

Limited cognition and clustered asset prices: evidence from betting markets*

Alasdair Brown[†]

Fuyu Yang[‡]

University of East Anglia

University of East Anglia

*We would like to thank Tarun Chordia, the editor, and an anonymous referee for very helpful comments and suggestions. We are also grateful to Kathryn Clark for assistance with the copy-editing. Financial support from UEA is gratefully acknowledged. The research presented in this paper was carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia.

[†]Corresponding Author. School of Economics, University of East Anglia, Norwich NR4 7TJ, U.K. Email: alasdair.brown@uea.ac.uk Tel: 44 1603 591131

[‡]School of Economics, University of East Anglia, Norwich NR4 7TJ, U.K. Email: fuyu.yang@uea.ac.uk
Tel: 44 1603 591058

Abstract

Asset prices tend to cluster at round numbers. We examine betting exchange data on U.K. horse races to establish whether limited cognition is partially responsible for this clustering. The key tool in this study is the stark increase in cognitive load faced by traders during races compared to prior to races. Using an approach that is part regression discontinuity and part difference-in-differences, we find that traders exhibit a substantially higher propensity to quote round numbers, rather than the nearest non-round numbers, during races. This result is robust to a series of placebo tests.

JEL classification: G02, G12, G14

Keywords: Limited cognition, Price clustering, Regression discontinuity, Difference-in-differences

1 Introduction

There is long-standing evidence that stock prices cluster at round numbers (e.g., Niederhoffer, 1965, 1966; Harris, 1991; Ahn, Cai, and Cheung, 2005; Sonnemans, 2006; Ikenberry and Weston, 2007; Bhattacharya, Holden, and Jacobsen, 2012; Kuo, Lin, and Zhao, 2015). It is unlikely that the fundamental values of stocks cluster at round numbers, and therefore — at least recently, as models of bounded rationality (Simon, 1955) have become more popular in financial economics — price clustering has been attributed to limited cognition on the part of traders. The story is that attention-constrained traders latch onto cognitive reference points, such as round numbers, when they do not have the mental capacity to make finer price distinctions.

Yet, if limited cognition is indeed causing price clustering, we should observe greater (lower) levels of clustering when traders are more (less) cognitively constrained. In this paper, we use a betting exchange in the United Kingdom as a unique laboratory to test this hypothesis, and therefore more firmly establish the role of limited cognition in the clustering of asset prices.

We examine Betfair betting exchange trading on the two major horse race meetings in the U.K: Royal Ascot and the Cheltenham Festival. We compare the propensity for bettors to quote odds at a range of round numbers with their propensity to quote odds at the nearest non-round number. A round number and its nearest neighbor reflect almost identical win probabilities: the only difference is that one price is arguably a cognitive reference point, and therefore more likely to be quoted by the cognitively-constrained bettor, and the other is not. Importantly, we compare these propensities prior to races, when there is little new

information arriving, with the same propensities during races, when important information on the relative positions of horses arrives in rapid succession. The idea is that bettors are more likely to be cognitively constrained during races, when the cognitive load is greater. We should therefore observe more round number quotes during races.

We find that bettors are substantially more likely to quote a round number than the nearest non-round number during races, compared to prior to races. This result is perhaps most vividly captured by Figure 1. This figure contains a simple bar chart of quoted odds (including the stake) at Royal Ascot in 2011, both pre-race (in Panel A) and during races (in Panel B). While there is some evidence that certain round numbers are quoted more frequently than other odds prior to races, all round numbers are substantially more likely to be quoted once races begin. This suggests that the cognitively constrained trader, forced to process the rapid unfolding of each race, employs round numbers as a heuristic method for pricing assets.

The logic of the empirical approach in this paper can be thought of as part regression discontinuity (RD) design, and part difference-in-differences (DID). The decision to quote a round number, or not, can be thought of as the "running" or "assignment" variable in a RD. In common with RD analysis, we assume that horses quoted at a round number, or just below, have almost identical characteristics (in this case, win probabilities). Therefore, in the absence of cognitive constraints, the bettor should be equally likely to quote a round number as its nearest neighbor. We are then interested in the extent to which the assignment decision (i.e., quoting a round number) is determined by the treatment. In our setting, the treatment is the running of each race and the increase in cognitive load faced by bettors. As with DID analysis, we compare the propensity to quote a round number relative to its nearest neighbor

prior to races (pre-treatment), with the same propensity during races (during the treatment). We observe that the propensity to quote a round number compared to its nearest neighbor substantially increases as the cognitive load of the bettor increases.

The setting for this study is chosen primarily for reasons of internal validity. To be specific, the separation of activity into pre-race and in-running betting creates stark variation in the cognitive load facing traders. Such variation would be more difficult to identify in many core financial markets. That being said, the results obtained on the betting exchange can still carry significant external validity. Betfair is the world's largest betting exchange, with 11.8 million GBP in bets placed on a single race (The Grand National) in 2012 (Betfair Annual Report 2012). Secondly, the market structure is a limit order book (more on this in Section 3), which is similar to that used in the majority of financial markets. Finally, the betting exchange has been in operation since 2000, and while this does not guarantee that subjects are experienced, this does at least give bettors the opportunity to have become familiar with the market structure.

As asset price clustering is a form of mispricing, this work is related to a growing literature on the effect of limited cognition on asset mispricing. This literature includes, among others, DellaVigna and Pollet (2009) and Hirshleifer, Lim, and Teoh (2009). These studies present evidence to suggest that mispricing (i.e., the post-earnings announcement drift) is more pronounced when traders are otherwise distracted, either because the weekend is imminent, or because there are a large number of announcements on the same day.

In using betting exchange data — and the concentration of information arrival during sporting events — we use a similar approach to that in Brown (2014). In that paper, Wimbledon tennis betting data were used to establish the role that information processing constraints

play in asset mispricing. Mispricing between two equivalent assets — a bet on a player to win in one market, and a replicating portfolio of bets on the same player in another market — increases substantially as information arrives during each tennis match. Part of the reason for this increase is that information is not synchronously incorporated into the two prices, thereby suggesting that information processing constraints are binding. In this paper, we are interested in the effect of limited cognition on mispricing between one asset and its fundamental value, rather than relative mispricing between two equivalent assets. "Absolute" mispricing, considered in this paper, can be measured by the extent to which asset prices cluster at round numbers.

While limited cognition is now arguably the dominant explanation for the clustering of asset prices at round numbers, ours is the first paper (to our knowledge) to show that the frequency of round number quotes increases when traders are subject to a greater cognitive load. In part, we are able to make this contribution due to the great contrast between the amount of information that needs to be processed during horse races, compared to prior to races. While earnings announcements and other announcements may increase the cognitive load on financial market traders, the contrast between these trading periods and other periods is less stark, as financial information arrives in an almost continuous fashion.

In addition, we also argue that our setting allows us to distinguish between different channels through which limited cognition may lead to round number prices. On the one hand, round numbers may arise due to the cognitive limitations of the trader who originally proposes the price. Kuo, Lin, and Zhao (2015) show that traders on the Taiwan Futures Exchange who frequently quote round numbers are amongst the worst performers (in terms of trading profits). This suggests that traders' own cognitive limitations lie behind their

predilection for round number quotes. On the other hand, round numbers may be used as a "trick" by one set of traders to induce orders from other cognitively constrained traders. Bhattacharya, Holden, and Jacobsen (2012) find pronounced selling at round numbers and pronounced buying just below these round numbers, suggesting a "left-digit bias" (see Section 2) that can be exploited by traders if they quote round numbers. To distinguish between these two channels, the variation in cognitive load — pre-race versus in-running — in our setting is again useful. During horse races (when the cognitive load is greater), we observe a greater propensity to quote round numbers, but not a greater propensity to order at round numbers. Our analysis therefore suggests that round number quotes arise more often because of the cognitive limitations of the trader who proposes the price, rather than the trader that takes the price.

The rest of the paper is ordered as follows. In Section 2 we provide some psychological foundations for the analysis and survey the relevant economic literature. In Section 3 we describe the data, and in Section 4 we present the main analysis. In Section 5 we describe a series of robustness tests, and we conclude in Section 6.

2 Psychological foundations

In this paper we test whether an increase in the cognitive load facing traders — stimulated by the arrival of information during each race — increases the probability that traders quote round numbers. The designation of round numbers as "cognitive reference points," likely to be latched on to by traders when they are cognitively constrained, is motivated by the work of Rosch (1975).

