilde{q})$ over the interval $[0,1]$ (with probability $p_{close-call}$) and a deterministic mass point at zero (with probability $1-p_{close-call}$).

$F(\widetilde{q})$ is an important object, and we make the following assumption about it, which we maintain throughout the analysis.

ASSUMPTION 1. *$F(\widetilde{q})$ is invariant with respect to (s,c).*

We view Assumption 1 as quite natural, although one should recognize that it entails important restrictions. For example, Assumption 1 rules out the possibility that the actual distribution of ball locations in the end of points varies as the match progresses or that the actual distribution is the same but players' perceptual uncertainty changes over the course of the match, perhaps because the players get more tired and their vision is not as sharp.[22] Equipped with Assumption 1, we can now define the optimal challenge probability as

$$\Pr{}^*(challenge \mid s,c) = \Pr(\widetilde{q} > q^*(s,c)) = p_{close-call}[1-\widetilde{F}(q^*(s,c))]. \tag{7}$$

A relationship that will play a crucial role in the analysis below relates the marginal success probability q^* to the expected success probability. This is simply a property of $\widetilde{F}(\widetilde{q})$ and is given

[21] A simplification that we maintain throughout the article is that we do not consider a strategic component. A more complete model may make player i take into account that his chance of winning subsequent points in the set could depend on his opponent's number of remaining challenges. Because challenges are fairly rare, and successful challenges are even less frequent, ignoring this indirect effect is inconsequential. We verify this point in the context of our results.

[22] We also note that this or some other similar assumption is necessary for any type of empirical analysis, because if $F(\widetilde{q})$ is allowed to flexibly vary with (s,c), one can interpret any pattern in the data as driven by this variation.

by

$$\Pr(success \mid q^*) \equiv E(\widetilde{q} \mid \widetilde{q} > q^*) = \frac{\int_{q^*}^{1} \widetilde{q} d\widetilde{F}(\widetilde{q})}{1 - \widetilde{F}(q^*)}, \tag{8}$$

where, because $\widetilde{q} \in [0, 1]$, it also implies that $\lim_{x \to 1} \Pr(success \mid q^* = x) = 1$ and $\Pr(success \mid q^* = 0) = E(\widetilde{q} \mid close - call) = E_{\widetilde{F}(\cdot)}(\widetilde{q})$. Combining Equation (8) with the optimal challenge strategy in Equation (7) gives rise to the challenge success rate that is implied by optimal behavior:

$$\Pr{}^*(success \mid s, c) = \Pr(success \mid q^*(s, c)). \tag{9}$$

In other words, the challenge success rate implied by optimal behavior at a given state in the match depends on the optimal cutoff for this state and is given by the expectation over the truncation of the distribution $\widetilde{F}(\widetilde{q})$, where the truncation point is given by the cutoff policy $q^*(s, c)$. It is this last equation, Equation (9), that we can take to the data, to which we now turn.

5. QUALITATIVE ANALYSIS: IS OBSERVED BEHAVIOR CONSISTENT WITH OPTIMAL BEHAVIOR?

5.1. *Basic Approach.* If players follow optimal decision rules in their challenging behavior, we expect them to have a greater propensity to challenge in points that are more important and when the option value of challenging is lower. However, we do not directly observe the propensity to challenge. We only have information on points that were actually challenged, creating a selection, or a missing data, problem. Instead, we use an alternative approach by examining the success rate of a challenge.

Specifically, if we observe that certain variables are associated with *greater* success rate, we infer that this must mean that these same variables are associated with *lower* propensity to challenge. This analysis is straightforward given our data, and it incorporates the fact that challenges are endogenous. This is a direct inference that relies on Assumption 1 and on the fact that expected success rate is a monotone increasing function of the cutoff q^*, the marginal success rate (Equation (8)). Intuitively, we expect that in less important points and when the option value of challenging is higher, players would be more conservative in their use of challenges and will use them only when they are sufficiently certain of overturning the point, thereby increasing their success rates (conditional on a challenge).

Consistent with our empirical approach, Whitney et al. (2008) solve the selection problem by focusing on a small set of points (4,457 points in 57 matches) and watching all of them on television. They report that although 94 of those points (2.1%) were "close calls" (as defined by the television presenting a close-up replay), only 85 of them were challenged, implying that 9 out of 94 (9.6%) close calls are not challenged, arguably due to dynamic considerations.[23]

5.2. *Preliminary Evidence.* We start with Table 1, which provides suggestive evidence on the strategic use of challenges. Table 1 reports challenge outcomes from three pairs of scenarios that we classified a priori. Each pair is such that the first scenario represents a point in the match that is intuitively less important than the point represented by the second scenario. The first pair considers points played late in a set (defined by a point at which at least one of the players has won more than three games) as more important than points played earlier in the set. The second pair considers points in a close game, when the game score is 40–40 (or "deuce," or 3–3 using our scoring notation) as more important than points played in a game in which one of the

[23] Reassuringly, both the probability of a call being challenged and the success rate of challenges in the data covered by Whitney et al. (2008) are in the same ballpark as the corresponding statistics in our data.

TABLE 1
PRELIMINARY EVIDENCE: COMPARISON OF MEANS

	Observation (1)	% Won (2)	Difference [p-value] (3)	Mean Distance[a] (4)	Difference [p-value] (5)
Early in the set[b]	1,007	40.5	5.5 [0.010]	−12.5	7.9 [0.003]
Late in the set[b]	1,001	34.9		−20.4	
Game score is 40–0 or 0–40[c]	55	56.4	23.0 [0.003]	0.1	20.2 [0.032]
Game score is 40–40[c]	147	33.3		−20.1	
Regular games	1,842	38.7	12.2 [0.002]	−14.7	21.5 [<0.001]
Tiebreaks	166	26.5		−36.1	

NOTES: The table reports challenge success rates and mean (sign adjusted) distance for three pairs of scenarios. In each pair, the top row represents a set of challenges that occurred in points that a priori seem less important, and the bottom row represents a set of points that a priori seem more important. For each pair, columns (3) and (5) report the differences in success rates and mean distance, respectively, and the p-value (associated with the null that the two are equal) in squared brackets.
[a]As in the rest of the article, the distance (of the ball from the line) is measured in millimeters and is "sign adjusted" to be negative for failed challenges and positive for successful ones. See Figure 1 for more details.
[b]We define points to be "early in the set" if they occur before either player won more than three games in the set. All other points are defined as "late in the set."
[c]In our scoring notation that we use throughout the article, game scores of 40–0, 0–40, and 40–40 are equivalent to game scores of 3–0, 0–3, and 3–3, respectively.

players has a strong lead of 40–0 (or 3–0 using our notation). Finally, the third pair considers points played in a tiebreak (the decisive game of a set that is tied at 6–6) as more important than other points.

If players challenged every time they saw a close call regardless of the importance of the point, Assumption 1 would imply that success rates should have been statistically indistinguishable within (and across) each pair. Table 1, however, shows a clear pattern. In all pairs, success rates (and the corresponding average distances) are higher in the first (less important) scenario than in the second (more important) scenario. All these differences are quite large and statistically significant. This pattern is consistent with the hypothesis that players use challenges strategically. As discussed earlier, although we do not have direct information about challenge propensity, higher success rates in certain points imply more conservative use of challenges and lower propensity to challenge in these points.

The second pair in Table 1 (40–40 vs. 40–0) is particularly useful, as it is more likely to isolate the effect of importance. The other two pairs we compare are more difficult to interpret because they are associated with cases in which the more important points also occur later in the set, thus confounding the effect of greater importance with the effect, which goes in the same direction, of lower option value of challenging. Much of our analysis in the rest of this section attempts to separate these two effects in a more systematic way.

