DO EXPECTATIONS REFLECT INFORMATION RELIABILITY? EVIDENCE FROM ODDS ON TENNIS MATCHES*

CONSTANTINOS ANTONIOU AND CHRISTOS P. MAVIS

ABSTRACT

We examine whether people form expectations by placing a larger weight on more reliable signals. To test this notion, we analyze subjective probabilities inferred from odds offered on the outcomes of men's tennis matches, exploiting exogenous variation in information reliability related to whether a tennis match is played in a long or short format. The premise of our tests is that higher-skilled players, who are more likely to win any single point, will win more often in longer matches, where more points are generally played. This notion, which is confirmed in the data, suggests that skill-related signals are relatively more reliable in longer matches, and should thus affect odds in those matches more strongly. However, we find that the likelihood of higher-ranked players winning in longer matches is *under-estimated*. This result is robust to inferring expectations from odds offered by professional bookmakers, or odds achieved on a person-to-person betting exchange. The resulting biases in expectations are costly. Results from various robustness tests, including a laboratory experiment and a placebo test using women's tennis data where all matches are played in the same length, support our conclusions. Overall, our analysis suggests that *information reliability neglect* influences expectations and outcomes in real-world markets.

Keywords: Bayes Rule; Decision Heuristics; Information Reliability; Under-reaction.

JEL Classifications: D8, G1

---

* Initial Version: March 2016. This version: April 2018. Antoniou (corresponding author): Warwick Business School, University of Warwick, U.K. E-mail: constantinos.antoniou@wbs.ac.uk. Mavis: Surrey Business School, University of Surrey, U.K., E-mail: c.mavis@surrey.ac.uk. We have benefited from discussions with Nick Chater, Lauren Cohen, David Comerford, Jerker Denrell, Arie Gozluklu, Glenn Harrison, Andrea Isoni, Nick Karampatsas, Alok Kumar, John List, Ragan Petrie, Daniel Read, Chris Starmer, Neil Stewart, Nathaniel Wilcox, Jared Williams and particularly Graham Loomes and Roman Kozhan, as well as conference and seminar participants at the Decision Research at Warwick Forum, the 54th Edwards Bayesian Research Conference, the 2016 Foundations of Utility and Risk Conference, the 2017 American Economic Association Meetings, University of Nottingham, Warwick Business School, Surrey Business School, Henley Business School, University of Stirling, and the University of Sussex. All remaining errors are our own. The paper was previously titled "Bayesian Reasoning: Evidence from the Field."

I. INTRODUCTION

The diagnostic value of information depends both on its content and its reliability. To illustrate, suppose two professional Formula 1 drivers will race identical cars in a 1-mile straight line. A punter, who is considering placing a bet on this race, observes an additional "signal", that the driver of a specific car is the highest ranked driver in the world. In this case, the punter should revise his probability that this driver wins the race upward, but not by much since skill is less relevant in such a straight-line race. In a different scenario, suppose that the race will take place in a 20-mile bendy track. In this case, the signal is much more reliable, since the race allows the higher-ranked driver to utilize his skills, thus it should influence the punter's expectations more strongly.

Several scholars have examined whether people respond optimally to information reliability in controlled laboratory experiments using abstract tasks. The results from these tests suggest that people neglect variations in information reliability, which results to systematic errors in expectations (e.g., Tversky and Kahneman, 1974; Griffin and Tversky, 1992). Much less research, however, has explored whether this bias influences expectations in real-world markets, where experienced agents are operating in their natural habitat, and where the incentives to "get it right" are much higher. We examine whether expectations reflect information reliability (henceforth IR) in the field, analyzing subjective probabilities inferred from odds offered by bookmakers on the outcomes of tennis matches. To conduct the test we exploit *exogenous* variation in IR, which is related to the *length* that different matches are played.

Specifically, men's singles tennis matches are played in two formats: A best-out-of-three set format (BO3), where a player must win two out of possible three sets to win a match, and a best-out-of-five set format (BO5), where a player must win three out of possible five sets to win a match.[1] In our sample, the BO5 matches are called "Grand Slams" (GS), and the BO3 matches are called "ATP World Tour Masters 1000" (MS). In each of these match types, bookmakers set their odds after estimating the probability that a player wins the match, using signals that relate to a player's skill, such as the player's official ranking.[2]

The premise of our tests is that the length of the match (i.e., BO3 vs. BO5) is related to the reliability of skill-related signals such as rankings. This point is based on the following reasoning: the more skillful player is expected to win any given point with a higher probability. Therefore, in the longer BO5 matches, where more points are generally played, the overall probability that the higher-ranked player wins the match is higher than in the shorter BO3 matches.[3] A second reason that longer matches may favor higher-skilled players, is that they allow more flexibility to try out new strategies during the game. In line with this logic, we find that higher-ranked players are 7.2% *more likely* to win in longer GS matches. This result shows that ranking is a more reliable signal for GS, therefore bookmaker's probabilities for higher-ranked players winning GS matches should be adjusted accordingly upward. Our objective is to test whether such an adjustment takes place.

For our main analysis, we use subjective probabilities inferred from fixed decimal odds offered by several major betting houses on professional men's tennis matches for the period 2005 to 2014. For each match, we infer the subjective probability that the

[1] Each set is comprised by individual games, and to win one set a player must win at least 6 games. To win one game a player must win at least four points. For more information on the rules of tennis see http://www.atpworldtour.com.

[2] Official rankings are based on the amount of points accumulated by a player during their past 52 weeks from a total of 18 tournaments. For more information on how rankings are calculated see http://www.atpworldtour.com/en/corporate/ rulebook. For each match, we refer to the player with the highest ranking (i.e., a smaller ranking number), as the higher-ranked player.

[3] An implicit assumption here is that the outcomes of different points are independent and identically distributed. Klaassen and Magnus (2014) analyse data from the Wimbledon, and conclude that, even though this assumption is weakly rejected by the data, it still "provides a reasonable approximation in many applications" (Klaassen and Magnus, 2014, p. 171).

higher-ranked player wins the match from bookmaker odds, π, estimate the "objective" probability as the fitted value from a logit model, $\hat{p}$, and define *bias* as π - $\hat{p}$. We test our hypothesis based on the difference in the average *bias* between MS and GS (Δ_{bias}). If bookmakers adjust π for GS to reflect the increase in $\hat{p}$ due to the longer match format (BO5), then this difference should be 0.

Our alternative hypothesis is based on experimental findings that people are not fully sensitive to IR when forming probabilities, where IR is usually measured with the size of the sample that generates a particular signal. For example, in a seminal study, Griffin and Tversky (1992) observed that subjects do not appreciate the effect of sample size on the posterior probability, thus *under-react* to information from large samples.[4] Kahneman and Tversky (1972), suggest that this behavior generally arises because subjects are insensitive to the effect of sample size on sampling errors.

In our analysis, match length is an element of the information set that is related to the reliability of rankings, similar to the role of sample size in laboratory experiments.[5] If bookmakers are insensitive to variations in IR related to match format, they will *under-estimate* the likelihood that the higher-ranked player wins a GS match. Therefore, under the hypothesis of *IR-neglect*, *bias* should be lower for GS matches ($\Delta_{bias} < 0$).[6]

Our results show that Δ_{bias} is -3.3% and highly statistical significant. Given that $\hat{p}$ is 7.2% higher for GS matches, this result shows that bookmakers are adjusting their probabilities for GS matches in the correct direction, but stop roughly half-way from

[4]Benjamin, Rabin, and Raymond (2016) review the experimental literature on Bayesian Updating and conclude that expectations formed using information from large samples are always below the corresponding Bayesian posteriors.

[5] However, in the experimental literature variations in IR come from different signals, whereas in our tests such variations come from applying the same signal (i.e., rankings) in two environments, MS vs. GS.

[6] In Appendix A we use a simple model to illustrate the hypothesis.

the level implied by full Bayesian Reasoning. This is the central finding of our paper, which suggests that under-reaction due to information reliability neglect influences expectations in real-world markets.

Could our findings reflect strategic efforts by rational bookmakers to exploit "irrational" punters? In this setting bookmakers set the odds on *both* players, thus could "salt" them in opportune directions. For example, Levitt (2004) finds that, in spread-betting markets, bookmakers offer biased prices to exploit the tendency of punters to bet on the favorite.

To examine whether strategic behavior is driving our results, we re-do our tests using subjective probabilities inferred from odds which are set competitively in a person-to-person real-life market called Betfair. In this setting strategic incentives do not exist, since the odds for the two players are determined competitively in two different markets.[7] Analyzing tennis matches for the period 2009-2014, we find that Δ_{bias} is -2.6% and statistically significant.

To further analyze the role of strategic incentives, we examine whether bookmakers are increasing their profitability by offering *better than fair* odds on the higher-ranked player in GS, which amounts to a more negative *bias*. Note that, by analogy, this implies that the odds on the lower-ranked player for GS matches are *worse than fair*. Therefore, if the volume on the lower-ranked player is sufficiently high, it may be optimal to set the odds in this way.

Contrary to this notion, we find that bookmaker's profits per match (as a proportion of the total volume staked) are *lower* for GS matches by an average of -2.6%.[8] Moreover, we find that bookmaker's profitability is significantly higher in a

[7] In these markets, prices converge to satisfy no arbitrage conditions, i.e., the sum of the implied probabilities for each player winning are close to 1.

[8] This means that the volume that backs the lower-ranked player is not sufficiently higher in GS so that a negative *bias* for the higher-ranked player is optimal. This is not surprising, given the strong preference

counterfactual scenario, where their odds on are set according to $\hat{p}$ from the logit model. Overall, the analysis of profits suggests that IR-neglect is costly to bookmakers.

Our analysis implicitly assumes that MS and GS matches are the same in every respect, except from the length of the match. However, GS matches offer more ranking points, higher prize money to players, and attract more attention from punters. Could our findings reflect these differences?

To address this issue, we conduct a placebo test using data for professional women's tennis matches. For women, GS matches also offer more ranking points and prize money, and attract more betting volume but they *are played in a BO3 format, exactly like the MS tournaments*. Hence, women's matches preserve the key differences across match formats *except* for the change in the reliability of rankings across MS and GS. If the bias we document is driven by factors other than IR-neglect, *bias* should be lower for GS in the women's data as well.

Logistic analysis confirms that the type of match (MS vs GS) *does not* affect the probability that the higher-ranked player wins for women's matches, consistent with no difference in IR in rankings across match formats. Therefore, bookmakers should not adjust π upward for GS. When we compare *bias* calculated from bookmaker odds for MS and GS we find that the difference is *insignificant* ($\Delta_{bias} = 0$). When we do the same for *bias* calculated from betting exchange odds, we find that Δ_{bias} is *higher* in GS by 1.6%. Overall, the results obtained from the women's data are starkly different to those obtained from the men data, which suggests that our baseline results more likely reflect under-reaction due to IR-neglect.

of punters to bet on the favorite, since on average 86% (82%) of the volume in GS (MS) matches backs the higher-ranked player. This preference for the favorite is also shown in Levitt (2004).

We conduct various robustness checks with different specifications for *bias*, for example obtaining the objective probability using non-parametric methods, or from specifications where the higher-skilled player is identified based on odds instead of rankings. Our results continue to hold in these alternative specifications. We also discuss several alternative explanations for our results, such as probability weighting (Snowberg and Wolfers, 2010), biased priors, or a general bias of under-reaction toward information, and conclude that these explanations do not offer a parsimonious explanation for all our findings. The robustness tests and alternative explanations are discussed in detail Section III.F of the paper.

Our last robustness check is a laboratory experiment, which tests the IR-neglect hypothesis in a setting where market-related factors are irrelevant. Since our field analysis suggests a sharp hypothesis that can be tested easily in the laboratory, this is a useful validity test.

We invited university students who were involved in tennis activities to participate in the experiment. These students participated in two different sessions, and in each session they were asked to consider upcoming matches from two tournaments, one MS and the other GS, and to assign probabilities to each player winning. Their responses were incentivized using the quadratic scoring rule. Our objective was to test whether subjects adjust their probabilities for the effect of the length of the match on the probability that the higher-ranked players wins.

The results from the experiment reveal a significant bias of under-reaction due to IR-neglect, with Δ_{bias} equaling to a statistically significant -3.30%. The students' subjective probabilities of the higher-ranked player wining show *no upward adjustment* for GS, contrary to what we observe in the data for bookmakers and punters (where partial adjustment occurs). This suggests that field agents are more sophisticated.

Overall, the experimental analysis provides further support to the claim that IR-neglect influences expectations.

For our final test, we examine whether IR-neglect influences outcomes in other sports betting markets. The premise of this analysis is that the degree to which we can predict outcomes in different sports varies. For example, outcomes are relatively less predictable in low scoring sports like soccer, compared to higher scoring ones like basketball. This implies that the IR of skill-related signals also varies across sports, according to whether outcomes in these sports contain more or less "noise". The IR-neglect hypothesis predicts that *bias* is lower for sports that contain less noise.

To rank different sports according to the noise the contain we use the analysis in Mauboussin (2012), who shows that football is more noisy than soccer, and soccer is more noisy than basketball. Using odds data from Betfair, and following the same methodology as we did for the tennis analysis, we find that average *bias* is equal to 2.81% in football, 1.69% in soccer, and -0.44% in basketball, with the differences being statistically significant. This finding suggests that IR-neglect influences expectations in betting markets more broadly.

