

Innocenti, Alessandro; Nannicini, Tommaso; Ricciuti, Roberto

Article

The importance of betting early

Risks

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Innocenti, Alessandro; Nannicini, Tommaso; Ricciuti, Roberto (2021) : The importance of betting early, *Risks*, ISSN 2227-9091, MDPI, Basel, Vol. 9, Iss. 4, pp. 1-15, <https://doi.org/10.3390/risks9040067>

This Version is available at:

<https://hdl.handle.net/10419/258156>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Article

The Importance of Betting Early

Alessandro Innocenti ¹, Tommaso Nannicini ^{2,*} and Roberto Ricciuti ³ 

¹ Department of Social, Political and Cognitive Sciences, LabSi & Befinlab, University of Siena, 53100 Siena, Italy; alessandro.innocenti@unisi.it

² Department of Social & Political Sciences, Bocconi University, IGIER, CEPR & IZA, Via Rontgen 1, 20136 Milan, Italy

³ Department of Economics, University of Verona & CESifo, 37129 Verona, Italy; roberto.ricciuti@univr.it

* Correspondence: tommaso.nannicini@unibocconi.it

Abstract: We evaluate the impact of timing on decision outcomes when both the timing and the relevant decision are chosen under uncertainty. Sports betting provides the testing ground, as we exploit an original dataset containing more than one million online bets on games of the Italian Major Soccer League. We find that individuals perform systematically better when they place their bets farther away from the game day. The better performance of early bettors holds controlling for (time-invariant) unobservable ability, learning during the season, and timing of the odds. We attribute this result to the increase of noisy information on game day, which hampers the capacity of late (non-professional) bettors to use very simple prediction methods, such as team rankings or last game results. We also find that more successful bettors tend to bet in advance, focus on a smaller set of events, and prefer games associated with smaller betting odds.

Keywords: sports betting; decision timing; information overload; forecasting



Citation: Innocenti, Alessandro, Tommaso Nannicini, and Roberto Ricciuti. 2021. The Importance of Betting Early. *Risks* 9: 67. <https://doi.org/10.3390/risks9040067>

Academic Editor: Dominic Cortis

Received: 23 October 2020

Accepted: 9 March 2021

Published: 6 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Decision timing is a key ingredient of decision making in many settings. Whenever the effect of a choice depends on a future state of the world—e.g., betting, financial markets, firm's strategy—agents face the additional choice of whether taking their decision close to or far from the future event. On the one hand, waiting for a last-minute decision may allow them to improve their information set. On the other hand, if they cannot efficiently process all inputs accruing in proximity to the event, information overload may be detrimental.

We study this tradeoff in the context of sports betting for two reasons. First, in terms of internal validity, as we exploit large data on online bets, we can estimate the effect of the distance from the event on the probability of success without losing statistical accuracy, even if we control for unobservable heterogeneity and for a number of time-varying confounding factors. Second, in terms of external validity, as we focus on a population of non-professional bettors, we isolate behavioral regularities that may extend beyond our context.

To test our hypothesis that decision timing matters, we analyze the winning probability of bets placed in two different seasons of the Italian Major Soccer League (*Serie A*). The dataset contains more than one million online bets. The 7093 individuals in our dataset are non-experts, who bet small amounts of money on multiple events to increase their potential profits and only win if all of the events happen. Betting on soccer relies on the availability of objective information, such as team rankings and win-loss records, which represent reasonably good predictors of game outcomes. For these reasons, we believe that the distance from the game day is a significant factor among those determining how these non-professional bettors process and make use of the available information. The tradeoff highlighted above is clearly at work. Betting too early might force individuals to dismiss relevant information, such as players' injuries that happen close to the game. On the other hand, betting late faces individuals with a large amount of information, which increases

with the public relevance of the event, comes from multiple sources, and may not be easy to handle.

In our empirical strategy, we control for individual fixed effects, therefore accommodating for (time-invariant) unobservable ability. Indeed, when we refer to “early” vs “late” bettor, we mean the same individual placing different bets at a different time distance from the relevant event. As in some specifications we control for individual-times-team fixed effects, we also accommodate for the fact that individuals might systematically bet earlier or later on specific teams (e.g., their favorite one). To control for learning as individuals place more and more bets, we use flexible control functions. Finally, to control for potential time-varying sources of omitted variable bias, we include the betting odds—which capture the strategic interaction between bettors and the other side of the market—and other attributes of the bet (e.g., financial amount, event’s characteristics).

According to our empirical evidence, for the same bettor, the probability of making a correct forecast is higher when the bet is made on the days before the event: As opposed to bets on game day, the chance of winning increases by 1.3 percentage points (that is, by about 3% with respect to the average). The effect is larger when big teams or multiple bets are involved (about 5% in both cases). The relationship between betting early and winning is monotonic, as the probability of a correct forecast is larger the higher the number of days from the event, up to the maximum effect of 6.7 percentage points (about 15% with respect to the average) 5 days before the event. This evidence supports the hypothesis that information overload may occur; as the event becomes closer, individuals receive more information than they are able to properly digest, therefore increasing the probability of mistakes.

The estimated individual fixed effects show that successful (non-professional) bettors also tend to place their bets in advance. Furthermore, they are more selective, as they place a smaller number of bets in the same week, and tend to focus on events associated with lower betting odds, which are arguably easier to forecast.

The paper is organized as follows. Section 2 reviews the related literature. Section 3 describes our empirical strategy. The results are discussed in Section 4. Section 5 concludes.

2. Related Literature

Since the 1970s, sports forecasting has been the object of extensive research motivated by two main reasons: (i) To ascertain if betting markets are informationally efficient and enable learning processes, and (ii) to check if experts make more accurate predictions than non-experts. Both strands of the literature aimed at analyzing the conditions under which the availability of comprehensive information and professional advice is fully discounted by market prices (that is, betting odds) and rules out observable biases that could allow speculators to make higher-than-average returns. A large body of empirical evidence supports the view that bettors’ behavior does not conform to the rational decision model and is affected by some cognitive biases (Diecidue et al. 2004; Osborne 2001). First, bettors show a clear tendency to under-bet favorites and over-bet long shots (Golec and Tamarkin 1995; Paul and Weinbach 2005; Newall and Cortis 2021). Second, they exhibit decision biases such as confirmation, gambler’s fallacy, and overconfidence related to inaccurate information processing (Blavatskyy 2009; Palomino et al. 2009). Third, they adopt a series of heuristics whose suitability is context-dependent (Conlisk 1993; Kochman et al. 2015). Finally, they are not effective enough in discounting the effect of noisy and redundant information and in reducing the impact of information overload (Bleichrodt and Schmidt 2002; Kaufmann and Weber 2013). If information directly enters the agent’s utility function, it can create an incentive to avoid information, even when it is useful, free, and independent of strategic considerations. For a survey on the theoretical and empirical literature on avoiding information, see Golman et al. (2017).