5.3. *Empirical Strategy.* A possible concern with the results presented in Table 1 is that they rely on an a priori and likely imperfect classification of points as more or less important. For example, a 5–0 score probably makes points less important than a 3–3 score, despite the fact that the former is played later in the set. Thus, it would be desirable to obtain a cleaner and more neutral metric by which to measure the importance of points.

Fortunately, the challenge trade-off described in Section 4 provides useful guidance regarding such a metric. Following Equation (2), there are two objects that should affect the propensity to challenge. The first is the *importance* of the point as measured by $imp \equiv V(s\oplus, c) - V(s\ominus, c)$. The second is the *option value* associated with keeping an additional challenge, as measured by $opt \equiv V(s\ominus, c) - V(s\ominus, c-1)$. Whereas in the next section we make enough assumptions that allow us to compute these theoretical objects explicitly, in this section we use one variable to proxy for *imp* and two variables to proxy for *opt*, allowing us to test the qualitative predictions of the model.

The basic idea is similar to the one that motivated our interpretation of Table 1. Equation (2) implies that the cutoff probability $q^*(s, c)$ is increasing in the option value associated with the point and is decreasing in the importance of the point. Using Equation (7), this implies that the probability of challenging is inversely related to $q^*(s, c)$, and is therefore predicted to decrease in the option value and increase in the importance of the point.

With complete point-by-point data, one could observe Pr(*challenge*) and test these predictions directly. However, we only have data on points in which a player challenged the call. The key observation is to notice that instead of testing directly the prediction implied by Equation (7), we test the predictions implied by Equation (9). That is, we regress the challenge success rates on proxy variables that capture the importance of a point and the option value associated with it, and we test whether, as our model predicts, the expected success rate Pr(*success* | *challenge*) is increasing in the option value, *opt*, and decreasing in the importance of the point, *imp*.

5.4. *Variable Construction.* To construct the key proxy variable for the importance of the point, we rely on the statistical model of tennis described in Section 4.2. This model is closer to the more complete model we take to the data in the next section, but because it abstracts from the opportunity to challenge it can be used separately to consider the importance of a point. The statistical model allows us to simulate a complete game of tennis and to compute each player's probability of winning the entire match, denoted by $\widehat{V}(s; p^s, p^r)$, conditional on being at each possible point of the match. Recall from Equation (2) that $imp \equiv V(s\oplus, c) - V(s\ominus, c)$, so we simply approximate it by $100 \times \left[\widehat{V}(s\oplus; p^s, p^r) - \widehat{V}(s\ominus; p^s, p^r)\right]$ (we multiply by 100 for ease of presentation). That is, importance is measured as the difference in the probability of winning the match if the point is won relative to if it is lost. Two nice features of this proxy are that it is symmetric for both players and its units have a natural interpretation.

The variable varies within matches, depending on the score, and across matches, depending on the probabilities of winning a point, p^s and p^r. For example, if p^s and p^r are both very close to one (or to zero) then one player is very likely to win the match, and no specific point in the match is particularly important. To obtain p^s and p^r, we simply use the statistics from each match and define p^s and p^r as the realized fraction of points a player won out of the total number of points played on his own or his opponent's serve respectively.[24,25]

To proxy for the option value associated with a challenge, we use two variables. The first relies on the same statistical model used to construct the importance variable. We simply use the simulated game of tennis, with the match-specific probabilities p^s and p^r, to compute the expected number of remaining points left in the set. The more points that are expected to be played in the remainder of the set, the more likely it is that challenge opportunities will arise, making it more important to not "waste" challenges. Of course, "wasting" a challenge is likely to be more costly when a player has a single challenge remaining, so we use the number of challenges left (in the set) as an additional proxy for the option value.

Table 2 reports some summary statistics for the resultant variables. Across the entire sample (top row), reversing the umpire call accounts, on average, for a 5% change in the probability of winning the match. This is quite high, although one should recognize that, as we will show later, important points are more likely to lead to challenging (and thus to appear in our data set), so a random point in a match is not as important. The average importance measure of 5% hides

[24] Because the statistical model assumes that the realizations of points are independent, the realized fraction of points won is the maximum likelihood estimate for a binomial probability. Because every match in our data had at least 12 (and often many more) points played on each player's serve and because win probabilities are never too close to 0 or 1, this estimate is quite precise. Of course, an implicit assumption throughout is that players know these probabilities.

[25] This is in slight contrast to Paserman (2010), who uses the same proxy variable to capture the importance of points. Unlike us, he runs a predictive model of p^s and p^r, which uses information from other matches as well. Given that each match may have its idiosyncracies, the "true" p^s and p^r are presumably somewhere in between these predictive probabilities and the realized ones.

TABLE 2
SUMMARY STATISTICS OF THE CONSTRUCTED VARIABLES

		Importance (Difference in Match Win Probability)				Option Value (Expected Number of Points Left in Set)			
	Observation	Mean	Standard Deviation	10th Percentile	90th Percentile	Mean	Standard Deviation	10th Percentile	90th Percentile
Entire baseline sample	2,008	5.17	6.57	0.07	13.02	28.27	17.30	5.89	52.62
1st and 2nd sets	1,671	3.88	4.74	0.05	9.80	28.04	17.44	5.54	52.88
3rd set	337	11.55	9.86	1.43	24.26	29.42	16.55	8.02	51.96
Early in the set[a]	1,007	4.22	5.01	0.06	10.68	42.70	10.78	29.26	57.44
Late in the set[a]	1,001	6.11	7.72	0.11	15.69	13.75	7.93	3.60	24.50
Game score is 40–0 or 0–40[b]	55	3.04	4.96	0.00	10.77	26.53	16.68	2.04	48.09
Game score is 40–40[b]	147	5.79	5.14	0.11	13.65	30.48	17.50	7.88	53.15
Regular games	1,842	4.38	5.50	0.06	11.10	30.31	16.59	8.78	53.18
Tiebreak games	166	13.86	10.22	3.55	26.48	5.68	3.16	1.80	10.15
Multiple challenges remaining	1,647	4.68	5.85	0.06	12.01	30.51	17.21	7.62	53.92
One challenge remaining	361	7.38	8.86	0.22	19.12	18.07	13.66	3.94	37.20

NOTES: The table reports summary statistics for the two constructed variables that are used in the regressions of Table 3. The importance measure proxies for $imp \equiv V(s\oplus, c) - V(s\ominus, c)$, that is, for the difference in the match win probability (in percentage points) between winning and losing the challenged point. The option value captures the expected number of points that are will be played in the remainder of the set. See the text for more details. The top row presents statistics for the entire baseline sample, and the remaining rows present similar statistics for the different subsamples, including those featured in Table 1.

[a]We define points to be "early in the set" if they occur before either player won more than three games in the set. All other points are defined as "late in the set."

[b]In the scoring notation that we use throughout the article, game scores of 40–0, 0–40, and 40–40 are equivalent to game scores of 3–0, 0–3, and 3–3, respectively.