Our findings are consistent with theories of choice under uncertainty that incorporate constraints on attention or computing capacity, such as "*sparse thinking*" (Gabaix, 2014) or "*local thinking*" (Gennaioli and Shleifer, 2010). Sparse thinkers, for example, may neglect variations in some parameters that are more costly to monitor, setting them equal to a default value (Gabaix, 2014). Local thinkers, due to limited and selective memory recall, may under-emphasize certain aspects of the information set that stand out less in one's memory (Gennaioli and Shleifer, 2010). The finding that people are sensitive to IR when combining *spatial* information (e.g., Ernst and Banks, 2002), where variations in IR are more "tangibly" experienced, supports the view that IR-neglect in the domain of expectation formation reflects cognitive constraints.

Given the importance of Bayesian Updating for the paradigm of rational expectations, the experimental evidence from psychology that errors in expectations are systematic (e.g., Kahneman, Slovic, and Tversky, 1982) have received much attention from economists, especially in validating whether they emerge in experimental economics conditions (e.g., Grether, 1980; Charness, Karni, and Levin, 2010; Antoniou, Harrison, Lau, and Read, 2017). What is particularly valuable is to determine whether these phenomena influence decisions in the field with significant economic consequences. Along these lines, De Bondt and Thaler (1990) examine whether the earnings forecasts of sell-side analysts are influenced by representativeness, and Chen, Moskowitz, and Shue (2016) examine whether the decisions of asylum judges, loan officers and baseball umpires are affected by the gambler's fallacy.[9]

We contribute to this literature by conducting a field test of Bayesian reasoning with three attractive features: Firstly, subjective probabilities are inferred from the decisions of expert agents who are pricing securities in their natural habitat with significant monetary consequences. Secondly, uncertainty is fully resolved when the match is finished, which allows us to test for a bias in subjective probabilities with relatively weak assumptions. Thirdly, and most importantly, in the tennis data variation in IR is completely exogenous, in the sense that it is governed solely (and transparently) by the rules of the game. Our findings suggest that people cannot accurately determine the reliability of the signals they use, thus under-react when confronted with more reliable signals.

[9] In other related work scholars have examined whether decisions in the field are influenced by heuristics (e.g., Simonsohn and Loewenstein, 2006; Lacetera, Pope, and Sydnor, 2012), or loss aversion (Coval and Shumway, 2005; Pope and Schweitzer, 2011)

Because we test the same hypothesis in a very similar task using both field and laboratory data our work also contributes to the general literature on economic methodology, which debates whether laboratory findings carry-over into the field (e.g., List, 2003; Harrison and List, 2004; Levitt and List, 2008; Camerer, 2015 Al-Ubaydli, List, and Suskind, 2017). Our results show that errors in expectations in the experimental data are similar to those in the field data (i.e., of the same sign), which suggests that, in this context, the process of belief formation is broadly similar in the two domains. However, the results also show that laboratory subjects are less sophisticated, since their probabilities show no adjustment to the increase in the reliability of rankings for GS matches.

Finally, our analysis contributes to the behavioral finance literature, which has attributed various asset pricing anomalies to investor under-reaction to information (i.e., Ball and Brown, 1968; Jegadeesh and Titman, 1993; Loughran and Ritter, 1995; Ikenberry, Lakonishok, and Vermaelen, 1995; Michaely, Thaler, and Womack, 1995; Zhang, 2006; Jiang and Zhu, 2017). According to this explanation, investors fail to recognize that certain information signals are good indicators of fundamentals, and therefore do not price them in stocks as strongly as they should. With the passage of time, as the fundamentals are slowly revealed and expectations are corrected, prices drift toward their equilibrium values. However, a caveat of these interpretations is that they rely on assumptions related to *expected* equity returns, or what Fama (1998) calls the *bad model problem*. A notable exception is Moskowitz (2015), who tests for the existence of asset pricing anomalies in sports betting markets where risks are completely idiosyncratic, and therefore the Fama (1998) critique does not apply. In a similar spirit, we contribute, by showing that under-reaction is a real phenomenon that influences asset prices in real-world markets.

The next section describes our data, methods and hypothesis. The third section presents and discusses the results, and the fourth section concludes the paper.

II. DATA AND METHODS

II.A. Data

For our baseline analysis, we obtain data from www.tennis-data.co.uk.[10] For every match this database contains data the name of the tournament, the date of the match, the names of the two competing players, their official ATP rankings, the winner of the match, as well as fixed decimal odds from various international betting houses on both players.[11] In our analysis, we average the odds offered by the various bookmakers on the two players and then use these average odds to infer subjective probabilities.[12]

We include in our sample Grand Slam (GS) matches, which are played in a BO5 format, and ATP World Tour Masters 1000 matches (MS), which are played in a BO3 format. GS tournaments are the most prestigious, with the winner receiving 2,000 ranking points, and on average collecting $2.5 million dollars (in 2015). For comparison, the winner of an MS tournament earns 1,000 ranking points and, on average, $0.8 million dollars (in 2015). There are other tournaments that are played in a BO3 format, which yield, for example, 500 or 250 ranking points to the winner of the tournament, and offer less prize money. Such tournaments are significantly less prestigious, involving on average lower-ranked players, and attracting less attention from punters. To ensure that the BO3 matches are as similar as possible to the BO5 matches, we focus on the more prestigious tournaments from the BO3 class.

[10] Data from this database has been used previously by academic work on tennis matches (e.g., Forrest and McHale, 2007; Del Corral and Prieto-Rodríguez, 2010).
[11] This dataset contains odds from eight different bookmakers; Bet365, Centrebet, Expekt, Ladbrokes, Interwetten, Pinnacles Sports, Stan & James, and Unibet.
[12] The odds offered by any two bookmakers on the same match are very highly correlated (at least 0.92). Thus, we lose little information by averaging odds across bookmakers.

We apply the following criteria to the initial dataset (n = 10,790) to create our final sample: (i) we drop matches that were not completed, matches with missing rankings information, matches with no odds for either player, and matches that entail a negative *housetake*[13] (n = 10,266); (ii) we drop matches where the higher-ranked player is indicated as an outsider by bookmakers even though he is ranked by at least 15 places higher than his opponent at the beginning of the tournament (n = 9,230). Such cases are likely to reflect recent developments like injuries, which are not yet incorporated in the rankings, thus making them outdated indicators of skill.[14]

Our final sample consists of 9,230 tennis matches from 2005-2014. Table I breaks down the matches by year and tournament. Overall, we have data for 4 GS tournaments, and for 11 MS tournaments. Some MS tournaments are discontinued and others are introduced at various point in time.

[Insert Table I here]

II.B. Methods

Assume a tennis match between players X and Y. The bookmaker offers fixed decimals odds for player X to win equal to d_X, and for player Y equal to d_Y, where d_X and d_Y are greater than 1. To obtain subjective probabilities, we first invert the quoted odds for X, $O_X = \frac{1}{d_X}$, and for Y, $O_Y = \frac{1}{d_Y}$. In a perfectly competitive and frictionless market with a risk-neutral bookmaker, O_X and O_Y correspond to true subjective beliefs. However, typically $O_X + O_Y > 1$, which reflects the housetake, a form of commission collected by the bookie. To obtain subjective probabilities, we normalize O_X and O_Y to

[13] To obtain the housetake odds are inverted and added together. In bookmaker markets this sum is typically greater than 1, reflecting a form of commission collected by the bookmaker. Matches with negative housetake are likely to be data errors.

[14] Our conclusions remain the same if we do not impose this filter. This is discussed in more detail in Section III.F of the paper.

sum to 1, using $\pi_X = \frac{O_X}{O_X+O_Y}$ and $\pi_Y = \frac{O_Y}{O_X+O_Y}$. Thus, the housetake is split proportionally between the two players, depending on their relative odds.[15] Throughout the analysis we refer to π_i as the subjective probability that the higher-ranked player wins match *i*.

To examine whether the reliability of rankings changes across match format (MS vs. GS) we use the logistic model, shown below:

$$\Pr(\mathrm{Y_i} = 1|\mathrm{GS_i}, \mathrm{RSkill_i}) = \mathrm{F}(\alpha + \beta_i \mathrm{GS_i} + \beta_2 \mathrm{RSkill_i}) \qquad (1)$$

The dependent variable, Y_i is a binary indicator taking the value of 1 if the higher-ranked player wins match *i*, and 0 otherwise. GS_i is a dummy variable that equals 1 if match *i* is Grand Slam, and 0 otherwise. $RSkill_i$ captures differences in player rankings for match *i* and is calculated as log(lower-ranked player ranking) - log (higher-ranked player ranking), following the specification in Klaassen and Magnus (2001).[16] *F* is the logistic distribution.

As we discuss in more detail in the next section, GS matches entail more players in the draw, and therefore entail higher average ranking differences between the two players. The model in equation (1) captures the effect of the change in the reliability of rankings across MS and GS, whilst controlling for these ranking differences. That is, the coefficient on *GS* should be positive and significant, reflecting that rankings are more reliable predictors in GS matches.

[15] This method of recovering beliefs from odds is typical in the literature (e.g., Croxson and Reade, 2014; Smith, Paton, and Williams, 2009).

[16] As discussed in Klaassen and Magnus (2001) player skill in tennis resembles a pyramid (i.e., the difference in skill between players ranked #1 and #10 is higher than the skill difference between players ranked #80 and #90), therefore a logarithmic transformation is appropriate. Note also that the ranking of the lower-ranked player is a larger number than the ranking of the higher-ranked player. Our results are the same if we calculate *RSkill* on the basis of differences in the ranking points of the two competing players.

To estimate the bias in subjective probabilities we use $bias_i = \pi_i - \hat{p}_i$, where $\hat{p}_i$ is the fitted value from the logit model. We test our hypothesis based on the difference in average bias between MS and GS matches, $\Delta_{bias} = \overline{bias_{GS}} - \overline{bias_{MS}}$.

Under the null hypothesis of Bayesian reasoning, where subjective probabilities are properly adjusted according to variations in IR between MS and GS, $\Delta_{bias} = 0$. The alternative hypothesis of under-reaction due to information reliability neglect is $\Delta_{bias} < 0$. We test these hypotheses by examining the sign and statistical significance (using two-sided tests) of β_1 in the ordinary least square regression shown below:

$$\text{bias}_i = \alpha + \beta_1 \text{GS}_i + \beta_2 \text{RSkill}_i + \varepsilon_i \qquad (2)$$

II.C. Descriptive Statistics

Table II presents descriptive statistics for our main sample, separately for GS (Panel A) and MS (Panel B) matches. The average posted odds offered by bookmakers that the higher-ranked player wins the match (HR_{Odds}) are much lower than those for the lower-ranked player (LR_{Odds}) (1.35 vs. 6.05 for GS and 1.48 vs. 4.10 for MS), which shows that player ranking is indeed an indicator of skill that is used by bookmakers. The average *Housetake* is very similar in the two match formats (0.05 in GS and 0.06 in MS); moreover, *Housetake* has low volatility across matches, as noted by other authors (Forrest and McHale, 2007). For GS the average ranking of higher- and lower-ranked players (HR_{Rank} and LR_{Rank}) is 28 and 99 respectively, whereas for MS the corresponding rankings are 22 and 67. GS matches entail higher ranking differences because GS tournaments allow more players in the draw.[17] The larger ranking

[17] GS matches involve 128 players, whereas MS matches involve 64 players on average (96 players: Indian Wells and Miami; 56 players: Monte Carlo, Madrid, Rome, Montreal, Toronto, Cincinnati, Shanghai; 48 players: Paris). In all tournaments the "seeded" players are allowed to compete automatically in the tournament, for example the top 32 players in the world. The remaining positions are filled by lower-ranked players who earn their position by qualifying in a single elimination tournament prior to the main event. The number of seeded players varies by tournament, and is higher in GS than in MS.

differences in GS imply that the probability of the higher-ranked player winning a GS match is higher, therefore we test the hypothesis using the model in equation (2), which includes *RSkill* as a control variable.

The next two rows show the average estimate of the objective probability obtained as the fitted value from the model in equation (1), $\hat{p}$. As it can be seen, $\hat{p}$ is higher for GS than MS (0.78 vs 0.69), consistent with the view that higher-ranked players are more likely to win a GS match (we present the full results from the logit model in section III.A). The penultimate row in the table shows the average subjective probability that the higher-ranked player wins the match (π), which is higher for GS compared to MS (0.74 vs. 0.68). This means that bookmakers are adjusting their subjective probabilities for GS matches relative to MS matches in the direction predicted by Bayes Rule. However, this adjustment seems insufficient, i.e., it does not completely reflect the increase in $\hat{p}$ for GS. This result, shown in Figure I, provides some early support to the IR-neglect hypothesis, which we formally test in the next section.

[Insert Table II and Figure I here]

III. ANALYSIS

III.A. Changes in Information Reliability in MS vs GS

We start our analysis by examining whether the higher-ranked player is more likely to win a GS as opposed to an MS match using the model shown in equation (1).

The marginal effects associated with each variable are shown in Table III. From column (1), we observe that the marginal effect associated with GS is 9.5% and highly statistically significant, indicating that higher-ranked players are more likely to win a GS match. Once we control for *RSkill* in column (2) the marginal effect associated with GS reduces to 7.2%, but remains highly statistically significant. *RSkill* is also positive and highly significant. In column (3), we add surface (clay vs. hard), [18] year, and round fixed effects (1 to 7), and obtain very similar results.

Overall, this analysis shows that higher-ranked players are more likely to win GS matches that are played in a BO5 format.

[Insert Table III here]

III.B. Biases in Subjective Probabilities

Table IV shows our main results. In a univariate setting in Panel A, we find that $\overline{bias_{GS}}$ is equal to -3.7% and $\overline{bias_{MS}}$ is equal to -0.4%, making Δ_{bias} equal to -3.3%, and highly statistically significant.