A major strand of research concerns horse-race betting, which is a naturally occurring asset market in which the transmission of information from informed to uninformed traders is not typically smooth. This betting market is efficient if it aggregates less-than-perfect

information owned by all the participants and disseminates it to all bettors, through the publicly available information given by track and bookmakers' odds and handicappers' picks (e.g., see [Snyder 1978](#); [Figlewski 1979](#); [Hausch et al. 1981](#); [Asch et al. 1984](#)). Baseball, basketball, football, and soccer are sports in which the sources of insider information are less relevant than in racetrack. [Pope and Peel \(1989\)](#) analyze the fixed odds offered by bookmakers and the forecasts made by professional tipsters on UK soccer league games; they argue that betting markets are efficient in preventing bettors to gain abnormal returns based on public information, but odds do not fully reflect all the available information. This finding is confirmed by [Forrest and Simmons \(2000\)](#), who consider newspaper tipsters offering professional advice on English and Scottish soccer games.

The fact that the condition of being experts is not necessarily associated with a high degree of forecasting accuracy is extensively discussed by [Camerer and Johnson \(1991\)](#) for various domains (medical, financial, academic). They conclude that experts' superiority in processing information is not strictly related to performance superiority, which is crucially affected by the matching of experts' cognitive abilities with "environmental demands" ([Camerer and Johnson 1991](#), p. 213). An interpretation of this finding can be traced back to the paper by [Oskamp \(1965\)](#), who argues that the extent of collected information cannot be directly related to predictive accuracy. While predictive ability reaches a ceiling once a limited amount of information has been collected, confidence in the ability to make accurate decisions continues to grow proportionally ([Davis et al. 1994](#); [Kaufmann and Weber 2013](#)). In the context of geopolitical questions, [Atanasov et al. \(2020\)](#) find that high-skill forecasters that make frequent, small updates outperform low-skill forecasters, who tend to confirm their initial judgments or make infrequent, large revisions. Therefore, small-increment updating is seen as a signal of early accuracy.

[Gigerenzer et al. \(1999\)](#), [Benartzi and Thaler \(2001\)](#), [Martignon and Hoffrage \(2002\)](#), [Rieskamp and Otto \(2006\)](#), and [Gigerenzer and Goldstein \(2011\)](#) argue that decision making can be better explained by models of heuristics rather than by the standard rational decision model. [Anderson et al. \(2005\)](#) use the recognition heuristics to account for non-experts' performance in soccer betting. According to [Newell and Shanks \(2004\)](#), recognition heuristics is assumed to demand little time, information, and cognitive effort, and exploits the relationship between a criterion value (e.g., success in home win) and its predictors (e.g., team rank position).

Heuristics perform quite well in environments affected by noisy and redundant information such as sports forecasting. Noisy information is defined as an information structure in which not only can one signal indicates several states, but also several signals can occur in the same state ([Bichler and Butler 2007](#); [Crawford and Sobel 1982](#)). In [Dieckmann and Rieskamp \(2007\)](#), redundant information is defined as information composed by pieces highly correlated with each other and supporting the same prediction (positive redundancy), or that contradict each other and suggest incompatible predictions (negative redundancy).

By again quoting [Oskamp \(1965\)](#), if bettors are provided with a very rich source of information without activating a costly search process, confidence increases in relation to the beliefs that they had before. For example, [Bettman et al. \(1993\)](#) provide support for the notion that people also select strategies adaptively in response to information redundancy; they show that participants choosing between gambles search only for a subset of the available information when they encounter a redundant environment with positively correlated attributes. Negatively correlated attributes, in contrast, give rise to search patterns consistent with compensatory strategies that integrate more information. This cognitive bias is known as the illusion of knowledge, according to which beyond a threshold more information on the event increases self-confidence more than accuracy ([Barber and Odean 2002](#)). In recent years, computational intelligence applications to sports have boomed; see [Fister et al. \(2015\)](#) for a survey. Computational intelligence involves algorithms for solving real-world problems somewhat intelligently as similar problems are solved by natural systems. Its

ability to fastly and efficiently adapt to a changing environment is promising in the field of betting, in contrast with the biases shown by humans.

The condition of “information overload” characterizes media information on Italian soccer, which provides the ground for our empirical analysis. The amount of information to be processed is greatly increased by the variety of communication systems on TV, the internet, and newspapers. Furthermore, much of the information is not original and watchers continuously process information received from other sources but differently presented. The introduction of online betting causes a further increase in the availability of information, which is also diffused by online betting sites. Our dataset, which is described in the next section, includes small bets, generally evenly distributed across individuals. Therefore, it can be safely assumed that the individuals contained in our dataset are “non-expert bettors.”

3. Empirical Strategy and Data

Based on the literature surveyed in the previous section and on the available data, we test the following hypothesis.

Hypothesis 1 (H1). (information overload) *As soon as the event approaches and the amount of noisy information available to bettors increases, their winning ability decreases.*

At the same time, we control for the following confounding hypothesis.

Hypothesis 2 (H2). (learning) *Bettors improve their performance over time, as they get more acquainted with the environment and the relative strength of the soccer teams.*

We use a unique (large) dataset of online bets from a provider specialized in the field. The company is located in Southern Italy, but bets are made from all over the country. Users have to register and then bet online through credit card payments. We were provided with bets on all games of 20 game weeks of the Italian Soccer Major League (*Serie A*), namely, the last 10 weeks of the 2004–2005 season and the first 10 weeks of the 2005–2006 season. Our dataset includes 1,205,597 single bets made by 7093 registered users. A large study by [Buhagiar et al. \(2018\)](#) analyses a total sample of 163,992 soccer odds from 41,003 matches for ten leagues over twelve seasons. Single bets may also be part of multiple bets including more than one event and may concern several events (e.g., which team wins, draw, goals scored, goals scored in the first half, and so forth). Multiple bets increase potential profits and are won only if all of the events happen at the same time. In our analysis, we focus on the simplest events: 1, X, 2, 12, 1X, and X2 (where 1 stands for home win, X for draw, and 2 for away win). These types of events account for 85% of all bets. Using all bets does not affect our results (available upon request).

The fact that bettor j correctly forecasts event i at game week t (W_{ijt}) is modeled as:

$$W_{ijt} = \gamma_j + g(D_{ijt}) + f(t) + X_{ijt}\beta + Z_{ijt}\alpha + \varepsilon_{ijt} \quad (1)$$

where γ_j are individual fixed effects (capturing all time-invariant characteristics of bettor j , including her intrinsic level of sophistication and ability); X_{ijt} is a vector of time-varying attributes linked to bettor j (such as the amount of money bet at game week t , or the number of other events linked to event i in a multiple bet); Z_{ijt} is a vector of time-varying attributes of event i (such as whether the home team or the favorite team won the game, and the day-by-day odds decided by the provider); $g(\cdot)$ is a flexible function of the distance from the day individual j places the bet to the day event i occurs (D_{ijt}); $f(\cdot)$ is a flexible function of game week t ; and ε_{ijt} is an idiosyncratic error clustered at the event level.