TABLE 3
QUALITATIVE EVIDENCE: REGRESSION ANALYSIS

	100 if Challenge Won, 0 if Lost			Adjusted Distance (in mm)[a]		
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
Importance						
Difference in match win probabilities[b]	−0.630***	−0.531***	−0.550**	−0.748***	−0.688**	−0.710**
	(0.163)	(0.195)	(0.240)	(0.253)	(0.282)	(0.313)
Option value						
Challenges remaining in the set	−4.729*	−4.819*	−14.813***	0.054	−0.170	−6.134
	(2.434)	(2.465)	(2.997)	(3.142)	(3.135)	(3.828)
Expected number of points left in set[b]	0.177***	0.141*	0.075	0.247***	0.166*	0.144
	(0.060)	(0.079)	(0.125)	(0.072)	(0.095)	(0.137)
Additional controls[c]	no	yes	yes	no	yes	yes
Match fixed-effects (479 unique matches)	no	no	yes	no	no	yes
Number of observations	2,008	1,973	1,973	2,008	1,973	1,973
R-squared	0.013	0.024	0.032	0.014	0.037	0.039

NOTES: The table reports regression results. The unit of observation is a challenge. The challenge outcome is regressed on our proxy variables for the importance and option value associated with the challenged point. Standard errors (clustered at the match level) are reported in parentheses. Stars reflect the usual p-values: *** implies <0.01, ** implies <0.05, and * implies <0.1.
[a]See Figure 1 and Table 1 for more details about the average and variation in the distance variable. Distance is "sign adjusted" to be negative for failed challenges and positive for successful ones.
[b]See the text for details about the construction of the key regressors, and see Table 2 for their summary statistics.
[c]Additional controls include the log of challenger's and opponent's ATP rank, an indicator for whether the challenger was serving the current point, an indicator for whether the player stopped playing the point before challenging (as opposed to challenged after the point naturally ended), the elapsed time since the start of the match, an indicator for whether the call was "in" or "out," and two dummy variables for which set the point occurred in. In all the specifications that include controls we lose 35 observations due to missing data (mainly due to the elapsed time variable).

significant heterogeneity, with challenges in our data made in points that range from essentially unimportant points (measured at zero) to 10% of the challenged points that have importance measure greater than 13%. The other rows of Table 2 attempt to "validate" the importance measure by presenting similar statistics for pairs of scenarios, including those analyzed in Table 1. Indeed, scenarios that seem a priori more important obtain higher importance measures. For example, the average challenged point in a tiebreak game is associated with importance which is three times greater than other challenged points, and the average challenged point in a third set is three times more important than the average challenged point in an earlier set.

Table 2 also presents similar statistics for the expected number of points left in the set that is one of our proxy variables for the option value. Across the sample, the average expected number of points left to be played in the set is 28, with a range from 5 to 50. To have some rough idea for the importance of the option value, a simple back-of-the-envelope may be useful. If the probability of a "close call" is 2.1% (per our earlier discussion of Whitney et al., 2008) and a player has only a single challenge remaining, then a lost challenge would mean that at the average challenged point (28.3 points left to be played) there is a 45% chance that there would later be at least one "close call" that the player would not be able to challenge. This chance is only 12% when the expected number of points is at its 10th percentile (5.9), but is 67% when it is in its 90th percentile. The corresponding effect is much smaller when there is more than a single challenge left.

5.5. *Results.* Table 3 reports the results of our qualitative analysis, which is based on regressing the challenge outcome on the proxy variables for the importance of the point and the option value associated with remaining challenges. We also include additional controls and match fixed effects in some of the specifications. The dependent variable in columns (1)–(3) is a variable that is equal to 100 if the challenge is successful and 0 otherwise. Columns (4)–(6) use the (adjusted) distance variable in which greater values mean "greater" success.

The results are consistent with the predictions of optimal challenge behavior, as derived in Section 4. The results are qualitatively stable across specifications and, especially when the dependent variable is the success dummy, are almost always statistically significant at conventional levels. The results suggest that success rates are significantly lower for more important points. The coefficient estimates imply that one standard deviation increase in the importance of the point reduces the success rate by 3–4%, and moving from the 10th to the 90th percentile of the importance distribution (Table 2) reduces success rates by 6–8%. This implies that players' propensity to challenge is significantly greater when the point is more important.

The results regarding option value are also consistent with the theory. Success rates are 5–15% lower when a player has two challenges remaining rather than one, and success rates increase by about 1% for every 7–12 additional points that are expected to be played in the set. Again, the inference is that players are much less likely to use their opportunity to challenge when they have a single challenge remaining in the set, and they also tend to increase their propensity to challenge as the set gets closer to its end (when "saving" challenges for future points is less valuable).[26]

5.6. *Additional Cost of Unsuccessful Challenges.* Although the results in this section are highly consistent with the qualitative predictions of optimal behavior (as prescribed by the model), we note that we also find some suggestive evidence for deviations from it. Specifically, inspecting Equation (2), one can observe that the optimal challenge strategy is invariant with respect to rescaling of $V(\cdot)$. Because the number of remaining challenges resets every set, this observation implies that—within a match (where p_s and p_r are held fixed)—optimal challenge behavior should be the same across sets. Yet, regressing challenge rate on indicator variables for second and third sets, as well as match fixed effects (as in column (3) of Table 3), reveals that players seem to challenge more often in later sets (although the differences are not statistically significant at conventional levels).[27] Similarly, optimal strategy responds to the importance of the point $imp \equiv V(s\oplus, c) - V(s\ominus, c)$, but this importance is largely symmetric for losers and winners and primarily captures close versus not-so-close matches. Yet, when we include an indicator for whether the player who challenges is losing or winning the match, we find that players who are winning the match tend to have greater success rates, presumably because their propensity to challenge is lower.

One way to rationalize these patterns is that players have some additional costs of challenging (beyond the option value associated with it) that are absolute in nature and invariant to the actual stakes. This would make them, all else equal, more likely to challenge when the stakes are greater (e.g., in the third set) or less likely to challenge when challenging a point may interfere with a perception of sportsmanship. Indeed, players are sometimes criticized for making too many unsuccessful challenges.[28] If players care about such criticism, this may lead them to deviate from optimal behavior and restrain them from challenging when the stakes are small. Moreover, much of the popular wisdom (as reflected in the press) views a high challenge success rate as a successful use of challenges, thus making failed challenges look bad. We should note that our analysis (and findings) suggest that, contrary to this popular wisdom, a player with greater success rate does not indicate better utilization of the Hawk-Eye technology. Because players recognize that what they see is a noisy signal of where the ball really landed, the decision

[26] It is important to note that points that are closer to the end of a set are more likely to be important and, at the same time, are associated with lower opportunity cost of challenging. These two forces affect the propensity to challenge in opposite directions. It is therefore important to recognize that the regression holds constant the importance of a point.

[27] The base category is the first set of the match. The estimated coefficient (standard error) on the second and third set dummy variables are −4.02 (2.75) and −6.67 (4.01), respectively.

[28] For example, Lamont (2009) quotes a BBC analyst: "[Roger Federer] will make [challenges] when the ball is three or four inches in; it can be embarrassing." Later in the article the same analyst says about a particular challenge made by Marat Safin in his game against Roger Federer in Wimbledon 2008: "Federer, the umpire and 15,000 fans sniggered when Safin challenged a serve in last year's semis: the ball was almost 40cm in.... We all agree this was The Worst Challenge Of All Time."

to challenge takes into account the expected cost and benefit of challenging. A player who has higher success rates may simply be under using the Hawk-Eye technology by only challenging more obvious missed calls by the umpire.[29]

6. QUANTITATIVE ANALYSIS: HOW CLOSE IS OBSERVED BEHAVIOR TO OPTIMAL BEHAVIOR?

6.1. *Description of The Exercise.* Although the previous section presented evidence that observed behavior is consistent with the qualitative predictions implied by optimal behavior, it is natural to ask to what extent the observed behavior *quantitatively* matches optimal challenge strategy. For example, although players may be more likely to challenge in important points than in other points, perhaps they do this too much or too little.

The model we developed in Section 4 provides a natural framework through which we can investigate this question. This quantification exercise, which goes beyond testing for whether observed behavior is consistent with optimal behavior, requires us to extend the empirical approach used in the last section and to make an additional explicit assumption on the perceptual uncertainty, $F(\widetilde{q})$, which is the distribution of probability of succeeding in a challenge from which players draw when they lose a point.

