

Explaining the Favorite-Longshot Bias: Is it Risk-Love, or Misperceptions?

Erik Snowberg
Stanford GSB
snowberg_erik@gsb.stanford.edu

Justin Wolfers
The Wharton School, U.Penn
CEPR, IZA, & NBER
jwolfers@wharton.upenn.edu
www.nber.org/~jwolfers

Abstract

The favorite-longshot bias presents a challenge for theories of decision making under uncertainty. This longstanding empirical regularity has documented that betting odds provide biased estimates of the probability of a horse winning, and longshots are overbet, while favorites are underbet. Neoclassical explanations have rationalized this puzzle by appealing to rational gamblers who overbet longshots due to risk-love, market structures, or alternatively information asymmetries. Competing behavioral explanations emphasize the role of misperceptions of probabilities. We provide a novel empirical test that can differentiate these competing theories, focusing on the pricing of compound or “exotic” bets. We test whether the model that best explains gamblers’ choices in one part of their choice set (betting to win) can also rationalize decisions over a wider choice set, including betting in the win, exacta, quinella or trifecta pools. We have a new large-scale dataset ideally suited to test these predictions and find evidence in favor of the view that misperceptions of probability drive the favorite-longshot bias, as suggested by Prospect Theory. Along the way we provide more robust evidence on the favorite-longshot bias, falsifying the conventional wisdom that the bias is large enough to yield profit opportunities (it isn’t) and that it becomes more severe in the last race (it doesn’t).

This draft: October 24, 2006

We would like to thank David Siegel of Equibase for supplying the data, and Scott Hereld and Ravi Pillai for their valuable assistance in managing the data. Jon Bendor, Larry Brown, Bruno Jullien, Marco Ottaviani, Bernard Salanié, Peter Norman Sørensen, Betsey Stevenson, George Wu, Matthew White and William Ziemba provided helpful feedback and insights. Wolfers gratefully acknowledges a Hirtle, Callaghan & Co. – Arthur D. Miltenberger Research Fellowship, and the support of the Zull/Lurie Real Estate Center, the Mack Center for Technological Innovation, and Microsoft Research.

1. Introduction

The racetrack provides a natural laboratory for economists interested in understanding decision-making under uncertainty. The most persistent empirical regularity in this literature is the so-called “favorite-longshot bias”. That is, equilibrium market prices (betting odds) provide biased estimates of the probability of each horse winning. Specifically, bettors value longshots more than one might expect given how rarely they win, and they value favorites too little given how often they actually win. As such, the rate of return to betting horses 100/1 or greater is about -61%; betting randomly yields average returns of -23%; while betting the favorite in every race yields losses of only around 5.5%.

The literature documenting these biases is voluminous, and covers both bookmaker- and pari-mutuel markets. The bias was first noted by Griffith in 1949, and has persisted in racetrack betting data around the world, with very few exceptions.¹

Roughly speaking, two broad sets of theories have been proposed to explain the favorite-longshot bias. First, standard neoclassical theory suggests that the prices that bettors are willing to pay for various gambles can be used to recover their utility function. While betting at any odds is actuarially unfair, the data suggest that this is particularly acute for longshots – which are also the riskiest investment. Thus, the neoclassical approach can reconcile both gambling and the longshot bias only by positing (at least locally) risk-loving utility functions, as in Friedman and Savage (1948).

Alternatively, behavioral theories suggest that cognitive errors play a role in market mispricing. These theories generally point to laboratory studies by cognitive psychologists suggesting that people are systematically poor at discerning between small and tiny probabilities, yielding the implication that they will price each similarly. Further, people exhibit a strong preference for certainty over even extremely likely outcomes, leading highly probable gambles to be under-priced. These results form a key part of Kahneman and Tversky’s Prospect Theory (1979). These theories can rationalize the purchase of sometimes extremely unfavorable lottery tickets, and the violations of expected utility theory such as Allais Paradox.

Our aim in this paper is to test whether the risk-love or misperceptions model best fits the data. While there exist many specific models of the favorite-longshot bias, in section 3 we will argue that they each yield implications for the pricing of gambles equivalent to our stark baseline

¹ Thaler and Ziemba (1988), Sauer (1998) and Snowberg and Wolfers (2006) survey the relevant literature.

models of either a risk-loving representative agent, or a representative agent who bases her decisions on a set of decision weights that diverge from true probabilities. As such, the “risk love” versus “misperceptions” distinction is not so much a sharp dividing line between two competing theories, but rather a taxonomy for organizing the two sets of theories. More formally, we ask whether the favorite-longshot bias reflects a non-linear response to the potential proceeds of a winning wager, or to the probability of winning that wager.²

We combine new data with a novel econometric identification strategy to differentiate between these two classes of theories. Our data include all 6 million horse race starts in the United States from 1992 to 2001. These data are an order of magnitude larger than any other dataset previously examined, and allow us to be quite precise in establishing the relevant stylized facts. Our econometric strategy relies on compound gambles to distinguish between theories based on risk-love or misperceptions. While previous authors have relied on average rates of return to win bets to describe the favorite-longshot bias, such data cannot separate the two theories without imposing arbitrary functional form assumptions regarding either the utility function or types of misperceptions. That is, the general tendency to overbet longshots and underbet favorites can be *fully* rationalized by appealing to a standard rational-expectations expected-utility model, with lower rates of return from betting favorites due to the different slopes of the utility function over different potential payoffs. Equally the favorite-longshot bias can be *fully* explained by appealing to expected wealth maximizing agents who are subject to a set of misperceptions that causes them to overweight small probabilities and underweight large probabilities in a specific way. That is, without parametric assumptions (which we are unwilling to make), the two theories yield observationally equivalent implications for average rates of return to betting horses at different odds. Our research question is most similar to Jullien and Salanié (2000) who provide the most sustained attempt at differentiating preference- and perception-based explanations of the favorite-longshot bias. Their approach emphasizes an identification strategy based on differences in the extent of the bias across different races.

Our innovation is to argue that compound lotteries (called “exotic bets” at the racetrack) can be used to derive testable restrictions that differentiate these theories. For example, an “exacta” requires one to bet on both which horse will come first and which will come second.

² Or adopting a neoclassical versus behavioral distinction, we follow Gabriel and Marsden (1990) in asking: “are we observing an inefficient market or simply one in which the tastes and preferences of the market participations lead to the observed results?”

Our approach is to ask whether the specific forms of preferences and perceptions that rationalize the favorite-longshot bias (based on win betting data) can also explain the pricing of exactas, and other compound lotteries. By expanding the choice set under consideration (to correspond with the bettor's actual choice set!), we have the opportunity to use each theory to derive unique testable restrictions. Rossett (1965) provides a related analysis in that he considers not only win bets, but also combinations of win bets as being present in the bettors' choice set, and authors such as Ali (1979) and Asch and Quandt (1987) have tested the efficiency of compound lottery markets. We believe that we are the first to use these prices to distinguish between competing theories of possible market (in)efficiency. Of course the idea is much older: Friedman and Savage (1948) noted that a hallmark of expected utility theory is "that the reaction of persons to complicated gambles can be inferred from their reaction to simple gambles."

To demonstrate the application of this idea to our data, note that the rate of return to betting horses between around 3/1 and 10/1 is approximately constant (at -18%), and close to the average (Figure 1). Thus, under the misperceptions model, one would infer that this is a range over which bettors are equally well calibrated, and hence that betting on combinations of outcomes among such horses should yield similar rates of return. That is, betting on an exacta with the 3/1 horse to win and the 10/1 horse to come second should yield similar expected returns to betting on the reverse ordering (albeit at different odds). The risk love model suggests that bettors have different preferences over betting at different odds, and hence the expected returns to these alternative exactas will differ. To see this, note that the more likely exacta (3/1 then 10/1) is about a 30/1 chance, while the reverse ordering exacta is about a 40/1 chance. Given that the risk-loving bettor prefers the opportunity to win big, they will be willing to accept a larger risk penalty (or negative risk premium) for betting on the less likely exacta, decreasing its rate of return in equilibrium.