To test H1 (information overload), we consider three specifications of $g(\cdot)$: Linear function of D_{ijt} (“betting distance”); dummy equal to one if the bet is placed before the game day and zero otherwise (“betting early”); non-parametric specification including a set of dummies for each value of D_{ijt} (which varies from zero for bets on game day to a

maximum of 5 days). To control for H2 (learning), we use three specifications of $f(\cdot)$: Linear trend; quadratic trend; game week dummies (with t varying from 1 to 20 across the two seasons in our dataset). The inclusion of individual fixed effects accommodates for all time-invariant bettors' characteristics correlated with both the outcome and the treatment. The inclusion of betting odds in Z_{it} controls for the decision of the other side of the market, that is, the betting company, which might strategically adjust the timing of the odds as the event approaches. The inclusion of the event's characteristics identified as relevant by the previous literature, such as the victory of the home or favorite team, controls for the fact that bettors might bet earlier on events easier to forecast. We also estimated specifications including an interaction term between the betting distance and the amount of money bet by the user, so as to partly account for overconfidence, but the coefficient was never statistically different from zero and therefore we excluded the interaction term from the baseline estimations. Finally, we also estimated specifications including a set of interactions between the individual fixed effects and the home or away team, so as to account for the fact that bettors might adopt different timing strategies with respect to different teams, such as their favorite one; results are again unchanged and available upon request.

Specifically, among the covariates related to event i , we consider the dummy "main teams," equal to one if the bet concerns at least one of the four leading teams during our sample period (*F.C. Internazionale*, *Juventus F.C.*, *A.C. Milan*, and *A.S. Roma*); the dummy "strong team wins," equal to one if the stronger team (measured by the relative ranking position in the league) wins; the dummy "home team wins," equal to one if the home team wins. Among the time-varying attributes of each bettor j 's decision, we consider the amount spent by the user in each game week ("amount by user"); the number of the other single bets associated with i within a multiple bet ("other events"); and the official evaluation that the betting company gives to each event when the bet i is placed by individual j (betting "odds"). To capture any systematic difference between the two seasons in our dataset, we also include a dummy for the 2005-06 season.

Table 1 reports the descriptive statistics of our variables. In our data, 45% of single bets are successful. This does not mean that bettors have such a high winning rate, because single bets may be part of multiple bets and some of them may be wrong. Indeed, the winning rate in multiple bets is just 5% on average. Most bettors place their play on the game day, while early bettors are about 32%. The average amount spent per bettor in a game week is 211 Euros, again with a large standard deviation. Almost 40% of bets are made on the main four teams.

Table 1. Descriptive statistics.

	Mean	Median	S.d.	Min	Max
Correct forecast	0.451	0.000	0.498	0.000	1.000
Betting distance	0.443	0.000	0.778	0.000	5.000
Betting early	0.317	0.000	0.465	0.000	1.000
Other events	5.151	5.000	2.061	0.000	13.000
Amount by user	0.211	0.153	0.242	0.003	6.018
Main teams	0.399	0.000	0.490	0.000	1.000
Home team wins	0.394	0.000	0.489	0.000	1.000
Strong team wins	0.366	0.000	0.482	0.000	1.000
Odds	2.216	1.900	1.033	1.050	18.000
2004/05 season	0.507	1.000	0.500	0.000	1.000

Notes. The number of observations is 1,205,575 for all variables. *Betting distance* is measured in days. *Betting early* is a dummy equal to one if the bet is placed before the game day, and zero otherwise. *Other events* captures the number of events associated with the single bet in a multiple bet. *Amount by user* is the amount bet by the user in the game week and is measured in thousands of Euros. All the other variables except *Odds* are dummies.

Table 2 provides information on the above variables and bettors' socio-economic characteristics by betting distance. We also test whether means are different between bets placed on game day and bets placed before. Thanks to the large sample size, many differences are statistically significant, although most of them are economically small. Early

bets tend to be placed on stronger teams, and to be associated with a larger number of multiple bets.

Table 2. Conditional means by betting distance.

	Conditional Means by Betting Distance						Conditional Means and <i>t</i> -Test		
	0	1	2	3	4	5	Distance > 0	Distance = 0	<i>p</i> -Value
Home team wins	0.392	0.396	0.408	0.420	0.368	0.355	0.398	0.392	0.000
Strong team wins	0.362	0.370	0.390	0.405	0.412	0.349	0.376	0.362	0.000
Other events	5.01	5.29	5.70	6.20	6.75	6.77	5.46	5.01	0.000
Amount by user	0.208	0.212	0.240	0.234	0.209	0.179	0.218	0.208	0.000
Main teams	0.384	0.421	0.445	0.472	0.461	0.379	0.429	0.384	0.000
Odds	2.24	2.19	2.11	2.03	1.92	1.92	2.16	2.24	0.000
2004/05 season	0.525	0.475	0.459	0.412	0.430	0.528	0.468	0.525	0.000
Female	0.135	0.132	0.143	0.134	0.118	0.084	0.133	0.135	0.003
Age	34.7	34.8	35.1	36	34.8	34.1	34.9	34.7	0.000
Lawyer	0.044	0.041	0.038	0.039	0.036	0.026	0.040	0.044	0.000
Bank employee	0.120	0.123	0.124	0.124	0.124	0.092	0.123	0.120	0.000
Engineer & programmer	0.047	0.046	0.049	0.044	0.045	0.048	0.046	0.047	0.365
Architect	0.035	0.033	0.032	0.024	0.036	0.034	0.032	0.035	0.000
Clerk	0.032	0.033	0.043	0.033	0.029	0.036	0.035	0.032	0.000
Unemployed	0.006	0.006	0.004	0.005	0.009	0.007	0.005	0.006	0.000
Other profession	0.062	0.060	0.064	0.076	0.080	0.070	0.062	0.062	0.579
Missing profession	0.653	0.660	0.648	0.656	0.642	0.688	0.657	0.653	0.000
Observations	820,966	277,989	72,331	20,614	8623	3052	382,609	822,966	
Share	0.683	0.231	0.060	0.017	0.007	0.003	0.317	0.683	

Notes. The number of observations is 1,205,575 for all variables. The last column reports *p*-values referring to *t*-tests of the equality of means in the subsamples of bets placed in the game day (*Distance* = 0) vs. before the game day (*Distance* > 0). Due to the accuracy granted by the large sample size, means tend to be statistically different between each other, although differences are small in many cases.

4. Empirical Results and Discussion

Tables 3–5 report our baseline specifications as in Equation (1). In the first three columns, we do not control for individual fixed effects, whereas this is done in the last three columns. The latter represents our preferred specifications, but it is instructive to compare results with and without fixed effects. As discussed above, to control for possible learning we use three specifications: linear trend in game week (columns 1 and 4); quadratic trend (columns 2 and 5); and a full set of game week dummies (columns 3 and 6). The difference between the three tables concerns how we model betting distance: linearly in Table 3; with the dummy “betting early” in Table 4; and with a full set of dummies for each value of the betting distance, which is measured in days, in Table 5.

Table 3 shows very similar results across all specifications. The coefficient of betting distance is significantly positive and very stable: The farther away from the event date the bet is, the higher the probability of winning. On average and for the same bettor, betting one day earlier increases the chance of winning by about 0.8 percentage points, that is, by about 1.8% with respect to the average probability of a correct forecast. This provides evidence of possible information overload. Moreover, as long as the season goes on, bettors worsen their performance, as highlighted by the significantly negative coefficients for the game week trend in both the linear and quadratic specifications.