One approach to model perceptual uncertainty is to write down a model of the players' and umpire's actual vision (Mather, 2008, follows this approach). For example, if the actual location of the ball is given by $\theta \sim N\left(0, \sigma_\theta^2\right)$, one could assume that the umpire's (unbiased) call is based on seeing $r = \theta + \varepsilon$, and the player's information is based on seeing $x = \theta + u$, where $\varepsilon \sim N\left(0, \sigma_\varepsilon^2\right)$, $u \sim N\left(0, \sigma_u^2\right)$, and $cov(\varepsilon, u) = 0$. Such a model would define $\widetilde{q} = \Pr(\theta > 0 \,|\, r < 0, x)$, and $F(\widetilde{q})$ could then be derived from the joint distribution of θ, r, and x.

But unless one's main interest is in perceptual uncertainty, it is more efficient to make assumptions directly about the (statistical) object that enters the model, $F(\widetilde{q})$, rather than about the deeper (physiological) primitives that give rise to it.[30] Moreover, because $F(\widetilde{q})$ enters our model through the relationship between marginal and average success rates to which it gives rise (see Equation (8)), we make our parametric assumptions directly on this relationship, rather than on $F(\widetilde{q})$. As will become clear below, because this relationship is much closer to empirical objects of interest, this strategy makes it easier to evaluate the model's implications and to use the data as a guide for plausible parameter values.

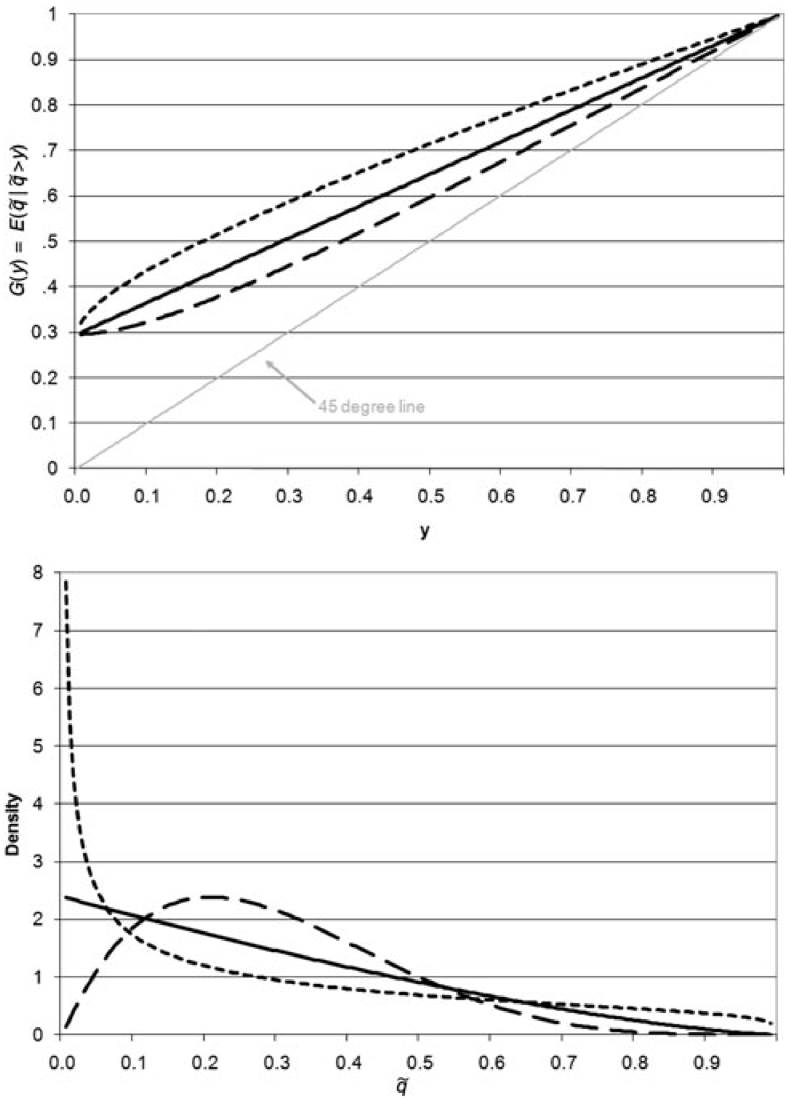
Specifically, we define the function G: $[0, 1) \to [0, 1)$, such that

$$G(y) = E(\widetilde{q} \,|\, \widetilde{q} > y) = \frac{1}{1 - \widetilde{F}(y)} \int_y^1 y d\widetilde{F}(y). \tag{10}$$

That is, $G(y)$ is the expected value of $\widetilde{q}$ conditional on $\widetilde{q}$ being greater than y. In the context of our exercise, if a player uses a cutoff strategy y and challenges every time that $\widetilde{q} > y$, then his expected success rate is given by $G(y)$. By inspecting Equation (8), it is easy to verify that there is a one-to-one mapping from $\widetilde{F}(\cdot)$ to $G(\cdot)$. One attractive feature of this transformation is that it is closer to what one might hope to observe in the data. For example, we know that $G(\cdot)$ is monotonically increasing, that $G(y) > y$ for all y, that $\lim_{x\to 1} G(x) = 1$, and the value of $G(0)$ can be pinned down by (naturally) assuming that players always challenge the last point of a set, if it is a close call (that is, if $\widetilde{q} > 0$). Thus, all we need is to make assumptions about the curvature of $G(\cdot)$ (and the probability of "close calls," $p_{close-call}$). Figure 3 illustrates this point

[29] Indeed, Roger Federer may be ahead of the pack on this dimension of tennis as well. Quoting Lamont (2009) again: "Federer is very hard to work out in regard to his challenges. . . . I think in his heart of hearts, he believes the Hawk-Eye system is poor—and uses it, begrudgingly, because he thinks it's fallible. In other words, even if the line call is right it might gift him the point by mistake."

[30] To see this, consider the model of perceptual uncertainty that we sketched in the preceding paragrpah. The model would require us to make assumptions about two distributions (from which ε and u are drawn), whereas, in the context of our exercise, all we would care about is the single distribution $F(\widetilde{q})$ to which they give rise.



NOTES: The figure illustrates the key parametric assumption that is required to implement our quantitative exercise. The assumption we make is about the function $G(y) = E(\widetilde{q} \mid \widetilde{q} > y)$ (see Equation (10)). Using the data, we calibrate the value of $G(0)$ (which in the figure we assume to be 0.295) and we know that $\lim_{x\to 1} G(x) = 1$. We then make an assumption about the curvature of $G(y)$; the top panel illustrates three such assumptions: linear (our baseline specification), convex, and concave. Because there is a unique mapping between $G(y)$ and $\widetilde{F}(\widetilde{q})$, an assumption about $G(y)$ can be mapped into a specific $\widetilde{F}(\widetilde{q})$. The bottom panel therefore illustrates the different densities for $\widetilde{q}$ that are implied by the different assumed curvatures in the top panel. The solid line is implied by the solid (linear) $G(y)$, the dashed line is implied by the dashed (convex) $G(y)$, and the dotted line is implied by the dotted (concave) $G(y)$.

FIGURE 3

ILLUSTRATING THE KEY PARAMETRIC ASSUMPTION ABOUT $G(Y)$

by plotting three different assumptions about the curvature of $G(\cdot)$ and their implications for the shape of $\widetilde{F}(\cdot)$.[31]

Equipped with this additional assumption, we can now quantitatively evaluate the optimality of observed challenge behavior. Specifically, we solve the model backwards and derive the

[31] To generate the figure we assume that $\widetilde{F}(\cdot)$ follows a $Beta(\alpha, \beta)$ distribution, which allows us to incorporate various degrees of curvature into $G(\cdot)$ and at the same time provides a closed-form mapping from $\widetilde{F}(\cdot)$ to $G(\cdot)$ and vice versa.

optimal challenge strategy given by $q^*(s, c; p^s, p^r)$. We then compare this optimal challenge behavior with the behavior observed.

