

# Explaining the Favorite-Longshot Bias: Is it Risk-Love or Misperceptions?\*

**Erik Snowberg**  
Stanford GSB

snowberg@stanford.edu  
[www.stanford.edu/~esnowber](http://www.stanford.edu/~esnowber)

**Justin Wolfers**  
The Wharton School, U.Penn  
CEPR, IZA, & NBER  
jwolfers@wharton.upenn.edu  
[www.nber.org/~jwolfers](http://www.nber.org/~jwolfers)

November 28, 2007

## Abstract

The favorite-longshot bias presents a challenge for theories of decision making under uncertainty. This longstanding empirical regularity is that betting odds provide biased estimates of the probability of a horse winning—longshots are overbet, while favorites are underbet. Neoclassical explanations focus on rational gamblers who overbet longshots due to risk-love. The competing behavioral explanations emphasize the role of misperceptions of probabilities. We provide novel empirical tests that can discriminate between these competing theories by focusing on the pricing of compound bets. We test whether the models that explain gamblers' choices in one part of their choice set (betting to win) can also rationalize decisions over a wider choice set, including compound bets in the exacta, quinella or trifecta pools. Using a new, large-scale dataset ideally suited to implement these tests we 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).

---

\*We thank David Siegel of Equibase for supplying the data, and Scott Hereld, Ravi Pillai and Stephanie Wang for their valuable assistance in managing the data. Jon Bendor, Betsey Stevenson, Matthew White and William Ziemba provided useful feedback. Snowberg gratefully acknowledges the SIEPR Dissertation Fellowship through a grant to the Stanford Institute for Economic Policy Research. 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 discussed empirical regularity in gambling is the *favorite-longshot bias*. That is, equilibrium market prices (betting odds) are biased estimates of the probability of a horse winning. Specifically, bettors value longshots more than expected given how rarely they win, and they value favorites too little given how often they actually win. Quantitatively, as shown in Figure 1, the rate of return to betting on horses with odds of 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  $5.5\%$ .

Since the favorite-longshot bias was first noted by Griffith in 1949, it has been found in racetrack betting data around the world, with very few exceptions. The literature documenting this bias is voluminous, and covers both bookmaker- and pari-mutuel markets.<sup>1</sup>

Two broad sets of theories have been proposed to explain the favorite-longshot bias. First, 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, this is particularly acute for longshots—which are also the riskiest investments. 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 and misperceptions of probabilities play a role in market mis-pricing. These theories incorporate laboratory studies by cognitive psychologists that show that people are systematically poor at discerning between small and tiny probabilities, and hence price both similarly. Further, people exhibit a strong preference for certainty over extremely likely outcomes, leading highly probable gambles to be under-priced. These results form an important foundation of Prospect Theory

---

<sup>1</sup>Thaler and Ziemba (1988), Sauer (1998) and Snowberg and Wolfers (2007) survey the literature. The exception to the favorite-longshot bias is found in Busche and Hall (1988) which finds that the favorite-longshot bias is not evident in data on 2,653 Hong Kong races; Busche (1994) confirms this finding on a further 2690 races in Hong Kong, and 1738 races in Japan.

(Kahneman and Tversky, 1979), which can rationalize the purchase of sometimes extremely unfavorable lottery tickets, and violations of expected utility theory such as the Allais (1953, 1979) Paradox.

Our aim in this paper is to test whether a *risk-love model* or *misperceptions model* best fits the data. While there exist many specific models of the favorite-longshot bias, we show in Section 3 that each yields implications for the pricing of gambles equivalent to stark models of either a risk-loving representative agent, or a representative agent who bases her decisions on biased perceptions of true probabilities. Thus, the risk-love versus misperceptions distinction is a taxonomy for two sets of theories, rather than a sharp dividing line between two competing theories. More formally, we ask whether the favorite-longshot bias reflects a non-linear response to the potential proceeds of a winning wager, or a non-linear response to the probability of winning that wager.<sup>2</sup>

We combine new data with a novel econometric identification strategy to discriminate between these two classes of theories. Our data include all 6.4 million horse race starts in the United States from 1992 to 2001. These data are an order of magnitude larger than any dataset previously examined, and allow us to be extremely precise in establishing the relevant stylized facts. Our econometric strategy relies on compound gambles to distinguish between theories based on risk-love or misperceptions.