Consistently with the previous literature, we find very strong effects for both “home team wins” and “strong team wins” (equal to 40.8% and 60.9%, respectively, with respect to the average outcome). The ability of winning is positively and significantly affected by the monetary amount that each player bets, with a large effect with respect to the average outcome (37.4% for an increase of the amount bet equal to its standard deviation). Both betting for the main teams and on more than one event increase the probability of winning. Columns 2 and 5 include the variable game week squared. We do not report its value since it is extremely small (in the order of four decimals); therefore the linear specification shows a fairly good fit. As we would also expect, higher odds are associated with a lower probability of winning (on average by −46.0% for an increase of the odds equal to the standard deviation).

Table 3. The impact of betting distance: baseline specifications.

	(1)	(2)	(3)	(4)	(5)	(6)
Betting distance	0.008 *** (0.001)	0.008 *** (0.001)	0.008 *** (0.001)	0.008 *** (0.001)	0.008 *** (0.001)	0.008 *** (0.001)
Home team wins	0.184 *** (0.002)	0.184 *** (0.002)	0.182 *** (0.002)	0.184 *** (0.002)	0.184 *** (0.002)	0.182 *** (0.002)
Strong team wins	0.283 *** (0.002)	0.283 *** (0.002)	0.298 *** (0.002)	0.284 *** (0.002)	0.284 *** (0.002)	0.297 *** (0.002)
Game week	−0.003 *** (0.000)	−0.007 *** (0.000)		−0.004 *** (0.000)	−0.007 *** (0.000)	
Other events	0.008 *** (0.000)	0.008 *** (0.000)	0.008 *** (0.000)	0.007 *** (0.000)	0.007 *** (0.000)	0.007 *** (0.000)
Amount by user	0.004 (0.003)	0.006 (0.003)	0.002 (0.003)	0.008 (0.006)	0.010 * (0.006)	0.006 (0.004)
Main teams	0.048 *** (0.002)	0.047 *** (0.002)	0.046 *** (0.002)	0.049 *** (0.002)	0.048 *** (0.002)	0.047 *** (0.002)
Odds	−0.096 *** (0.001)	−0.096 *** (0.001)	−0.096 *** (0.001)	−0.096 *** (0.001)	−0.096 *** (0.001)	−0.095 *** (0.001)
2004/05 season	−0.011 *** (0.004)	−0.029 *** (0.004)		−0.014 *** (0.005)	−0.031 *** (0.005)	
Game week squared		0.000 *** (0.000)			0.000 *** (0.000)	
Game week dummies	No	No	Yes	No	No	Yes
Individual fixed effects	No	No	No	Yes	Yes	Yes
No. of observations	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575
No. of individuals	7093	7093	7093	7093	7093	7093

Notes. Dependent variable: probability of correctly forecasting a single event. Estimated model: linear probability model as in equation (1); standard errors clustered at the event level are reported in parentheses. Significance at the 10% level is represented by *, and at the 1% level by ***. The main coefficient of interest (*Betting distance*) shows that the probability of a correct forecast increases by 0.8 percentage points (1.8% with respect to the average) for each day by which individuals decides to bet earlier.

Table 4. The impact of betting early: baseline specifications.

	(1)	(2)	(3)	(4)	(5)	(6)
Betting early	0.013 *** (0.001)	0.013 *** (0.001)	0.011 *** (0.001)	0.013 *** (0.001)	0.013 *** (0.001)	0.011 *** (0.001)
Home team wins	0.184 *** (0.002)	0.184 *** (0.002)	0.182 *** (0.002)	0.184 *** (0.002)	0.184 *** (0.002)	0.182 *** (0.002)
Strong team wins	0.283 *** (0.002)	0.283 *** (0.002)	0.298 *** (0.002)	0.284 *** (0.002)	0.284 *** (0.002)	0.297 *** (0.002)
Game week	−0.003 *** (0.000)	−0.007 *** (0.000)		−0.004 *** (0.000)	−0.007 *** (0.000)	
Other events	0.008 *** (0.000)	0.008 *** (0.000)	0.008 *** (0.000)	0.007 *** (0.000)	0.008 *** (0.000)	0.007 *** (0.000)
Amount by user	0.004 (0.004)	0.006 (0.003)	0.002 (0.003)	0.008 (0.006)	0.010 * (0.006)	0.006 (0.004)
Main teams	0.048 *** (0.002)	0.047 *** (0.002)	0.046 *** (0.002)	0.049 *** (0.002)	0.048 *** (0.002)	0.047 *** (0.002)
Odds	−0.096 *** (0.001)	−0.097 *** (0.001)	−0.096 *** (0.001)	−0.096 *** (0.001)	−0.096 *** (0.001)	−0.095 *** (0.001)
2004/05 season	−0.011 *** (0.004)	−0.029 *** (0.004)		−0.014 *** (0.005)	−0.031 *** (0.005)	
Game week squared		0.000 *** (0.000)			0.000 *** (0.000)	
Game week dummies	No	No	Yes	No	No	Yes
Individual fixed effects	No	No	No	Yes	Yes	Yes
No. of observations	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575
No. of individuals	7093	7093	7093	7093	7093	7093

Notes. Dependent variable: probability of correctly forecasting a single event. Estimated model: linear probability model as in Equation (1); standard errors clustered at the event level are reported in parentheses. Significance at the 10% level is represented by *, and at the 1% level by ***. The main coefficient of interest (*Betting early*) shows that the probability of a correct forecast increases by 1.3 percentage points (2.9% with respect to the average) if the individuals bet earlier than in the event date.

Table 5. The impact of betting distance: non-parametric specifications.