6.2. *Implementation and Results.* In our baseline exercise, we calibrate $G(0) = 0.295$. We arrive at this number by focusing on the 37 (challenged) points in our data that are not associated with an option value, and observing that 11 (29.7%) of these challenges were successful. That is, these are points in the end of sets, in which the player who challenged would have lost the set otherwise. The assumption is that in such cases players would challenge every close call, as optimal behavior would prescribe. Of course, this assumption is true only if players actually follow optimal behavior (a question that we address in this exercise), but it seems natural to assume that optimal behavior in this particular set of points is sufficiently straightforward.

In addition to calibrating the value of $G(0)$, we also assume that $G(\cdot)$ is linear, and—based on our earlier discussion of Whitney et al. (2008)—that the probability of a "close call," $p_{close-call}$, is 2%, thus implying a mass point of 98% at $\widetilde{q} = 0$. Below we show that the key results are not too sensitive to reasonable modifications of these assumptions.

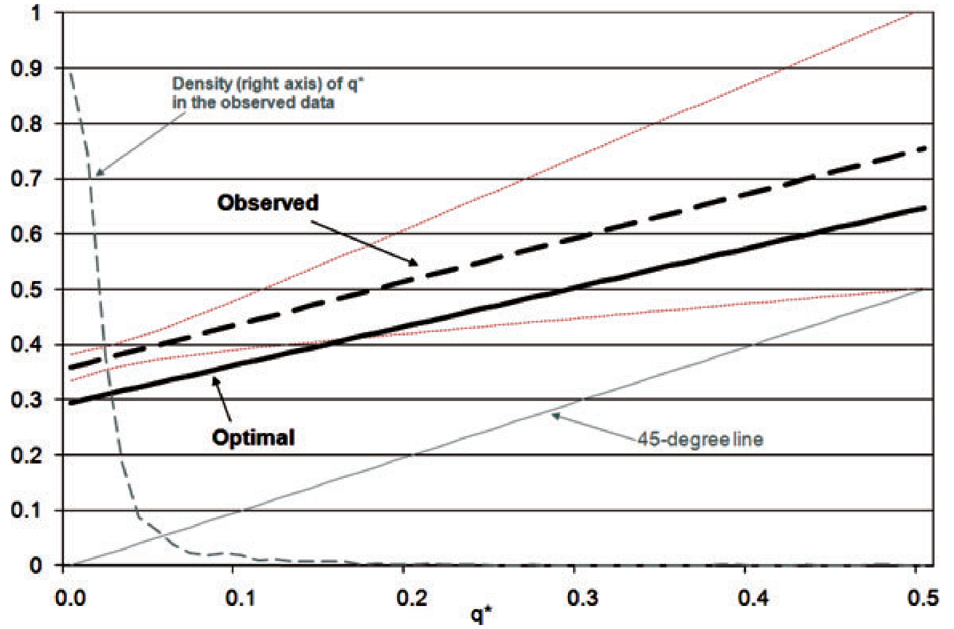
Equipped with the function $G(\cdot)$ and a value for $p_{close-call}$, we then use the distribution $F(\cdot)$ (that is implied by $G(\cdot)$ and $p_{close-call}$) to compute the optimal challenge policy in each match in our data. We are then able to derive the optimal challenge strategy $q^*(s, c; p^s, p^r)$ for every value of $(s, c; p^s, p^r)$,[32] and assign an optimal challenge strategy, q^*, to each data point in our baseline sample. To do this, we use the same assumptions as in Section 5. In particular, we obtain p^s and p^r by using the statistics of each match, and define p^s and p^r as the realized fraction of points a player won out of the total number of points played on his own or his opponent's serve, respectively. Once we obtain the optimal challenge strategy, we can again use the assumed $G(\cdot)$ function in order to compare the observed challenge behavior with the challenge success rate at each point as implied by optimal behavior.

Figure 4 reports the main results of this exercise. The (optimal) marginal probability of challenging, q^*, is on the horizontal axis. The expected success rate is on the vertical axis. The thick solid line reports the expected success rate implied by optimal behavior, which is, by construction, equal to the function $G(q^*)$. The dashed line reports the linear projection of the observed success rate, which is estimated by regressing a dummy variable for whether a challenge is successful on the optimal cutoff challenge strategy, q^*, associated with each point in our baseline sample. One can clearly see that the observed rate is higher than the optimal rate, implying that at least on average players are too conservative and tend to challenge too little.

It is important to recognize that whereas the "optimal" line is driven by our assumption regarding $G(\cdot)$, the same assumption also drives the implied q^* associated with each point in our data. Therefore, it is not necessarily the case that one could narrow the gap between the optimal and observed behavior by simply changing the assumption about the function $G(\cdot)$. For example, if we assumed that $G(\cdot)$ is higher—that is, by calibrating a higher value for $G(0)$—the implied optimal strategy would change too, shifting both curves in the same direction. Nevertheless, we report below several robustness checks that verify that the key results remain similar even after modifying some of the assumptions regarding $G(\cdot)$. We also investigated whether observed challenge behavior is closer to optimal behavior at different points or for different players, but could not find any systematic patterns (possibly due to lack of statistical power).

Our final exercise aims to use the model to quantify the value of the opportunity to challenge and the extent to which observed behavior is different from optimal behavior. To do so, we use a 25-point grid of p^s and p^r: For each point on this grid, we compute the value function at the beginning of the match under four scenarios. The first, $V_{no\ challenge}(p^s, p^r)$, is when no challenges are used (this replicates our construction of the proxy variable in Section 5 for the case of a challenge in the first point of the match). The second, $V_{optimal}(p^s, p^r)$, is when we use the optimal challenge strategy. The third, $V_{observed}(p^s, p^r)$, is when we use the observed challenge

[32] That is, we compute $q^*(s, c; p^s, p^r)$ for every possible value of s and c, and for a 25-point grid of p^s and p^r. To obtain $q^*(s, c; p^s, p^r)$ for a given pair of p^s and p^r we use the nearest values on the grid and interpolate.



NOTES: The figure reports the key results from our quantification exercise in Section 6. The horizontal axis represents q^*, the marginal probability above which one should optimally challenge a given point. The solid line (labeled as "optimal") represents the average success rate implied by our assumption about $G(q^*)$ (which is assumed to be linear in our baseline specification). The dashed line (labeled "observed") represents the success rate implied by observed behavior. To construct the latter, we use the model to compute the optimal challenge strategy (defined by q^*) at each point of our baseline sample and use a linear regression to predict the observed success rate as a function of the value of q^* associated with every challenged point. The dotted lines represent a 95% confidence interval around this prediction; the interval is much wider for high values of q^* as our results suggest that most of the challenged points observed in the data are associated with a relatively small value of q^* (as can be seen by the kernel density plot [dashed gray] in the figure).

FIGURE 4

ACTUAL VERSUS OPTIMAL BEHAVIOR

strategy (as implied by Figure 4).[33] The fourth, $V_{random}(p^s, p^r)$, assumes that players' overall propensity to challenge is the same as in the observed behavior, but that they are not more likely to challenge important points over less important ones. One way to think about the latter scenario is that players have some implicit costs (psychic costs or perhaps reluctance to annoy the fans; see our related discussion in the end of Section 5) that make them reluctant to use the challenge opportunity "too often" (and fail "too much"). By construction, it is always the case that $V_{no\ challenge}(p^s, p^r) < V_{random}(p^s, p^r) < V_{observed}(p^s, p^r) < V_{optimal}(p^s, p^r)$. The "random" strategy then allows us to assess how much players benefit, conditional on the same level of usage, from a "strategic" use of these challenge opportunities. By strategic we mean that instead of randomly challenging close calls, players are more likely to challenge important points (as shown to be the case in Section 5).