Rosch (1975) conducted a series of experiments, one of which showed that non-round numbers are evaluated in the context of their relation to round numbers. The experimental method was to give subjects a statement with blanks, such as $_$ is essentially $_$. Subjects would then be given two numbers to fill the blanks, one a round number such as 1000, and the other a non-round number such as 996. If 1000 was seen as the cognitive reference point, we should observe more subjects returning the statement "996 is essentially 1000" than the opposite statement. This was indeed the result that was found. To ensure the robustness of the results, Rosch also asked the same question with two non-round numbers and found that the two numbers were equally likely to be placed in either space.¹

Later in this paper we attempt to establish whether round numbers are used by cognitively constrained traders as a heuristic method for pricing assets, or are used to take advantage of the cognitive limitations of their potential trading counterparty. For example, those ordering at quotes already in the limit order book may exhibit what is termed, at least in the economics literature if not the psychology literature, the "left-digit bias." The identification of this bias is based on the work of Hinrichs, Yurko, and Hu (1981) and Poltrock and Schwartz (1984), who found that subjects evaluate numbers sequentially, focusing on the left digit first. A cognitively constrained subject, therefore, may not get around to evaluating the second and third digits. The implication for our setting is that there may be a jump in the probability that a bettor places an order at odds of 2, compared to 1.99, but no such jump in the

¹We are not aware of any study in experimental psychology that examines whether the propensity to choose the round number as the cognitive reference point is affected by the amount of time pressure that subjects are under. Even if there were such a study, we may wish to see whether this effect also occurs in an economic setting (e.g., the betting market in our project).

probability of an order at odds of 2.02 compared to 2. It is possible that bettors quote a higher proportion of round numbers during races as this is when their potential counterparty is subject to a greater cognitive load, and therefore more likely to exhibit such a left-digit bias. We will test this hypothesis in Section 5.

There is in fact quite prominent evidence consistent with a left-digit bias in other economic settings. For example, one explanation for the prevalence of 99 cents (pence) pricing is that consumers concentrate on the left digit, so, for example, a \$1.99 product is regarded as disproportionately cheaper than a \$2 product (Ginzberg, 1936). More recently, Lacetera, Pope, and Syndor (2012) and Busse et al. (2013) identified a left-digit bias in the used car market. They observed large discontinuities in the prices of cars depending on the number of miles driven. For example, there were large drops in the price of a car once it hit 10,000 miles. Lacetera, Pope, and Syndor (2012) also conducted a placebo test using Canadian data, where distance is calculated in kilometers, and found a similar discontinuity at the 10,000 kilometer mark. Car owners, in the U.S. at least, appear to respond to this bias by bringing a sizeable number of cars with just less than 10,000 miles on the odometer to the market.

3 Data

The data in this study are taken from Betfair, a betting exchange based in the U.K. The exchange allows bettors to bet on, or against, a horse to win. Bets on a horse to win are labelled "back" bets, and bets on a horse to lose are labelled "lay" bets. (A bettor laying a horse is allowing their counterparty to back the horse.) Bettors are also able to wager on a horse to place (a horse places if they do not win but finish in a high-ranking position).

The Betfair exchange operates as a standard limit order book. This means that bettors can provide quotes (where they specify the odds), or execute their bets at odds quoted by others. When a bettor provides a quote, this is labelled a "limit order," and when a bettor trades at a quote provided by another bettor, this is labelled a "market order." These betting options reflect a widening of choice for punters. Prior to the inception of these types of exchanges, bettors could only take long positions on a horse to win or place, and could only execute at prices quoted by a bookmaker. Now each bettor can submit quotes themselves.²

Figure 2 provides a screenshot of the Betfair limit order book. This example is taken from the 13.45 race at Wolverhampton on April 29, 2015, a race outside of our sample. All back bet quotes (wagering on a horse to win) can be found on the left, with all lay bet quotes (betting on the horse to lose) on the right. The best back and lay quotes are nearest the center. Prices on the exchange are quoted in the form of odds. Odds of 7, as quoted on Black Truffle for example, would return 7 GBP for every 1 GBP staked if the horse in question won. If the horse did not win, the trader that quoted that limit order would pocket the amount staked by their counterparty.

Horses are listed in each market from the expected favorites (those with the lowest odds) at the top, to the expected longshots at the bottom. Bettors who wish to provide quotes can specify the direction of their bet (back or lay), the odds that they wish to offer (or request), and the volume involved in the bet. If these specific requirements are not fully met by existing quotes, then the punter's quote enters the book as a limit order.

²Betfair also operate a parallel service, the Sportsbook, where they operate as a traditional bookmaker. The service is intended for less popular sporting events, where liquidity on the main exchange may be limited. Historical data for Betfair's Sportsbook pricing are not available.

The pricing grid, which determines the precision of quotes, becomes more coarse for higher odds. For example, odds can be quoted in increments of 0.01 between 1 and 2 (i.e., 1.01, 1.02 and so on), but only in increments of 0.02 between 2 and 3.³ The increased coarseness of the pricing grid for higher odds is one reason why we choose fewer round numbers for higher odds (more on this in Section 4). If a bettor enters an impermissible quote, such as 2.01 for example, then the exchange automatically rounds up the quote to the nearest permissible number (in this case, 2.02). As with quotes at permissible odds, the bettor can then decide whether to proceed with the bet or not.⁴

Once a quote is in the book, other bettors can take the other side of the bet. However, if a bet has not yet been taken, the bettor providing the quote is free to cancel or update the price and volume that they wish to bet. This freedom is particularly important during races, as quotes will quickly become outdated.

The data in this paper are sourced from Fracsoft, a third-party vendor of Betfair limit order book data. We have chosen to study two prominent horse race meetings across two

³Odds are quoted in increments of 0.05 between 3 and 4, of 0.1 between 4 and 6, of 0.2 between 6 and 10, of 0.5 between 10 and 20, of 1 between 20 and 30, of 2 between 30 and 50, of 5 between 50 and 100, and of 10 between 100 and 1000.

⁴It is possible that Betfair's procedure for dealing with impermissible odds contributes to the increase in the frequency of round numbers quoted during races. If bettors are more likely to make mistakes in-running — and quote impermissible odds — then we will see quotes of 2.99, for example, rounded up to 3 (our designated round number) rather than rounded down to 2.98 (our designated nearest neighbor). However, this potential confound should not be such an issue when we designate 2 as our round number, and 1.99 as our nearest neighbor. Without going beyond 2 decimal places, and therefore expecting payment in a fraction of pennies, it is not possible for bettors to enter an impermissible quote in this range. And, as we shall see in Table 6, quotes of 2 still become much more likely than quotes of 1.99 during races.

years. We examine the premier flat racing meeting in the U.K., Royal Ascot, which takes place in June each year, and also the premier jump racing event, the Cheltenham Festival, which takes place in March each year. Royal Ascot lasts for five days, with six races per day, while the Cheltenham Festival lasts for four days, with seven races per day.⁵ At least 394 horses take part in each of these meetings.

For this study, we examine the odds available on a horse to win and also the odds on a horse to place. These odds are displayed in separate markets. For every horse, the best-priced quotes are sampled every second for 20 minutes prior to each race (the pre-running phase), and every second for the full duration of each race (the in-running phase). As there is a 30-40 minute gap between the start times of each race, pre-race bettors are not distracted by the running of a previous race. Immediately prior to the beginning of each race, the market is suspended for a few seconds and all quotes are canceled by the exchange. The market is then reopened as the race begins, and quotes can then begin to fill the book once more. This means that there are no stale pre-race quotes leftover during the in-running period.

4 Analysis

Betfair allow for quotes from 1.01 (i.e., 1 penny returned for each pound staked), to 1000 (999 pounds returned for each pound staked). We classify the round numbers in this study as 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, and 50. This includes the majority of horses that are in contention, but not so certain to win or place that pricing their bets becomes a

⁵Data are available on the Fracsoft database for all days, except June 17, 2011 and June 19, 2012, both at Ascot.

simple exercise. The round numbers reflect implied probabilities from 0.5 (odds of 2), to 0.02 (odds of 50). We pick fewer round numbers from 10 onwards as the pricing grid becomes significantly more coarse at this point.

For back bets we classify the price immediately below each of these round numbers as the nearest neighbors. These odds are 1.99, 2.98, 3.95, 4.9, 5.9, 6.8, 7.8, 8.8, 9.8, 14.5, 19.5, 29, 38, and 48. We pick the odds just below the round number as a control price because quotes just above the round number could become prevalent because of undercutting. To be specific, a cognitively constrained trader may quote a round number because of its role as a cognitive reference point but be immediately undercut by another (perhaps algorithmic) trader simply using the initial quote as a reference point and improving upon it. In Figure 1 there is visible tapering of the frequency of quotes to the right of a round number, particularly during races, which suggests that systematic undercutting does take place. For the nearest neighbors of lay bets, we consider the odds immediately above the round number, as undercutting on lay bets will work in the opposite direction. This means odds of 2.02, 3.05, 4.1, 5.1, 6.2, 7.2, 8.2, 9.2, 10.5, 15.5, 21, 32, 42, and 55.