	(1)	(2)	(3)	(4)	(5)	(6)
1 day before	0.013 *** (0.001)	0.013 *** (0.001)	0.009 *** (0.001)	0.013 *** (0.001)	0.013 *** (0.001)	0.010 *** (0.001)
2 days before	0.010 *** (0.003)	0.010 *** (0.003)	0.011 *** (0.002)	0.010 *** (0.003)	0.010 *** (0.003)	0.012 *** (0.003)
3 days before	0.021 *** (0.005)	0.021 *** (0.005)	0.021 *** (0.005)	0.018 *** (0.005)	0.018 *** (0.005)	0.018 *** (0.005)
4 days before	0.029 *** (0.009)	0.030 *** (0.009)	0.029 *** (0.008)	0.024 *** (0.009)	0.024 *** (0.008)	0.024 *** (0.008)
5 days before	0.067 *** (0.015)	0.067 *** (0.015)	0.072 *** (0.013)	0.062 *** (0.015)	0.062 *** (0.014)	0.069 *** (0.013)
Home team wins	0.184 *** (0.002)	0.184 *** (0.002)	0.182 *** (0.002)	0.184 *** (0.002)	0.184 *** (0.002)	0.182 *** (0.002)
Strong team wins	0.283 *** (0.002)	0.283 *** (0.002)	0.298 *** (0.002)	0.284 *** (0.002)	0.284 *** (0.002)	0.297 *** (0.002)
Game week	−0.003 *** (0.000)	−0.007 *** (0.000)		−0.004 *** (0.000)	−0.007 *** (0.001)	
Other events	0.008 *** (0.000)	0.008 *** (0.000)	0.008 *** (0.000)	0.007 *** (0.000)	0.008 *** (0.000)	0.007 *** (0.000)
Amount by user	0.004 (0.004)	0.006 * (0.003)	0.002 (0.003)	0.008 (0.006)	0.011 * (0.006)	0.006 (0.004)
Main teams	0.048 *** (0.002)	0.047 *** (0.002)	0.046 *** (0.002)	0.049 *** (0.002)	0.048 *** (0.002)	0.047 *** (0.002)
Odds	−0.096 *** (0.001)	−0.096 *** (0.001)	−0.096 *** (0.001)	−0.096 *** (0.001)	−0.096 *** (0.001)	−0.095 *** (0.001)
2004/05 season	−0.010 *** (0.004)	−0.029 *** (0.004)		−0.014 *** (0.005)	−0.031 *** (0.005)	
Game week squared		0.000 *** (0.000)			0.000 *** (0.000)	
Game week dummies	No	No	Yes	No	No	Yes
Individual fixed effects	No	No	No	Yes	Yes	Yes
No. of observations	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575
No. of individuals	7093	7093	7093	7093	7093	7093
1 day = 2 days	0.466	0.400	0.515	0.414	0.353	0.582
2 days = 3 days	0.074	0.071	0.075	0.214	0.204	0.219
3 days = 4 days	0.435	0.390	0.396	0.578	0.532	0.536
4 days = 5 days	0.017	0.016	0.003	0.015	0.015	0.003

Notes. Dependent variable: probability of correctly forecasting a single event. Estimated model: linear probability model as in equation (1); standard errors clustered at the event level are reported in parentheses. Significance at the 10% level is represented by *, and at the 1% level by ***. In italics, *p*-values for Wald tests on the equality of the coefficients of the betting-distance dummies. The main coefficients of interest (i.e., the betting-distance dummies) show that the probability of a correct forecast increases if the individuals bet earlier than in the event date; in particular, at the maximum distance (5 days before), as opposed to betting on the event date, the probability of a correct forecast is larger by 6.7 percentage points (15% with respect to the average).

In Table 4 the regressor of interest is the dummy “betting early,” equal to one if the bet was placed on one of the 5 days preceding game day. This variable is significantly positive, meaning that the probability of making the correct forecast is higher when the bet is made in advance. On average and for the same bettor, the chance of winning increases by 1.3 percentage points (that is, by 2.9% with respect to the average). All of the other variables confirm their behavior from both a qualitative and a quantitative point of view.

Table 5 includes a full set of dummies for each value of betting distance. The effect of the distance from the event on the probability of winning is monotonic, as it increases to its maximum when individuals bet 5 days in advance (the largest distance from the event day that is allowed in this betting context). At this maximum distance, as opposed to betting on the event day, the probability of a correct forecast is larger by 6.7 percentage points (i.e., by about 15% with respect to the average). Wald tests on the equality of coefficients confirm the statistical significance of this increasing effect as we move away from game day. Again, all of the other variables confirm their behavior.

As further robustness checks aimed at assessing the validity of the mechanism on information overload, in Appendix A we address heterogeneity issues, that is, we assess whether the effect of betting distance is stronger in specific subsamples. Specifically, in Table 1, we distinguish between bets on one of the main teams and on all the other teams.

In Table 2, we discriminate between bets done on many events (that is, above the median of events associated with multiple bets) or lesser events. Table 3 distinguishes between “hard bets” (that is, bets whose amount is above the median value, where we consider the amount of the multiple bets made by the individual) and all the others. We find evidence of heterogeneity in the first two exercises. In particular, the effect of betting early is quantitatively larger for bets linked to main teams and for bets linked to other bets in a multiple play. This confirms our information-overload interpretation of the positive effect of betting early.

Finally, the estimated individual fixed effects allow us to shed light on additional behavioral patterns in our data. Figure A1 in Appendix A shows that more successful bettors (that is, those with a larger fixed effect) also tend to bet in advance. This regularity, of course, does not affect the estimates discussed above, as they accommodate for unobservable heterogeneity, but it is an interesting finding per se. More skilled bettors seem to anticipate information overload and place their bets in advance. They are also more selective, as they place a smaller number of bets per game week, focus on bets associated with smaller betting odds, and tend to spend lower amounts.

5. Conclusions

We find that betting timing matters. From the analysis of more than 1,250,000 online bets, we detect a statistically significant and stable difference in the winning probability of early versus late bettors. The estimated effect controls for time-invariant unobservable heterogeneity, learning, betting odds, and observable characteristics of the event. Therefore, when we refer to “late” versus “early” bettors we are comparing the same individual making bets at different distances from each event. The poorer forecasting performance of late bettors is attributed to inefficient processing of information, also consistent with the heterogeneity results that we are able to disclose thanks to the richness of our data. The late bettors’ decision process is affected by various cues that, unknown to the earlier bettors, have scarce relevance for predicting the outcomes. The excess of noisy information (especially harsh if the same individual decides to bet on the main teams or on multiple events) reduces the possibility of using very simple prediction methods, such as team rankings or home team winning. The use of these criteria and cues greatly improves the possibility of placing a winning bet. Some skilled bettors partly anticipate the issue, as individuals with larger fixed effects tend to bet from 3 to 5 days in advance.

We acknowledge two limitations of our results. First, they are based on small stakes and we cannot rule out that when stakes are higher information processing could become more efficient. Second, we cannot rule out the fulfillment of other emotional objectives rather than standard profit maximization. We leave to future research the generalization of our results to betting contexts characterized by a larger degree of sophistication. We also leave to future research a more direct test of the information-overload mechanism that we indirectly disclose while estimating the causal effect of betting early on forecasting accuracy.

Author Contributions: Conceptualization, A.I., T.N. and R.R.; methodology, A.I. and T.N.; software, T.N.; validation, A.I., T.N. and R.R.; formal analysis, T.N.; investigation, A.I., T.N. and R.R.; data curation, T.N.; writing—original draft preparation, A.I., T.N. and R.R.; writing—review and editing, R.R.; project administration, T.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data can be found at: <https://data.mendeley.com/datasets/8vhr4kz822/1>.

Acknowledgments: We thank seminar participants at IGIER-Bocconi, University of East Anglia, and University of Siena, and three anonymous reviewers for insightful comments. Paolo Donatelli, Nicola Pierri, and Vincenzo Scrutinio provided outstanding research assistance. We also thank Microgame.it

and in particular Fabrizio D'Aloia for providing us with the data. Errors are ours and follow a random walk.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The appendix reports the results for further heterogeneity analysis. In Table 1, betting distance is always significantly positive, but the size of its coefficient is about three times larger when only the main teams are involved in the bet. This is consistent with our interpretation of the positive impact of betting early, because information overload on the event date is expected to be even more relevant for major teams. Compared with the previous estimations, another relevant variable changes its behavior: game week is usually positive in the linear specification when the main teams are included, and negative otherwise. Therefore, we observe some positive learning when the main teams—which are usually under the spotlight of newspapers—are involved.