Table 4 reports the results. We report the results for each value of (p^s, p^r) on the grid, and then average the results using weights that are proportional to the frequency of challenges in the baseline sample, for which the pair (p^s, p^r) is the closest on the grid. Throughout the discussion, recall that the value function is the ex ante probability of winning a match. Depending on p^s and

[33] To get the observed policy, we use the inverse of the function $G(\cdot)$ to convert the observed average success rate to the underlying challenge stategy or the marginal success rate that would give rise to it. To illustrate this graphically, using Figure 4, the difference between the optimal strategy q and the observed strategy q' can be found by starting at q, drawing a vertical line up to the "observed" (dashed) line, then drawing a horizontal line to the "optimal" (solid) line, and then drawing a vertical line down to the x-axis.

TABLE 4
THE VALUE FROM OPTIMAL AND OBSERVED CHALLENGE BEHAVIOR

(p_s, p_r) (1)	Number of Points[a] (2)	$V_{no\ challenge}$ (3)	$V_{optimal}$ (4)	$V_{observed}$ (5)	$\frac{V_{observed}-V_{no\ challange}}{V_{optimal}-V_{no\ challange}}$ (6)	$\frac{V_{random}-V_{no\ challange}}{V_{optimal}-V_{no\ challange}}$ (7)
(0.1, 0.1)	0	0.0000	0.0000	0.0000	–	–
(0.1, 0.3)	4	0.0000	0.0000	0.0000	–	–
(0.1, 0.5)	0	0.0000	0.0000	0.0000	–	–
(0.1, 0.7)	0	0.0173	0.0208	0.0207	96.6%	72.9%
(0.1, 0.9)	0	0.5000	0.5289	0.5280	96.9%	74.5%
(0.3, 0.1)	0	0.0000	0.0000	0.0000	–	–
(0.3, 0.3)	0	0.0000	0.0000	0.0000	–	–
(0.3, 0.5)	0	0.0053	0.0067	0.0067	96.6%	72.7%
(0.3, 0.7)	0	0.5000	0.5270	0.5261	96.9%	74.6%
(0.3, 0.9)	0	0.9827	0.9848	0.9847	97.0%	75.5%
(0.5, 0.1)	27	0.0000	0.0000	0.0000	–	–
(0.5, 0.3)	384	0.0053	0.0067	0.0067	96.5%	72.7%
(0.5, 0.5)	156	0.5000	0.5307	0.5298	96.9%	75.0%
(0.5, 0.7)	0	0.9947	0.9956	0.9955	97.2%	76.3%
(0.5, 0.9)	0	1.0000	1.0000	1.0000	–	–
(0.7, 0.1)	70	0.0173	0.0208	0.0207	96.6%	72.9%
(0.7, 0.3)	971	0.5000	0.5270	0.5261	96.7%	74.6%
(0.7, 0.5)	321	0.9947	0.9956	0.9955	97.0%	76.3%
(0.7, 0.7)	0	1.0000	1.0000	1.0000	–	–
(0.7, 0.9)	0	1.0000	1.0000	1.0000	–	–
(0.9, 0.1)	0	0.5000	0.5289	0.5280	96.9%	74.5%
(0.9, 0.3)	55	0.9827	0.9848	0.9847	96.9%	75.5%
(0.9, 0.5)	19	1.0000	1.0000	1.0000	–	–
(0.9, 0.7)	1	1.0000	1.0000	1.0000	–	–
(0.9, 0.9)	0	1.0000	1.0000	1.0000	–	–
Weighted average[b]		0.4781	0.4941	0.4936	96.8%	74.5%

NOTES: The table reports the value of the match—that is, the probability of winning the match as of the start of the match—for different types of matches using different challenging strategies. Column (1) defines the type of match using a grid of values of the probabilities of winning a point on and against a serve, (p_s, p_r). Column (3) reports the value of the match under the assumption of no challenge opportunities, column (4) reports the value assuming optimal challenge behavior, and column (5) applies the estimated observed behavior. In column (6) we compute how much of the potential value from challenges is obtained by the estimated observed behavior. As an additional benchmark, in column (7) we compute a similar statistic, assuming that the propensity to challenge is the observed one, but the sorting of challenges across close calls is random (rather than skewed toward more desirable points).

[a]Column (2) reports the number of points in our baseline sample (of 2,008 challenged points) that are closest to each table value of (p_s, p_r), where p_s and p_r are computed for each point based on match statistics (as described in the text), and are then matched to the row in the table that is closest (using a simple Euclidean metric).

[b]The weighted average uses the empirical frequencies of challenged points (from column (2)) as weights. For columns (6) and (7) we only average over those rows that have values in the table; these rows cover 1,957 (97.4%) of the entire points in the baseline sample.

p^r, the absolute value of the challenge technology varies. If, for example, both p^s and p^r are high (or both are low), the player is very likely to win (or lose) the match with or without the ability to challenge the umpire's calls. However, when the match is more even (as most of the matches in our data are), the value of the challenge technology is quite meaningful. Overall, we find that the (unilateral) opportunity to challenge the umpire calls increases a player's probability of winning a match by about 1.6 percentage points, on average.

Our exercise suggests that players take advantage of this opportunity quite well. Although—as shown in Figure 4—players seem to use challenges too little, making the observed behavior suboptimal, their challenge strategy quantitatively generates most of the payoffs that could be captured. To see this, we report in column (6) of Table 4 the ratio

TABLE 5
ROBUSTNESS

Modeling Assumptions			Key Results		
G(·)'s Shape (1)	Probability of a "Close Call" (2)	G(0) (3)	Excess Propensity to Challenge (4)	Value of "Observed" Strategy (5)	Value of "Random" Strategy (6)
Linear	0.02	0.295	0.064	96.8%	74.5%
Concave	0.02	0.295	0.036	99.7%	77.0%
Convex	0.02	0.295	0.074	89.1%	68.1%
Linear	0.01	0.295	0.072	96.0%	74.3%
Linear	0.03	0.295	0.055	97.4%	74.5%
Linear	0.02	0.25	0.110	89.0%	57.7%
Linear	0.02	0.35	0.008	99.9%	92.0%

NOTES: The table reports the key results of the model under alternative assumptions. Columns (1)–(3) describe the three assumptions that are explored. Columns (4)–(6) report the key results. In column (4), we report the average (weighted by the density of observed points) distance between observed and optimal challenge behavior. Graphically, this is the (weighted) average of the vertical distance between the "observed" and "optimal" lines in Figure 4, and this is one way to quantify the extent to which observed behavior implies that players challenge too little. In columns (5) and (6), we report the share of the value extracted by the observed and random behavior, respectively, which is parallel to the last two figures reported in the bottom row of Table 4 for the baseline specification. The top row describes the results for the baseline specification, while the other rows change each assumption at a time. To compute concave and convex $G(\cdot)$, we assume that the corresponding distribution $\widetilde{F}(\cdot)$ follows a Beta distribution $Beta(\alpha, \beta)$ with α equal to 0.5, 1, or 2, which imply concave, linear, and convex $G(\cdot)$, respectively. The value of β is then set so that $G(0)$ obtains its assumed value.

$$\frac{V_{observed}(p^s, p^r) - V_{no\ challenge}(p^s, p^r)}{V_{optimal}(p^s, p^r) - V_{no\ challenge}(p^s, p^r)}. \tag{11}$$

The denominator captures the "money on the table" from optimal use of the challenge technology. The numerator captures the payoffs obtained using the observed utilization strategies. On average, players' observed behavior captures almost 97% of the value, which seems quite high. This is also high with respect to the random-challenging alternative described above, which (see column (7)) captures only 75% of the challenge value on average.[34]

Finally, in Table 5, we report some robustness analysis of our key results with respect to our assumptions about the function $G(\cdot)$. We investigate three assumptions, about the value of $G(0)$ (the intercept in Figure 4), about the curvature of $G(\cdot)$, and about the probability of close calls $p_{close-call}$. Table 5 reports the key results, changing one assumption at a time. Overall, our qualitative results are quite stable. Under all the specifications, we find that players challenge too little (column (4), in which the numbers could have been negative if players challenged too much). Nevertheless, players seem to capture most of the value of the opportunity to challenge (ranging from 89% to almost 100%), with a random strategy doing significantly worse (capturing 57% to 92% of the potential value). One notable case (reported in the bottom row of Table 5) is the specification in which $G(0) = 0.35$ (relative to 0.295 in the baseline specification), which makes observed behavior extremely close to optimal.