Previous authors have only used data on win bets, but these cannot separate the two theories without imposing arbitrary functional form assumptions on preferences or types of misperceptions. That is, the favorite-longshot bias in win bets can be fully rationalized by a standard rational-expectations expected-utility model, with lower rates of return to betting on favorites due to the shape of the utility function. Equally, the bias can be fully explained by appealing to an expected wealth maximizing agent who overweights small probabilities and underweights large probabilities. Thus, without parametric assumptions (which we are

---

<sup>2</sup>Or adopting a behavioral versus neoclassical 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?”

unwilling to make), the two theories are observationally equivalent in win bet data.

Our innovation is to differentiate these theories by deriving testable predictions about compound lotteries (called *exotic bets* at the racetrack). For example, an *exacta* is a bet on both which horse will come first and which will come second. Essentially, we ask whether the preferences and perceptions that rationalize the favorite-longshot bias (in win bet data) can also explain the pricing of exactas, *quinellas*—a bet on two horses to come first and second in either order—and *trifectas*—a bet on three horses to come in first, second and third in order. By expanding the choice set under consideration (to correspond with the bettor’s actual choice set!), we use each theory to derive unique testable predictions.

To demonstrate the application of this idea to our data, note that betting on horses with odds between 3/1 and 10/1 has an approximately constant rate of return (at  $-18\%$ , see Figure 1). Thus, under the misperceptions model we infer bettors are equally well calibrated over this range, and hence betting on combinations of outcomes among such horses will yield similar rates of return. That is, betting on an exacta with a 3/1 horse to win and a 10/1 horse to come second will yield similar expected returns to betting on the exacta with the reverse ordering (although the odds of the two exactas will differ). In contrast, under the risk-love model 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. As the risk-loving bettor prefers the opportunity to win big, she 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.

Our research question is most similar to Jullien and Salanié (2000) and Gandhi (2007) who differentiate between preference- and perception-based explanations of the favorite-longshot bias using only win bet data. The results of the former favor perception-based explanations and the results of the later favor preference-based explanations. Rosett (1965) conducts a related analysis in that he considers both win bets and combinations of win bets as present

in the bettors' choice set. Ali (1979) and Asch and Quandt (1987) test 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) notes 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.”

## 2 Stylized Facts

Our data contains all 6,403,712 horse starts run in the United States between 1992 and 2001. These data are official jockey club data; the most precise data available. Data of this nature are extremely expensive, which presumably explains why previous studies have used substantially smaller samples. Appendix A provides more detail about the data.

We summarize our data in Figure 1. We calculate the rate of return to betting on every horse at each odds, and use Lowess smoothing to take advantage of information from horses with similar odds. Data are graphed on a log-odds scale so as to better show their relevant range. Figure 1 shows the 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  $-61\%$ . This is the favorite-longshot bias. Figure 1 also shows the same pattern for the 206,808 races (with 1,485,112 horse starts) 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 from the South Coast Database, and 380,000 starts in Great Britain from flatstats.co.uk. The favorite-longshot bias appears equally evident in these countries, despite the fact that odds are determined by a bookmaker-dominated

market in the United Kingdom, and bookmakers competing with a state-run pari-mutuel market in Australia. Figure 2 also includes historical estimates of the favorite-longshot bias, showing that it has been stable since it was first noted in Griffith (1949).

The literature suggests two other empirical regularities that we can explore. First, Thaler and Ziemba (1988) suggest that there are positive rates of return to betting extreme favorites, perhaps suggesting limits to arbitrage. This is not true in any of our datasets, providing a similar finding to Levitt (2004): 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. These studies have come to be widely cited despite being based on small samples. Kahneman and Tversky (1979) and Thaler and Ziemba (1988) interpret these results as consistent with loss aversion: most bettors are losing at the end of the day, and 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 differ slightly, these differences are not statistically significant. If there was evidence of loss aversion in earlier data, it is no longer evident in 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 on longshots;
- The bias has been persistent for fifty years; and
- The bias occurs across bookmaker, pari-mutuel and combination markets.