In Panel B, we test the hypothesis whilst controlling for the effect of *RSkill*. In column (1) it is shown that the coefficient on *GS* is essentially the same, equal to -3.1% and highly statistically significant. In column (2), we estimate an additional specification that includes surface, year, and round fixed effects, and find that the coefficient on *GS* remains negative and significant in this specification.

The coefficient of *RSkill* is also negative and significant, indicating that bookmakers are under-estimating the probability that the higher-ranked player wins, when the ranking differential is high. However, this result will not be robust throughout all our tests.

Overall, the results in Table IV show that bookmakers are under-estimating the probability that the higher-ranked player wins a GS match, consistent with the hypothesis of IR-neglect.

[Insert Table IV here]

[18] One GS match is played on grass, one on clay, and two on hard courts. In terms of MS tournaments, seven are played on hard courts and four are played on clay courts. In our sample, the only tournament played on grass is the Wimbledon (GS), and the only tournament played on carpet is Paris (MS) between 2005 and 2008. Because in terms of speed, grass and carpet surfaces are more similar to hard than clay surfaces, we include them in the hard court category.

III.C. Betting Exchange Data

Do the results in the previous section reflect the strategic incentives of rational bookmakers to exploit punters? To address this question, we test our hypothesis using subjective probabilities inferred from odds achieved on a person-to-person betting exchange called Betfair.[19] In this setting, strategic incentives do not exist because the odds on the two players are set in two different markets. Our sample contains 4,893 observations for the period 2009-2014. Descriptive statistics for this sample are shown in Table A. 1 of the Appendix. One noteworthy feature of this data is that because odds are determined in a market setting and thus do not entail a housetake, the sum of "raw" implied probabilities is closer to 1.[20]

The results are shown in Table V. In a univariate setting in Panel A, we find that $\overline{bias_{GS}}$ is equal to -1.8% and $\overline{bias_{MS}}$ is equal to 0.5%, making Δ_{bias} equal to -2.3%, and highly statistically significant. As shown in Panel B, controlling for *RSkill* and additional fixed effects does not change our findings.[21]

Overall, this analysis shows that the bias IR-neglect also exists in the betting exchange data, where strategic incentives do not exist.

[Insert Table V here]

[19] Betfair is the largest person-to-person betting exchange, with almost one million active users (Croxson and Reade, 2014). The Betfair dataset was purchased from Fracsoft available at http://www.fracsoft.com. The dataset has incomplete coverage of the Australian Open (GS tournament), with no observations for 2009 and 2010. Moreover, it does not include data for matches that were completed in more than one day (for example due to rain delays). We use "back" odds on the two players to calculate subjective probabilities (i.e., odds available to punters who want to bet that a specific player wins).

[20] However, to ensure consistency and comparability to our previous results, we normalize the inverted odds to add to 1.

[21] The coefficient on *RSkill* is insignificant in this setting, which indicates that the market is pricing correctly ranking differentials.

III.D. Profits

In this section, we examine whether the negative *bias* documented in Section III.B is costly to bookmakers. To set the stage for this test, equation (3) below shows the per unit return on bets for the higher-ranked player, where d_{HR} are the odds offered on the higher-ranked player by bookmakers, and D_{HR} is an indicator that equals to one if the higher-ranked player won the match.

$$r_{HR} = \begin{cases} 1 & if\ D_{HR} = 0 \\ -(d_{HR} - 1) & if\ D_{HR} = 1 \end{cases} \tag{3}$$

We define the per-unit return on bets on the lower-ranked player, r_{LR}, in the same way. Columns (1) and (2) in Table VI show that, due to the negative Δ_{bias}, bookmakers in GS matches are earnings less on bets on the higher-ranked player, and more on bets on the lower-ranked player. Thus, if a sufficiently large proportion of volume backs the lower-ranked player in GS, a negative *bias* actually increases bookmakers' *total* profitability.[22]

We estimate bookmaker's total profits per match, as a proportion of the total volume staked for match i, Π_i, using the equation below:[23]

$$\Pi_i = 1 - RVol_{HR} d_{HR} D_{HR} - RVol_{LR} d_{LR} D_{LR} \tag{4}$$

$RVol_{HR}$ is the proportion of the total volume that backs the higher-ranked player calculated using data from the betting exchange market;[24] d_{HR} and D_{HR} are defined as in equation (3). $RVol_{LR}$, d_{LR} and D_{LR} are analogously defined for the lower-ranked

[22] We discuss in more detail the conditions that a negative *bias* is optimal in Appendix B.
[23] To conduct this test we use the volume information from the betting exchange. Therefore, an implicit assumption is that that betting volume behaves similarly in bookmaker and betting exchange markets.
[24] Therefore, an implicit assumption here is that that betting volume behaves similarly in bookmaker and betting exchange markets.

player. Because the data for the betting exchange start in 2009, the sample used for this test is smaller.

From the third column of Table VI, we observe that Π_i is equal to 2.5% for GS matches, and 5.10% for MS matches, for a significant difference of -2.6%.[25] This result implies that the volume backing the lower-ranked player in GS is not sufficiently higher to justify a negative *bias*;[26] hence bookmakers are earning a smaller proportion of the volume staked in GS compared to MS matches.

For our final test, we construct the *hypothetical* profit $H\Pi_i$ collected by bookmakers, in a counterfactual scenario where subjective probabilities for the higher-ranked player for GS matches are set equal to $\hat{p}$ from the logit model in equation (1). This involves adjusting the odds offered for the two players, as well as *RVol*.[27] As shown in column (4) of Table VI average $H\Pi_i$ for GS matches is equal to 5.1%, significantly different from the actual average Π_i of 2.5%.[28] This result suggests that bookmakers' profitability would have been higher, if they had used the logit model to set their odds.[29]

Overall the analysis in this section suggests that IR-neglect is costly to bookmakers.

[Insert Table VI here]

[25] Our results hold in a multivariate setting, when we regress Π_i on a *GS* dummy and *RSkill*, including round, time, and surface fixed effects.
[26] This finding is in line with Levitt (2004), who finds that punters in pari-mutual markets show an excessive preference toward the favourite. This preference is also shown in our data, since 86% of the volume in GS backs the higher-ranked player, who wins about 78% of the matches (Table A. 1, Panel A).
[27] We discuss in detail how we adjust odds and *RVol* in Appendix B.
[28] Moreover, the cross-sectional standard deviation of $H\Pi_i$ for GS matches is 32%, and of Π_i 33%. So this adjustment does not change materially the volatility of profits in GS for bookmakers.
[29] This finding is in line with the finding that simple, statistical algorithms can outperform human judgement (e.g., Dawes, Faust, and Meehl, 1989).

III.E. A Placebo Test

Our tests implicitly assume that the only difference across MS and GS matches is match format, i.e., BO3 vs. BO5. However, as mentioned, GS tournaments are more prestigious than MS tournaments, offering more prize money and ranking points, and attracting higher betting volumes. To examine whether our findings are affected by these differences, we conduct a placebo test using data for women's tennis matches. This data provide an ideal setting for such a test as they preserve the key differences across MS and GS that are observed in the men's sample (i.e., GS are more prestigious, offer more prize money and ranking points, and attract more betting volume), but for women there is *no change* in the reliability of rankings, because *both* MS and GS are played in a BO3 format. Therefore, if our results do not reflect IR-neglect, Δ_{bias} should continue to be negative and significant in the women data.

We construct the women's sample using an approach similar to that used in our baseline analysis with the men's data. GS tournaments are the same for women as for the men, and for MS tournaments we again focus on the more prestigious tournaments.[30] After applying the same filters to the initial sample as those for the men we end up with 2,527 MS matches and 3,624 GS matches from 2007-2014 for the bookmaker sample, and 1,425 MS matches and 2,259 GS matches from 2009-2014 for the betting exchange sample.

Table A. 2 and Table A. 3 in the Appendix present descriptive statistics for the women samples for MS and GS, and show that the housetake and ranking differentials

[30] Pre-2009, the top tier masters tournaments were called Tier-I and post-2009 they are called Premier Mandatory. For the bookmaker sample, we have 12 MS tournaments prior to 2009 (Berlin, Charleston, Doha, Indian Wells, Miami, Montreal, Moscow, Rome, San Diego, Tokyo, Toronto, and Zurich) and the 4 Premier Mandatory tournaments after 2009 (Beijing, Indian Wells, Madrid, Miami). The betting exchange sample starts in 2009 so we have data on the 4 Premier Mandatory tournaments. The GS matches for women are the same as for the men (Australian Open, French Open, Wimbledon, and US Open).

are very similar as those found in the men data. From Table A. 3, we observe that *TotVol* is lower for women than for men, whereas *RVol* is the same (85% in GS and 82% in MS). Overall, the women samples are broadly similar to the men samples.

In Table VII, we present results from the logistic regressions shown in equation (1) using the women data (Panel A: bookmaker; Panel B: betting exchange). As shown in column (2), the coefficient on the *GS* dummy is insignificant once we control for *RSkill*, consistent with no change in IR in GS matches. Thus, bookmakers *should not* upwardly adjust their subjective probabilities of the higher-ranked player winning a GS match.

[Insert Table VII here]

The results from the model in equation (2) with women data are shown in Table VIII. From column (2) in Panel A (bookmaker), we observe that the coefficient on *GS* is negative but statistically *insignificant*. The corresponding results in Panel B (betting exchange) show that the coefficient on *GS* is *positive* and significant, equal to 1.6%, reflecting the fact that punters are increasing their subjective probabilities for the higher-ranked player winning a GS match, even though this increase is unwarranted.

Overall, the placebo test with the women's data produces results that are in stark contrast from those obtained with the men's data. This suggests that our baseline results in Table IV and Table V more likely reflect biases due to information reliability neglect.

[Insert Table VIII here]

III.F. Robustness and Alternative Explanations

III.F.I. Robustness Checks

In this section, we conduct various tests of robustness, presenting the results in Table IX. In Panel A, we define *bias* using probabilities inferred from bookmaker odds and in Panel B using betting exchange odds.

In our baseline analysis, *bias* depends on the specification of the logit model used to calculate $\hat{p}$. In column (1), we test the hypothesis using a non-parametric specification for the objective probability. We first rank our sample according to match format (MS and GS) and then rank again within each format into deciles based on *RSkill*. In each decile, we use the proportion of matches won by the higher-ranked player, p_{HR}, as the estimate of the objective probability, and define bias as π - p_{HR}. The results are robust in this specification, as the coefficient of GS is equal to -3.4% in Panel A and -2.6% in Panel B, both statistically significant.

In column (2), we avoid altogether estimating an objective probability, and define *bias* as π - D_{HR}, where D_{HR} is a dummy that equals 1 if the higher-ranked player won the match, and 0 otherwise. The coefficient on GS continues to be negative and significant.

Our procedure of estimating $\hat{p}$ uses full sample information, which may introduce some look-ahead bias. To make sure that such a bias does not influence our findings in column (3), we estimate the logit model using only backward-looking information. For example, for matches played in 2006 (2007) we calculate $\hat{p}$ estimated from a logit model that only uses data from 2005 (2005 and 2006), etc. The coefficient of GS is equal to -4.2% in Panel A and -2.8% in Panel B, both statistically significant.

Because GS matches offer more prize money it is possible that higher-ranked players "time" their form to peak at GS matches. Such timing effects could influence our findings. To address this issue in column (4), we add an additional control variable in our models (including the logit model used to estimate $\hat{p}$), *RStreak*, calculated as the difference in the proportion of matches won by the higher- and lower-ranked player in the two previous tournaments. The results show that the coefficient on *GS* remains negative and significant in this specification, equaling -2.4% in both panels.[31]

The information signal we use is the ranking of the players at the start of the tournament. However, bookmakers are observing other information besides rankings before setting their odds. We examine whether our findings hold when we define as the higher-skilled player the one who is favored by the bookmakers (as shown by the odds). The results from this test are shown in column (5), whereby the coefficient on *GS* is negative and significant (-3.0% in Panel A and -2.8% in Panel B).[32]

For our baseline results, we control for *RSkill* and test our hypothesis by observing the coefficient on the intercept dummy *GS.* An alternative method is to use a slope dummy, interacting *RSkill* with *GS,*[33] expecting that the coefficient on the interaction is negative and significant. The results shown in column (6) show that this coefficient is indeed negative and significant, and that the total effect of *RSkill* on *bias* is negative for GS matches (-0.003 - 0.010 = -0.013 in Panel A and 0.005 - 0.007 = -0.002 in Panel B).

In untabulated analysis, available from the authors on request, we find that our results hold even when we do not impose filter (*ii*) from Section II.A. In this sample, the coefficient of GS is -2.41% for bookmaker data, and -1.98% for betting exchange data, statistically significant in both cases. Moreover, to control for the effect of any outliers, we estimate equation (2) using a quantile regression model, expressing the

---

[31] This model entails a smaller number of observations due to missing values in *RStreak*, which occur when a player did not compete in the two previous tournaments.
[32] For this test we have more observations because we do not impose filter (*ii*) from Section II.A.
[33] In this specification, $\hat{p}_i = Pr(Y_i = 1 | RSkill_i, RSkill_i \times GS_i) = F(\alpha + \beta_1 RSkill_i + \beta_2 RSkill_i \times GS_i)$.

median of the conditional distribution of *bias* as a linear function of the independent variables. The coefficient of GS is negative and significant in this specification.

Overall, the results in this section show that our conclusions are robust to different *bias* and sample specifications.