7. CONCLUSIONS

We use a large newly assembled data set on tennis challenges to investigate the optimality of decision making of professional tennis players. Overall, we find that players' qualitative behavior is consistent with what optimal behavior would prescribe, and that although quantitatively we

[34] We note that an "always challenge" strategy—that is, challenge *every* close call until running out of challenges—would do quite well given our estimates, implying that the option value of remaining challenges is relatively low. However, as mentioned earlier and discussed in greater length in the end of Section 5, one can imagine that players bear certain costs of challenging unsuccessfully, and therefore prefer to challenge less.

find that players should challenge more often than they actually do, they appear to capture most of the gains from the challenge technology. Overall, we view our findings as largely consistent with optimal behavior.

From a methodological perspective, our application illustrates how a data selection problem can be overcome using outcome data and an independence assumption (Assumption 1). This idea applies more broadly. Indeed, our qualitative analysis in Section 5 is closely related to a similar selection problem that has been the focus of recent literature that tries to detect racial and gender bias in police enforcement (Knowles et al., 2001; Anwar and Fang, 2006) and medical treatments (Balsa and McGuire, 2003; Chandra and Staiger, 2010). Just as we try to back out the propensity to challenge from challenge outcomes, this literature tries to back out the propensity to search a motorist or to treat a heart attack from the observed success rate of the police search or the medical procedure, respectively. That is, the researchers observe the outcome of an endogenously selected action and use differences (or lack thereof) in average outcomes to back out the unobservable decision rules used by the decision maker.[35,36]

Our subsequent analysis in Section 6 adds to this literature by illustrating what additional assumptions are required in order to apply the same basic idea, move beyond testing, and use the analysis for quantitative predictions and counterfactual exercises.[37] In the context of the quantitative exercise, we benefit from the relative simplicity and the dynamic structure of the game of tennis, which allow us to use subsequent realizations of points and fairly standard dynamic programming techniques in order to impose more structure on agents' objective functions at the time of decision making. Imposing such a structure in other contexts, such as police enforcement and medical treatments, may be more challenging, as the objective function of the decision makers may be more difficult to characterize, and observed data on subsequent outcomes may only imperfectly reflect the set of considerations taken into account by decision makers.

Beyond methodology, our study could have implications for other contexts. The assumption that economic agents behave optimally is central to most economic models. Although a large literature in behavioral and experimental economics suggests that individuals exhibit substantial and systematic biases from optimal behavior, our article suggests a setting where individuals' behavior appears close to optimal.

Indeed, large deviations from optimality have been shown even in the contexts of similar settings in other sports, where optimality is arguably expected to hold. For example, Romer (2006) studies the choice of American football coaches as to whether to kick on a fourth down or try for a first down, and found "overwhelmingly statistically significant departures from the decisions that would maximize teams' chances of winning." Similarly, Bar-Eli et al. (2007), in the context of penalty kicks in soccer, also provide evidence of strong deviations from optimal behavior, documenting that goalkeepers almost always jump right or left although optimal behavior is often to stay in the center. In both examples, the authors interpret their results as

[35] One complicating factor in the police enforcement case is that the motorists are also economic agents, who may respond to the police officer decision rule, so this has to be taken into account. This is less of a concern in the context of medical treatments, where it seems reasonable to view the referring physician as the decision maker and the surgeon who performs the procedure as color/gender blind. It is certainly not a concern for us, where outcomes are mediated via the Hawk-Eye technology, which clearly does not respond to the challenger decision rules.

[36] Another issue that complicates the analysis in these other contexts is the inframarginality problem emphasized by Anwar and Fang (2006). The researcher observes the average outcome and tries to draw inference regarding the marginal outcome; thus the results may be sensitive to differences across groups in treatment effectiveness for inframarginal cases. This is less of a concern in our setting because tennis players only challenge close calls, so our analysis only focuses on cases that are much closer to the margin. That is, suppose players play important points more conservatively and are less likely to hit balls close to the line. This would mean very different behavior of inframarginal cases in important points. But the option to challenge would only be useful and empirically relevant for close calls, so the nature of our exercise allows us to effectively condition out these inframarginal cases, and this is precisely the rationale for Assumption 1.

[37] This more quantitative exercise has commonalities with the idea of Levitt and Porter (2001), who try to quantitatively infer the extent of drunk driving from the endogenously selected fatal crashes that involve two cars.

either a failure to optimize or as driven by more complex objective functions, for example, due to agency cost. An important difference between these contexts and the setting we study is that tennis is an individual sport, so agency or team-production arguments are less likely to be important. Thus, our finding that tennis players' decisions appear close to optimal may suggest that earlier findings are driven by misspecified objectives rather than by a failure to optimize.

What can we learn beyond sports from the case study of tennis? Although any attempt to extrapolate to other contexts should be taken with appropriate caution, we think that our study of the decisions made by tennis players may be analogous to decisions taken by other individual decision makers. Consider, for example, the case study reported in Levitt (2006) about the decision of a bagel delivery company regarding the quantity and price of bagels delivered. Just like tennis challenges, the quantity decision (but not the price) is made frequently and feedback is abundant, and Levitt (2006) indeed finds it to be quite close to optimal. Unlike tennis challenges, however, the quantity decision is quite simple, so one could have been worried that a more complex decision would reveal larger deviations from optimality. Our results may suggest that this may not be the case.

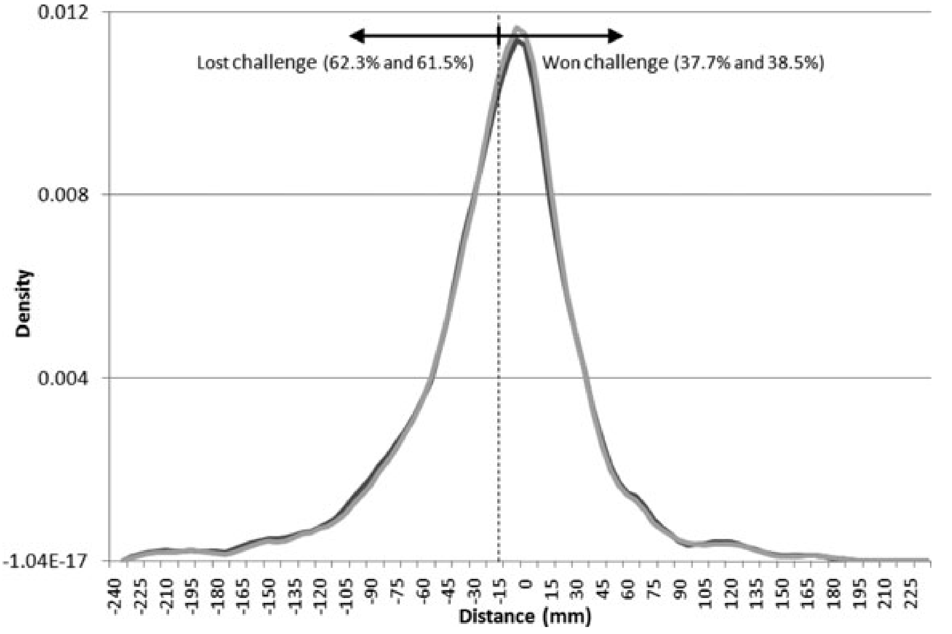
More broadly, to identify other contexts in which our findings may apply, we note that key features of the decisions we study are that they are made with high frequency and are associated with quick and informative feedback, meaning they provide considerable opportunity to learn. In addition, decision makers in our setting must decide quickly and under time pressure; this feature may not be so conducive to optimal behavior, perhaps making the small deviations from optimality that we document appear more impressive. These features also characterize the decisions of other professional decision makers, such as medical surgeons, police officers, pilots, and air traffic controllers. Thus, our findings may shed light on the optimality of decisions made in such contexts, although such extrapolations should be taken with extreme caution.