The rest of this paper proceeds as follows. In section two, we review the empirical literature, and establish a set of robust stylized facts. Section three provides a mapping of the theories proposed in the favorite-longshot bias literature into our "risk-love" versus "misperceptions" taxonomy. We then lay out the implications of each theory for the pricing of exotic bets, formalizing the intuition offered above. To preview our findings, the pricing function implied by the misperceptions models better matches the observed prices of exactas,

quinellas and trifectas.³ Section five reviews the robustness of this result, and section six concludes. Our key finding is that rationalizing prices of *both* win bets and compound lotteries requires a utility function that is non-linear in probabilities, and the relevant probability weighting function resembles that proposed in Kahneman and Tversky's Prospect Theory.

2. Stylized Facts

Our data contains all 6,301,016 horse starts run in the United States between 1992 and 2001. These data are official jockey club data, and hence are the most precise data available. Data of this nature are prohibitively expensive, which presumably explains why previous studies have used substantially smaller samples. While we have a vast database on every horse and every race, jockey, owner, trainer, sire and dam, we will only exploit the betting data, and whether or not a horse finished first, second or third in each race. The betting odds are determined by a pari-mutuel betting system. Appendix A further describes the data.

We summarize our data in Figure 1. We group horses according to their odds and calculate the rate of return to betting on every horse in each group. Data are graphed on a log-odds scale so as to better show the relevant range of the data. Figure 1 shows the actual rate of return to betting on horses in each category. The average rate of return for betting favorites is about -5.5%, while for horses at a mid-range of 3/1 to 15/1 yield a rate of return of -18%, and real longshots – horses at 100/1 or more – yield much lower returns of -61%. It is this finding that we refer to as the “favorite-longshot bias.” Figure 1 also shows the same pattern for the 201,685 races for which the jockey club recorded payoffs to exacta, quinella or trifecta bets. Given that much of our analysis will focus on this smaller sample, it is reassuring to see a similar pattern of returns.

Figure 2 shows the same rate of return calculations for several other datasets. We present new data from 2,725,000 starts in Australia using data from South Coast Database, and 380,000 starts in Great Britain, using data from flatstats.co.uk. The favorite-longshot bias appears equally evident in these countries, despite the fact that these odds are from a bookmaker-dominated market in the United Kingdom, and bookmakers competing with a state-run pari-

³ Exactas are a bet on two horses to finish first and second in a particular order. A quinella is a bet on two horses to come first and second in either order, and a trifectas is a bet on three horses to come in first, second and third in order.

mutuel in Australia. Figure 2 also includes historical estimates of the favorite-longshot bias, showing that it has been largely stable since it was first noted in Griffith (1949).⁴

The literature has suggested two other empirical regularities that we can explore. First, Thaler and Ziemba (1988) have suggested that there are in fact positive rates of return to betting extreme favorites, perhaps suggesting limits to arbitrage. However, as the confidence intervals in Figure 1 show, there is substantially greater statistical uncertainty about returns on extreme favorites, and in none of our datasets are there statistically significant gains to betting extreme favorites. This is similar to Levitt's (2004) finding that despite significant anomalies in the pricing of bets, there are no profit opportunities from simple betting strategies.

Second, McGlothlin (1956), Ali (1977) and Asch, Malkiel and Quandt (1982) argue that the rate of return to betting moderate longshots falls in the last race of the day. While these conclusions were based on small samples, these studies have come to be widely cited. Kahneman and Tversky (1979) and Thaler and Ziemba (1988) interpreted these previous results as consistent with loss aversion: most bettors are losing at the end of the day, and the “break even stakes”—as bettors call the last race—provides them with a chance to recoup their losses. Thus, bettors underbet the favorite even more than usual, and overbet horses at odds that would eliminate their losses. The dashed line in Figure 1 separates out data from the last race; while the point estimates of the longshot bias differ, these differences are not statistically significantly different from earlier races. (Given that this sample is about one-ninth as large as the full sample, the relevant confidence interval is about three times wider.) If there was evidence of loss aversion in earlier data, it no longer appears evident in more recent data, even as the favorite-longshot bias has persisted.

As such, we propose that a satisfactory theory must be compatible with the following robust stylized facts:

- Rates of return to betting fall as the odds rise. Returns are slightly negative on extreme favorites, moderately negative on mid-range horses and extremely negative for longshots;
- The bias has been persistent for fifty years; and
- The bias occurs across bookmaker, pari-mutuel and combination markets.

⁴ The most notable exception is Busche and Hall's (1988) finding that the favorite-longshot bias was not evident in data on 2,653 Hong Kong races; Busche (1994) confirms this finding on a further 2,690 races in Hong Kong, and 1,738 races in Japan.

In section three we will argue that these facts are not sufficient to separate risk-love from misperception-based theories. As such, we propose a fourth test: that a theory developed to explain equilibrium odds of horses winning should also be able to explain the equilibrium odds in the exacta, quinella and trifecta markets.

3. Two Models of the Favorite-Longshot Bias

We start with two extremely stark models, each of which has the merit of simplicity. Both are representative agent models, but as we suggest below, can be usefully expanded to incorporate heterogeneity. Aggregate price data cannot separately identify more complex models from these representative agent models.

The Risk-Love Model

Following Weitzman (1965), we postulate an expected utility maximizer with unbiased beliefs. In equilibrium, bettors must be indifferent between betting on the favorite horse, A at odds of $O_A/1$, with a probability of winning of p_A , and betting on a longshot B at odds of $O_B/1$, with probability of winning, p_B :

$$p_A U(O_A) = p_B U(O_B) \quad (\text{normalizing utility to zero, if the bet is lost}).^5 \quad [1]$$

Given that we observe the odds (O_A, O_B) and the probabilities (p_A, p_B) of horses in each odds-group winning, these data identify the representative bettor's utility function (up to a scaling factor).⁶ In order to fix a scaling, throughout this paper we normalize so that utility is zero if the bet loses, and utility is one if you choose not to bet. Thus, if the bettor is indifferent as to whether to accept a gamble that wins with probability p , offering odds of $O/1$, then $U(O)=1/p$. The left panel of Figure 3 performs precisely this analysis, backing out the utility function required to fully rationalize the choices shown in Figure 1.

As can be seen, a risk-loving utility function is required in order to rationalize bettors accepting lower average returns on long shots, even as they are riskier bets. The utility function shown fully explains all of the variation in Figure 1 (by construction). The chart also shows that a CRRA utility function also explains the data reasonably well.

⁵ Following Jullien and Salanié (2000), we also assume that each bettor chooses to bet on only one horse in a race, and that they bet a constant amount.

⁶ See Weitzman (1965), Ali (1977), Quandt (1986) and Jullien and Salanie (2000) for prior examples.

Several other theories of the favorite-longshot bias have also been proposed that yield implications that are observationally equivalent to a simple risk-loving representative agent model. For instance, Thaler and Ziemba (1988) argue that “bragging rights” accrue from winning a bet at long odds. Formally, this suggests agents maximize expected utility, where utility is the sum of the felicity of wealth, y , and the felicity of bragging rights or the thrill of winning, b , and hence the expected utility of a gamble at odds O which wins with probability p , can be expressed: $EU(O) = p [y(w_0+O) + b(O)] + (1-p) y(w_0-1)$

As in the representative agent model, bettors will be prepared to accept lower returns on riskier wagers (betting on longshots) if $U'' > 0$. This is possible if either the felicity of wealth is sufficiently convex, or bragging rights are increasing in the payoff at a sufficiently increasing rate. More to the point, revealed preference data do not allow us to separately identify effects operating through y , rather than b , and this is the sense in which the model is observationally equivalent with the simple representative agent who is risk loving. A similar argument applies to Conlisk (1993) in which the mere purchase of a ticket on a longshot may confer some utility.

The Misperceptions Model

Alternatively, under the perceptions-based approach, we postulate a risk-neutral subjective expected utility maximizer, whose subjective beliefs, $\pi(p)$, are systematically biased estimates of the true probabilities.⁷ In equilibrium, bettors must believe that the rates of return to betting on any pair of horses A and B are equal, and that there are no unexploited profit opportunities:

$$\pi(p_A) (O_A+1) = \pi(p_B) (O_B+1) = 1 \quad [2]$$

Consequently data on the odds of each horse (O_A, O_B) and the probabilities of horse in each odds class winning (p_A, p_B) reveal the “decision weights” of the representative bettor. The right panel of Figure 3 shows the probability weighting function implied by the data in Figure 1. The low rates of return to betting longshots are thus rationalized by the assertion that bettors tend to bet as though horses “tiny” probabilities are actually “moderate” probabilities. Beyond this,

⁷ While we term the divergence between π and p “misperceptions”, in non-expected utility theories, π can be interpreted as a preference over types of gambles. Under either interpretation our approach is valid, in that we test whether gambles are motivated by nonlinear functions of wealth, or utility.

the specific shape of the declining rates of return identifies the decision weights at each point.⁸ Interestingly, this function shares some of the features of the decision weights in Kahneman and Tversky’s (1979) Prospect Theory, and the figure shows that the one-parameter decision weighting function suggested by Prelec (1998) fits the data quite closely.