Before we proceed, we should clarify why we are interested in the prevalence of round numbers relative only to their nearest neighbors, rather than relative to all other numbers. The main issue is that the distribution of prices (odds) changes greatly between pre-race and in-running periods. Prior to the races in our four meetings, 80.4% of back quotes can be found in the interval between 2 and 50, our smallest and largest designated round numbers. Contrast this with the in-running period, when only 68.7% of back quotes can be found in the interval between 2 and 50. As each race nears completion, the imminent winners will be quoted at 1.01 (if at all), and the imminent losers will be quoted at 1000 (if at all). This

means that non-round numbers, which are outside of our range between 2 and 50, will likely increase in prevalence during races, simply because horses drift to extreme odds and the fundamental value of the bet either becomes very low or very high.

With this change in the distribution of odds in mind, we wanted to find appropriate odds to compare to our round numbers. We therefore choose only the odds closest to the round numbers, their nearest neighbors. Round numbers and their nearest neighbors reflect almost identical win probabilities and therefore, in the absence of cognitive constraints, we should be just as likely to observe a round number as its nearest neighbor. This argument applies both pre-race and in-running. The result is that if a round number becomes more prevalent than its nearest neighbor during races, after controlling for any differences pre-race, we can more reliably apportion this effect to limited cognition.

By comparing the incidence of round numbers only to their nearest neighbors, we also reduce the importance of our choice of round numbers. We choose every whole number from 2 to 10, as many horses will find themselves quoted in this range, and then taper the number of round numbers from then on (15, 20, 30, 40 and 50), as the pricing grid becomes more coarse for higher odds. One might ask, however, why not 25? Any why not 100, or 200? And going the other way, why not 1.1 through to 1.9, as these numbers are round compared to 1.11 and 1.91? A case could be put for each of these numbers, but the important thing is that by comparing our round numbers only to their nearest neighbors, we negate the effect of any drift in fundamental values during races. A quote of 50, for example, could perhaps become more likely as races progress and horses drop out of contention, but, in the absence of cognitive constraints, such a price is almost equally likely as a price of 48 or 55. Similarly a quote of 2 may become more likely as a horse nears the finish line in the lead, but, again,

in the absence of cognitive constraints, such a quote should be just as likely as a quote of 1.99 or 2.02.

In Table 1 we provide summary statistics on the four race meetings separately. We examine the percentage of seconds where there is at least one quote, and the percentage of seconds where there is a round number quote, or a nearest neighbor quote. These percentages are compared for back and lay bets, and for pre-race and in-running phases.

First, it is apparent that liquidity declines, particularly on the lay side, during races. For example, we only have lay quotes 54.96% of the time for the win market at Royal Ascot in 2011 during races. Second, across both markets, and across all events, the pre-race proportion of round numbers and nearest neighbors is approximately 10%. Once we evaluate in-running data, however, there are marked differences between the two markets. While the proportion of round numbers and nearest neighbors declines slightly in the win market, there are substantial increases in round numbers in the place market. As argued above, round numbers and nearest neighbors can be found in the central mass of the implied probability distribution. Therefore, once a race begins, the probability that a horse wins drifts towards the tails — either as victory becomes likely, or defeat inevitable — thereby leaving fewer observations at round numbers or nearest neighbors. This does not necessarily occur in the place market, where a greater proportion of horses may remain in contention for much of the race. These summary statistics illustrate why it is sensible to compare the frequency of round numbers, not with all other quotes, but only with their nearest neighbors.

The analysis that follows therefore excludes all quotes except the round numbers and their nearest neighbors. This means discarding a great deal of data. For example, the Royal Ascot 2011 place market data includes 720,880 observations, which is reduced to 114,967

when we consider only round numbers and their nearest neighbors. By focusing on this small set of numbers, we are only picking quotes that reflect almost identical win or place probabilities. For example, odds of 4 in the win market imply a win probability of 0.25, while the nearest neighbor odds of 3.98 reflect an implied win probability of 0.251. Therefore, we assume that "nature" is almost equally likely to draw a horse with a win probability of 0.25 as one with a probability of 0.251. The key difference is that one of the odds is a round number — and therefore, we hypothesise, more likely to be latched onto by the cognitively constrained bettor — and one is not. To ensure that limited cognition is indeed behind the use of round numbers, we then exploit variation in the cognitive load facing traders. This variation is created by the running of each race, and the resultant need to process rapidly arriving information and place bets in a short space of time.

Table 2 presents the results of the main analysis. We regress an indicator variable equaling 1 if a round number (previously defined) is quoted, on an indicator variable equaling 1 if the race was in-running at that stage. The results for back quotes are in Panels A and B. Only data on round numbers, and the odds immediately below (nearest neighbors) are included. A logit specification is used, and random effects are included for each horse (bet) in the sample.⁶ In Panel A we include win market data, with place market data in Panel B. These regressions are carried out for Royal Ascot in 2011 and 2012, and the Cheltenham Festival in 2011 and 2012. The coefficient associated with the intercept term suggests that there is

a small baseline prevalence for quoting round numbers for back bets. The treatment effect,

⁶In addition to this specification, we also checked the results with two-way clustering of standard errors, albeit without random effects, when the two clusters were the bet and the race, and also when the two clusters were the bet and the time. Our results are qualitatively similar to those that we present.

however, captured by the in-running indicator, is of greater importance. We find that as races begin there is a large and significant increase in the propensity to quote a round number, for place markets, and a smaller increase for win markets (all significant at the 0.1% level). This is the case across all four horse race meetings. The coefficients imply an odds ratio of between 1.29 (Royal Ascot 2012, win market) and 5.25 (Cheltenham 2012, place market) of observing a round number quote in-running relative to prior to races. This treatment effect suggests that the clustering of quotes at round numbers is indeed a manifestation of limited cognition.⁷

In the bottom half of Table 2, we show the results of an analysis for quotes on the lay side of the bet. We regress an indicator variable equaling 1 if a lay bet was quoted at a round number, on an indicator variable equaling 1 if the race was in-running at that point. Win (place) market data are in Panel C (Panel D), a logit specification is used, and random effects for each horse (bet) are included. The intercept term suggests that lay quotes are more likely to be nearest neighbors than round numbers prior to races. If you have a back bet at a round number, the lay bet cannot be at the same round number (otherwise the two bets will execute with each other, and immediately disappear from the book). Therefore, if a bettor quotes the minimum spread (as is common prior to races), we are likely to observe many lay bets at nearest neighbors, purely as a result of the observed tendency to quote back

⁷At this stage we should clarify why there are differences in the number of random effects (bets) in win and place markets for the same meeting (e.g., Ascot, 2011). As we are only using the quotes at round numbers or nearest neighbors, a bet is only incorporated in our analysis if, at some point pre-race or in-running, it is priced at one of these figures. We are likely to see more bets in our place market analyses than our win market analyses, as place odds are more likely to be in the middle of the implied win probability distribution, where most of our round numbers and nearest neighbors are found.

bets at round numbers. Once races begin however, and spreads widen, this mechanical effect is not present. For all four events, and both markets, a pronounced in-running propensity to quote round numbers is observed (with statistical significance at the 0.1% level).

In the next section, we examine the robustness of our results, and whether there are convincing alternative explanations for the prevalence of round number quotes during races.

5 Robustness tests

Our first concern at this stage of our analysis is that, due to the prevalence of undercutting, for back bets we have exclusively chosen nearest neighbors that are priced just below the round number. Therefore, we may be confounding the limited cognition effect with subtle drifts in the odds that occur when a race begins. As the majority of horses do not win or place, most will see their odds increase during the in-running period. Therefore a comparison of the frequency of, for example, odds of 2.98 and 3 could find slightly more 3s when races are under way, as horses that were hitherto priced at 2.98 slowly drift out of contention.

To deal with this problem, we run a series of placebo tests on the four race meetings. For these tests, we define the original nearest neighbors — 1.99, 2.98, 3.95, 4.9, 5.9, 6.8, 7.8, 8.8, 9.8, 14.5, 19.5, 29, 38, and 48 — as the "placebo" round numbers for our analysis of back quotes. We then create new nearest neighbors just one price down. These new nearest neighbors are 1.98, 2.96, 3.9, 4.8, 5.8, 6.6, 7.6, 8.6, 9.6, 14, 19, 28, 36, and 46. If our results for back bets can be apportioned to a general drift in the odds during races, we should observe a similar effect when we repeat the analysis with these odds. We also conduct placebo tests for lay quotes. Here we potentially have the opposite issue. Perhaps there is a drift down in odds

during races (this may be more of an issue in place markets as more horses place than win), which we have attributed to limited cognition because our round numbers are lower than our nearest neighbors for lay quotes. We therefore reclassify our nearest neighbors for lay quotes — 2.02, 3.05, 4.1, 5.1, 6.2, 7.2, 8.2, 9.2, 10.5, 15.5, 21, 32, 42, and 55 — as "placebo" round numbers, and create new nearest neighbors at the following odds: 2.04, 3.1, 4.2, 5.2, 6.4, 7.4, 8.4, 9.4, 11, 16, 22, 34, 44, and 60. If our lay quote results are due to a drift down in odds as races progress, we should observe a similar effect here.