In Section 3 we argue that these three facts are not sufficient to separate risk-love from misperception-based theories. We propose a fourth test: 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 models where all agents have the same preferences and perceptions, 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.

#### 3.1 The Risk-Love Model

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

$$p_A U(O_A) = p_B U(O_B) \quad (\text{normalizing utility to zero, if the bet is lost}).^3 \quad (1)$$

Given that we observe the odds  $(O_A, O_B)$  and the probabilities  $(p_A, p_B)$  of horses at each odds winning, these data identify the representative bettor's utility function up to a scaling factor.<sup>4</sup> To fix a scaling we normalize utility to zero if the bet loses, and to one if the bettor chooses not to bet. Thus, if the bettor is indifferent between accepting or rejecting a gamble that wins with probability  $p$ , offering odds of  $O/1$ , then  $U(O) = \frac{1}{p}$ . The left panel of Figure 3 performs precisely this analysis, backing out the utility function required to fully explain

---

<sup>3</sup>We also assume that each bettor chooses to bet on only one horse in a race.

<sup>4</sup>See Weitzman (1965), Ali (1977), Quandt (1986) and Jullien and Salanié (2000) for prior examples.

all of the variation in Figure 1.

As can be seen from Figure 3, a risk-loving utility function is required to rationalize the bettor accepting lower average returns on longshots, even as they are riskier bets. Figure 3 also shows that a CARA utility function fits the data reasonably well.

Several other theories of the favorite-longshot bias yield implications that are observationally equivalent to a risk-loving representative agent model. Some of these theories are clearly equivalent—such as Golec and Tamarkin (1998), which argues that bettors prefer skew rather than risk—as they are theories about the shape of the utility function. Other theories require more discussion. 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,  $y(\cdot): \mathbb{R} \rightarrow \mathbb{R}$ , and the felicity of bragging rights or the thrill of winning,  $b(\cdot): \mathbb{R} \rightarrow \mathbb{R}$ . Hence the expected utility to a bettor with initial wealth  $w_0$  of a gamble at odds  $O$  that wins with probability  $p$  can be expressed as:  $E(U(O)) = p[y(w_0 + O) + b(O)] + [1 - p]y(w_0 - 1)$

As before, bettors will 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  $y(\cdot)$ , rather than  $b(\cdot)$ . In this sense the model of Thaler and Ziemba (1988) is observationally equivalent to a risk-loving representative agent. A similar argument applies to Conlisk (1993) in which the mere purchase of a bet on a longshot may confer some utility.

### 3.2 The Misperceptions Model

Alternatively, under the misperceptions model we postulate risk-neutral subjective expected utility maximizers, whose subjective beliefs are given by the *probability weighting function*  $\pi(p): [0, 1] \rightarrow [0, 1]$ . In equilibrium, bettors must believe that the subjective rates of return to betting on any pair of horses  $A$  and  $B$  are equal so that there are no unexploited opportunities

for subjective gain:

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

Consequently, data on the odds of each horse ( $O_A$ ,  $O_B$ ) and the probabilities of the horse at each odds winning ( $p_A$ ,  $p_B$ ) reveal the misperceptions of the representative bettor.<sup>5</sup>

The right panel of Figure 3 shows the probability weighting function  $\pi(p)$  implied by the data in Figure 1. The low rates of return to betting longshots are rationalized by bettors who bet as though horses with tiny probabilities of winning actually have moderate probabilities of winning. The specific shape of the declining rates of return identifies the probability weighting function at each point.<sup>6</sup> This function shares some of the features of the decision weights in Prospect Theory (Kahneman and Tversky, 1979), and the figure shows that the one-parameter probability weighting function in Prelec (1998) fits the data quite closely.