While the assumption of risk-neutrality is clearly too stark, as long as bettors gamble small proportions of their wealth, the relevant risk premia are also second-order.⁹ Moreover while we have presented a very sparse model, a number of richer theories have been proposed that also yield similar implications. For instance, Henery (1985) and Williams and Paton (1997) argue that bettors discount a constant proportion of the gambles in which they bet on a loser, possibly due to a self-serving bias in which losers argue that conditions were atypical. Because longshot bettors lose more often, this discount yields perceptions in which betting on longshots seems more attractive.

Implications for Pricing Compound Lotteries

We now turn to showing how our two families of models—while each just-identified based on the prices of win bets—yield different implications when pricing exotic bets. As such, our approach partly responds to Sauer’s (1998, p.2026) call for research that provides “equilibrium pricing functions from well-posed models of the wagering market.”

We start by showing the example of the exacta in detail (picking the first two horses in order). As before, we price these bets by considering indifference conditions. Pricing an exacta requires data on the perceived likelihood of the pick for first actually winning, and conditional on that occurring, the likelihood of the pick for second coming second, as well as the bettor’s utility function. As such, a bettor will be indifferent between betting on an exacta with horses *A* then *B* paying odds of $O_{AB}/1$ and not betting (which yields no change in wealth, and hence a utility of one), if:

⁸ There remains one minor issue: As figure 1 shows, horses never win as often as suggested by their win odds because of the track-take. Thus we follow the convention in the literature and adjust the odds-implied probabilities by a factor of one minus the track take, so that they are on average unbiased; our results are qualitatively similar whether or not we make this adjustment.

⁹ For instance, assuming log utility, if the bettor is indifferent over betting $x\%$ of their wealth on horse *A* or *B*, then: $\pi(p_A) \log[w+wxO_A] + [1-\pi(p_A)] \log[w-wx] = \pi(p_B) \log[w+wxO_B] + [1-\pi(p_B)] \log[w-wx]$, which under the standard approximation simplifies to: $\pi(p_A) (O_A+1) \approx \pi(p_B) (O_B+1)$, as in equation [2].

Risk-Love Model

(Risk-lover, Unbiased expectations)

$$p_A p_{B/A} U(O_{AB}) = 1$$

Noting $p = 1/U(O)$ from equation [1]

$$\Rightarrow O_{AB} = U^{-1}(U(O_A)U(O_{B/A})) \quad [3]$$

Misperceptions Model

(Biased expectations, Risk-neutral)

$$\pi(p_A)\pi(p_{B/A})(O_{AB} + 1) = 1$$

Noting $\pi(p) = 1/(O+1)$ from equation [2]

$$\Rightarrow O_{AB} = (O_A + 1)(O_{B/A} + 1) - 1 \quad [4]$$

Thus under the perceptions model, the odds of an exacta are a simple function of the odds of horse A winning, and conditional on this, on the odds of B coming second. The preferences model is more demanding, requiring that we estimate the utility function. We estimated the utility function based on the pricing of win bets (in Figure 3), and thus we can invert this to compute unbiased win probabilities from the betting odds.¹⁰

Our empirical tests simply test which of equations [3] and [4] better fit the pricing of exacta bets. We also apply an analogous approach to the pricing of quinella and trifectas bets; the intuition remains the same, but the mathematical details are described in Appendix B.

Two Digressions

Coding of Compound Lotteries

As in Prospect Theory, the frame the bettor adopts in trying to assess each gamble is a key issue, particularly in the misperceptions model. Specifically, equation [4] assumes that bettors first attempt to assess the likelihood of horse A winning, $\pi(p_A)$, and then assess the likelihood of B coming second, $\pi(p_{B/A})$, where $p_{B/A}$ denotes the probability of horse B coming second given that horse A is the winner. An alternative frame might suggest that bettors directly assesses the likelihood of first-and-second combinations, $\pi(p_A p_{B/A})$. Given a non-linear weighting function, these different frames yield different implications.

There is a direct analogy in the literature on the assessment of compound lotteries: does the bettor separately assess the likelihood of winning an initial gamble (picking the winning horse) which yields a subsequent gamble as its prize (picking the second-placed horse), or does she consider the reduced-form compound lottery? Analysis by Camerer and Ho (1994) suggests that the accumulated experimental evidence is more consistent with subjects failing to reduce

¹⁰ Note from figure 1 that we do not have sufficient data to estimate the utility of winning bets at odds greater than 132/1, and so we do not attempt to price bets whose odds would be longer than 132/1; this limitation is most binding for our analysis of trifectas bets.

compound lotteries into their simple lottery equivalent, providing a potential rationale for our treatment in equation [4].

Alternatively, we could choose not to defend either assumption, leaving it as a matter for empirical testing. Interestingly, if gamblers' adopt a frame consistent with the reduction of compound lotteries into their equivalent simple lottery form, this yields a pricing rule for the misperceptions model that is equivalent to that implied by the risk love model.¹¹ Thus, evidence consistent with what we are calling the risk-love model accommodates either risk-love by unbiased bettors, or non-risk-loving but biased bettors, whose bias affects their perception of an appropriately reduced compound lottery. By contrast, the competing "misperceptions model" not only relies on falsification of the reduction of compound lotteries, but also posits a specific form for this violation (shown as equation [4]).

This discussion implies that results consistent with our risk-love model are also consistent with a richer set of models emphasizing choices over simple gambles, including models based on the utility of gambling, information asymmetry or limits to arbitrage, such as Ali (1977), Conlisk (1993), Shin (1992), Hurley and McDonough (1995), Ottaviani and Sørensen (2003), Manski (2004). That is, any theory that prescribes a specific bias in a market for one form of simple gamble (win betting) will yield similar implications in a related market for compound gambles if gamblers assess their equivalent simple gamble form. By implication, rejecting the risk-love model substantially narrows the set of plausible theories of the favorite-longshot bias.

Conditional Probabilities

Note that both equations [3] and [4] suggest that pricing an exacta bet requires data on $O_{B/A}$ – the odds of B coming second, conditional on A coming first; however the odds of this bet are not directly observed. We begin by inferring the conditional probability $p_{B/A}$ (and hence $\pi(p_{B/A})$ and $O_{B/A}$) from win odds, thereby assuming that bettors believe in conditional independence. That is, we apply the so-called Harville (1973) formula: $\pi(p_{B/A}) = \pi(p_B) / (1 - \pi(p_A))$, where $\pi(p) = p$ under the risk-love model. This assumption is akin to thinking about the race for

¹¹ To see this, note that the indifference condition for the reduced compound lottery is: $\pi(p_A p_{B/A}) (O_{AB} + 1) = 1$, and hence $O_{AB} = 1 / \pi(p_A p_{B/A}) - 1$. The risk-love pricing model can be expressed: $O_{AB} = U^{-1}(p_A p_{B/A})$. Because identical data (from Figure 1) is used to construct the utility and decision weight functions respectively, and because each is constructed to rationalize the same set of choices over simple lotteries, each also rationalizes the same set of choices over compound lotteries if choices in both models obey the reduction of compound lotteries into equivalent simple lotteries.

second as a “race within the race” (Sauer, 1998). With this assumption in hand, we can explore how either the utility function or decision weights depicted in Figure 3 yield different implications for pricing of exactas. While relying on the Harville formula is standard in the literature—see for instance Asch and Quandt, 1987—in section 5 we show that our results are robust to dropping this independence assumption, and estimating this conditional probability from the data.

4. Results

Figure 4 shows the pricing functions implied by the risk-love and misperception models, respectively; the x- and y-axes show the odds on each horse, and the z-axis shows the equilibrium exacta odds implied by each model.