[Insert Table IX here]

III.F.II. Alternative Explanations

Snowberg and Wolfers (2010) show that the long-shot bias in the odds for horse races is driven by errors in expectations related to probability weighting (i.e., Quiggin, 1982). To examine whether probability weighting affects our findings, we estimate bookmaker's probabilities assuming the probability weighting function of Prelec (1998). We use two different values for the probability weighting parameter, one used in Snowberg and Wolfers (2010) (0.928) and the other proposed by Kahneman and Tversky (0.65). As shown in Table X, we find that the coefficient on GS does not change materially from our baseline specification.[34]

[Insert Table X here]

Our analysis implicitly assumes that bookmakers start from a correct prior, which they update after observing rankings information and considering the match format (MS vs GS).[35] However, what if bookmakers start their calculations from a *biased prior*, which they update correctly for both MS and GS? Due to the non-linearity of the

[34] Probability weighting does not significantly change the results, because the difference in probabilities between the favorite and the long-shot is relatively small in tennis data, and thus the shape of the probability weighting function does not change drastically in this region. Probability weighting, on the other hand, is more important for horse races, where the long-shot in each race is a very low probability event (i.e., odds of 200 to 1).

[35] Before any information on the two players is observed, it is logical to assume that each player is equally likely to win the match.

adjustment, for a specific set of biased priors,[36] we can end up with $\Delta_{bias} < 0$, even though bookmakers responded to variations in IR correctly.

The main challenge to the *biased priors* explanation for the negative Δ_{bias} for men's matches, is that it requires biases in priors of *different* nature for women's matches, since Δ_{bias} is different in these markets. It is not clear a priory why biases in priors should be so different in the two datasets. Moreover, in the case of a biased prior, the sign of *bias* in MS and GS should be the same. However, in the men's betting exchange data, *bias* is positive in MS and negative in GS (Table V, Panel A).

Notwithstanding these caveats, we can make an adjustment to the probabilities to examine whether a biased prior is sufficient to explain our results. Specifically, we can back-out the probability that the higher-ranked player wins a *single set*. Under the null, Δ_{bias} calculated from single-set probabilities should be equal to 0, even in the presence of biased priors. Under the alternative of IR-neglect $\Delta_{bias} < 0$.

Specifically, assuming that the outcomes of different sets in each match are independently and identically distributed, we can express the observed probabilities as:

$$MS\colon \pi_3 = \binom{3}{2} \times \pi_{1,3}^2 \times \left(1-\pi_{1,3}\right) + \binom{3}{3} \times \pi_{1,3}^3 \qquad (5)$$

In the above expression, π_3 is the probability that the higher-ranked player wins an MS match, and $\pi_{1,3}$ is the corresponding probability that the same player wins only one set. Similarly, we can re-write the probabilities for GS, π_5, as:

$$GS\colon \pi_5 = \binom{5}{3} \times \pi_{1,5}^3 \times \left(1-\pi_{1,5}\right)^2 + \binom{5}{4} \times \pi_{1,5}^4 \times \left(1-\pi_{1,5}\right) + \binom{5}{5} \times \pi_{1,5}^5 \qquad (6)$$

We can express both subjective and objective probabilities according to equations (5) and (6), back out the corresponding $\pi_{1,3}$ and $\pi_{1,5}$, re-calculate *bias* and test whether

[36] Specifically, $\Delta_{bias} < 0$ if the subjective prior is greater than the correct prior, and vice-versa.

it is equal across MS and GS. The results, shown in Table X column (3), show that the coefficient on GS is negative and significant, which suggests that biased priors cannot explain our results.[37]

Another possibility is that our results are driven by a more *general bias* in expectations, and not IR-neglect. To illustrate, assume that an agent receives two normally distributed signals, S_{MS} and S_{GS}, with their reliability indexed by their variance, and $\sigma_{MS}^2 > \sigma_{GS}^2$. The means of the two signals are the same. However, the agent mis-perceives their variances by the same proportion κ as: $\sigma_{Perceived,MS}^2 = \kappa\sigma_{MS}^2$ and $\sigma_{Perceived,GS}^2 = \kappa\sigma_{GS}^2$, where κ >1 (<1) leads to underreaction (overreaction). Due to the fact that the effect of κ on the posterior probability depends on the variance of the signal, Δ_{bias} is negative when κ <1. There are three challenges to this explanation: First, the sign of the biases in MS and GS must be the same, if the same κ is applied to all signals. However, in the betting exchange data, the bias is positive in MS and negative in GS (Table V, Panel A). Secondly, κ <1 implies overreaction, and therefore a positive *bias* in both MS and GS, which again is not what we observe in the data. And thirdly, it must be the case that the general bias is of *different* nature for women's matches, and it is not clear why should this be the case.

Overall, the alternative explanations we have considered do not seem to offer a parsimonious explanation for our findings, across bookmaker and betting exchange markets, for both men's and women's tennis matches.

[37] In a different scenario, biases in priors may be different for GS and MS. We cannot empirically exclude this possibility.

III.F.III. A Laboratory Experiment

Our final robustness test is a laboratory experiment which allows us to test our hypothesis in a setting that removes potential confounds arising from market-related factors.

The experiments were conducted using students at the University of Surrey, U.K. To make sure that our subjects knew about tennis, and could therefore appreciate the effect of match length on the probability that the higher-ranked player wins, we invited students that were involved in the Tennis Society to participate. Our experiment attracted 17 students,[38] who participated in two different sessions lasting roughly an hour each. The second session was conducted two weeks after the first one.[39] For both sessions students received a show-up fee of £10 and a performance-related payment, which depended on the accuracy of their predictions. The average total payment for each session was £22.

In the experiment, subjects were asked to provide a probability that each player wins for several upcoming matches. In the first session the 33 matches were from the Rome ATP World Tour Masters 1000 (MS; BO3) and in the second the 64 matches were from the French Open (GS; BO5). Both tournaments are played on the same surface (clay). For each match we provided to subjects the name and ranking of each player at the start of the tournament. Moreover, our instructions in the beginning of each session explicitly mentioned the format that these matches were played (BO3 or BO5). We used the quadratic scoring rule to incentivize subjects' choices, and the

[38] Our sample is smaller than what is usually observed in experiments because we were only interested in subjects who knew about tennis. Moreover, our sessions could only occur at a specific day and time, after the first-round matches were determined but not yet played. This time constraint also contributed to the small sample size. However, because each subject made a number of choices in each session, the total number of observations is quite large. In our models, we control for correlated responses as we cluster standard errors on the subject level.

[39] This gap was necessary because students had to make their choices after the draw for each tournament is made, and before the games are actually played.

random lottery procedure to determine their performance-related payment in each session.[40] Our full instructions, additional details about our experimental protocol, and various descriptive statistics are provided in Table A. 4 of the Appendix.

In the first session, prior to the main task, subjects where firstly asked to consider the extent to which the expression "*I follow tennis closely*" applies to them (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). The mean response to this question was 3.64, indicating that on average our subjects did indeed follow tennis. To confirm this statement, we also gave students 8 tennis-related quiz questions (4 in each session), and found that on average students answered 5.4 of those questions correctly. The correlation between the response to the first question and the number of correctly answered quiz questions was 0.7. These findings suggest that our subjects were knowledgeable about tennis.

The results are shown in Table XI. We used the same procedure as with the field data to calculate *bias* for MS and GS. In Panel A, in a univariate setting, we see that π and $\hat{p}$ for the MS matches are equal to 66.6%, and *bias* is equal to 0. For GS, we observe that $\hat{p}$ increases to 73%, but π only increases to 69.8%. Δ_{bias} is -3.2% and statistically significant, consistent with our field results from bookmakers and the betting exchange. In Panel B, we conduct a regression that controls for *RSkill*. The coefficient on *GS* remains negative at -3.3% and statistically significant.

Although the magnitude of Δ_{bias} in the experiment is very similar to the one found in the field data, closer inspection reveals that the extent of the bias is potentially larger in the laboratory. In unreported analysis, when we regressed π on *GS* and *RSkill,* we found that the coefficient on *GS* is insignificant. This means that subjects do not

[40] The quadratic scoring rule assumes subjects are risk neutral. However, experimental work has shown that subjects are risk averse in the laboratory (e.g., Holt and Laury, 2002). Because our tests are based on within-subject responses across the two sessions, assuming risk neutrality is unlikely to bias our results (assuming that risk attitude does not change systematically across the two sessions).

consider the effect of match length at all, and only increase their π's for GS matches responding to the higher *RSkill* in those matches.

One concern with our experimental design is that because there is a two week lag between the first and second session, some feedback effects may be influencing our results. For instance, it may be the case that our results are driven by "disappointed" subjects that performed poorly in the first task, who provided more conservative estimates for π in the second session.[41] To address this concern, in column (3) of Panel B, we include the payment of the subject in the first session as an additional explanatory variable.[42] The results in this specification are unchanged from those presented in column (2). Finally, in column (4) of Panel B, to capture any effect related to unobserved heterogeneity between subjects, we include subject fixed effects. The results are again unchanged in this specification.

At the end of the second session, after all the data were collected, we asked students to list the major factors that influenced their responses in order of significance (see Figure II). Player rankings was the most important factor by nearly all the students, whereas *no student mentioned match format* (MS vs. GS).

Overall, the results from the experiment suggest the IR- neglect induces biases in expectations.

[Insert Table XI and Figure II here]

[41] If the subject expresses a probability that is closer to 0.5, the ex-ante dispersion in final payment outcomes is reduced (Andersen, Fountain, Harrison, and Rutström, 2014).

[42] We stress however that the performance-related payment of the subject for the first session was actually paid at the end of the second session. However, the subject after the first session could have checked the outcome of the game that would determine her payment, and estimate the payment for the first session.

III.F.IV. Information Reliability Neglect in Other Sports Betting Markets

For our final test, we examine whether IR-neglect influences outcomes in other sports betting markets. The premise of this test is that outcomes in some sports are less noisy, which makes skill-related signals for such sports more reliable. The IR-neglect hypothesis predicts that *bias* is lower for less noisy sports.

To conduct the test we use the analysis in Mauboussin (2012) who ranks five sports according to the noise they contain, producing the following ranking (from most to least noisy): 1. Ice hockey, 2. Football, 3. Baseball, 4. Soccer and 5. Basketball (figure 1-1, pp. 23).[43] We collect data from Betfair for three of these sports (football, soccer, and basketball),[44] follow the procedure previously explained in the paper, and examine whether *bias* is higher in the more noisy sports.

The results are shown in Table XII.[45] We that the higher-skilled team wins 80.2% of the time in basketball, and 78.3% in soccer and 64.9% in football, which supports the analysis of Mauboussin (2012), that basketball is less noisy, followed by soccer and then football.

In line with the IR-neglect hypothesis, we find that average *bias* is equal to 2.8% in soccer, 1.7% in soccer and -0.4% in basketball. The hypothesis that average *bias* is equal across the three sports is strongly rejected in the data.[46]

Overall, these results suggest that IR-neglect influences outcomes in sports betting markets more broadly.

[43] To perform this calculation the author calculates empirically the degree of mean reversion in these sports. A higher mean reversion implies that luck plays a relatively more important role. For more details see the Appendix in Mauboussin (2012).
[44] Betfair contains odds data for a very small number of ice hockey and baseball games (<100), thus we do not use them in the analysis.
[45] The methodology used for this test is similar to that in column (1) of Table IX. From more details see the caption of Table XII.
[46] The results continue to hold in a multivariate regression framework, which controls for the odds difference in each match.

[Insert Table XII here]

IV. CONCLUDING REMARKS

We conduct a field test of Bayesian reasoning by examining whether agents form expectations by placing a larger weight on more reliable cues. Our results are robust to inferring probabilities from odds offered by professional bookmakers, or odds achieved in a person-to-person betting exchange. Several tests of robustness, including a placebo test using women's matches where all matches are played in the same format, and a laboratory experiment support our conclusions. Overall, our findings suggest under-reaction due to *"information reliability neglect"* influences expectations and outcomes in real-world markets.

APPENDIX A. A MODEL OF DECISION-MAKING UNDER RELIABILITY NEGLECT

Assume an agent who correctly identifies that the prior distribution for variable *m* takes the form $m \sim N(h, \sigma_P^2)$. Starting from this prior the agent receives two signals on two different occasions, $S_1 \sim N(g, \sigma_1^2)$ and $S_2 \sim N(g, \sigma_2^2)$, where $\sigma_1^2 > \sigma_2^2$ indicates their difference in reliability. To draw a parallel with the tennis analysis, *m* is a variable that maps onto the probability that a player wins the match, and *S* is the signal related to his relative skill, i.e., whether he is the higher-ranked player. The agent then observes the signal and updates his probability. Because the higher-ranked player is on average more skillful (i.e., $g > h$) the adjustment is upward for the higher-ranked player (and, conversely, downward for the lower-ranked player).

To simplify the illustration, we can set $h = 0$ without losing any generality. Under the null of Bayesian updating the agent's posterior expectation for *m* after observing S_1 is:

$$E(m|S_1) = \frac{g(\sigma_1^2)^{-1}}{(\sigma_1^2)^{-1} + (\sigma_P^2)^{-1}} \tag{1}$$

And, equivalently, his posterior expectation for *m* after observing S_2 is:

$$E(m|S_2) = \frac{g(\sigma_2^2)^{-1}}{(\sigma_2^2)^{-1} + (\sigma_P^2)^{-1}} \tag{2}$$

In this case *bias* (i.e., difference between the conditional expectation after each signal and the corresponding Bayesian posterior) is equal to 0 in both cases, and $\Delta_{bias} = bias_2 - bias_1$ is also 0.