Table 3 presents the results of a regression where an indicator variable equaling 1 if a placebo round number is quoted, is regressed on an indicator variable equaling 1 if the race in question was in-running at that stage. Only placebo round numbers, and the odds immediately below them (for back quotes) or above (for lay quotes), are included in the regressions. As before, a logit specification is used, and random effects for each horse (bet) are included. Back quote analysis is in the top half, with lay quote analysis in the bottom half. Win (place) market data can be found in Panels A and C (Panels B and D). Firstly, we will discuss the back quote results. Judging by the eight regression results in the top half of Table 3, one for each market in each race meeting, there is little evidence that the in-running effect we observed in Table 2 is due to a subtle drift upwards in the odds in-running. In fact, the coefficients associated with the in-running indicator are negative and significant, albeit quite small. This indicates that the round numbers absorb such a high proportion of their nearest neighbors that these neighbors are slightly less likely to arise than other non-round numbers. For the lay quote analysis, there are two instances when the placebo effect is positive — for the place market for Ascot 2011 and 2012 — but overall the majority of placebo estimates are negative. These results re-affirm the impression that limited cognition

was driving the increase in round number quotes during races observed in Table 2, rather than drifts in the distribution of odds.

At this stage we have accounted for any subtle drifts in the distribution of odds during races, but perhaps there are more drastic shifts in both the mean and variance of the distribution of odds that are affecting our results. Ideally we would like to compare the frequency of round number quotes pre-race and in-running, while broadly holding the distribution of odds fixed. Unfortunately, this is difficult, as odds drift to the extremes (1.01 and 1000) during races as uncertainty is resolved and horses near victory or defeat. (Incidentally, this is one reason why we did not choose 1000 as a round number as odds would mechanically gravitate to this number during races, as it is the highest.)

There is, however, a matching approach that we can take to deal with this issue. Firstly, we calculate the mean and standard deviation of the odds, for each market, for each race, in each second. This gives us two moments of the distribution of the odds at a given time. We then match every pre-race odds observation (at a round number or nearest neighbor), with an in-running observation (at a round number or nearest neighbor), which was quoted when the distribution of odds (including all prices at that time-point) were similar. We then ask whether round numbers are more common pre-race or in-running, amongst this matched sample. In other words, we are comparing the frequency of round number quotes, relative to their nearest neighbors, while (almost) holding the distribution — or at least the first two moments of the distribution — fixed. This matching procedure, which is described in Abadie and Imbens (2011), matches on the basis of the Mahalanobis distance between the matching variables (in our case the mean and standard deviation of odds at that timepoint), and is

known as nearest neighbor matching.⁸ However, we will refer to it as simply "matching" so as not to confuse it with our use of the term "nearest neighbors" (the prices closest to round numbers).

In order to understand the effect that a changing distribution of odds may be having on our results, we also estimate a linear probability model, where the estimates are more directly comparable to the matching estimates than our earlier logit regressions. For the linear probability model, estimated by OLS, we regress an indicator variable equaling 1 if a round number is quoted, on an indicator variable equaling 1 if the race was in-running at that stage. As before, only round numbers and their nearest neighbors are included in the regressions.

Our matching and linear probability model estimation results are in Table 4. We display only the coefficient associated with the in-running indicator. We breakdown the analysis, by event, market, and back/lay quote. While we use all quoted odds to calculate the mean and standard deviation of the odds at a given time, to give us a sense of the distribution of odds at the time, we still only incorporate round numbers and their nearest neighbors to estimate the in-running effect. We find that even after accounting for changes in the distribution of odds as races begin, there is still a heightened tendency to quote round numbers during the races.

⁸Ideally, we would like to match on higher moments of the distribution of odds as well, particularly skewness, but the more matching variables we include the more difficult it becomes to find appropriate pre-race observations to compare with our in-running observations. Similarly, we would like to ensure that we match only within the same bet (i.e., one observation for a given bet pre-race matched with another observation for the same bet during the race) but this is problematic as there may not be times during races when the distribution looks at all similar to how it did pre-race. We therefore have to allow for matches across bets.

As an example, it is worth examining the results for back quotes in the win market for Ascot 2011, the first results in Panel A. Using the linear probability model, we estimate that round number quotes increase in frequency by 10.7% more than nearest neighbors (one price down) during races compared to pre-race. Yet, once we match pre-race and in-running observations according to the distribution of odds at the time, we find that round numbers still increase in frequency by 6.8% more than their nearest neighbors. A similar pattern emerges for the other markets, events, and quotes. Only in one case — back quotes in the win market for Ascot 2012 — is the in-running effect deemed insignificant when the changes in the distribution of odds are accounted for. In three cases, the in-running effect is actually larger once changes in the distribution of odds are acknowledged. It is likely that by considering only round numbers or nearest neighbors in our earlier analysis, we have largely mitigated any problems that may emerge due to changes in the distribution of odds during races.

We next examine the drivers of increased clustering during races. The interpretation taken throughout the paper is that a cognitively constrained trader uses round numbers as a shortcut pricing method. If more time and mental capacity had been available, a more precise quote may have been provided. An alternative interpretation is that the trader submitting the limit order (i.e., providing the quote) submits a round number to take advantage of the limited processing capabilities of other traders during races. In other words, limit order traders believe that market order traders will spy relative value in a back quote of 2, compared to 1.99, when races go in-running, because their limited cognition does not allow them to spot the similarities between the two prices. The empirical prediction of such a hypothesis is that the frequency of orders should jump at round numbers, relative to nearest neighbor non-round numbers, and jump specifically during races when traders are cognitively constrained.

This argument only applies to back quotes, as round numbers and nearest neighbors have the same left-digit for lay quotes (e.g., in the case of 2 and 2.02).⁹

Table 5 presents the results of a regression where an indicator variable equaling 1 if a (back) order was observed in the last second, is regressed on an indicator variable equaling 1 if a round number was quoted in the previous second, an indicator variable equaling 1 if the race was in-running, and an interaction between the two aforementioned indicators. The interaction term is crucial as this captures the extent to which market order traders fail to identify the similarities between round number and nearest neighbor quotes, specifically during races. Only data on round numbers, and the odds immediately below (back quote nearest neighbors) are included in this regression. As before, a logit specification is used and random effects are included for each horse (bet). This regression is carried out for win and place market data for the four horse race meetings.

There are three results from Table 5 to be discussed. Firstly, in all eight cases, quotes at round numbers are slightly more likely to attract an order. This is consistent with a left-digit bias, but if this was the result of interest, a placebo test would need to be conducted to ensure that this is not simply picking up the fact that higher odds attract more frequent back orders. The second result is that orders decline as races begin. Finally, and most importantly, there is no evidence that the propensity for orders to arrive for round number quotes increases during

⁹The left-digit bias is not a potential explanation for the prevalence of round number lay quotes during races. If a limit order trader suspects that market order traders did not process the digits of a figure beyond its first left-digit, then they would likely quote 2.02, for example, rather than 2, in order to extract higher odds from their cognitively constrained counterparty. However, we observe that quotes of 2 are more common, during races, than 2.02, when it comes to lay quotes.

racers. The coefficient is negative and significant for five of the regressions, and insignificant for the other three. In other words, rather than a tactic to induce orders from constrained counterparties, the round number appears to be used as a heuristic method for pricing bets when finer information processing is not feasible.

We next examine the breadth of the in-running effect documented throughout the paper. While the effect is robust across markets, across events, and across time (at least from 2011 to 2012), it would be instructive to examine whether the effect is robust across all of the round numbers we have specified. To this end, we pool the win and place market data, as well as the data for Royal Ascot and the Cheltenham Festival in both years. This gives the following tests substantial statistical power. We repeat the analysis of Table 2 with back quotes and lay quotes for each individual round number and present the results in Table 6. When looking at the in-running effect at the round number of 3, only quotes of 3 and 2.98 are included for back quotes, and only quotes of 3 and 3.05 are included for lay quotes. The results show that the propensity to quote a round number, relative to a nearest neighbor non-round number, increases as races go in-running for all 14 round numbers used for back quotes, and for 13 of the 14 for lay quotes (with significance at the 0.1% level). The number 15 is the exception for lay quotes. Although the size of the effect is not monotonic with the size of the odds, there is some evidence that the greatest propensity to quote round numbers — both pre-race and in-running, and for both back and lay quotes — is found for lower round numbers, particular those from 2 to 5.

One argument that may be put forward is that the increased incidence of round numbers during races is not due to limited cognition, but is instead due to another factor that arises due to time-pressure. Bettors may quote round numbers during races as, in most cases, round

numbers require the inputting of fewer digits. In other words, rather than enter 1.99 as the quote for a back bet, a punter enters 2, which is almost the same price, to save entering the two digits that follow the decimal place. Such a shortcut allows the bettor to update their quotes, and secure trade, much more quickly. While speed is not so important prior to races, it is during races.