Our test of the two models simply involves estimating which of the pricing functions shown in Figure 4 better fits the data. In Table 1 (as in Figure 4) we convert the odds into the price of a contingent contract that pays \$1 if the chosen exacta wins: $Price = 1/(Odds+1)$. We test the ability of each economic model to predict this price by regressing the price of the winning exacta against the prices implied by preference model (column 1), the perceptions model (column 2) and then put them both in horse-race regression (excuse the pun) in column 3.

Comparing columns 1 and 2, the explanatory power of the perceptions model is substantially greater, and the regression in column 3 confirms this, showing that when the regression is allowed to choose optimal weights on the implications of each theory, it strongly prefers the perceptions model.

Panels B and C of Table 1 repeat this analysis, but this time extending our test to see which model can better explain the pricing of quinella and trifecta bets; the intuition is similar to the exacta test. Appendix B contains further mathematical detail. Each of these tests across all three panels suggests that the misperceptions model fits the data better than the risk-love model.

We have also re-run these regressions a number of other ways to test for robustness, and our conclusions are unaltered by: whether or not we include constant terms in the regressions; whether or not we weight by the size of the betting pool; whether we drop observations where the models imply very long odds; whether or not we adjust the perceptions model in the manner described in footnote 8; and different functional forms for the price of a bet, including the natural log price of a \$1 claim, the odds, or log-odds.

An immediate question that arises is why the presence or potential entry of unbiased bettors has not undone the price effects of bettors whose probability assessments are biased. The persistence of the bias in this context may reflect the large track take (equivalent to a bid-ask spread in financial markets), which ensures that the misperceptions model yields almost no exploitable profit opportunities in any of the betting pools. This is not to say that these misperceptions are not costly: As Figure 1 shows, betting on longshots is around eleven times more costly than betting on favorites, and this finding carries through to compound lotteries.

5. Robustness and Conditional Independence

Recall that we observe all of the inputs to both pricing models except the odds of horse B finishing second, conditional on horse A winning. While we used the convenient assumption of conditional independence to assess the likely odds of this bet, there may be good reason to doubt this assumption. For instance, if a heavily favored horse does not win a race, this may reflect the fact that it was injured during the race, which then implies that it is very unlikely to come second. That is, the win odds may provide useful guidance on the probability of winning, but conditional on not winning, may be a poor guide to the race to come second. We now turn to both testing this assumption, and then derive two further tests can distinguish between the risk-love and misperceptions models even if conditional independence fails.¹²

We test the conditional independence assumption by asking whether the Harville formula provides a sufficient statistic for whether a horse will come second. We compute the Harville statistic as $p_{APB}/(1-p_A)$, where p_A and p_B reflect the probability that horses at odds of A/I and B/I , respectively, win the race. We then run a linear probability model where the dependent variable is an indicator variable for whether horse B runs second. Probits yield similar results.

The first column of Table 2 shows that the Harville formula is an extremely useful predictor of the probability of a horse finishing second. To provide a useful yardstick for thinking about the explanatory power, note that this is about four-fifths as high as the R^2 one gets when trying to explain which horse wins the race, using the betting odds as the regressor. In columns two and three however, we find compelling evidence that the Harville formula is not a sufficient statistic. In column two we add dummy variables representing the odds of the first

¹² A qualifier: Even if conditional independence fails, it is not immediately obvious that it yields errors that are correlated in such a way as to drive our main results. Even so, this is an issue for empirical testing.

place horse and the odds of the second placed horse (we use 74 odds groupings in each case). F-tests clearly reveal these fixed effects to be statistically significant. In column three we include a full set of interactions of these fixed effects, estimating the conditional probability non-parametrically from the odds of the first and second placed horses; this regression is equivalent to estimating a large table showing the proportion of runners at odds of B/I who won the race for second, given the winner was at odds of A/I .

We now use these non-parametrically estimated probabilities as a robustness check on our earlier results in Table 1. That is, rather than inferring $p_{B/A}$ (and hence $\pi(p_{B/A})$ and $O_{B/A}$) from the Harville formula, we simply apply the empirical probabilities estimated in the equation shown in column 3 of Table 2. We implement this exercise in Table 3, calculating the price of exotic bets under the risk-love and misperception models, but adapting our earlier approach to, so that $p_{B/A}$, is derived from the data. Again we run a horse race between the competing theories.¹³

The results in Table 3 are consistent with those in Table 1. For both exacta and quinella bets, the misperceptions model has greater explanatory power than the risk-love model, and in a horse race, is strongly preferred.¹⁴

Relative Pricing of Exactas and Quinellas

Our final test of the two models is even more non-parametric, and relies only on the relative pricing of exacta and quinella bets. The power of this test comes from simultaneously considering exacta and quinella bets as both being present in the bettor's choice set.¹⁵ As before, we derive predictions from each model and test which better explains the observed data. The advantage of focusing only on comparisons between the first two horses is that these tests are – by construction – conditionally independent of the characteristics of all other horses in the race and hence the assumptions required for identification are even weaker.

To see the relevant intuition, consider the pricing of a both an exacta and a quinella involving both a horse A, at odds A/I , and horse B, at odds B/I . The exacta $A-B$ ($A-B$ represents

¹³ Because the precision of our estimates of $p_{B/A}$ vary greatly, WLS weighted by the product of the squared standard error of $p_{B/A}$ and p_A might be appropriate. Such an estimation procedure produces qualitatively identical results.

¹⁴ Unfortunately we cannot extend this method to pricing Trifecta bets, because we cannot estimate the conditional probability of a third-placed finish, conditional on the odds of the first two with any real accuracy.

¹⁵ Note that these tests are distinct from the work by authors such as Asch and Quandt (1987) and Dolbear (1993), who test whether exacta pricing is arbitrage-linked to win pricing. Instead, we ask whether the same model that explains pricing of win bets can *jointly* explain the pricing of exacta and quinella bets.

A winning and B coming second) occurs with probability $p_A * p_{B/A}$; the B-A exacta occurs with probability $p_B * p_{A/B}$. By definition the corresponding quinella pays off when the winning exacta is either A-B or B-A and hence occurs with probability $p_A * p_{B/A} + p_B * p_{A/B}$. If horse A is the favorite, the exacta A-B is more likely than B-A and hence less risky. This implies that under the risk-love model the equilibrium the rate of return to exactas putting the favorite first will be *higher* than that on the reverse ordering. By contrast, the misperceptions models is linear in beliefs, implying relative payoffs to the two bet types are proportional to their *perceived* occurrence. As such, under the misperceptions model there are values of A and B in which the rate of return to exactas putting the favorite first will be *lower* than the reverse ordering. While we only observe the prices of the winning exacta, we also observe the winning quinella, which effectively bundles both the A-B and B-A exacta, and hence each model yields unique implications for the relative prices of the winning exacta and quinella bets. Specifically, consider the A-B exacta at odds of $E_{AB}/1$, and the corresponding quinella at $Q/1$.

Risk-Love Model

(Risk-lover, Unbiased expectations)

$$Exacta : p_A p_{B/A} U(E_{AB}) = 1$$

$$\Rightarrow p_{B/A} = \frac{U(A)}{U(E_{AB})}$$

Misperceptions Model

(Biased expectations, Risk-neutral)

$$Exacta : \pi(p_a)\pi(p_{B/A})(E_{AB} + 1) = 1$$

$$[5] \quad \pi(p_{B/A}) = \frac{A+1}{E_{AB} + 1} \Rightarrow p_{B/A} = \pi^{-1}\left(\frac{A+1}{E_{AB} + 1}\right) \quad [6]$$

$$Quinella : [p_A p_{B/A} + p_B p_{A/B}] U(Q) = 1 \quad Quinella : [\pi(p_A)\pi(p_{B/A}) + \pi(p_B)\pi(p_{A/B})] (Q+1) = 1$$

$$\Rightarrow p_{B/A} = U(A) \left(\frac{1}{U(Q)} - \frac{1}{U(E_{BA})} \right) \quad [7] \quad \pi(p_{A/B}) = (B+1) \left(\frac{1}{Q+1} - \frac{1}{E_{AB} + 1} \right) \Rightarrow p_{A/B} = \pi^{-1} \left(\frac{(B+1)(E_{AB} - Q)}{(E_{AB} + 1)(Q+1)} \right) \quad [8]$$

Hence, from [1], [5] and [7]:

$$\frac{p_B p_{A/B}}{p_B p_{A/B} + p_A p_{B/A}} = \frac{U(Q)}{U(E_{BA})} \quad [9] \quad \frac{p_A p_{B/A}}{p_A p_{B/A} + p_B p_{A/B}} = \frac{\pi^{-1}\left(\frac{1}{A+1}\right)\pi^{-1}\left(\frac{A+1}{E_{AB} + 1}\right)}{\pi^{-1}\left(\frac{1}{A+1}\right)\pi^{-1}\left(\frac{A+1}{E_{AB} + 1}\right) + \pi^{-1}\left(\frac{1}{B+1}\right)\pi^{-1}\left(\frac{(B+1)(E_{AB} - Q)}{(E_{AB} + 1)(Q+1)}\right)} \quad [10]$$

Equations [9] and [10] shows that for any pair of horses at win odds $A/1$ and $B/1$ with quinella odds $Q/1$, each model yields different implications for how frequently we expect to observe the A-B exacta winning, relative to the B-A exacta. In a simple regression predicting which of the top two horses is the winner, the misperceptions model yields a robust and significant positive correlation with actual outcomes (coefficient = 0.70; standard error = 0.015),

while the misperceptions model is negatively correlated with outcomes (coefficient = -0.41; standard error = 0.014).