In Table 2, interestingly, the effect of betting early is quantitatively larger for bets linked to other bets in a multiple play. Again, in these circumstances, information overload is likely to exacerbate fallacies in decision making and to reduce the probability of winning. In Table 3, instead, we do not detect statistically significant differences in the size of coefficients for “hard bets” versus the others. Interestingly, registered users that place the 50% of bets that we code as hard (1352) are just one-fifth of all bettors (7093). This means that only a fraction of sophisticated bettors places higher-than-median bets, but their behavior in terms of informational patterns is not significantly different from the behavior of the other, less sophisticated, bettors. In the last row of each table, we report the *p*-value of the Wald-test on the equality of the estimated coefficients of betting distance for each pair of subsamples. The subsample coefficients are statistically different between each other only in the case of “many events” and in some estimates for “main teams.”

Figure A1 shows that more successful bettors (that is, those with a larger fixed effect) also tend to bet in advance, from 3 to 5 days before the event takes place. This means that more skilled bettors seem to anticipate information overload and place their bets in advance. They are also more selective, as they place a smaller number of bets per game week (upper right panel); focus on bets associated with smaller betting odds (lower left panel), which are arguably easier to forecast; and tend to spend lower amounts (lower right panel), although this correlation is less strong than the others. Finally, there is no clear pattern of association between ability and age or other observable bettors' characteristics (available upon request).

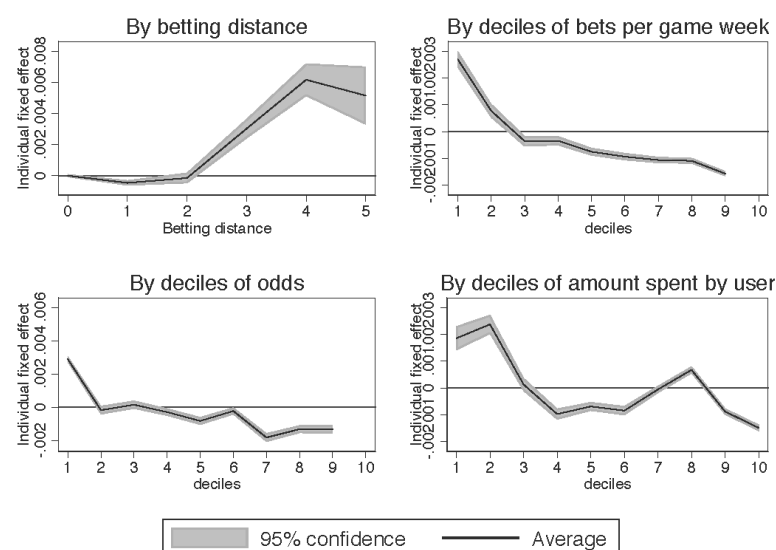


Figure A1. Individual fixed effects by various dimensions.

Table 1. Heterogeneity: main teams vs. others.

	Main Teams		Main Teams		Main Teams		Main Teams		Main Teams		Main Teams	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Betting distance	0.012 ** (0.001)	0.005 *** (0.001)	0.012 *** (0.001)	0.005 *** (0.001)	0.010 *** (0.001)	0.006 *** (0.001)	0.013 *** (0.001)	0.003 *** (0.001)	0.013 *** (0.001)	0.003** (0.001)	0.012 *** (0.001)	0.004 *** (0.001)
Home team wins	0.016 *** (0.002)	0.244 *** (0.002)	0.016 *** (0.002)	0.244 *** (0.002)	0.125 *** (0.002)	0.240 *** (0.003)	0.115 *** (0.002)	0.244 *** (0.003)	0.115 *** (0.002)	0.244 *** (0.003)	0.124 *** (0.002)	0.240 *** (0.003)
Strong team wins	0.478 *** (0.003)	0.100 *** (0.003)	0.480 *** (0.003)	0.098 *** (0.003)	0.485 *** (0.003)	0.091 *** (0.003)	0.479 *** (0.003)	0.099 *** (0.003)	0.481 *** (0.003)	0.098 *** (0.003)	0.486 *** (0.003)	0.091 *** (0.003)
Game week	0.004 *** (0.000)	−0.007 *** (0.000)	−0.002 *** (0.000)	−0.012 *** (0.001)			0.004 *** (0.000)	−0.007 *** (0.000)	−0.003 *** (0.001)	−0.012 *** (0.001)		
Other events	0.002 *** (0.000)	0.013 *** (0.000)	0.002 *** (0.000)	0.013 *** (0.000)	0.002 *** (0.000)	0.012 *** (0.000)	0.002 *** (0.000)	0.012 *** (0.000)	0.002 *** (0.000)	0.012 *** (0.000)	0.002 *** (0.000)	0.011 *** (0.000)
Amount by user	−0.003 (0.005)	0.006 (0.005)	−0.000 (0.005)	0.008 * (0.005)	0.000 (0.004)	0.003 (0.003)	−0.000 (0.007)	0.009 (0.007)	0.005 (0.008)	0.013** (0.006)	0.003 (0.005)	0.006 (0.005)
Odds	−0.103 *** (0.001)	−0.088 *** (0.001)	−0.103 *** (0.001)	−0.088 *** (0.001)	−0.103 *** (0.001)	−0.088 *** (0.001)	−0.102 *** (0.001)	−0.086 *** (0.001)	−0.102 *** (0.001)	−0.086 *** (0.001)	−0.102 *** (0.001)	−0.086 *** (0.001)
2004/05 season	0.196 *** (0.004)	−0.150 *** (0.005)	0.160 *** (0.006)	−0.173 *** (0.005)			0.204 *** (0.006)	−0.151 *** (0.006)	0.162 *** (0.007)	−0.171 *** (0.007)		
Game week squared			0.000 *** (0.000)	0.000 *** (0.000)					0.000 *** (0.000)	0.000 *** (0.000)		
Game week dummies	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Individual fixed effects	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	480,534	725,041	480,534	725,041	480,534	725,041	480,534	725,041	480,534	725,041	480,534	725,041
No of individuals	6814	6905	6814	6905	6814	6905	6814	6905	6814	6905	6814	6905
Wald test <i>p</i> -value	0.000		0.000		0.005		0.000		0.000		0.000	

Notes. Dependent variable: probability of correctly forecasting a single event. Estimated model: linear probability model as in Equation (1) in separate subsamples (events involving main teams vs. events involving other teams). Standard errors clustered at the event level are reported in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***. The Wald test *p*-value captures the significance of the difference between the coefficients of betting distance in the two subsamples.

Table 2. Heterogeneity: bets linked to many events vs. others.