Now suppose that the agent is not fully sensitive to variations in reliability across signals. A simple way to capture this behavior is to assume that the agent starts the calculation from a default level of reliability $\bar{\sigma}^2$, equal to, for example, the average reliability from the two signals. The agent then adjusts this default value according to the reliability in each observed signal, however this adjustment is insufficient. The cognitive mechanism driving this behavior could be "sparse thinking" as discussed by Gabaix (2014): "*To supply the missing elements....people rely on defaults-which are typically the expected values of variables...When taking into account some information, agents anchor on the default and do a limited adjusted toward the truth, as in Tversky and Kahneman's (1974) "anchoring and adjustment"* (Gabaix, 2014 pp.1663).

This mechanism leads to IR-neglect. The agent is under-estimating the variance of S_1 and over-estimating the variance of S_2, such that the perceived reliabilities of the two signals are $\bar{\sigma}_1^2 = \kappa\sigma_1^2$ and $\bar{\sigma}_2^2 = \lambda\sigma_2^2$, where $k < 1$ and $\lambda > 1$.

The agent's posterior expectation for *m* conditional on S_1 is:

$$E(m|S_1) = \frac{g(\kappa\sigma_1^2)^{-1}}{(\kappa\sigma_1^2)^{-1}+(\sigma_P^2)^{-1}} \tag{3}$$

And similarly, the agent's posterior expectation for *m* conditional on S_2 is:

$$E(m|S_2) = \frac{g(\lambda\sigma_2^2)^{-1}}{(\lambda\sigma_2^2)^{-1}+(\sigma_P^2)^{-1}} \tag{4}$$

In this Δ_{bias} is equal to:

$$\Delta_{bias} = \frac{g\sigma_P^2\sigma_2^2(1-\lambda)}{(g+\lambda\sigma_2^2)(\sigma_2^2+\sigma_P^2)} - \frac{g\sigma_P^2\sigma_1^2(1-\kappa)}{(g+\kappa\sigma_1^2)(\sigma_1^2+\sigma_P^2)} \tag{5}$$

Due to IR-neglect (i.e., $\lambda > 1$ and $\kappa < 1$), the agent *overreacts* to S_1 and *underreacts* to S_2, therefore $\Delta_{bias} < 0$.

APPENDIX B. BIAS AND EXPECTED PROFITS

Bias and Expected Profits

The overall profitability of bookmakers depends on how the objective probability of the higher-ranked player wining compares to the relative volume that backs this player. To illustrate, assume that the objective probability that the higher-ranked player wins a match is p_{HR}, and correspondingly for the low ranked player is $p_{LR} = 1 - p_{HR}$. The bookmaker starts from this probability, adjusts to reflect a housetake, *v*, and a bias *b*, arriving at p_{HR}^{*}:

$$p_{HR}^{*} = (p_{HR} + b)(1 + v) \quad (1)$$

Correspondingly, the adjusted probability for the low ranked player is:

$$p_{LR}^{*} = (1 - p_{HR} - b)(1 + v) \quad (2)$$

The odds offered for the higher- and lower-ranked player are thus $d_{HR} = \frac{1}{p_{HR}^{*}}$ and $d_{LR} = \frac{1}{p_{LR}^{*}}$, respectively. Furthermore, Vol_{HR} and Vol_{LR} are the volumes that back the higher- and lower-ranked players respectively, and $TotVol = Vol_{HR} + Vol_{LR}$ is the total volume staked.

The profit for the bookmaker for each match is therefore:

$$\Pi(\$) = TotVol - Vol_{HR}d_{HR}D_{HR} - Vol_{LR}d_{LR}D_{LR} \quad (3)$$

Dividing through by *ToVol* leads to equation (3) in the paper, where $RVol_{HR} = \frac{Vol_{HR}}{TotVol}$ and $RVol_{LR} = \frac{Vol_{LR}}{TotVol}$:

$$\Pi(\%) = 1 - RVol_{HR}d_{HR}D_{HR} - RVol_{LR}d_{LR}D_{LR} \quad (4)$$

To illustrate the conditions that a negative bias ($b < 0$) is optimal it is useful to express the above equation in expectation form, replacing the ex-post indicators D_{HR} and D_{LR} with the corresponding probabilities:

$$\mathrm{E}(\Pi) = 1 - RVol_{HR}d_{HR}p_{HR} - RVol_{LR}d_{LR}p_{LR} \tag{5}$$

Expressing the odds as a function of p_{HR}, b and v:

$$E(R\Pi) = 1 - \frac{p_{HR}RVol_{HR}}{(p_{HR}+b)(1+v)} - \frac{(1-p_{HR})(1-RVol_{HR})}{(1-p_{HR})(1+v)} \tag{6}$$

In Figure A. 1 in this section of the Appendix, we plot $E(\Pi)$ for different values of b and $RVol_{HR}$, setting $p_{HR} = 0.79$ and $v = 0.05$. From the top panel of Figure A. 1, starting with the black solid line where $RVol_{HR} = p_{HR}$, we observe that $E(\Pi)$ is inverse U-shaped, with a unique maximum at $b = 0$ equal to v. When $RVol_{HR} < p_{HR}$, $E(\Pi)$ increases as b decreases, and this relationship is steeper as $RVol_{HR}$ becomes smaller. Conversely, when $RVol_{HR} > p_{HR}$, $E(\Pi)$ decreases as b becomes smaller, and this relationship is steeper as $RVol_{HR}$ becomes larger.

In the bottom Panel of Figure A. 1, we plot the corresponding standard deviation of these profits, $\sigma(\Pi)$. If $RVol_{HR} = p_{HR}$, and $b = 0$ the bookmaker faces no risk. When $RVol_{HR} < p_{HR}$ a negative bias lowers $\sigma(\Pi)$, whereas the opposite is true when $RVol_{HR} > p_{HR}$.

This analysis shows that bookmakers increase their expected profit by allowing a negative b *only* if punters under-bet the higher-ranked player ($RVol_{HR} < p_{HR}$). If, on the other hand, punters over-bet the higher-ranked player ($RVol_{HR} > p_{HR}$), then the bookmaker is better off by increasing b, receiving higher expected profits with lower standard deviation.

[Insert Figure A. 1 here]

Hypothetical Odds

To calculate hypothetical profits $H\Pi_i$ in each GS match, we set π_{HR} equal the fitted value from the logit model, $\hat{p}$, (and π_{LR} =1-$\hat{p}$), obtain the new hypothetical odds as per equations (1) and (2) from the previous section in the Appendix, $HO_{HR,i}$ and $HO_{LR,i}$, and calculate the hypothetical profits, $H\Pi$ following equation (4) in Section III.D of the paper.

However, before we use the hypothetical odds to determine profits we need to consider how changing the odds would influence *RVol* in this counterfactual scenario. To obtain an estimate for the hypothetical $RVol_{HR}$ we use the model below, estimated using the betting exchange data for GS matches:

$$RVol_{HR,i} = \alpha + \beta O_{HR,i} + \epsilon_i$$

We find that α is equal to 1.43 and β is equal to -0.42, both highly statistically significant. This shows that when the odds on the higher-ranked player increase (i.e., when he becomes less of a favorite), $RVol_{HR}$ decreases. This result reflects the preference of punters to bet on the favorite (e.g., Levitt, 2004). We obtain the hypothetical $RVol_{HR}$ for each GS match by calculating $\hat{\alpha} + \hat{\beta} HO_{HR,i}$. We use hypothetical volumes and odds, and the actual game outcomes (as indicated by D_{HR}) to calculate profits in this counterfactual scenario using equation (4) in Section III.D of the paper.

APPENDIX C. LABORATORY EXPERIMENT

Instructions and Procedures

In this section, we include the full-instructions seen by the students, shown below. The statements in brackets in italics are clarifying additions, not seen by the students.

Welcome to the experiment!

This experiment is about predicting the outcome of professional tennis matches. The matches that you will be predicting in this session are matches that will be played at the ATP 1000 Masters event that is happening right now in Rome. The winner is decided in a best-of-three-set match. Before moving to the predictions please answer the questions below. For each of the quiz questions in Section 3 that you answer correctly you will receive an additional 50 pence (£0.50) at the end of the session.

[*For the second session, the second and third sentence where replaced with:* The matches that you will be predicting in this session are matches that will be played at the French Open that is happening right now in Paris. The winner is decided in a best-of-five-set match.*]*

[*Students then provided the following information: Name, student ID, and Degree title*.]

[*Students then were asked to circle the answer that most applied to them in relation to the statement "I follow tennis closely", from "strongly agree" to "strongly disagree" (5 levels)*]

[*Students then answered 4 tennis related trivia questions, like "How many Grand Slam titles has Roger Federer won?*]

[*Students then proceeded to the next part of the instructions*]

About probabilities

For each match we will tell you the players involved and their corresponding rankings. Your task is to think how likely each player is to win the match, and express this likelihood as a probability between 0% and 100%. Of course, we do not know yet the outcome of the match, so there is *uncertainty*.

To provide you with some background about probabilities, suppose we have a normal six-sided die, and you want to calculate the probability of rolling the die and getting a three (3). To calculate this probability you need to count how many outcomes lead to a "win", and divide that by how many outcomes are possible. In this case, a win occurs with only one outcome, which is to roll the die and get a three (3). There are six (6) possible outcomes (numbers 1 to 6) and, therefore, the probability of rolling a three (3) in one roll is 1/6 (or roughly 17%). If instead you want to calculate the probability of rolling the die and getting an even number (2, 4, 6), then this probability is higher because you win with three outcomes instead of one, therefore the probability of winning in one roll is 3/6 (or 50%).

A similar logic can be applied when thinking about probabilities of players winning a tennis match. For example, suppose that we have a match between players A and B. To calculate the probability of each player winning this match you need to estimate how many matches each player would win, if they hypothetically played 100 matches in total. For example, suppose that you think that B is a better player than A, so B would win 70 matches out of 100 (and correspondingly A would win 30 matches out of 100). This means that your estimated probability of B winning the match is 70/100 (or 70%) and of A 30/100 (or 30%).

Your task is to think along these lines for each match and provide your best estimate of the probability of each player winning the match. Note that the

probabilities you express for the two players in each match must equal to 100% when added together! (in the above example 70%+30%=100%)

Please note that in our analysis we will use your predictions in the experiment **anonymously**.

[Students first read the above silently, and then it is read aloud by the experiment asking for clarifying questions.]

[Students then proceeded to the next part of the instructions]

How you will get paid for your predictions.

Apart from the money that you will be paid for showing up and for answering correctly the quiz questions, you will also be paid a **performance-based payment**. This payment will be based on the accuracy of your predictions using a procedure that is commonly used in economics experiments, known as the *quadratic scoring rule*.

[*Students received £0.5 for each quiz question answered correctly*]

To illustrate how this works, let's return to our hypothetical example between players A and B where your best guess of the probability that B wins is 70% (or 0.7). In this case, your payment would be calculated using the formula below:

$$Payment = £13 - £13(D - p)^2$$

The symbol $\boldsymbol{p}$ in the equation stands for your estimate of the probability that player B wins, in this case 0.7. $\boldsymbol{D}$ is an indicator, which takes the value of 1 if B has won the match, or 0 if A has won the match. Of course you do not know $\boldsymbol{D}$ when you make your choices, this will be revealed after the match is played and we know the winner.

Assume, for example, that the match is now finished and player B has won. This means your prediction in this case was good in the sense that you assigned a higher probability to B winning. Your payment in this case would be:

$$Payment = £13 - £13(1 - 0.7)^2 = £13 - £13(0.3)^2 = £13 - £1.17 = £11.83$$

[*This equation is written on the board and explained by the experimenter, when he read the instructions aloud*]

If you had assigned a higher probability to B winning, say 0.8, your payment would be even higher and equal to:

$$£13 - £13(1 - 0.8)^2 = £12.48$$

However, you need to be careful, because the higher the probability you give for B winning, *the lower your payment will be if A wins.*

For example, when your probability of B winning is 0.7, your payment if A wins is:

$$£13 - £13(0 - 0.7)^2 = £13 - £13(-0.7)^2 = £13 - £6.37 = £6.63$$

When your probability of B winning is 0.8, your payment if A wins is even lower:

$$£13 - £13(0 - 0.8)^2 = £13 - £13(-0.8)^2 = £13 - £8.32 = £4.68$$

In general, with this payoff scheme, the largest possible payment is £13, which you will receive if you assign a 100% chance to B wining, and B wins. However, if you choose to assign 100% probability to B but A wins, your payment will be the lowest possible, equal to £0.

Since your prediction is made before the match is played and therefore you do not know what will actually happen, **the best thing you can do to maximize the expected payoff is to simply state what is your best guess for the probability that**

B wins the match. Any other prediction will *decrease* the amount you can expect to earn.

You will assign probabilities to players for several matches, and at the end we **will randomly choose one of these matches** to calculate your performance payment. This random choice will be made in your presence, at the end of this session. Of course we won't know the outcome of the match at the end of the session, so we will calculate your payment in a few days after the winner is announced. You will collect your performance-payment for this session after the second session of this experiment is completed. The details on how to collect your performance-based payment will be given to you at the end of the second session.

The second session will be held on Sunday 28th May at 10:00 a.m.