While the assumption of risk-neutrality may be too stark, as long as bettors gamble small proportions of their wealth the relevant risk premia are second-order.<sup>7</sup> Moreover, while we have presented a very sparse model, a number of richer theories have been proposed that also yield similar implications. For instance, Ottaviani and Sørensen (2003) show that initial information asymmetries between bettors may lead to misperceptions of the true probabilities of horses winning. Moreover, 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.

---

<sup>5</sup>While we term the divergence between  $\pi(p)$  and  $p$  misperceptions, in non-expected utility theories,  $\pi(p)$  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 probability. In (2) we implicitly assume that  $\pi(1) = 1$ , although we allow  $\lim_{p \rightarrow 1} \pi(p) \leq 1$ .

<sup>6</sup>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.

<sup>7</sup>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 (2).

### 3.3 Implications for Pricing Compound Lotteries

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

We discuss the pricing of exactas (picking the first two horses in order) in detail. Prices for these bets are constructed from: the bettors’ utility function, indifference conditions as in (1) and (2), data on the perceived likelihood of the pick for first  $A$  actually winning ( $p_A$  or  $\pi(p_A)$ , depending on the model), and conditional on  $A$  winning, the likelihood of the pick for second  $B$  coming second ( $p_{B|A}$  or  $\pi(p_{B|A})$ ). A bettor will be indifferent between betting on an exacta on horses  $A$  then  $B$  in that order, paying odds of  $E_{AB}/1$ , and not betting (which yields no change in wealth, and hence a utility of one), if:

#### Risk-Love Model

(Risk-lover, Unbiased expectations)

$$p_A p_{B|A} U(E_{AB}) = 1$$

Noting  $p = \frac{1}{U(O)}$  from (1)

$$E_{AB} = U^{-1}(U(O_A)U(O_{B|A})) \quad (3)$$

#### Misperceptions Model

(Biased expectations, Risk-neutral)

$$\pi(p_A)\pi(p_{B|A})(E_{AB} + 1) = 1$$

Noting  $\pi(p) = \frac{1}{O + 1}$  from (2)

$$E_{AB} = (O_A + 1)(O_{B|A} + 1) - 1 \quad (4)$$

Thus, under the misperceptions model the odds of the exacta  $E_{AB}$  are a simple function of the odds of horse  $A$  winning  $O_A$ , and conditional on this, on the odds of  $B$  coming second  $O_{B|A}$ . The preferences model is more demanding, requiring that we estimate the utility function. The utility function is estimated from the pricing of win bets (in Figure 3), and can be inverted to compute unbiased win probabilities from the betting odds.<sup>8</sup>

---

<sup>8</sup>Our econometric method imposes continuity on the utility and probability weighting functions; the data mandate that both be strictly increasing. Together this is sufficient to ensure that  $\pi(\cdot)$  and  $U(\cdot)$  are invertible. As in Figure 1, we do not have sufficient data to estimate the utility of winning bets at odds greater than 200/1. This prohibits us from pricing bets whose odds are greater than 200/1, which is most binding for our

Our empirical tests simply determine which of (3) or (4) better fit the actual prices of exacta bets. We apply an analogous approach to the pricing of quinella and trifecta bets; the intuition is the same, and the mathematical details are described in Appendix B.

Note that both (3) and (4) require  $O_{B|A}$ , which is not directly observable. In Section 4 we infer the conditional probability  $p_{B|A}$  (and hence  $\pi(p_{B|A})$  and  $O_{B|A}$ ) from win odds by assuming that bettors believe in conditional independence. That is, we apply the Harville (1973) formula:  $\pi(p_{B|A}) = \frac{\pi(p_B)}{1 - \pi(p_A)}$ ; replacing  $\pi(p)$  with  $p$  in the risk-love model. This assumption is akin to thinking about the race for second as a “race within the race” (Sauer, 1998). 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 assumption and estimating this conditional probability  $p_{B|A}$  directly from the data.