	Many Events		Many Events		Many Events		Many Events		Many Events		Many Events	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Betting distance	0.009 *** (0.001)	0.006 *** (0.001)	0.008 *** (0.001)	0.006 *** (0.001)	0.009 *** (0.001)	0.005 *** (0.001)	0.008 *** (0.001)	0.006 *** (0.001)	0.008 *** (0.001)	0.006 *** (0.001)	0.009 (0.001)	0.005 *** (0.001)
Home team wins	0.274 *** (0.003)	0.134 *** (0.002)	0.274 *** (0.003)	0.134 *** (0.002)	0.270 *** (0.003)	0.133 *** (0.002)	0.274 *** (0.003)	0.134 *** (0.002)	0.274 *** (0.003)	0.134 *** (0.002)	0.271 *** (0.003)	0.133 *** (0.002)
Strong team wins	0.371 *** (0.003)	0.233 *** (0.003)	0.372 *** (0.003)	0.233 *** (0.003)	0.385 *** (0.004)	0.246 *** (0.003)	0.373 *** (0.003)	0.233 *** (0.003)	0.374 *** (0.003)	0.233 *** (0.003)	0.385 *** (0.004)	0.246 *** (0.003)
Game week	−0.001 *** (0.000)	−0.004 *** (0.000)	−0.008 *** (0.001)	−0.008 *** (0.000)			−0.001 *** (0.000)	−0.004 *** (0.000)	−0.008 *** (0.001)	−0.008 *** (0.001)		
Main teams	0.012 *** (0.003)	0.064 *** (0.002)	0.010 *** (0.003)	0.063 *** (0.002)	0.011 *** (0.003)	0.063 *** (0.002)	0.014 *** (0.003)	0.065 *** (0.002)	0.013 *** (0.003)	0.064 *** (0.002)	0.013 *** (0.003)	0.063 *** (0.002)
Amount by user	0.006 (0.005)	0.002 (0.004)	0.000 * (0.005)	0.004 (0.004)	0.000 (0.004)	0.002 (0.003)	0.014 * (0.008)	0.003 (0.007)	0.018 * (0.008)	0.006 (0.007)	0.008 *** (0.001)	0.004 (0.001)
Odds	−0.120 *** (0.002)	−0.092 *** (0.001)	−0.120 *** (0.002)	−0.092 *** (0.001)	−0.119 *** (0.002)	−0.091 *** (0.001)	−0.117 *** (0.002)	−0.091 *** (0.001)	−0.117 *** (0.002)	−0.091 *** (0.001)	−0.116 *** (0.002)	−0.091 *** (0.001)
2004/05 season	0.047 *** (0.006)	−0.032 *** (0.004)	0.017 *** (0.006)	−0.051 *** (0.005)			0.044 *** (0.008)	−0.033 *** (0.006)	0.016 * (0.008)	−0.051 *** (0.006)		
Game week squared			0.000 *** (0.000)	0.000 *** (0.000)					0.000 *** (0.000)	0.000 *** (0.000)		
Game week dummies	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Individual fixed effects	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	408,489	798,086	408,489	798,086	408,489	798,086	408,489	798,086	408,489	798,086	408,489	798,086
No of individuals	5326	6696	5326	6696	5326	6696	5326	6696	5326	6696	5326	6696
Wald test <i>p</i> –value	0.160		0.196		0.025		0.202		0.213		0.013	

Notes. Dependent variable: probability of correctly forecasting a single event. Estimated model: linear probability model as in Equation (1) in separate subsamples (bets linked to a higher-than-median number of multiple bets vs. others). Standard errors clustered at the event level are reported in parentheses. Significance at the 10% level is represented by *, and at the 1% level by ***. The Wald test *p*-value captures the significance of the difference between the coefficients of betting distance in the two subsamples.

Table 3. Heterogeneity: bets involving larger-than-median-amount vs. others.

	High Amount		High Amount		High Amount		High Amount		High Amount		High Amount	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Betting distance	0.008 *** (0.001)	0.008 *** (0.001)	0.008 *** (0.001)	0.008 *** (0.001)	0.007 *** (0.001)	0.008 *** (0.001)	0.008 *** (0.001)	0.009 *** (0.001)	0.008 *** (0.001)	0.009 *** (0.001)	0.007 *** (0.001)	0.009 *** (0.001)
Home team wins	0.184 *** (0.003)	0.184 *** (0.002)	0.184 *** (0.003)	0.184 *** (0.002)	0.182 *** (0.003)	0.182 *** (0.002)	0.184 *** (0.003)	0.184 *** (0.002)	0.184 *** (0.003)	0.184 *** (0.002)	0.182 *** (0.003)	0.183 *** (0.002)
Strong team wins	0.293 *** (0.004)	0.273 *** (0.003)	0.293 *** (0.004)	0.273 *** (0.003)	0.312 *** (0.004)	0.284 *** (0.003)	0.295 *** (0.004)	0.274 *** (0.003)	0.295 *** (0.004)	0.274 *** (0.003)	0.311 *** (0.004)	0.284 *** (0.003)
Game week	−0.003 *** (0.000)	−0.004 *** (0.000)	−0.007 *** (0.001)	−0.007 *** (0.001)			−0.003 *** (0.000)	−0.004 *** (0.000)	−0.008 *** (0.001)	−0.007 *** (0.001)		
Main teams	0.043 *** (0.003)	0.053 *** (0.002)	0.042 *** (0.003)	0.052 *** (0.002)	0.041 *** (0.003)	0.052 *** (0.002)	0.043 *** (0.003)	0.055 *** (0.002)	0.042 *** (0.003)	0.054 *** (0.002)	0.041 *** (0.003)	0.053 *** (0.002)
Amount by user	0.006 (0.005)	−0.017 (0.021)	0.008 (0.005)	−0.017 (0.021)	0.001 (0.004)	−0.010 (0.016)	0.003 (0.007)	−0.097 ** (0.039)	0.006 (0.007)	−0.195 *** (0.039)	0.004 (0.004)	−0.049 (0.031)
Other events	0.007 *** (0.000)	0.009 *** (0.000)	0.007 *** (0.000)	0.009 *** (0.000)	0.007 *** (0.000)	0.008 *** (0.000)	0.006 *** (0.000)	0.008 *** (0.000)	0.007 *** (0.000)	0.009 *** (0.000)	0.007 *** (0.000)	0.008 *** (0.000)
Odds	−0.095 *** (0.001)	−0.098 *** (0.001)	−0.095 *** (0.001)	−0.098 *** (0.001)	−0.094 *** (0.001)	−0.097 *** (0.001)	−0.094 *** (0.001)	−0.097 *** (0.001)	−0.094 *** (0.001)	−0.097 *** (0.001)	−0.094 *** (0.001)	−0.096 *** (0.001)
2004/05 season	0.007 (0.006)	−0.030 *** (0.005)	0.015 (0.006)	−0.046 *** (0.005)			0.008 (0.009)	−0.038 *** (0.007)	−0.019** (0.009)	−0.051 *** (0.007)		
Game week squared			0.000 *** (0.000)	0.000 *** (0.000)					0.000 *** (0.000)	0.000 *** (0.000)		
Game week dummies	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Individual fixed effects	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	596,037	609,538	596,037	609,538	596,037	609,538	596,037	609,538	596,037	609,538	596,037	609,538
No of individuals	1352	7014	1352	7014	1352	7014	1352	7014	1352	7014	1352	7014
Wald test <i>p</i> -value	0.933		0.999		0.364		0.594		0.543		0.148	

Notes. Dependent variable: probability of correctly forecasting a single event. Estimated model: linear probability model as in Equation (1) in separate subsamples (bets linked to higher-than-median amount vs. others). Standard errors clustered at the event level are reported in parentheses. Significance at the 5% is represented level by **, and at the 1% level by ***. The Wald test *p*-value captures the significance of the difference between the coefficients of betting distance in the two subsamples.