[*Students first read the above silently, and then it is read aloud by the experiment asking for clarifying questions.*]

[*Students then proceeded to provide their choices for the different matches*]

[*When all the students finished, each student approached the experimenter individually to receive the show-up fee and to select the match that would determine their performance-related payoff. The performance-related payoff for the first session was paid at the end of the second session, and the performance related payment for the second session was paid a few days after the end of that session when the games were completed.*]

[*In our sessions we also asked students to make predictions for the women's matches in the corresponding tournaments, and found results that are in line with those in the paper for the betting exchange (i.e., $\Delta_{bias} > 0$). To conserve space we do not report these results here, but they are available from the authors on request. The performance-*

based payment for each student in each session was calculated on the basis of one randomly selected match].

REFERENCES

Al-Ubaydli, O., List, J. A., Suskind, D. L., 2017. What can we learn from experiments? Understanding the threats to the scalability of experimental results. American Economic Review 107, 282-286.

Andersen, S., Fountain, J., Harrison, G. W., Rutström, E. E., 2014. Estimating subjective probabilities. Journal of Risk and Uncertainty 48, 207-229.

Antoniou, C., Harrison, G. W., Lau, M. I., Read, D., 2017. Information characteristics and errors in expectations: Experimental evidence. Journal of Financial and Quantitative Analysis 52, 737-750.

Ball, R., Brown, P., 1968. An empirical evaluation of accounting income numbers. Journal of Accounting Research 159-178.

Benjamin, D. J., Rabin, M., Raymond, C., 2016. A model of nonbelief in the law of large numbers. Journal of the European Economic Association 14, 515-544.

Camerer, C. F., 2015. The promise and success of lab-field generalizability in experimental economics: A critical reply to levitt and list. In: G. Fréchette, Schotter, A. (Ed.), Handbook of experimental economic methodology. Oxford University Press, pp.

Charness, G., Karni, E., Levin, D., 2010. On the conjunction fallacy in probability judgment: New experimental evidence regarding linda. Games and Economic Behavior 68, 551-556.

Chen, D. L., Moskowitz, T. J., Shue, K., 2016. Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires. Quarterly Journal of Economics 131, 1181-1242.

Coval, J. D., Shumway, T., 2005. Do behavioral biases affect prices? Journal of Finance 60, 1-34.

Croxson, K., Reade, J. J., 2014. Information and efficiency: Goal arrival in soccer betting. Economic Journal 124, 62-91.

Dawes, R. M., Faust, D., Meehl, P. E., 1989. Clinical versus actuarial judgment. Science 243, 1668-1674.

De Bondt, W. F. M., Thaler, R. H., 1990. Do security analysts overreact? American Economic Review 52-57.

Del Corral, J., Prieto-Rodríguez, J., 2010. Are differences in ranks good predictors for grand slam tennis matches? International Journal of Forecasting 26, 551-563.

Ernst, M. O., Banks, M. S., 2002. Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415, 429-433.

Fama, E. F., 1998. Market efficiency, long-term returns, and behavioral finance. Journal of Financial Economics 49, 283-306.

Forrest, D., McHale, I., 2007. Anyone for tennis (betting)? European Journal of Finance 13, 751-768.

Gabaix, X., 2014. A sparsity-based model of bounded rationality. Quarterly Journal of Economics 129, 1661-1710.

Gennaioli, N., Shleifer, A., 2010. What comes to mind. Quarterly Journal of Economics 125, 1399-1433.

Grether, D. M., 1980. Bayes rule as a descriptive model: The representativeness heuristic. Quarterly Journal of Economics 95, 537-557.

Griffin, D., Tversky, A., 1992. The weighing of evidence and the determinants of confidence. Cognitive Psychology 24, 411-435.

Harrison, G. W., List, J. A., 2004. Field experiments. Journal of Economic Literature 42, 1009-1055.

Holt, C. A., Laury, S. K., 2002. Risk aversion and incentive effects. American Economic Review 92, 1644-1655.

Ikenberry, D., Lakonishok, J., Vermaelen, T., 1995. Market underreaction to open market share repurchases. Journal of Financial Economics 39, 181-208.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. Journal of Finance 48, 65-91.

Jiang, G. J., Zhu, K. X., 2017. Information shocks and short-term market underreaction. Journal of Financial Economics 124, 43-64.

Kahneman, D., Slovic, P., Tversky, A., 1982. Judgement under uncertainty: Heuristics and biases. Cambridge University Press, Cambridge.

Kahneman, D., Tversky, A., 1972. Subjective probability: A judgment of representativeness. Cognitive Psychology 3, 430-454.

Klaassen, F., Magnus, J. R., 2014. Analyzing wimbledon: The power of statistics. Oxford University Press, Oxford.

Klaassen, F. J. G. M., Magnus, J. R., 2001. Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. Journal of the American Statistical Association 96, 500-509.

Lacetera, N., Pope, D. G., Sydnor, J. R., 2012. Heuristic thinking and limited attention in the car market. American Economic Review 102, 2206-2236.

Levitt, S. D., 2004. Why are gambling markets organised so differently from financial markets? Economic Journal 114, 223-246.

Levitt, S. D., List, J. A., 2008. Homo economicus evolves. Science 319, 909-910.

List, J. A., 2003. Does market experience eliminate market anomalies? Quarterly Journal of Economics 118, 41-71.

Loughran, T., Ritter, J. R., 1995. The new issues puzzle. Journal of Finance 50, 23-51.

Mauboussin, M. J., 2012. The success equation: Untangling skill and luck in business, sports, and investing. Harvard Business Press,

Michaely, R., Thaler, R. H., Womack, K. L., 1995. Price reactions to dividend initiations and omissions: Overreaction or drift? Journal of Finance 50, 573-608.

Moskowitz, T. J., 2015. Asset pricing and sports betting. University of Chicago.

Pope, D. G., Schweitzer, M. E., 2011. Is tiger woods loss averse? Persistent bias in the face of experience, competition, and high stakes. American Economic Review 101, 129-157.

Prelec, D., 1998. The probability weighting function. Econometrica 497-527.

Quiggin, J., 1982. A theory of anticipated utility. Journal of Economic Behavior & Organization 3, 323-343.

Simonsohn, U., Loewenstein, G., 2006. Mistake# 37: The effect of previously encountered prices on current housing demand. Economic Journal 116, 175-199.

Smith, M. A., Paton, D., Williams, L. V., 2009. Do bookmakers possess superior skills to bettors in predicting outcomes? Journal of Economic Behavior & Organization 71, 539-549.

Snowberg, E., Wolfers, J., 2010. Explaining the favorite–long shot bias: Is it risk-love or misperceptions? Journal of Political Economy 118, 723-746.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. Science 185, 1124-1131.

Zhang, F. X., 2006. Information uncertainty and stock returns. Journal of Finance 61, 105-137.

TABLE I

Observations by Tournament and Year

	Year										
Tournament	05	06	07	08	09	10	11	12	13	14	Total
Panel A: Grand Slam (GS)											
Australian Open	111	112	111	114	112	112	108	108	117	99	1,104
French Open	107	102	101	111	109	115	114	114	112	116	1,101
US Open	109	108	107	117	109	109	104	114	108	109	1,094
Wimbledon	105	107	98	101	103	107	107	108	105	108	1,049
Other	6	6	1	0	0	0	0	0	0	0	13
Total	438	435	419	443	433	443	433	444	442	432	4,361
Panel B: Masters (MS)											
Cincinnati	57	59	47	43	46	45	47	51	46	51	492
Hamburg	53	53	48	47	0	0	0	0	0	0	201
Indian Wells	81	78	85	78	80	79	72	80	74	85	792
Madrid	40	36	43	41	43	45	49	50	53	50	450
Miami	75	69	77	79	81	79	77	82	73	76	768
Monte Carlo	50	59	46	45	50	48	48	50	47	48	491
Montreal	55	0	49	0	50	0	50	0	49	0	253
Paris	37	38	43	38	41	43	41	42	43	40	406
Rome	54	46	46	42	51	49	45	49	51	48	481
Shanghai	0	0	0	0	44	47	48	52	49	50	290
Toronto	0	51	0	52	0	47	0	44	0	51	245
Total	502	489	483	465	486	482	477	500	485	499	4,869

Notes. This table shows a breakdown of the sample used in the analysis sorted by tournament and year. The data are retrieved from www.tennis-data.co.uk. Panel A contains data for Grand Slam (GS) matches which are played in a best-out-of-five format, and Panel B for ATP World Tour Masters 1000 (MS) matches, which are played in a best-out-of-three format. We drop non-completed matches, matches with missing information or negative housetake, and matches where the higher-ranked player is indicated as an outsider by the odds even though he is ranked by at least 15 places higher than then lower-ranked player. Our final sample contains 9,230 matches from 2005-2014.

TABLE II

Descriptive Statistics

Variable	Mean	σ	Min	Q1	Median	Q3	Max
			Panel A: GS				
HR_{Odds}	1.35	0.42	1.00	1.10	1.24	1.48	6.79
LR_{Odds}	6.05	5.73	1.09	2.62	4.01	7.17	60.00
Housetake	0.05	0.01	0.00	0.05	0.06	0.06	0.11
HR_{Rank}	28	29	1	7	19	38	279
LR_{Rank}	99	96	2	48	80	119	1,370
$\hat{p}$	0.78	0.10	0.60	0.70	0.78	0.86	0.99
π_{HR}	0.74	0.15	0.14	0.64	0.76	0.87	0.98
Prize Money ($)	2,525,000						
			Panel B: MS				
HR_{Odds}	1.48	0.46	1.01	1.19	1.38	1.61	7.45
LR_{Odds}	4.10	3.25	1.09	2.27	3.01	4.58	32.80
Housetake	0.06	0.01	0.01	0.05	0.06	0.06	0.09
HR_{Rank}	22	22	1	6	15	30	414
LR_{Rank}	67	83	2	29	49	78	1,517
$\hat{p}$	0.69	0.12	0.50	0.58	0.67	0.78	0.98
π_{HR}	0.68	0.15	0.13	0.59	0.69	0.79	0.97
Prize Money ($)	790,000						

Notes. This table shows descriptive statistics for the main variables used in our analysis. HR_{Odds} (LR_{Odds}) are the average decimal odds offered by the betting houses that the higher-ranked (lower-ranked) player wins a match. *Housetake* is the housetake, which is obtained by summing the inverse of the odds for the higher- and lower-ranked players and subtracting one. HR_{Rank} and LR_{Rank} are the rankings for the higher- and lower-ranked player, respectively. $\hat{p}$ is the estimated "objective" probability, obtained from averaging the predicted values from the logit model shown in equation (1) for MS and GS. π_{HR} is the subjective probability that the higher-ranked player wins the match, as derived from the bookmaker odds. The sample consists of 9,230 matches that satisfy the criteria outlined in Table I. The last row in the table shows the average prize money collected by the winner of an *MS* and a *GS* tournament using 2015 prize money.

TABLE III

Likelihood that Higher-ranked Player Wins

Variable	(1)	(2)	(3)
GS	0.095***	0.072***	0.069***
	(0.009)	(0.009)	(0.009)
RSkill		0.122***	0.124***
		(0.005)	(0.005)
Surface F.E.	NO	NO	YES
Year F.E.	NO	NO	YES
Round F.E.	NO	NO	YES
N	9,230	9,230	9,230
pseudo-R^2	0.010	0.071	0.072

Notes. This table presents results from logit models, where the dependent variable takes the value of 1 if the higher-ranked player has won the match, and 0 otherwise. *GS* is a dummy variable that equals 1 if the match is GS, and 0 otherwise. *RSkill* is calculated as the log (lower-ranked player ranking) - log (higher-ranked player ranking). The table reports marginal effects associated with each of the independent variables. The sample consists of 9,230 matches that satisfy the criteria outlined in Table I. The robust standard errors shown in brackets are calculated using the Huber-White estimator. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

TABLE IV

Biases in Subjective Probabilities

Panel A: Univariate

	MS	GS	GS – MS
π	0.681	0.743	
$\hat{p}$	0.686	0.780	
bias	-0.004	-0.037	-0.033***
			(0.002)
N	4,869	4,361	

Panel B: Multivariate

Variable	(1)	(2)
GS	-0.031***	-0.034***
	(0.002)	(0.002)
RSkill	-0.007***	-0.005***
	(0.001)	(0.001)
Surface F.E.	NO	YES
Year F.E.	NO	YES
Round F.E.	NO	YES
N	9,230	9,230
R^2	0.027	0.050

Notes. This table reports biases in subjective probabilities for GS and MS tennis matches. In Panel A, we present univariate analysis, and in Panel B multivariate analysis. In Panel A, π and $\hat{p}$ denote the average subjective and the average (estimated) objective probabilities that the higher-ranked player wins an MS or a GS match. $\hat{p}$ is obtained by averaging the predicted values from the logit model shown in equation (1) for MS and GS, and π is derived from the average odds on the higher-ranked player to win the match offered by the bookmakers. *bias* is the average difference between π and $\hat{p}$ for MS and GS. In Panel B, we present results from OLS regressions with an intercept. The dependent variable is *bias* and the independent variables are *GS* and *RSkill*, defined as in Table III. The sample consists of 9,230 matches from 2005 to 2014 that satisfy the criteria outlined in Table I. In Panel B, the robust standard errors shown in brackets are calculated using the Huber-White estimator. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

TABLE V

Biases in Subjective Probabilities – Betting Exchange Data

Panel A: Univariate

	MS	GS	GS – MS
π	0.701	0.764	
$\hat{p}$	0.696	0.782	
bias	0.005	-0.018	-0.023***
			(0.003)
N	2,751	2,142	