To deal with such a concern, we take advantage of the fact that a subset of the round numbers in our analysis has the same number of digits as their nearest neighbors. For 15, 20, 30, 40 and 50, the nearest neighbors are 14, 19, 28, 36, and 46 respectively. In other words, it takes just as much time to input the round number price as the nearest neighbor. In contrast, round numbers of 10 and below involve fewer inputs than their nearest neighbors. (In the case of 10, there are two characters inputted for the round number, and 3 characters for the nearest neighbor, 9.6.) Therefore, we argue that the treatment effect for higher round numbers — from 15 to 50 — can be more reliably attributed to limited cognition.

With this in mind, in Table 7 we present the results of four further regressions; two for back quotes (in Panel A) and two for lay quotes (in Panel B). We retain the pooled data from the Table 6 analysis, to ensure the greatest statistical power, and break the full sample down into low round numbers (2 to 10) and high round numbers (15 to 50). We then repeat the regressions of Table 2. We find that both low and high round numbers are more prevalent than their nearest neighbors, particularly during races. The running of each race does increase the incidence of lower round numbers to a greater extent than high round numbers, suggesting that limited time to input digits may well be a factor. However, the effect for high round numbers is still highly significant, leading us to believe that limited cognition is a contributor to the clustering of asset prices at round numbers.

We have argued that bettors latch on to round numbers when the cognitive load is greater during races. This implicitly assumes that round numbers are more frequently quoted than their nearest neighbors, rather than simply being more persistent.¹⁰ However, in all of our analysis thus far, we have considered the best back and lay quotes each second, irrespective of whether the quote was also present in the previous second. For our final analysis, we therefore repeat the main analysis, but this time only consider quotes if they are "new," (i.e., were not the best quote in the previous second). (We cannot observe each individual limit order placed, so this is the best way, in our opinion, to classify a new order.) The results are in Table 8. We also display the results for the full sample including both new and old quotes for comparison. In the case of back and lay quotes, the effect is slightly larger for all quotes than for just new quotes, thereby hinting that round number quotes are more persistent than nearest neighbors, but the in-running effect remains large and highly significant (at the 0.1% level) even when we consider only new quotes. This applies to both back and lay quotes, and gives us confidence that cognitively constrained bettors are indeed latching on to round numbers more frequently than nearest neighbor prices.

6 Conclusion

Asset prices tend to cluster at round numbers. This clustering is found in stock prices, currency prices (Sopranzetti and Datar, 2002), bond prices (Gwilym, Clare, and Thomas, 1998a), and option prices (Gwilym, Clare, and Thomas, 1998b). Furthermore, clustering

¹⁰This may itself be a sign of limited cognition (i.e., cognitively constrained traders are more likely to quote round numbers, and are also slower to update these quotes, and hence round numbers are more persistent).

appears to be resistant to changes in market design. Ikenberry and Weston (2007) find little decrease in clustering after the decimalization of prices (i.e., prices had previously been quoted in coarser dollar fractions) on the New York Stock Exchange and the NASDAQ in 2001. This would suggest that the use of round numbers is a "hard-wired" reflex action for the cognitively constrained trader.

Missing from this story, however, is any evidence that the level of clustering varies with the cognitive load facing traders. If limited cognition is truly behind price clustering, we should observe particularly pronounced levels of clustering when traders are under an intense cognitive load. Rather than test such an idea in the laboratory, where the experience of subjects may be in some doubt, we use data arising from horse race betting on a U.K. betting exchange to test this hypothesis. Information in this market is produced in such a way that there is stark variation in the cognitive load facing traders. Prior to races there is little new information to process, and bettors can therefore make fine price distinctions. Once the race begins, however, and horses switch positions frequently, information must be processed — and bets newly priced — in a very short space of time. We hypothesise that round numbers will become more prevalent in this latter period.

To tease out such an effect, we use an empirical approach based partly on regression discontinuity (RD) methods, and partly on difference-in-differences (DID) methods. We compare the propensity for traders to quote round numbers with their propensity to quote the nearest non-round number. These two prices reflect almost identical values and therefore, if bettors had unlimited cognition, each price should be equally likely to be quoted (or "assigned," to use the RD parlance). Unlike RD analysis, we are not interested in the outcome for the assigned and unassigned variables, but instead we are interested in how the assignment

decision varies across pre-race (pre-treatment) periods, and in-running (treatment) periods. If bettors display a greater propensity to quote round numbers rather than their nearest neighbors during races — when the cognitive load is greater — this puts limited cognition front-and-center as an explanation for the clustering of asset prices.

We find a significant and substantial increase in the proportion of round number quotes during races. The results are robust to a series of placebo tests (and matching exercises) to deal with changes in the distribution of odds as races progress. The results also apply when inputting a round number requires as much effort as entering a nearest neighbor (i.e., when the round number has the same number of digits as its neighbor). In short, we find that limited cognition does indeed lead to asset price clustering.

References

- Abadie, A., Imbens, G., W., (2011). Bias-Corrected Matching Estimators for Average Treatment Effects. *Journal of Business and Economic Statistics*, 29, 1-11.
- Ahn, H., J., Cai, J., Cheung, Y., L., (2005). Price Clustering on the Limit-Order Book: Evidence from the Stock Exchange of Hong Kong. *Journal of Financial Markets*, 8, 421-451.
- Bhattacharya, U., Holden, C., W., Jacobsen, S., (2012). Penny Wise, Dollar Foolish: Buy-Sell Imbalances On and Around Round Numbers. *Management Science*, 58, 413-431.
- Brown, A., (2014). Information Processing Constraints and Asset Mispricing. *Economic Journal*, 124, 245-268.
- Busse, M., R., Lacetera, N., Pope, D., G., Silva-Risso, J., Sydnor, J., R., (2013). Estimating the Effect of Salience in Wholesale and Retail Car Markets. *American Economic Review*:

Papers and Proceedings, 103, 575-579.

DellaVigna, S., and Pollet, J. M., (2009). Investor Inattention and Friday Earnings Announcements, *Journal of Finance*, 64, 709-749.

Ginzberg, E., (1936). *Customary Prices*. *American Economic Review*, 26, 296.

Gwilym, O., P., Clare, A., Thomas, T., (1998a). Price Clustering and Bid-Ask Spreads in International Bond Futures. *Journal of International Financial Markets, Institutions and Money*, 8, 377-391.

Gwilym, O., P., Clare, A., Thomas, T., (1998b). Extreme Price Clustering in the London Equity Index Futures and Options Markets. *Journal of Banking and Finance*, 22, 1193-1206.

Harris, L., (1991). Stock Price Clustering and Discreteness. *Review of Financial Studies*, 4, 389-415.

Hinrichs, J., V., Yurko, D., S., Hu, J., M., (1981). Two-Digit Number Comparison: Use of Place Information. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 890-901.

Hirshleifer, D., Lim, S., S., Teoh, S., H., (2009). Driven to Distraction: Extraneous Events and Underreaction to Earnings News. *Journal of Finance*, 64, 2289-2325.

Ikenberry, D., L., Weston, J., P., (2007). Clustering in US Stock Prices after Decimalisation. *European Financial Management*, 14, 30-54.

Kuo, W., Y., Lin, T., C., Zhao, J., (2015). Cognitive Limitation and Investment Performance: Evidence from Limit Order Clustering. *Review of Financial Studies*, 28, 838-875.

Lacetera, N., Pope D., G., Syndor, J., R., (2012). Heuristic Thinking and Limited Attention in the Car Market. *American Economic Review*, 102, 2206-2236.

Niederhoffer, V., (1965). Clustering of Stock Prices. *Operations Research*, 13, 258-265.

-
- Niederhoffer, V., (1966). A New Look at the Clustering of Stock Prices. *Journal of Business*, 39, 309-313.
- Poltrock, S., E., Schwartz, D., R., (1984). Comparative Judgements of Multidigit Numbers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 32-45.
- Rosch, E., (1975). Cognitive Reference Points. *Cognitive Psychology*, 7, 532-547.
- Simon, H., A., (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, 69, 99-118.
- Sonnemans, J., (2006). Price Clustering and Natural Resistance Points in the Dutch Stock Market: A Natural Experiment. *European Economic Review*, 50, 1937-1950.
- Sopranzetti, B., J., Datar, V., (2002). Price Clustering in Foreign Exchange Spot Markets. *Journal of Financial Markets*, 5, 411-417.