Equations 9 and 10 also yield distinct predictions of the winning exacta even *within* any set of apparently similar races (those whose first two finishers are at A/I and B/I with the quinella paying Q/I). Thus, we can include a full set of fixed effects for A , B , Q and *their interactions* in our statistical tests of the predictions of each model.¹⁶ The residual after differencing out these fixed effects is the predicted likelihood that A beats B , relative to the average for all races in which a horses at odds of A/I and B/I fill the quinella at odds Q/I . That is, for all races we compute the predictions of each model for the likelihood that exacta $A-B$ occurs, relative to $B-A$, and subtract the baseline $A*B*Q$ cell mean to yield the model predictions, relative to the fixed effects. The results, summarized in Figure 5 are remarkably robust to the inclusion of these multiple fixed effects (and interactions): the coefficient on the misperceptions model declines slightly (and insignificantly), while the risk-love model maintains a significant but perversely negative correlation with outcomes.

Given the presence of these $A*B*Q$ fixed effects, it should be clear that this test of our two theories differs from our earlier tests; specifically by focusing only on the relative rankings of the first two horses, this test entirely eliminates parametric assumptions about “the race for second place.” It is also clearly that the preference-based model does a much better job in predicting the winning exacta, given horses that finish in the top two positions (and their odds).

These tests imply that while a risk-love model can be constructed to account for the pricing of win bets, it yields inaccurate implications for the *relative* pricing of exacta and quinella bets. By contrast, the perceptions-based model is consistent with the pricing of exacta, quinella and trifecta bets, and as this section showed, also consistent with the relative pricing of exacta and quinella bets. Moreover, these results are robust to a range of different approaches to testing the theory.

6. Conclusion

Employing a new and much larger dataset, we document a set of stylized facts concerning rates of return to betting on horses. As with other authors, we note a substantial

¹⁶ Because the odds A , B and Q are actually continuous variables, I include fixed effects for each percentile of the distribution of each variable (and a full set of interactions of these fixed effects).

favorite-longshot bias. Naturally, the term “bias” is somewhat misleading here. That the rate of return to betting on horses at long odds is much lower than the average return to betting on favorites simply falsifies a model that bettors maximize a function that is linear in probabilities and linear in payoffs. Thus the pattern of pricing can be reconciled either by positing concavity in the utility function, or a non-expected utility function employing nonlinear probability weights that violate the reduction of compound lotteries. For compactness, we referred to the former as explaining the data with “risk-love”, while we refer to the latter as explaining the data with “misperceptions”. Neither label is particularly accurate, because each category includes a wider range of competing theories.

We show that these models can be separately identified based on aggregate data by demanding that models that can explain choices over betting on different horses to win can also explain choices over compound bets, such as exactas, quinellas and trifectas. Because the underlying risk, or set of beliefs (depending on the relevant theory) is traded in both the win and compound betting markets, we can derive unique testable implications of both sets of theories. Our results are more consistent with the favorite-longshot bias being driven by misperceptions rather than risk-love. Indeed, while each model is individually quite useful for pricing compound lotteries, in a horse race the misperceptions model strongly dominates the risk-love model. These results are robust to a range of alternative approaches to testing the theories.

These biases likely persist in equilibrium because the misperceptions are not sufficiently large as to generate profit opportunities for unbiased bettors. That said, the cost of this bias is also very large, and de-biasing an individual bettor could reduce the cost of their gambling substantially.

While noting that our misperceptions-based model fits the full set of bettors’ choices over simple and compound bets, rather than stating a strong conclusion, we would simply argue that our results suggest it seems likely that non-expected utility theories are the more promising candidate for explaining racetrack bettor behavior. As such, this provides some cause for optimism that misperceptions may also explain anomalies in other domains of decision-making under uncertainty.

References

- Ali, Mukhtar (1977), "Probability and Utility Estimates for Racetrack Bettors", *Journal of Political Economy*, 85(4), pp.803-815.
- Ali, Mukhtar (1979), "Some Evidence of the Efficiency of a Speculative Market", *Econometrica*, 47(2), pp.387-392.
- Asch, Peter, Burton Malkiel and Richard Quandt (1982), "Racetrack Betting and Informed Behavior", *Journal of Financial Economics*, 10, 187-194.
- Asch, Peter and Richard Quandt (1987), "Efficiency and Profitability in Exotic Bets", *Economica*, 54, 289-298.
- Busche, Kelly and Christopher D. Hall (1988), "An Exception to the Risk Preference Anomaly", *Journal of Business*, 61: 337-46.
- Busche, Kelly (1994), "Efficient Market Results in an Asian Setting", in D. Hausch, V. Lo, and W.T. Ziemba (eds), *Efficiency of Racetrack Betting Markets*, New York: Academic Press, pp. 615-16.
- Conlisk (1993), "The Utility of Gambling", *Journal of Risk and Uncertainty*, 6(3), 255-275.
- Dolbear, Trener Jr. (1993), "Is Racetrack Betting on Exactas Efficient?", *Economica*, 60(237), pp.105-111.
- Friedman, Milton and L.J. Savage (1948), "The Utility Analysis of Choices Involving Risk", *Journal of Political Economy*, 56(4), 279-304.
- Gabriel, Paul and James Marsden (1990), "An Examination of Market Efficiency in British Racetrack Betting", *Journal of Political Economy*, 98(4), pp.874-885.
- Griffith, R. (1949), "Odds Adjustment by American Horse Race Bettors", *American Journal of Psychology*, 62, 290-294.
- Harville, David A. (1973), "Assigning Probabilities to the Outcomes of Multi-entry Competitions", *Journal of the American Statistical Association*, 68, 312-316.
- Henery, Robert J. (1985), "On the Average Probability of Losing Bets on Horses with Given Starting Price Odds", *Journal of the Royal Statistical Society. Series A (General)*, 148(4), 342-349.
- Hurley, William and Lawrence McDonough (1995), "A Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Parimutuel Betting", *American Economic Review*, 85(4), 949-955.

- Jullien, Bruno and Bernard Salanié (2000), “Estimating Preferences Under Risk: The Case of Racetrack Bettors”, *Journal of Political Economy*, 108(3), 503-530.
- Kahneman, Daniel and Amos Tversky (1979), “Prospect Theory: An Analysis of Decision under Risk”, *Econometrica*, 47(2), 263-292.
- Levitt, Steven (2004), “Why are Gambling Markets Organised so Differently from Financial Markets?”, *The Economic Journal*, 114, 223-246
- Manski, Charles (2004), “Interpreting the Predictions of Prediction Markets”, *NBER Working Paper*, #10359, March 2004.
- McGlothlin (1956), “Stability of Choices Among Uncertain Alternatives”, *American Journal of Psychology*, 69(4), 604-615.
- Ottaviani, Marco and Peter Norman Sørensen (2003), “Late Informed Betting and the Favourite-Longshot Bias”, *CEPR Discussion Paper #4092*.
- Prelec, Drazen (1998), “The Probability Weighting Function”, *Econometrica*, 66(3), 497-527.
- Quandt, Richard E. (1986), “Betting and Equilibrium”, *Quarterly Journal of Economics*, 101(1), 201-208.
- Rosett, Richard N. (1965), “Gambling and Rationality”, *Journal of Political Economy*, 73, 595-607.
- Sauer, Ray (1998), “The Economics of Wagering Markets”, *Journal of Economic Literature*, 36(4), 2021-2064.
- Shin, Hyung Song (1992), “Prices of State Contingent Claims with Insider Traders, and the Favourite Longshot Bias”, *Economic Journal*, 102, 426-435.
- Snowberg, Erik and Justin Wolfers (2006), “The Favorite-Longshot Bias: Understanding a Market Anomaly”, in preparation for Donald Hausch and William Ziemba (eds.), *Efficiency of Sports and Lottery Markets* (Elsevier: Handbooks in Finance series)
- Thaler, Richard and William Ziemba (1988), “Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries”, *Journal of Economic Perspectives* 2(2): 161-174.
JPE (1991)
- Weitzman, Martin (1965), “Utility Analysis and Group Behavior: An Empirical Study”, *Journal of Political Economy*, 73(1), 18-26.
- Williams, Leighton Vaughan and David Paton (1997), “Why is There a Favorite-Longshot Bias in British Racetrack Betting Markets”, *Economic Journal*, 107(440), pp.150-158.