Panel B: Multivariate

Variable	(1)	(2)
GS	-0.022***	-0.026***
	(0.003)	(0.003)
RSkill	-0.001	0.001
	(0.001)	(0.002)
Surface F.E.	NO	YES
Year F.E.	NO	YES
Round F.E.	NO	YES
N	4,893	4,893
R^2	0.010	0.023

Notes. This table reports biases in subjective probabilities for GS and MS tennis matches. The analysis in Panel A and in Panel B for columns (1) and (2) is the same as in Table IV. We drop non-completed matches, matches with missing information or housetake greater than 0.05, and matches which do not satisfy criterion (ii) in Section II.A of the paper. The sample consists of 2,751 MS matches and 2,142 GS matches from 2009-2014. In Panel B, the robust standard errors shown in brackets are calculated using the Huber-White estimator. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

TABLE VI

Bookmaker Profits

Variable	r_{HR} (1)	r_{LR} (2)	Π (3)	HΠ (4)	HΠ - Π (4) – (3)
MS	0.046***	0.101***	0.051***		
	(0.013)	(0.031)	(0.007)		
	2,745	2,745	2,745		
GS	0.013	0.278***	0.025***	0.051***	0.025***
	(0.013)	(0.039)	(0.007)	(0.007)	(0.005)
	2,142	2,142	2,142	2,142	
Δ(GS - MS)	-0.031*	0.178***	-0.026**		
	(0.019)	(0.050)	(0.010)		

Notes. This table reports average values for different metrics of profitability using bookmaker odds. The top row shows the variable analyzed. r_{HR} and r_{LR} are the per unit profit earned by the bookmaker for the higher- and lower-ranked player, respectively, and is calculated using equation (3). Π is an estimate of the actual profit earned by the bookmakers, using volume information from the betting exchange, and is calculated using equation (4). $H\Pi$ is the average hypothetical profit earned by bookmakers for GS matches, in a hypothetical world where for each GS match bookmakers raised their subjective probability on the higher-ranked player and set equal to $\hat{p}$ from the logit model. The procedure used for this test is explained in Appendix B. The final column presents the difference between $H\Pi$ and Π for GS matches. *Δ(GS – MS)* shows the difference between MS and GS. The standard errors are shown in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

TABLE VII

Likelihood that Higher-ranked Player Wins – Women's Data

	Panel A: Bookmaker		**Panel B: Betting Exchange**	
	(1)	(2)	(3)	(4)
GS	0.043***	0.017	0.028*	0.004
	(0.012)	(0.011)	(0.015)	(0.015)
RSkill		0.128***		0.123***
		(0.006)		(0.008)
Surface F.E.	NO	YES	NO	YES
Year F.E.	NO	YES	NO	YES
Round F.E.	NO	YES	NO	YES
N	6,151	6,151	3,684	3,684
R^2	0.002	0.065	0.001	0.060

Notes. This table presents results from logit models, where the dependent variable takes the value of 1 if the higher-ranked player has won the match, and 0 otherwise. All variables are defined as in Table III and Table IV. In columns (1) and (2), we estimate subjective probabilities using the bookmaker odds, and in columns (3) and (4) using the betting exchange odds. The robust standard errors shown in brackets are calculated using the Huber-White estimator. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

TABLE VIII

Biases in Subjective Probabilities – Women's Data

	Panel A: Bookmaker		**Panel B: Betting Exchange**	
	(1)	(2)	(3)	(4)
GS	0.001	0.003	0.016***	0.016***
	(0.002)	(0.003)	(0.004)	(0.003)
RSkill	-0.008***	-0.008***	-0.000	-0.000
	(0.001)	(0.001)	(0.002)	(0.002)
Surface F.E.	NO	YES	NO	YES
Year F.E.	NO	YES	NO	YES
Round F.E.	NO	YES	NO	YES
N	6,151	6,151	3,684	3,684
R^2	0.007	0.017	0.005	0.008

Notes. This table reports biases in subjective probabilities for GS and MS women's tennis matches, presenting results from OLS regressions that include an intercept. The dependent variable is *bias* and the independent variables are *GS* and *RSkill*, defined as in Table IV. In columns (1) and (2), we present results when subjective probabilities are derived from the bookmaker odds, and in columns (3) and (4) when subjective probabilities are derived from the betting exchange odds. The robust standard errors shown in brackets are calculated using the Huber-White estimator. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

TABLE IX

Robustness Checks

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Bookmaker						
GS	-0.034***	-0.030***	-0.042***	-0.024***	-0.030***	
	(0.002)	(0.009)	(0.002)	(0.003)	(0.000)	
RSkill	-0.002*	-0.007*	-0.004	-0.007***		-0.003**
	(0.001)	(0.004)	(0.001)	(0.001)		(0.001)
RStreak				-0.017***		
				(0.004)		
ROdds					-0.001***	
					(0.000)	
RSkill x GS						-0.010***
						(0.001)
N	9,230	9,230	8,290	6,362	10,266	9,230
R^2	0.048	0.003	0.058	0.069	0.533	0.038
Panel B: Betting Exchange						
GS	-0.026***	-0.021*	-0.028***	-0.024***	-0.028***	
	(0.003)	(0.012)	(0.004)	(0.004)	(0.001)	
RSkill	0.003**	0.002	0.007***	0.005**		0.005***
	(0.002)	(0.005)	(0.002)	(0.002)		(0.002)
RStreak				-0.020***		
				(0.006)		
ROdds					0.0002***	
					(0.000)	
RSkill x GS						-0.007***
						(0.001)
N	4,893	4,893	4,164	3,524	5,406	4,893
R^2	0.023	0.005	0.074	0.078	0.259	0.051

Notes. This table reports various robustness checks. In column (1), we estimate the objective probability by sorting the sample for each format by *RSkill* in deciles, and obtaining the frequency that the higher-ranked player wins in each decile, $p_{HR.}$ We then calculate *bias* as π - p_{HR}. In column (2), we estimate bias as π - D, where D is a dummy that equals 1 if the higher-ranked player has won the match and 0 otherwise. In column (3), we estimate $\hat{p}$ with a logit model that uses only backward-looking information (i.e., for matches in 2006 we use observations in 2005, for matches in 2007 we use observations in 2005 and 2006, etc.). In column (4), we add *RStreak* as an additional control variable, calculated as the difference in the proportion of matches won by the higher- and lower-ranked players in the previous two tournaments. In column (5), we define the higher-ranked player based on odds, and in column (6), we test our hypothesis by interacting *GS* with *RSkill*. More details for these tests are provided in Section III.F of the paper. All models include an intercept term and surface, year, and round fixed effects. In Panel A, we derive subjective probabilities using bookmaker odds and in Panel B using betting exchange odds. The robust standard errors shown in brackets are calculated using the Huber-White estimator. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

TABLE X

Alternative Explanations

Variable	(1)	(2)	(3)
	α=0.928	α=0.65	
Panel A: Bookmaker			
GS	-0.033***	-0.036***	-0.018***
	(0.002)	(0.003)	(0.002)
RSkill	-0.004***	-0.008***	-0.007***
	(0.001)	(0.001)	(0.001)
N	9,230	9,230	9,230
R^2	0.047	0.044	0.054
Panel B: Betting Exchange			
GS	-0.026***	-0.032***	-0.012***
	(0.003)	(0.004)	(0.002)
RSkill	0.001	-0.006***	0.002*
	0.002	(0.002)	(0.001)
N	4,893	4,893	4,893
R^2	0.022	0.028	0.018

Notes. This table presents estimates from OLS models that include an intercept term and surface, year, and round fixed effects. The dependent variable in all models is *bias*, which is however derived from different expressions for the subjective probabilities. In columns (1) and (2), we adjust subjective probabilities assuming the probability weighting function in Prelec (1998), using two different values for the probability weighting parameter (α). In column (3), we calculate *bias* using single-stage probabilities for GS and MS tennis matches, using equations (5) and (6) in Section III.F.II. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

TABLE XI

Laboratory Experiment

Panel A: Univariate

	MS	GS	GS – MS	
π	0.666	0.698		
$\hat{p}$	0.666	0.730		
bias	0.000	-0.032	-0.032**	
			(0.007)	
N	459	1,071		

Panel B: Multivariate

Variable	(1)	(2)	(3)	(4)
GS	-0.032**	-0.033**	-0.033**	-0.033**
	(0.012)	(0.012)	(0.012)	(0.012)
RSkill		0.003	0.003	0.003
		(0.007)	(0.007)	(0.007)
Pay			0.009**	
			(0.004)	
Subject FE	NO	NO	NO	YES
N	1,530	1,530	1,530	1,530
R^2	0.012	0.013	0.050	0.325

Notes. This table reports biases in subjective probabilities for GS and MS men’s tennis matches, presenting results from OLS regressions that include an intercept. The dependent variable is *bias* and the independent variables are *GS* and *RSkill*, defined as in Table III. Subjective probabilities, π, were obtained from a laboratory experiment, where 17 subjects expressed probabilities for 97 professional men's tennis matches, 33 played in an MS format and 64 played in a GS format. Objective probabilities,$\hat{p}$, where obtained using the logit model in equation (1) in Section II.B of the paper, when fitted to these matches. *RSkill* is calculated as in the previous tables. *Pay* is the amount of money (in £’s) won by the subject after the end of the first experimental session. The experiments were conducted at the University of Surrey, and included students who where involved in the Tennis Society. We incentivized students’ choices using the quadratic scoring rule, and used the random lottery procedure to determine the payment. Details on our experimental procedures and descriptive statistics are shown in Appendix C. In both panels, the standard errors shown in brackets are clustered on the subject level. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

TABLE XII

Biases in Football, Soccer and Basketball

	Football	Soccer	Basketball	Equality F-Test (p-value)
π	0.677	0.800	0.797	
p_{HR}	0.649	0.783	0.802	
bias	0.028	0.017	-0.004	0.000
N	1,108	163,583	10,042	

Notes. This table reports biases in subjective probabilities for football, soccer, and basketball matches. The data for this test were obtained directly from Betfair. We retain the last odds posted on each outcome before the match begins (indicated as out of play). We drop matches where the absolute value of the housetake is greater than 0.05. Our final sample consists of 1,108 football matches, 163,583 soccer matches and 10,042 basketball matches, for the period 2004-2016. To calculate *bias,* we follow the procedure in column (1) of Table IX, ranking all matches in each sport into deciles according to the difference in the odds (because football has three possible outcomes, we take the difference between the most and least favorite outcome). We use the proportion of times that the favorite outcome materializes in each decile as the estimate for the objective probability. In the last column, we present the p-value from an F-test that examines whether *bias* is equal across the three sports.

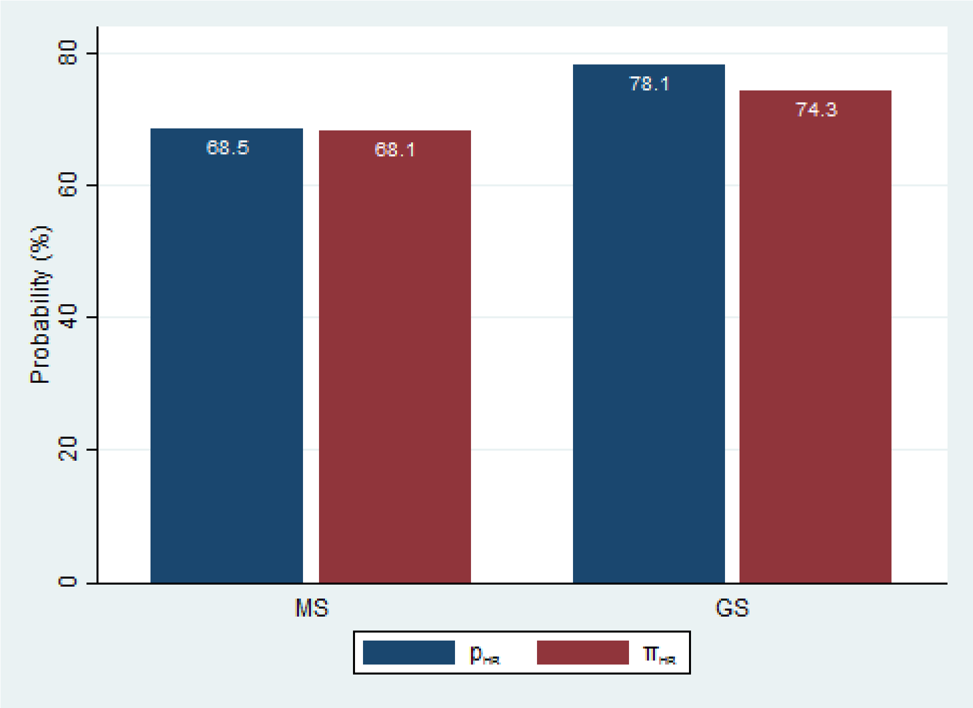

FIGURE I

Probabilities in MS vs GS

This figure depicts average objective probabilities (blue bars) and average subjective probabilities (red bars) for MS and GS matches.