Table 1: Summary Statistics				
Panel A: Royal Ascot 2011				
Win Market		Quotes	Round Number	Nearest Neighbor
Pre-Race	Back Odds	97.26%	12.28%	9.82%
	Lay Odds	97.18%	9.96%	12.14%
In-Running	Back Odds	98.15%	8.01%	4.07%
	Lay Odds	54.96%	6.66%	5.71%
Place Market		Quotes	Round Number	Nearest Neighbor
Pre-Race	Back Odds	97.26%	12.20%	9.18%
	Lay Odds	97.26%	10.80%	11.04%
In-Running	Back Odds	96.45%	20.20%	4.07%
	Lay Odds	76.45%	13.93%	10.60%
Panel B: Royal Ascot 2012				
Win Market		Quotes	Round Number	Nearest Neighbor
Pre-Race	Back Odds	96.08%	12.90%	9.99%
	Lay Odds	95.20%	10.21%	12.90%
In-Running	Back Odds	95.43%	10.98%	6.87%
	Lay Odds	76.22%	9.01%	6.82%
Place Market		Quotes	Round Number	Nearest Neighbor
Pre-Race	Back Odds	96.08%	10.95%	9.42%
	Lay Odds	96.08%	10.80%	10.17%
In-Running	Back Odds	94.65%	16.54%	4.76%
	Lay Odds	66.33%	9.99%	4.81%
Panel C: Cheltenham Festival 2011				
Win Market		Quotes	Round Number	Nearest Neighbor
Pre-Race	Back Odds	98.62%	12.34%	9.66%
	Lay Odds	98.36%	9.87%	12.33%
In-Running	Back Odds	98.38%	10.13%	6.69%
	Lay Odds	77.89%	8.76%	7.20%
Place Market		Quotes	Round Number	Nearest Neighbor
Pre-Race	Back Odds	98.58%	11.76%	10.23%
	Lay Odds	98.58%	12.39%	10.53%
In-Running	Back Odds	98.04%	18.79%	5.23%
	Lay Odds	70.24%	12.58%	6.67%
Panel D: Cheltenham Festival 2012				
Win Market		Quotes	Round Number	Nearest Neighbor
Pre-Race	Back Odds	98.34%	11.59%	9.37%
	Lay Odds	97.63%	9.59%	11.50%
In-Running	Back Odds	98.11%	10.02%	6.37%
	Lay Odds	77.31%	8.63%	7.46%
Place Market		Quotes	Round Number	Nearest Neighbor
Pre-Race	Back Odds	98.34%	10.35%	9.91%
	Lay Odds	98.34%	11.66%	9.53%
In-Running	Back Odds	97.72%	19.04%	4.23%
	Lay Odds	67.04%	14.18%	5.60%

The percentage of seconds with at least one quote, and the percentage of seconds with a quote at a round number or a nearest neighbor. The round numbers are defined as 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, and 50. The nearest neighbors are the odds immediately below (for a back quote), or immediately above (for a lay quote). These statistics are compared for pre-race and in-running trading periods. Data for the win and place market for Royal Ascot and the Cheltenham Festival in 2011 and 2012 are displayed.

Table 2: Main Test Results (Logit)				
Panel A: Win Market (Back Quotes)				
Dep. Var.: Odds=Round Number	Ascot 2011	Ascot 2012	Cheltenham 2011	Cheltenham 2012
Intercept	0.475*** (.066)	0.374*** (.072)	0.391*** (.065)	0.367*** (.057)
In-Running	0.442*** (.026)	0.253*** (.027)	0.281*** (.018)	0.255*** (.017)
No. of Observations	115,861	120,606	137,461	146,264
ρ	0.294	0.337	0.308	0.275
No. of Random Effects (Bets)	342	340	363	421
Panel B: Place Market (Back Quotes)				
Dep. Var.: Odds=Round Number	Ascot 2011	Ascot 2012	Cheltenham 2011	Cheltenham 2012
Intercept	0.435*** (.061)	0.278*** (.056)	0.25*** (.055)	0.288*** (.056)
In-Running	1.437*** (.028)	1.237*** (.027)	1.334*** (.017)	1.666*** (.017)
No. of Observations	114,967	110,030	143,222	153,083
ρ	0.283	0.248	0.259	0.293
No. of Random Effects (Bets)	375	374	409	475
Panel C: Win Market (Lay Quotes)				
Dep. Var.: Odds=Round Number	Ascot 2011	Ascot 2012	Cheltenham 2011	Cheltenham 2012
Intercept	-0.316*** (.062)	-0.275*** (.073)	-0.269*** (.068)	-0.249*** (.059)
In-Running	0.422*** (.025)	0.543*** (.028)	0.371*** (.018)	0.367*** (.017)
No. of Observations	116,033	120,723	137,335	146,558
ρ	0.27	0.334	0.32	0.285
No. of Random Effects (Bets)	331	325	351	404
Panel D: Place Market (Lay Quotes)				
Dep. Var.: Odds=Round Number	Ascot 2011	Ascot 2012	Cheltenham 2011	Cheltenham 2012
Intercept	0.033 (.06)	0.136* (.061)	0.281*** (.057)	0.418*** (.06)
In-Running	0.221*** (.023)	0.799*** (.03)	0.444*** (.017)	0.793*** (.017)
No. of Observations	117,309	109,829	141,700	153,074
ρ	0.278	0.279	0.275	0.327
No. of Random Effects (Bets)	372	369	408	472

The main test results in the paper. An indicator variable equaling 1 if a round number is quoted, is regressed on an indicator variable equaling 1 if the race is ‘in-running’ at that time-point. The round numbers are defined as 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, and 50. Only quotes at these odds, or those immediately below (for back quotes) or above (for lay quotes), are included in the regressions. Analysis is conducted separately for win and place market data. A logit specification is used, random effects for each bet are included (with ρ representing the variation in the dependent variable captured by these random effects), standard errors are in parentheses, and *** and * indicates significance at the 0.1% and 5% levels, respectively.

Table 3: Placebo Test Results (Logit)				
Panel A: Win Market (Back Quotes)				
Dep. Var.: Odds=Placebo Round Number	Ascot 2011	Ascot 2012	Cheltenham 2011	Cheltenham 2012
Intercept	-0.076 (.06)	0.084 (.058)	0.068 (.052)	0.011 (.045)
In-Running	-0.197*** (.028)	-0.244*** (.029)	-0.187*** (.019)	-0.063** (.018)
No. of Observations	101,833	100,659	116,905	128,050
ρ	0.253	0.239	0.21	0.183
No. of Random Effects (Bets)	334	330	350	401
Panel B: Place Market (Back Quotes)				
Dep. Var.: Odds=Placebo Round Number	Ascot 2011	Ascot 2012	Cheltenham 2011	Cheltenham 2012
Intercept	-0.048 (.056)	0.085 (.054)	0.047 (.051)	-0.099 (.056)
In-Running	-0.19*** (.034)	-0.205*** (.034)	-0.091*** (.02)	-0.081*** (.022)
No. of Observations	95,992	96,976	119,425	127,408
ρ	0.243	0.226	0.228	0.284
No. of Random Effects (Bets)	357	353	394	450
Panel C: Win Market (Lay Quotes)				
Dep. Var.: Odds=Placebo Round Number	Ascot 2011	Ascot 2012	Cheltenham 2011	Cheltenham 2012
Intercept	0.265** (.078)	0.482*** (.092)	0.266** (.08)	0.322 (.07)
In-Running	-0.068* (.026)	-0.665*** (.031)	-0.277*** (.02)	-0.210*** (.019)
No. of Observations	116,836	111,127	131,843	142,333
ρ	0.373	0.448	0.407	0.366
No. of Random Effects (Bets)	332	330	364	408
Panel D: Place Market (Lay Quotes)				
Dep. Var.: Odds=Placebo Round Number	Ascot 2011	Ascot 2012	Cheltenham 2011	Cheltenham 2012
Intercept	0.638*** (.096)	0.663*** (.113)	0.496*** (.083)	0.484*** (.095)
In-Running	0.412*** (.029)	0.147*** (.039)	-0.143*** (.021)	-0.271*** (.021)
No. of Observations	103,212	90,889	114,073	116,989
ρ	0.495	0.566	0.445	0.542
No. of Random Effects (Bets)	372	357	399	463

Placebo test results. An indicator variable equaling 1 if a placebo round number is quoted, is regressed on an indicator variable equaling 1 if the race is ‘in-running’ at that time-point. The placebo round numbers are 1.99, 2.98, 3.95, 4.9, 5.9, 6.8, 7.8, 8.8, 9.8, 14.5, 19.5, 29, 38, and 48 for back quotes, and 2.02, 3.05, 4.1, 5.1, 6.2, 7.2, 8.2, 9.2, 10.5, 15.5, 21, 32, 42, and 55 for lay quotes. Only quotes at these odds, or those immediately below (for back quotes) or above (for lay quotes), are included in the regressions. Analysis is conducted separately for win and place market data. A logit specification is used, random effects for each bet are included (with ρ representing the variation in the dependent variable captured by these random effects), standard errors are in parentheses, and ***, **, and * indicates significance at the 0.1%, 1%, and 5% levels, respectively.