Appendix A: Data

Our dataset consists of all horse races run in North America from 1992 to 2001. The data was generously provided to us by Axcis Inc., a subsidiary of the jockey club. The data record performance of every horse in each of its starts, and contains the universe of officially recorded variables having to do with the horses themselves, the tracks and race conditions.

Our concern is with the pricing of bets. Thus, our primary sample consists of the 6,301,016 observations in 763,238 races for which win odds and finishing positions are recorded. We use these data, subject to the data cleaning restrictions below, to generate Figures 1-3. We are also interested in pricing exacta, quinella and trifectas bets and have data on the winning payoffs in 314,977, 116,307 and 282,576 races respectively. (The prices of non-winning combinations are not recorded.)

Due to the size of our dataset, whenever observations were problematic, we simply dropped the entire race from our dataset. Specifically, if a race has more than one horse owned by the same owner, rather than deal with “coupled runners”, we simply dropped the race. Additionally, if a race had a dead heat for first, second or third place the exacta, quinella and trifecta payouts may not be accurately recorded and so we dropped these races. When the odds of any horse were reported as zero we dropped the race. Further if the odds across all runners implied that the track take was less than 15% or more than 22%, we dropped the race. After these steps, we are left with 5,608,281 valid observations on win bets from 679,049 races and 1,651,018 observations from 201,685 races include both valid win odds and payoffs for the winning exotic bets.

Finally, Figures 1-3 show the mapping between odds and the true probability of winning that we use throughout the paper. For prices that are relatively common (such as 4/1), we had enough observations that we could reliably estimate the probability of horses at those odds winning. At more unusual levels levels we had to group together horses with similar odds. Our grouping algorithm chose the width of each bin so as to yield a standard error on the estimated rate of return in that bin less than 2%; we include all starts above 100/1 in a single final grouping. We used a consistent set of data and odds groupings for all the results in our paper, and linearly interpolate between bins when necessary.

Appendix B: Pricing of Compound Lotteries using Conditional Independence

In the text we derived our pricing formulae for pricing exacta bets explicitly; this appendix extends that analysis to also include the pricing of quinella and trifecta betting. The following formulae, derived in the text, are central for the derivations in this section:

Risk-Love Model

(Risk-lover, Unbiased expectations)

$$U(O) = 1/p \quad [1]$$

$$\Rightarrow O = U^{-1}(1/p)$$

Misperceptions Model

(Biased expectations, Risk-neutral)

$$\pi(p) = 1/(1+O) \quad [2]$$

$$\Rightarrow O = 1/\pi(p) - 1$$

Our pricing of compound bets proceeds simply by noting that the expected utility of all bets should be equalized. An exacta requires the bettor to correctly specify the first two horses, in order. A quinella is a bet on two horses to finish first and second, but the bettor need not specify their order. A trifecta is a bet on the three horses to finish first, second and third, and the bettor must correctly specify their order. Thus the quinella and trifecta analogues to equations [3] and [4] in the main text fottttare:

Risk-Love Model

Quinella

$$(p_A p_{B|A} + p_B p_{A|B})U(Q_{AB}) = 1$$

$$\Rightarrow Q_{AB} = U^{-1}\left(\frac{U(O_A)U(O_{B|A})U(O_B)U(O_{A|B})}{U(O_A)U(O_{B|A}) + U(O_B)U(O_{A|B})}\right) \quad [3q]$$

Trifecta

$$p_A p_{B|A} p_{C|A,B} U(T_{ABC}) = 1$$

$$\Rightarrow T_{ABC} = U^{-1}(U(O_A)U(O_{B|A})U(O_{C|A,B})) \quad [3t]$$

Misperceptions Model

Quinella

$$[\pi(p_A)\pi(p_{B|A}) + \pi(p_B)\pi(p_{A|B})](Q_{AB} + 1) = 1$$

$$\Rightarrow Q_{AB} = \frac{(O_A + 1)(O_{B|A} + 1)(O_B + 1)(O_{A|B} + 1)}{(O_A + 1)(O_{B|A} + 1) + (O_B + 1)(O_{A|B} + 1)} - 1 \quad [4q]$$

Trifecta

$$\pi(p_A)\pi(p_{B|A})\pi(p_{C|A,B})(T_{ABC} + 1) = 1$$

$$\Rightarrow T_{ABC} = (O_A + 1)(O_{B|A} + 1)(O_{C|A,B} + 1) - 1 \quad [4t]$$

The odds data, O_A , O_B and O_C are directly observable. The utility (U) and probability weighting (π) functions that we use are shown in figure 3. Thus we have all the data necessary to price these compound bets except the conditional probabilities $O_{B|A}$, $O_{A|B}$ and $O_{C|A,B}$.

We provide two approaches to recovering these unobservables. In the first approach we assume conditional independence, as in Harville (1973). Thus, $p_{B|A} = p_B/(1-p_A)$, $p_{A|B} = p_A/(1-p_B)$ and $p_{C|A,B} = p_C/(1-p_A-p_B)$. Our second approach directly estimates $p_{B|A}$ from the data for each $O_A * O_B$ cell, as described in column 3 of Table 2. The same function also yields an estimate of $p_{A|B}$. Unfortunately we do not have enough data to estimate $p_{C|A,B}$ in the same way. Under both approaches these probability estimates are then fed into formulae [1] and [2], respectively to recover the relevant odds $O_{B|A}$, $O_{A|B}$ and $O_{C|A,B}$.