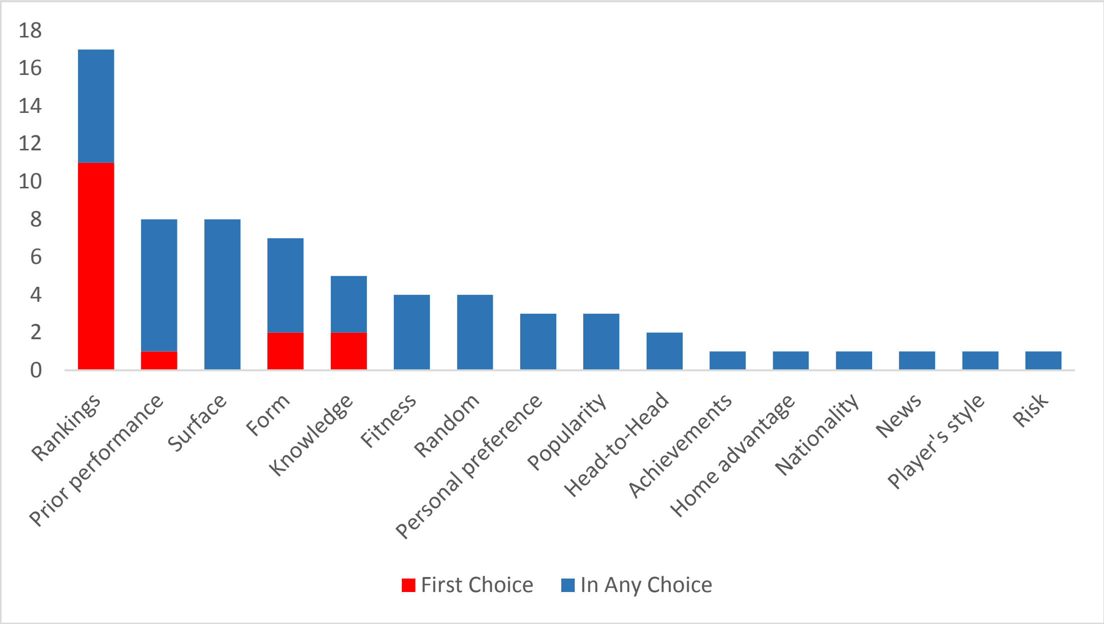

FIGURE II

Factors Affecting Probabilities in the Lab Experiment

This chart shows the factors that subjects considered when assigning probabilities in the experiment. Subjects were asked to rank factors according to how significant they were for their expectations. The red bar indicates how often subjects' indicated a specific factor as their first choice. The blue bar indicates how often subjects listed a specific factor in any of their choices.

TABLE A. 1

Descriptive Statistics – Men's Betting Exchange Data

Variable	Mean	σ	Min	Q1	Median	Q3	Max
Panel A: GS							
HR_{Odds}	1.39	0.49	1.01	1.09	1.26	1.52	8.60
LR_{Odds}	10.06	14.25	1.12	2.84	4.70	10.50	100.00
Housetake	0.01	0.01	0.00	0.00	0.01	0.01	0.05
HR_{Rank}	27	29	1	7	18	36	285
LR_{Rank}	96	91	2	45	78	118	1,120
TotVol	240	450	4	44	86	210	6,200
RVol	0.86	0.20	0.02	0.85	0.94	0.97	1.00
$\hat{p}$	0.78	0.11	0.58	0.70	0.79	0.87	0.99
π_{HR}	0.76	0.17	0.12	0.65	0.79	0.91	0.99
Panel B: MS							
HR_{Odds}	1.52	0.54	1.01	1.21	1.40	1.66	9.00
LR_{Odds}	5.69	7.37	1.12	2.46	3.40	5.60	90.00
Housetake	0.01	0.01	0.00	0.00	0.01	0.01	0.05
HR_{Rank}	21	23	1	5	14	29	414
LR_{Rank}	64	74	2	27	47	77	1,171
TotVol	170	220	2.3	39	84	200	2,200
RVol	0.82	0.23	0.02	0.81	0.91	0.96	1.00
$\hat{p}$	0.70	0.13	0.50	0.59	0.69	0.80	0.99
π_{HR}	0.70	0.16	0.11	0.60	0.71	0.82	0.99

Notes. This table shows descriptive statistics for the main variables used in the analysis with the betting exchange data. The variables are defined in Table II. *TotVol* is the total volume that backs the higher-ranked player ($000's), and *RVol* is the proportion of the total volume bet in the match on both players that backs the higher-ranked player. π_{HR} is the subjective probability that the higher-ranked player wins the match, as derived from the betting exchange odds. We drop non-completed matches, matches with missing information or housetake greater than 0.05, and matches which do not satisfy criterion (*ii*) in Section II.A. The sample consists of 2,751 MS matches and 2,142 GS matches from 2009-2014.

TABLE A. 2

Descriptive Statistics – Women's Bookmaker Data

Variable	Mean	σ	Min	Q1	Median	Q3	Max
Panel A: GS							
HR_{Odds}	1.38	0.46	1.01	1.12	1.27	1.51	9.40
LR_{Odds}	5.17	4.05	1.06	2.55	3.71	6.34	35.40
Housetake	0.06	0.01	0.00	0.05	0.06	0.06	0.08
HR_{Rank}	28	27	1	8	19	38	252
LR_{Rank}	94	84	2	49	80	113	1,208
$\hat{p}$	0.75	0.11	0.55	0.65	0.74	0.83	0.98
π_{HR}	0.73	0.15	0.10	0.63	0.74	0.85	0.97
Prize Money ($)	2,525,000						
Panel B: MS							
HR_{Odds}	1.46	0.42	1.01	1.19	1.36	1.59	5.47
LR_{Odds}	3.91	2.52	1.14	2.31	3.13	4.51	22.00
Housetake	0.06	0.01	0.02	0.05	0.06	0.06	0.09
HR_{Rank}	24	25	1	7	16	35	248
LR_{Rank}	73	83	2	32	54	86	1,132
$\hat{p}$	0.70	0.11	0.53	0.60	0.69	0.79	0.98
π_{HR}	0.69	0.14	0.17	0.59	0.70	0.79	0.96
Prize Money ($)	665,000						

Notes. This table shows descriptive statistics for the main variables used in the analysis of the women bookmaker data. The variables and the sample construction criteria are defined as in Table II. The sample consists of 2,527 MS matches and 3,624 GS matches from 2007 to 2014.

.

TABLE A. 3

Descriptive Statistics – Women's Betting Exchange Data

Variable	Mean	σ	Min	Q1	Median	Q3	Max
Panel A: GS							
HR_{Odds}	1.44	0.54	1.01	1.15	1.33	1.57	11.50
LR_{Odds}	6.88	8.62	1.09	2.68	3.90	7.20	100
Housetake	0.01	0.01	0.00	0.01	0.01	0.01	0.05
HR_{Rank}	27	27	1	8	19	37	252
LR_{Rank}	92	82	2	48	79	113	1,208
TotVol	120	250	0.4	19	43	110	2,900
RVol	0.85	0.21	0.01	0.84	0.93	0.97	1.00
$\hat{p}$	0.74	0.11	0.55	0.65	0.73	0.82	0.97
π_{HR}	0.73	0.16	0.09	0.63	0.75	0.86	0.99
Panel B: MS							
HR_{Odds}	1.52	0.47	1.02	1.24	1.41	1.67	6.20
LR_{Odds}	4.56	4.18	1.17	2.44	3.30	4.80	46.00
Housetake	0.01	0.01	0.00	0.01	0.01	0.01	0.05
HR_{Rank}	24	24	1	7	16	35	217
LR_{Rank}	71	77	2	32	53	82	1,004
TotVol	56	82	0.4	13	26	59	690
RVol	0.82	0.23	0.02	0.79	0.91	0.96	1.00
$\hat{p}$	0.71	0.11	0.54	0.61	0.70	0.79	0.97
π_{HR}	0.69	0.15	0.16	0.59	0.70	0.80	0.98

Notes. This table shows descriptive statistics for the main variables used in the analysis of the women's betting exchange data. The variables are defined in Table II. *TotVol* is the total volume that backs the higher-ranked player ($000's), and *RVol* is the proportion of the total volume bet in the match on both players that backs the higher-ranked player. π_{HR} is the subjective probability that the higher-ranked player wins the match, as derived from the betting exchange odds. We drop non-completed matches, matches with missing information or housetake greater than 0.05, and matches which do not satisfy criterion (*ii*) in Section II.A. The sample consists of 1,425 MS matches and 2,259 GS matches from 2009-2014.

TABLE A. 4

Descriptive Statistics – Laboratory Experiment

Variable	Mean	σ	Min	Q1	Median	Q3	Max
Panel A: Participants (N=17)							
Knowledge	3.65	1.06	1	4	4	4	5
% Correct	65	30	13	50	63	88	100
Payment (£'s)	44.23	3.76	33.52	43.29	44.81	46.62	49.18
Panel B: GS							
HR_{Rank}	41	31	1	17	33	59	129
LR_{Rank}	125	82	37	70	98	153	463
$\hat{p}$	0.73	0.10	0.59	0.63	0.75	0.79	0.95
π_{HR}	0.70	0.17	0.20	0.60	0.70	0.80	1.00
Panel C: MS							
HR_{Rank}	31	21	1	14	29	50	76
LR_{Rank}	83	86	21	32	52	87	403
$\hat{p}$	0.67	0.10	0.56	0.58	0.63	0.72	0.93
π_{HR}	0.67	0.17	0.10	0.55	0.65	0.80	1.00

Notes. This table shows descriptive statistics for the sample generated in the laboratory experiment. 17 subjects participated, who made choices for 33 MS matches and 64 GS matches. The GS matches were all first-round matches. The MS matches contained 7 qualifying matches, 17 first round matches and 9 second round matches. For each of the second-round matches, we asked students to make assessments for each of all possible combinations that could have resulted from the first round matches, which were unknown at the time that the MS session was held. The qualifying matches and the second round matches were included to increase the MS sample. In Panel A, *Knowledge* is the self-reported knowledge of subjects in tennis (1 to 5), *% Correct* is the percentage of correct answers given to the tennis quiz questions (8 in total), and *Payment* is the total payment collected by subjects for both sessions.

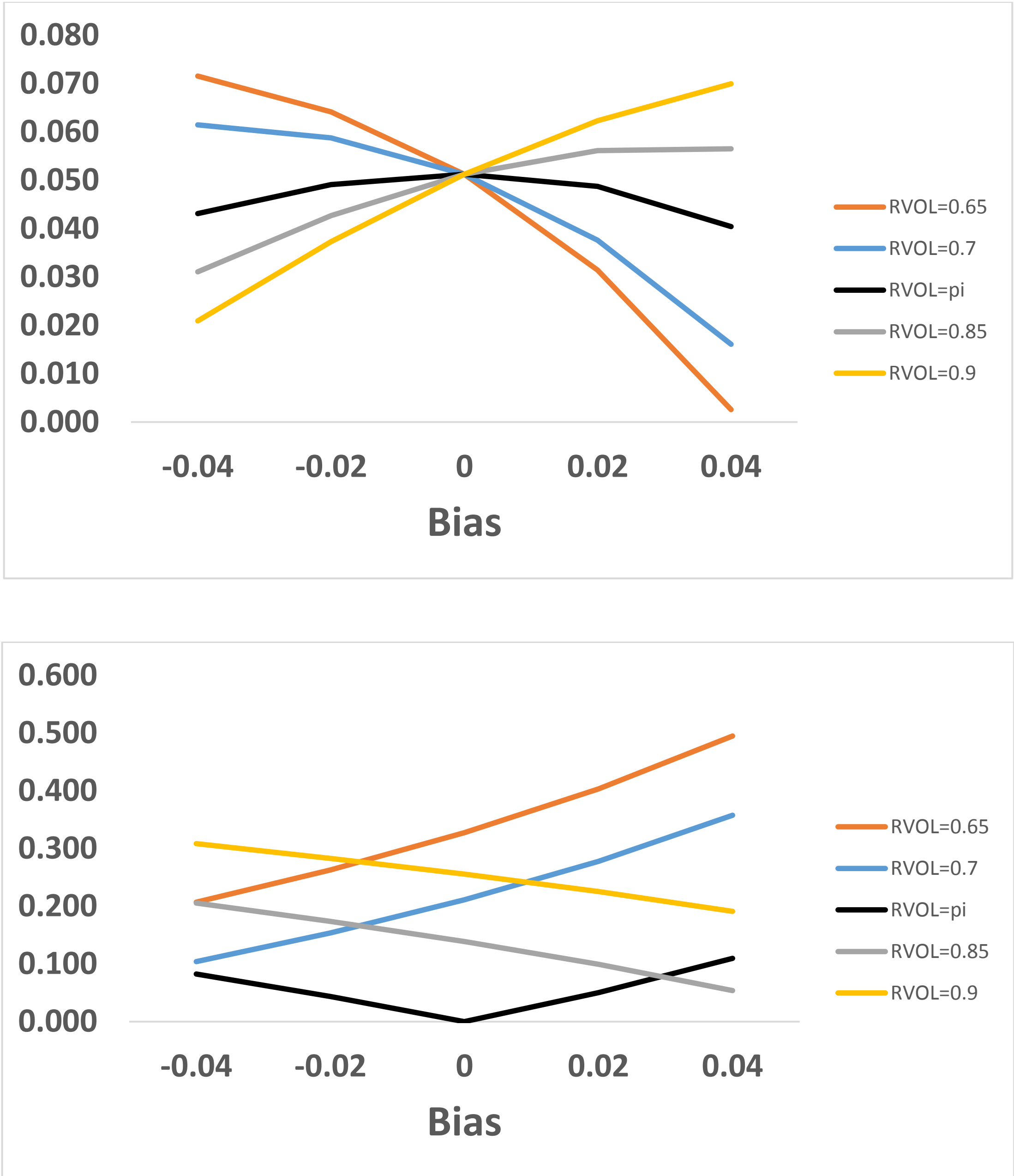

FIGURE A. 1

Expected Profit and Bias

This figure in the top panel depicts expected profits (Y-axis) for different levels of bias (X-axis), following the procedure explained in Appendix B. The figure in the bottom Panel depicts the corresponding standard deviations.