Table 4: Matching on Odds Distribution			
Panel A: Win Market (Back Quotes)			
Event	Matching In-Running Estimate	LPM In-Running Estimate	No. of Observations
Ascot 2011	0.068** (.025)	0.107*** (.005)	115,861
Ascot 2012	0.041 (.028)	0.051*** (.005)	120,606
Cheltenham 2011	0.054* (.022)	0.041*** (.003)	137,461
Cheltenham 2012	0.058*** (.015)	0.058*** (.003)	146,264
Panel B: Place Market (Back Quotes)			
Event	Matching In-Running Estimate	LPM In-Running Estimate	No. of Observations
Ascot 2011	0.245*** (.021)	0.261*** (.003)	114,967
Ascot 2012	0.196*** (.021)	0.238*** (.004)	110,030
Cheltenham 2011	0.287*** (.018)	0.247*** (.002)	143,222
Cheltenham 2012	0.291*** (.016)	0.307*** (.002)	153,083
Panel C: Win Market (Lay Quotes)			
Event	Matching In-Running Estimate	LPM In-Running Estimate	No. of Observations
Ascot 2011	0.081*** (.017)	0.087*** (.005)	116,033
Ascot 2012	0.097** (.03)	0.127*** (.005)	120,723
Cheltenham 2011	0.088*** (.015)	0.104*** (.003)	137,335
Cheltenham 2012	0.056*** (.015)	0.081*** (.003)	146,558
Panel D: Place Market (Lay Quotes)			
Event	Matching In-Running Estimate	LPM In-Running Estimate	No. of Observations
Ascot 2011	0.113*** (.017)	0.073*** (.004)	117,309
Ascot 2012	0.145*** (.022)	0.159*** (.005)	109,829
Cheltenham 2011	0.106*** (.017)	0.112*** (.003)	141,700
Cheltenham 2012	0.152*** (.012)	0.166*** (.002)	153,074

Matching and linear probability model (LPM) estimates of the ‘in-running’ effect on the propensity to quote round numbers. In the matching estimation, pre-race and in-running observations are matched with each other on the basis of the distribution (mean and standard deviation) of odds in that race at that time. Matching is carried out with replacement (i.e., an observation pre-race can be matched with more than one observation in-running, and vice-versa), and estimates are bias-corrected according to the method in Abadie and Imbens (2011). In the linear probability model, with estimation by OLS, an indicator variable equaling 1 if a round number is quoted, is regressed on an indicator variable equaling 1 if the race is in-running at that time-point. Only the coefficient associated with the in-running indicator, and not the intercept, is displayed. The round numbers are defined as 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, and 50. Only quotes at these odds, or those immediately below (for back quotes) or above (for lay quotes), are included in the two sets of estimations, but all observations are used to calculate the distribution (mean and standard deviation) of the odds in the matching procedure. All analyses are carried out separately for win and place markets, and back and lay quotes, across the four events. Standard errors are in parentheses, and ***, **, and * indicates significance at the 0.1%, 1%, and 5% levels, respectively.

Table 5: Left-Digit Bias Test Results (Logit)

Panel A: Win Market (Back Quotes)				
Dep. Var.: Order Indicator	Ascot 2011	Ascot 2012	Cheltenham 2011	Cheltenham 2012
Intercept	-0.544*** (.04)	-0.486*** (.041)	-0.377*** (.038)	-0.2*** (.039)
Odds=Round Number	0.14*** (.014)	0.139*** (.0144)	0.171*** (.014)	0.166*** (.013)
In-Running	-0.836*** (.045)	-0.699*** (.042)	-0.667*** (.026)	-0.848*** (.026)
Odds=Round Number*In-Running	0.06 (.055)	-0.137* (.054)	-0.102** (.034)	-0.064* (.032)
No. of Observations	115,861	120,606	137,461	146,264
ρ	0.124	0.129	0.123	0.141
No. of Random Effects (Bets)	342	340	363	421
Panel B: Place Market (Back Quotes)				
Dep. Var.: Order Indicator	Ascot 2011	Ascot 2012	Cheltenham 2011	Cheltenham 2012
Intercept	-1.957*** (.039)	-1.867*** (.04)	-1.903*** (.038)	-1.671*** (.039)
Odds=Round Number	0.143*** (.02)	0.087*** (.02)	0.125*** (.019)	0.162*** (.017)
In-Running	-0.601*** (.09)	-1.044*** (.099)	-0.81*** (.052)	-0.907*** (.054)
Odds=Round Number*In-Running	-0.33** (.1)	0.032 (.111)	0.011 (.059)	-0.226*** (.06)
No. of Observations	114,967	110,030	143,222	153,083
ρ	0.116	0.121	0.125	0.152
No. of Random Effects (Bets)	375	374	409	475

The results of tests for a left-digit bias amongst bettors. An indicator variable equaling 1 if a back order was placed in the last second, was regressed on an indicator variable equaling 1 if a round number had been quoted, an indicator equaling 1 if the race was ‘in-running’ at that time-point, and an interaction between the two aforementioned indicators. The interaction term would capture any left-digit bias that arises specifically during races. As before, the round numbers are defined as 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, and 50. Only quotes at these odds, or those immediately below, are included in the regressions. Win market data are in Panel A and place market data are in Panel B. A logit specification is used, random effects for each horse (bet) are included (with ρ representing the variation in the dependent variable captured by these random effects), standard errors are in parentheses, and ***, **, and * indicates significance at the 0.1%, 1%, and 5% levels, respectively.

Table 6: Individual Round Numbers (Logit)								
Panel A: Win & Place Markets (Back Quotes)								
Dep. Var.: Odds=Round Number	2	3	4	5	6	7	8	
Intercept	3.368*** (.335)	2.442*** (.167)	2.436*** (.148)	1.044*** (.103)	1.435*** (.091)	0.664*** (.095)	1.246*** (.099)	
In-Running	1.271*** (.128)	1.276*** (.067)	0.56*** (.058)	1.233*** (.044)	0.647*** (.04)	0.46*** (.04)	0.458*** (.038)	
No. of Observations	10,480	31,280	44,200	66,406	87,624	74,999	78,914	
ρ	0.678	0.614	0.726	0.622	0.578	0.631	0.645	
No. of Random Effects (Bets)	449	702	803	944	944	891	958	
Dep. Var.: Odds=Round Number	9	10	15	20	30	40	50	
Intercept	0.699*** (.082)	1.451*** (.084)	0.133. (.068)	1.182*** (.079)	0.677*** (.074)	0.149*** (.078)	0.966*** (.102)	
In-Running	0.125** (.039)	0.315*** (.038)	0.491*** (.036)	0.256*** (.036)	0.132*** (.033)	0.151*** (.034)	0.448*** (.036)	
No. of Observations	70,910	98,268	90,771	95,757	107,624	82,719	101,542	
ρ	0.553	0.565	0.487	0.547	0.535	0.534	0.663	
No. of Random Effects (Bets)	884	1,008	961	1,011	939	872	892	
Panel B: Win & Place Markets (Lay Quotes)								
Dep. Var.: Odds=Round Number	2	3	4	5	6	7	8	
Intercept	1.38*** (.205)	0.657*** (.132)	0.357** (.123)	0.669*** (.103)	0.017 (.091)	0.386*** (.095)	0.557*** (.092)	
In-Running	0.834*** (.093)	1.146*** (.052)	1.104*** (.052)	0.812*** (.038)	0.755*** (.033)	0.544*** (.039)	0.696*** (.04)	
No. of Observations	9,674	29,671	42,046	63,857	90,338	73,190	74,128	
ρ	0.609	0.61	0.663	0.624	0.610	0.615	0.610	
No. of Random Effects (Bets)	354	523	643	739	821	786	842	
Dep. Var.: Odds=Round Number	9	10	15	20	30	40	50	
Intercept	0.415*** (.087)	-0.018 (.086)	0.654*** (.085)	-0.052 (.084)	-0.238** (.076)	0.470*** (.081)	-0.733*** (.099)	
In-Running	0.689*** (.045)	0.817*** (.033)	-0.090** (.038)	0.326*** (.033)	0.384*** (.031)	0.211*** (.036)	0.439*** (.035)	
No. of Observations	68,278	103,051	87,689	99,867	111,799	80,733	108,240	
ρ	0.565	0.603	0.593	0.613	0.564	0.565	0.707	
No. of Random Effects (Bets)	790	920	887	980	962	903	1,075	

A repeat of the regressions used to generate the results in Table 2, but this time run for each individual round number, and including pooled win and place market data for all four meetings. Back quote results are in Panel A with lay quote results in Panel B. An indicator variable equaling 1 if a round number is quoted, is regressed on an indicator variable equaling 1 if the race is "in-running" at that time-point. The round numbers are defined as 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40 and 50. Only quotes at these odds, or those immediately below (for back quotes) or above (for lay quotes), are included in the regressions. A logit specification is used, random effects for each (horse) bet are included (with ρ representing the variation in the dependent variable captured by these random effects), standard errors are in parentheses, and ***, **, and . indicates significance at the 0.1%, 1%, and 10% levels, respectively.