References

- Anderson, Patric, Jan Edman, and Matias Ekman. 2005. Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting* 21: 565–76. [\[CrossRef\]](#)
- Asch, Peter, Burton G. Malkiel, and Richard E. Quandt. 1984. Market efficiency in racetrack betting. *Journal of Business* 57: 165–75. [\[CrossRef\]](#)
- Atanasov, Pavel, Jens Witkowski, Lyle Ungar, Barbara Mellers, and Philip Tetlock. 2020. Small steps to accuracy: Incremental belief updaters are better forecasters. *Organizational Behavior and Human Decision Processes* 160: 19–35. [\[CrossRef\]](#)
- Barber, Brad M., and Terrance Odean. 2002. Online investors: Do the slow die first? *Review of Financial Studies* 15: 455–87. [\[CrossRef\]](#)
- Benartzi, Shlomo, and Richard H. Thaler. 2001. Naive diversification strategies in retirement saving plans. *American Economic Review* 91: 79–98. [\[CrossRef\]](#)
- Bettman, James R., Eric J. Johnson, Mary Frances Luce, and John W. Payne. 1993. Correlation, conflict, and choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19: 931–51. [\[CrossRef\]](#)
- Bichler, Urs, and Monica Butler. 2007. *Information Economics*. London: Routledge.
- Blavatsky, Pavlo R. 2009. Betting on own knowledge: Experimental test of overconfidence. *Journal of Risk and Uncertainty* 38: 39–49. [\[CrossRef\]](#)
- Bleichrodt, Han, and Ulrich Schmidt. 2002. A Context-Dependent Model of the Gambling Effect. *Management Science* 48: 802–12. [\[CrossRef\]](#)
- Buhagiar, Ranier, Dominic Cortis, and Philip W. S. Newall. 2018. Why do some soccer bettors lose more money than others? *Journal of Behavioral and Experimental Finance* 18: 85–93. [\[CrossRef\]](#)
- Camerer, Colin F., and Eric J. Johnson. 1991. The process–performance paradox in expert judgment: How can experts know so much and predict so badly? In *Toward a General Theory of Expertise: Prospects and Limits*. Edited by K. Anders Ericsson and Jaqui Smith. New York: Cambridge Press, pp. 195–217.
- Conlisk, John. 1993. The utility of gambling. *Journal of Risk and Uncertainty* 6: 255–75. [\[CrossRef\]](#)
- Crawford, Vincent P., and Joel Sobel. 1982. Strategic information transmission. *Econometrica* 50: 1431–51. [\[CrossRef\]](#)
- Davis, Fred D., Gerald L. Lohse, and Jeffrey E. Kottmann. 1994. Harmful effects of seemingly helpful information on forecasts of stock earnings. *Journal of Economic Psychology* 15: 253–67. [\[CrossRef\]](#)
- Diecidue, Enrico, Ulrich Schmidt, and Peter P. Wakker. 2004. The Utility of Gambling Reconsidered. *Journal of Risk and Uncertainty* 29: 241–59. [\[CrossRef\]](#)
- Dieckmann, Anja, and Jörg Rieskamp. 2007. The influence of information redundancy on probabilistic inferences. *Memory & Cognition* 35: 1801–13.
- Figlewski, Stephen. 1979. Subjective information and market efficiency in a betting market. *Journal of Political Economy* 87: 75–88. [\[CrossRef\]](#)
- Fister, Iztok, Jr., Karin Ljubič, Ponnuthurai Nagarathnam Suganthan, Matjaž Perc, and Iztok Fister. 2015. Computational intelligence in sports: Challenges and opportunities within a new research domain. *Applied Mathematics and Computation* 262: 178–86. [\[CrossRef\]](#)
- Forrest, David, and Robert Simmons. 2000. Forecasting sport: The behaviour and performance of football tipsters. *International Journal of Forecasting* 16: 317–31. [\[CrossRef\]](#)
- Gigerenzer, Gerd, and Daniel G. Goldstein. 2011. The recognition heuristic: A decade of research. *Judgment and Decision Making* 6: 100–21.
- Gigerenzer, Gerd, Peter M. Todd, and The ABC Research Group. 1999. *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.
- Golec, Joseph, and Maury Tamarkin. 1995. Do bettors prefer long shots because they are risk-lovers, or are they just overconfident? *Journal of Risk and Uncertainty* 11: 51–64. [\[CrossRef\]](#)
- Golman, Russel, David Hagmann, and George Loewenstein. 2017. Information avoidance. *Journal of Economic Literature* 55: 96–135. [\[CrossRef\]](#)
- Hausch, Donald B., William T. Ziemba, and Mark Rubinstein. 1981. Efficiency of the market for racetrack betting. *Management Science* 27: 1435–52. [\[CrossRef\]](#)
- Kaufmann, Christine, and Martin Weber. 2013. Sometimes less is more—The influence of information aggregation on investment decisions. *Journal of Economic Behavior & Organization* 95: 20–33.
- Kochman, Lad, Randy Goodwin, and David Bray. 2015. So easy a caveman can beat the football betting market. *The American Economist* 60: 225–28. [\[CrossRef\]](#)
- Martignon, Laura, and Ulrich Hoffrage. 2002. Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision* 52: 29–71. [\[CrossRef\]](#)
- Newall, Philip W. S., and Dominic Cortis. 2021. Are sports bettors biased toward longshots, favorites, or both? A literature review. *Risks* 9: 22. [\[CrossRef\]](#)
- Newell, Ben R., and David R. Shanks. 2004. On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30: 923–35. [\[CrossRef\]](#)
- Osborne, Evan. 2001. Efficient Markets? Don't Bet on It. *Journal of Sports Economics* 2: 50–61. [\[CrossRef\]](#)

- Oskamp, Stuart. 1965. Overconfidence in case-study judgments. *Journal of Consulting Psychology* 2: 261–65. [[CrossRef](#)]
- Palomino, Frederich, Luc Rennebog, and Chendy Zhang. 2009. Information salience, investor sentiment, and stock returns: The case of British soccer betting. *Journal of Corporate Finance* 15: 366–87. [[CrossRef](#)]
- Paul, Rodney J., and Andrew P. Weinbach. 2005. Bettor Misperceptions in the NBA. *Journal of Sports Economics* 6: 390–400. [[CrossRef](#)]
- Pope, Peter F., and David A. Peel. 1989. Information, prices and efficiency in a fixed-odds betting market. *Economica* 56: 323–41. [[CrossRef](#)]
- Rieskamp, Jörg, and Phillip Otto. 2006. SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General* 135: 207–36. [[CrossRef](#)] [[PubMed](#)]
- Snyder, Wayne W. 1978. Horse racing: Testing the efficient markets model. *Journal of Finance* 33: 1109–18. [[CrossRef](#)]