Figure 1

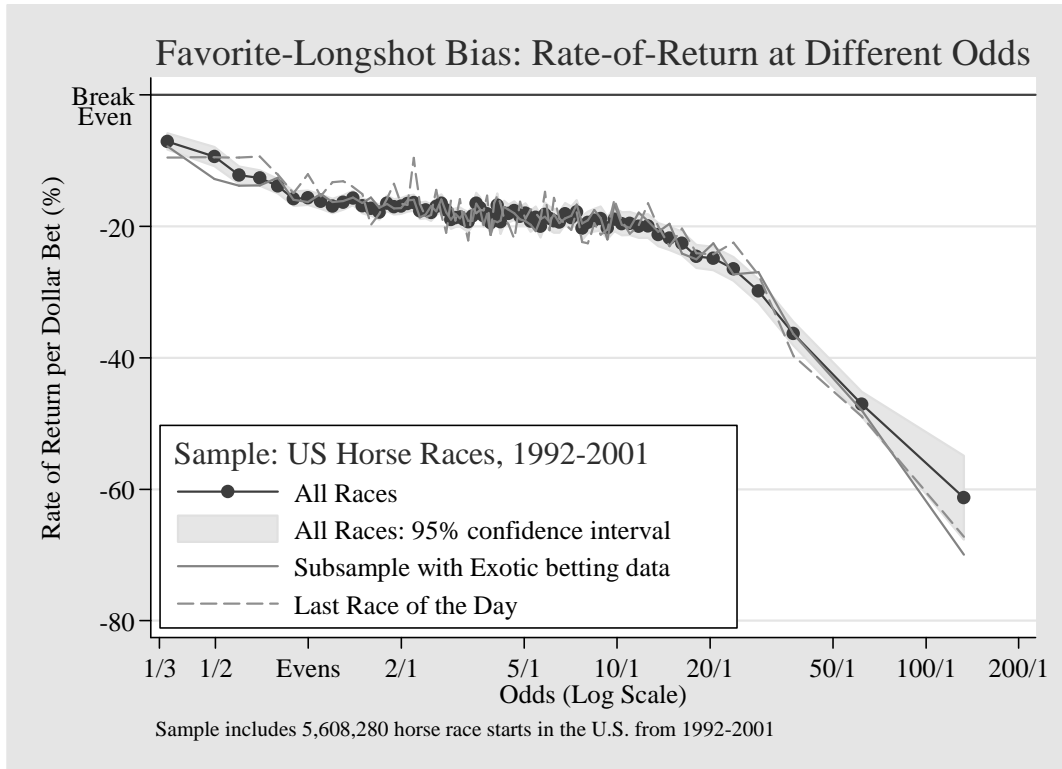


Figure 2

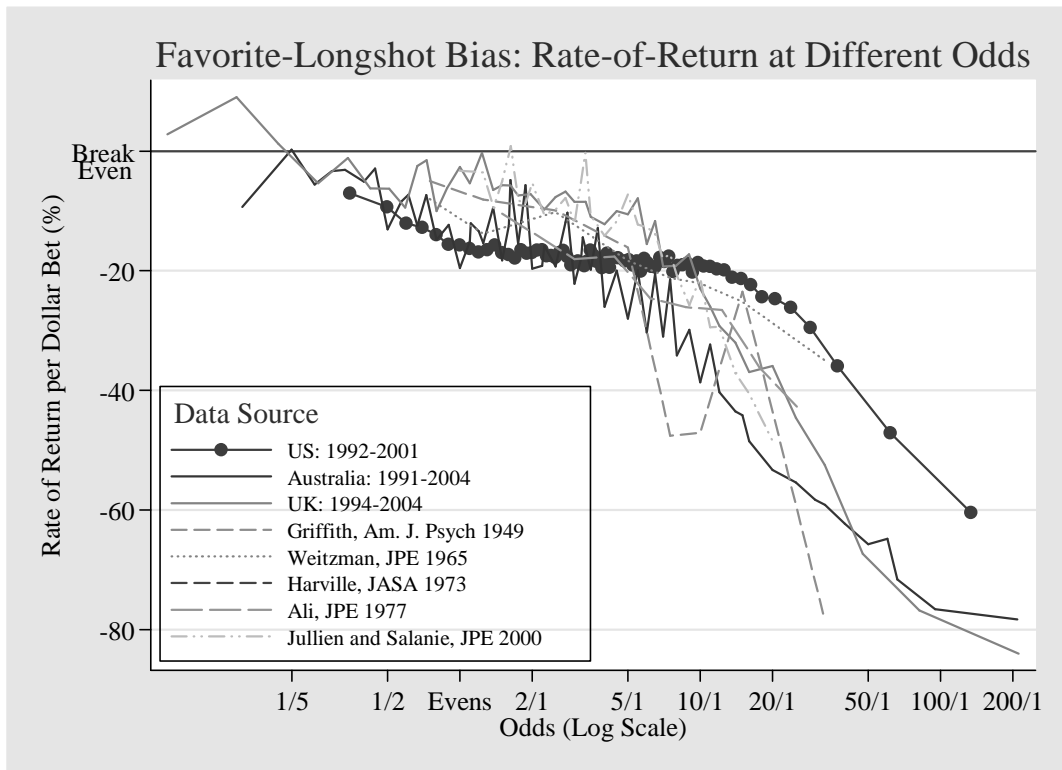


Figure 3: Rationalizing the Data

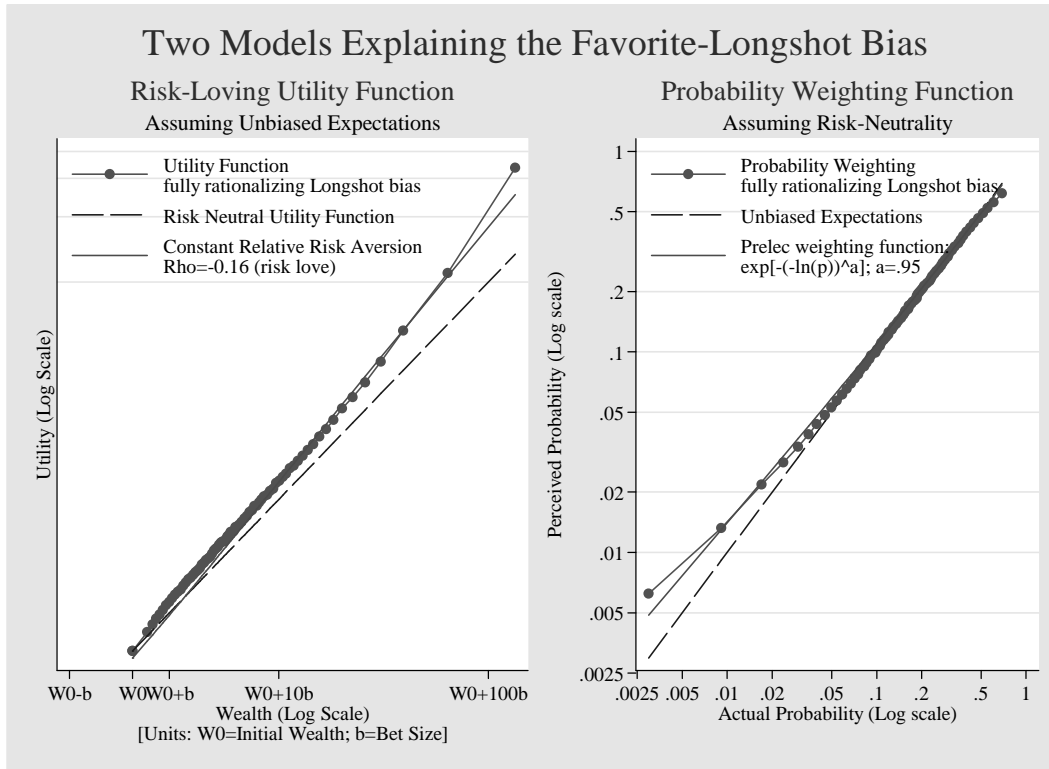


Figure 4: Predicted Exacta Pricing – Risk-Love and Misperception Models

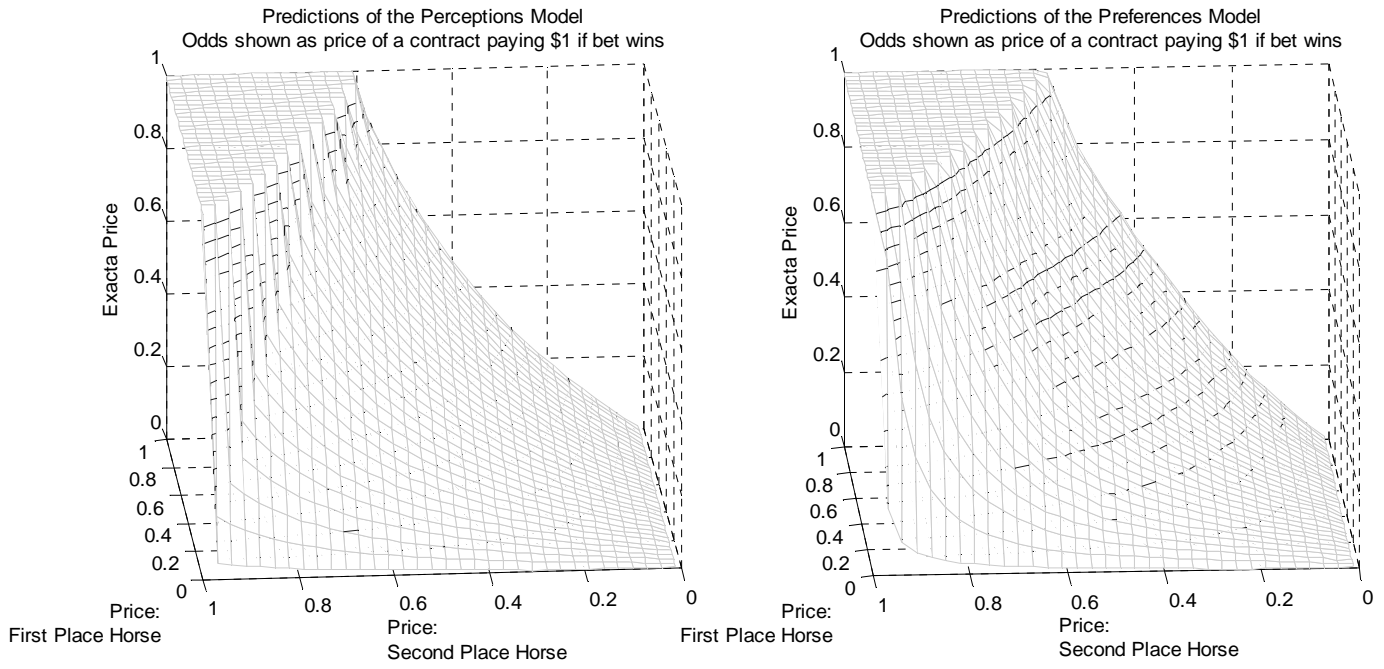
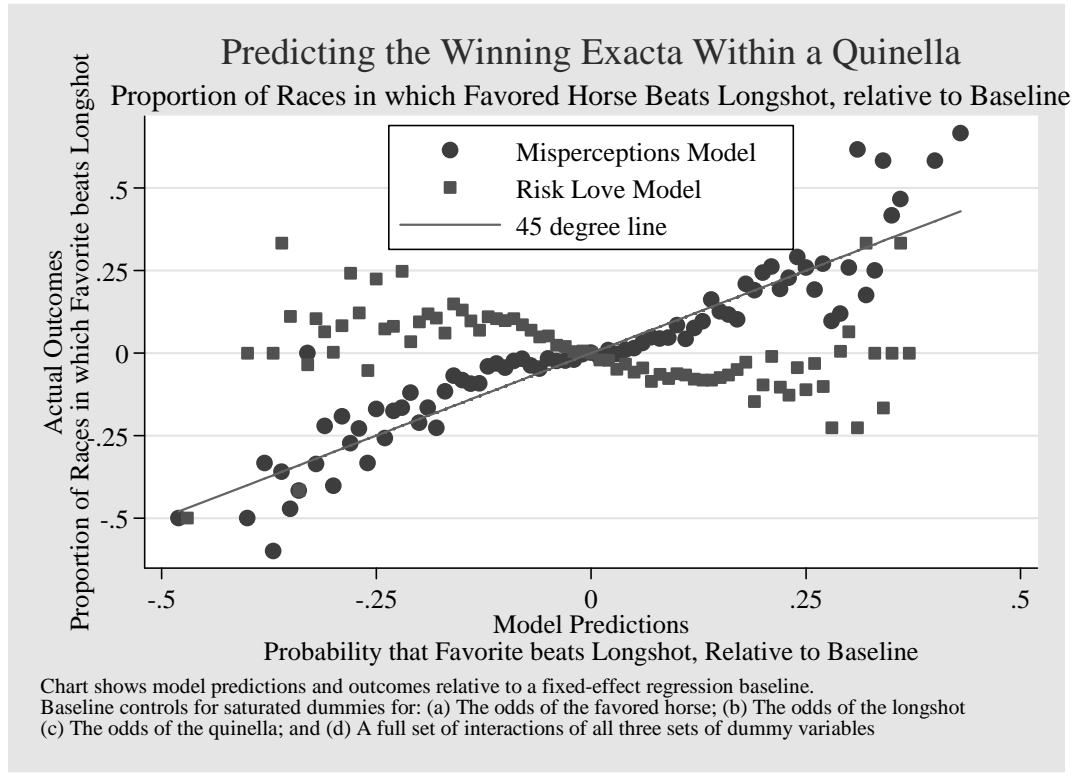


Figure 5: Dropping Conditional Independence



Notes: For each race we took the first two finishers in each race and computed the likelihood each was the winner, conditional on knowing the winning quinella. These predictions are made under the two models outlined in the text, using as inputs data on the odds of each horse (A/I , B/I), their quinella (Q/I) and the winning exacta (E/I). We then compute the mean predictions and outcomes for all races within the same $\{A, B, Q\}$ cell. Subtracting these means yields the model predictions and outcomes relative to these fixed effects. For the purposes of the plot, we round these residuals to the nearest percentage point (shown on the x-axis), and the y-axis shows actual win percentages for races in each bucket.

Table 1: Predicting the Price of Compound Bets

Dependent Variable: Price of a contract paying \$1 if the compound bet wins			
<i>Specification</i>	(1) <i>Misperceptions</i>	(2) <i>Risk-Love</i>	(3) <i>Horse Race</i>
Panel A: Exacta Bets ($n = 193,425$)			
<i>Misperceptions model predictions</i>	0.8830 (.0008)		0.8112 (0.0066)
<i>Risk-love model predictions</i>		0.7162 (0.0007)	0.0591 (0.0054)
<i>Constant</i>	0.0055 (.0001)	0.0092 (0.0001)	0.0057 (0.0001)
<i>Adjusted R²</i>	0.8499	0.8382	0.8500
Panel B: Quinella Bets ($n = 69,062$)			
<i>Misperceptions model predictions</i>	0.8443 (0.0015)		1.0175 (0.0120)
<i>Risk-love model predictions</i>		0.6108 (0.0011)	-0.1277 (0.0088)
<i>Constant</i>	0.0124 (0.0002)	0.0250 (0.0002)	0.0101 (0.0002)
<i>Adjusted R²</i>	0.8189	0.8005	0.8194
Panel C: Trifecta Bets ($n = 120,646$)			
<i>Misperceptions model predictions</i>	0.7912 (0.0014)		0.9566 (0.0131)
<i>Risk-love model predictions</i>		0.7333 (0.0013)	-0.1552 (0.0140)
<i>Constant</i>	0.0032 (0.0001)	0.0009 (0.0001)	0.0028 (0.0001)