With so much evidence in recent years emphasizing behavioral biases and other deviations from optimal behavior,[38] we think that it is useful to provide some evidence that at least sometimes basic economic assumptions are not so bad approximations, even when the decision problem is nontrivial. We do not claim that tennis players actually solve our model every time they decide whether to challenge, but at least they develop heuristics that lead to "near-rational" behavior, which entails only minor efficiency losses (as in, for example, Simon, 1978; Akerlof, 1979; Cochrane, 1989).

The famous argument by Friedman and Savage (1948), that expert billiard players play as if they knew the laws of physics, has become a standard way for us to convince our undergraduates that they should care about utility functions and maximization problems despite the fact that we do not see consumers solve maximization problems in the supermarket. Their example illustrates why the seemingly complicated models we use to study economic agents could actually make sense. One way to view our article is that it provides a concrete empirical content for Friedman and Savage's classic argument.

[38] For recent examples in the context of sports, see Garicano et al. (2005), Massey and Thaler (2010), Brown (2011), and Pope and Schweitzer (2011).

APPENDIX



NOTES: The figure replicates Figure 1 for the baseline sample (2,008 challenges, darker line) and the entire sample (2,741 challenges, lighter line). The distributions are almost identical. In the baseline sample, mean distance is −16.4 with a standard deviation of 59.7. Some percentiles from the distance distribution are as follows: −110 (5th), −80 (10th), −39 (25th), −10 (median), +12 (75th), +38 (90th), +62 (95th). In the entire sample, mean distance is −15.5 with a standard deviation of 57.5. Some percentiles from the distance distribution are as follows: −106 (5th), −75 (10th), −38 (25th), −10 (median), +12 (75th), +37 (90th), +60 (95th).

FIGURE A.1

THE DISTRIBUTION OF DISTANCE IN CHALLENGED POINTS (FULL SAMPLE)

TABLE A.1
PRELIMINARY EVIDENCE (FULL SAMPLE)

	Observation (1)	% Won (2)	Difference [p-value] (3)	Mean Distance[a] (4)	Difference [p-value] (5)
Early in the set[b]	1,376	41.1	5.7 [0.002]	−11.5	8.3 [<0.001]
Late in the set[b]	1,321	35.5		−19.8	
Game score is 40–0 or 0–40[c]	65	56.9	20.2 [0.004]	−0.7	17.7 [0.037]
Game score is 40–40[c]	196	36.7		−18.5	
Regular games	2,540	39.5	14.6 [<0.001]	−13.7	24.6 [<0.001]
Tiebreaks	201	24.9		−38.3	

NOTES: The table uses the entire sample (2,741 challenges) to reproduce the statistics reported in Table 1 for the baseline sample (2,008 challenges). The statistics are remarkably similar.
[a]As in the rest of the article, the distance (of the ball from the line) is measured in millimeters and is "sign adjusted" to be negative for failed challenges and positive for successful ones. See Figure 1 for more details.
[b]We define points to be "early in the set" if they occur before either player won more than three games in the set. All other points are defined as " late in the set."
[c] In the scoring notation that we use throughout the paper, game scores of 40–0, 0–40, and 40–40 are equivalent to game scores of 3–0, 0–3, and 3–3, respectively.

REFERENCES

AKERLOF, G. A., "Irving Fisher on His Head: The Consequences of Constant Threshold-Target Monitoring of Money Holdings," *Quarterly Journal of Economics* 93 (1979), 169–87.

ANWAR, S., AND H. FANG, "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence," *American Economic Review* 96 (2006), 127–51.

BALSA, A. I., AND T. G. MCGUIRE, "Prejudice, Clinical Uncertainty and Stereotyping as Sources of Health Disparities," *Journal of Health Economics* 22 (2003), 89–116.

BAR-ELI, M., O. H. AZAR, I. RITOV, Y. KEIDAR-LEVIN, AND G. SCHEIN, "Action Bias among Elite Soccer Goalkeepers: The Case of Penalty Kicks," *Journal of Economic Psychology* 28 (2007), 606–21.

BROWN, J., "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars," *Journal of Political Economy* 119 (2011), 982–1013.

CHANDRA, A., AND D. O. STAIGER, "Identifying Provider Prejudice in Healthcare," NBER Working Paper No. 16382, 2010.

CHANG, A., "Study: Tennis Players Should Challenge Out Calls," *USA Today*, October 28, 2008.

COCHRANE, J. H., "The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-Rational Alternatives," *American Economic Review* 79 (1989), 319–37.

FRIEDMAN, M., AND L. J. SAVAGE, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy* 56 (1948), 279–304.

GARICANO, L., I. PALACIOS-HUERTA, AND C. PRENDERGAST, "Favoritism under Social Pressure," *Review of Economics and Statistics* 87 (2005), 208–16.

KLAASSEN, F. J. G. M., AND J. R. MAGNUS, "Are Points in Tennis Independent and Identically Distributed? Evidence from a Dynamic Binary Panel Data Model," *Journal of the American Statistical Association* 96 (2001), 500–9.

KNOWLES, J., N. PERSICO, AND P. TODD, "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy* 109 (2001), 203–32.

LAMONT, T., "Man versus Machine," *The Guardian* (UK), July 26, 2009.

LEVITT, S. D., "An Economist Sells Bagels: A Case Study in Profit Maximization," NBER Working Paper No. 12152, 2006.

———, AND J. PORTER, "How Dangerous Are Drinking Drivers?" *Journal of Political Economy* 109 (2001), 1198–1237.

MASSEY, C., AND R. H. THALER, "The Loser's Curse: Overconfidence vs. Market Efficiency in the National Football League Draft," mimeo, University of Chicago, 2010.

MATHER, G., "Perceptual Uncertainty and Line-Call Challenges in Professional Tennis," *Proceedings of the Royal Society B: Biological Sciences* 275 (2008), 1645–51.

PASERMAN, M. D., "Gender Differences in Performance in Competitive Environments? Evidence from Professional Tennis Players," mimeo, Boston University, 2010.

POPE, D. G., AND M. E. SCHWEITZER, "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes," *American Economic Review* 101 (2011), 129–57.

ROMER, D., "Do Firms Maximize? Evidence from Professional Football," *Journal of Political Economy* 114 (2006), 340–65.

SCHWARTZ, A., "That Shot Was Out? A Clue on When to Challenge a Call," *The New York Times*, June 24, 2009.

SIMON, H. A., "Rationality as Process and as Product of Thought," *American Economic Review* 68 (1978), 1–16.

WALKER, M., AND J. WOODERS, "Minimax Play at Wimbledon," *American Economic Review* 91 (2001), 1521–38.

WHITNEY, D., N. WURNITSCH, B. HONTIVEROS, AND E. LOUIE, "Perceptual Mislocalization of Bouncing Balls by Professional Tennis Referees," *Current Biology* 18 (2008), R947–9.