Table 7: Low and High Round Numbers (Logit)

Panel A: Win & Place Markets (Back Quotes)

Dep. Var.: Odds=Round Number	Low Round Numbers	High Round Numbers
Intercept	0.581*** (.03)	0.358*** (.035)
In-Running	0.985*** (.009)	0.523*** (.011)
No. of Observations	563,081	478,413
ρ	0.315	0.369
No. of Random Effects (Bets)	1,994	1,775

Panel B: Win & Place Markets (Lay Quotes)

Dep. Var.: Odds=Round Number	Low Round Numbers	High Round Numbers
Intercept	0.002 (.03)	-0.04 (.041)
In-Running	0.781*** (.009)	0.17*** (.011)
No. of Observations	554,233	488,328
ρ	0.3	0.451
No. of Random Effects (Bets)	1,802	1,880

A repeat of the regressions used to generate the results in Table 2, but this time run separately for low and high round numbers, and including pooled win and place market data for all four meetings. Back quote results are in Panel A, with lay quote results in Panel B. An indicator variable equaling 1 if a round number is quoted is regressed on an indicator variable equaling 1 if the race is "in-running" at that time-point. The low round numbers are defined as 2, 3, 4, 5, 6, 7, 8, 9, and 10, and the high round numbers are 15, 20, 30, 40, and 50. Only quotes at these odds, or those immediately below (for back quotes) or above (for lay quotes), are included in the regressions. A logit specification is used, random effects for each horse (bet) are included (with ρ representing the variation in the dependent variable captured by these random effects), standard errors are in parentheses, and *** indicates significance at the 0.1% level.

Table 8: New Quotes (Logit)

Panel A: Win & Place Markets (Back Quotes)		
Dep. Var.: Odds=Round Number	All Quotes	New Quotes
Intercept	0.327*** (.022)	0.017 (.011)
In-Running	0.898*** (.007)	0.764*** (.015)
No. of Observations	1,041,494	101,214
ρ	0.278	0.032
No. of Random Effects (Bets)	2,732	2,732

Panel B: Win & Place Markets (Lay Quotes)		
Dep. Var.: Odds=Round Number	All Quotes	New Quotes
Intercept	0.000 (.023)	0.08*** (.01)
In-Running	0.496*** (.007)	0.284*** (.015)
No. of Observations	1,042,561	89,261
ρ	0.296	0.024
No. of Random Effects (Bets)	2,673	2,673

A repeat of the regressions used to generate the results in Table 2, but this time run only for new quotes (i.e., those not the best quote in the previous second), and including pooled win and place market data for all four meetings. Back quote results are in Panel A, with lay quote results in Panel B. An indicator variable equaling 1 if a round number is quoted is regressed on an indicator variable equaling 1 if the race is "in-running" at that timepoint. The round numbers are defined as 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, and 50. Only quotes at these odds, or those immediately below (for back quotes) or above (for lay quotes), are included in the regressions. In the first regression, all quotes are analyzed, but in the second regression only new quotes are analyzed. A logit specification is used, random effects for each horse (bet) are included (with ρ representing the variation in the dependent variable captured by these random effects), standard errors are in parentheses, and *** indicates significance at the 0.1% level.

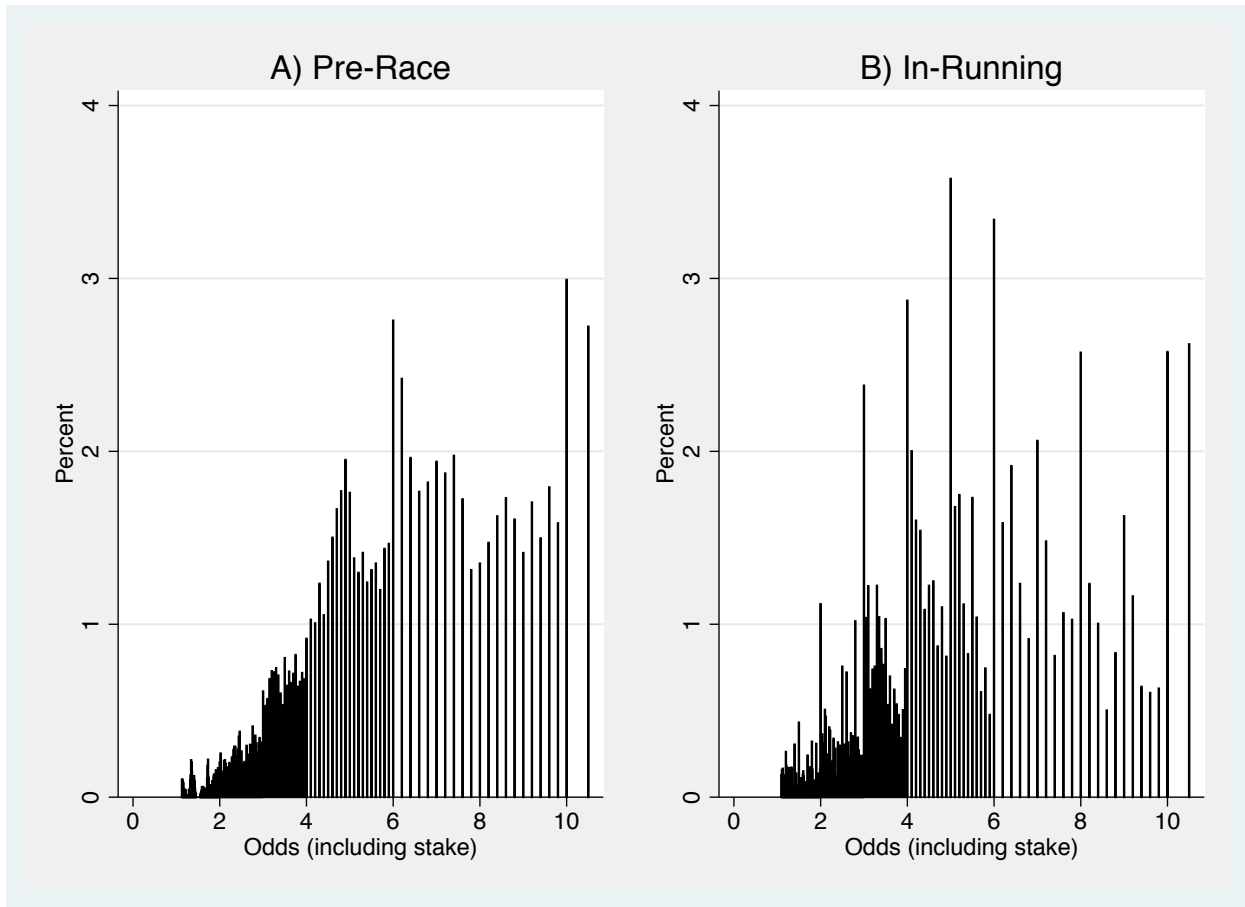


Figure 1: **Quoted Odds.** Bar charts of quoted win and place odds (back quotes) on horses in races at Royal Ascot in 2011. Panel A contains pre-race odds and Panel B contains in-running odds. The plot is restricted to odds (including the stake) of below 11.

13:45 Wolverhampton

Wed 29 Apr | 7f Hcap

Live Stream Radio Form & Results

Tote **Multiples**

Win
 Each Way
 Place
 3 Places
 More ▾

Going In-Play
 Cash Out
 Rules
 Betfair SP [?]
 Timeform
 Matched: **GBP 127,000**

7 selections		101.7%		Back all	Lay all	98.6%	
	5 (1) Rafaaf Rob Hornby	3.05 £838	3.1 £567	3.15 £392	3.2 £301	3.25 £567	3.3 £426
	1 (3) Hagree Luke Leadbitter	4.6 £193	4.7 £99	4.8 £117	4.9 £55	5 £128	5.1 £134
	2 (5) The Happy Hammer Martin Dwyer	5.8 £131	5.9 £241	6 £210	6.2 £106	6.4 £172	6.6 £109
	7 (4) Black Truffle Charlotte Jenner	6.6 £211	6.8 £259	7 £249	7.2 £94	7.4 £132	7.6 £99
	4 (6) Indian Affair Tom Queally	7 £335	7.2 £123	7.4 £113	7.8 £119	8 £132	8.2 £86
	6 (7) Kuanyao Ann Stokell	29 £18	30 £16	32 £14	36 £23	38 £40	40 £28
	3 (2) Paradise Spectre John Fahy	55 £2	60 £6	65 £3	75 £15	80 £31	85 £19

Figure 2: **Betfair Screenshot.** A pre-race screenshot from the Betfair limit order book for the 13.45 race at Wolverhampton on Wednesday April 29, 2015 (an example outside of our sample). To bet on a horse to win, the punter would place a "back" bet on the left, and to bet against a horse, the punter would place a "lay" bet on the right. The best back and lay quotes are displayed, for each horse, in the center. The volumes available at each price are indicated underneath the odds.