<i>Adjusted R²</i>	0.7223	0.7119	0.7226

Notes: Model predictions generated as per equations [3] and [4] in the text; utility functions and decision weights are generated using data from Figure 3. Both actual prices and model-generated prices are shown as the price of a bet paying \$1 if the bet wins.

Table 2: Models for Predicting the Conditional Probability of a 2nd Placed Finish
Dependent Variable: Indicator for whether a horse came in second
(Conditional on not winning)

<i>Specification</i>	<i>(1)</i>	<i>(2)</i>	<i>(3)</i>
<i>Prediction from conditional independence (Harville Formula)</i>	0.7972 (.0012)	0.8569 (.0078)	
<i>Odds of this horse and Odds of First horse (74 dummy variables for each)</i>		F=42.3 (p=0.00)	
<i>Full set of interactions: This horse * First horse (5476 dummy variables)</i>			F=80.15 (p=0.00)
<i>Adjusted R²</i>	0.0774	0.0786	0.0812

Notes: $n = 4,929,198$ starts (excludes all winning horses).

Table 3: Robustness to Relaxing Conditional Independence Assumption

Dependent Variable: Price of a contract paying \$1 if the compound bet wins			
<i>Specification</i>	(1)	(2)	(3)
	<i>Misperceptions</i>	<i>Risk-Love</i>	<i>Horse Race</i>
Panel A: Exacta Bets ($n = 195,921$)			
<i>Misperceptions model predictions</i>	0.9993 (.0009)		1.0299 (0.0058)
<i>Risk-love model predictions</i>		0.8002 (0.0008)	-0.0251 (0.0047)
<i>Constant</i>	0.0018 (.0001)	0.0066 (0.0001)	0.0017 (0.0001)
<i>Adjusted R²</i>	0.8513	0.8276	0.8513
Panel B: Quinella Bets ($n = 69,753$)			
<i>Misperceptions model predictions</i>	0.9970 (0.0018)		0.9603 (0.0136)
<i>Risk-love model predictions</i>		0.7290 (0.0013)	0.0272 (0.0100)
<i>Constant</i>	0.0030 (0.0002)	0.0176 (0.0002)	0.0035 (0.0003)
<i>Adjusted R²</i>	0.8230	0.8104	0.8231

Notes: Model predictions generated as per equations [3] and [4] in the text; utility functions and decision weights are generated using data shown in Figure 3. Estimates of conditional probabilities are generated using the regression in column 3 of table 2. Both actual prices and model-generated prices are shown as the price of a bet paying \$1 if the bet wins.