### 3.4 Failure to Reduce 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, (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 given that  $A$  is the winner  $\pi(p_{B|A})$ . An alternative frame might suggest that bettors directly assesses the likelihood of first-and-second combinations,  $\pi(p_A p_{B|A})$ . Unless the probability weighting function is a power function ( $\pi(p) = p^\alpha$ ), these different frames yield different implications (Aczél, 1966).

There is a direct analogy to 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 equivalent simple lottery (as in Samuelson (1952))? The accumulated experimental evidence is more consistent with subjects failing to reduce compound lotteries into simple lotteries, providing a rationale for our treatment in (4) (Camerer and Ho, 1994).<sup>9</sup>

---

analysis of trifecta bets.

<sup>9</sup>Additionally, note that (4) satisfies the compound independence axiom of Segal (1990).

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.<sup>10</sup> Thus, evidence consistent with what we are calling the risk-love model accommodates either risk-love by unbiased bettors, or risk-neutral but biased bettors, whose bias affects their perception of an appropriately reduced compound lottery. By contrast, the competing misperceptions model implies the failure to reduce compound lotteries and posits a specific form for this failure (shown as (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. These include 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), Manski (2006). Any theory that prescribes a specific bias in a market for a 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.

## 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 determining which of the pricing functions in Figure 4 better fits the data. In Table 1 (as in Figure 4) we convert the odds into the price

---

<sup>10</sup>To see this, note that identical data (from Figure 1) is used to construct the utility and decision weight functions respectively, so each is constructed to rationalize the same set of choices over simple lotteries. This implies each model also rationalizes the same set of choices over compound lotteries if preferences in both models obey the reduction of compound lotteries into equivalent simple lotteries.

of a contingent contract that pays \$1 if the chosen exacta wins:  $Price = \frac{1}{Odds + 1}$ . We test the ability of each economic model to predict this price by examining the mean absolute error of the predictions of both models from the actual prices of exotic bets (column 1). We further investigate which of the models produces predictions that are closer, observation-by-observation, to the prices that are actually observed (column 3). The explanatory power of the perceptions model is substantially greater. The misperceptions model is six percentage points closer to the actual prices of exactas (column 2) an improvement of 20% over the risk-love 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 in all three cases; Appendix B contains further mathematical detail. Each of these tests across all three panels shows that the misperceptions model fits the data better than the risk-love model.<sup>11</sup>

An immediate question 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 be due to the large track take (equivalent to a large bid-ask spread in financial markets), which ensures that the misperceptions model yields 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.

---

<sup>11</sup>We have also re-run these tests a number of other ways to test for robustness. Our conclusions are unaltered by: whether or not we weight observations by the size of the betting pool, whether we drop observations where the models imply very long odds, whether or not we adjust the models 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.

## 5 Robustness and Conditional Independence

Recall that we observe all of the inputs to both pricing models except  $O_{B|A}$ , the odds of horse  $B$  finishing second, conditional on horse  $A$  winning. In Section 4 we used the convenient assumption of conditional independence to assess the likely odds of this bet, but there may be good reason to doubt this assumption. For example, if a heavily favored horse does not win a race, this may reflect the fact that it was injured during the race, which would imply that it is very unlikely to come second. That is, the win odds may provide useful guidance on the probability of winning, but may be a poor guide to the race for second. In this section we test the assumption of conditional independence and derive two further tests can distinguish between the risk-love and misperceptions models even if this assumption fails.<sup>12</sup>

We test the conditional independence assumption by asking whether we can improve on the predictions of the Harville (1973) formula using other data. As in Section 3.3, the Harville formula is:  $p_{B|A} = \frac{p_B}{(1 - p_A)}$ , where  $p_A$  and  $p_B$  reflect the probability that horses  $A$  and  $B$ , respectively, win the race. We estimate linear probability model where the dependent variable is an indicator variable for whether horse  $B$  finished second. A probit specification yields similar results.

The first specification of Table 2 shows that the Harville formula is an extremely useful predictor of the probability of a horse finishing second. As a guide for thinking about the explanatory power of the Harville formula, note that the  $R^2$  of specification 1 is about four-fifths the  $R^2$  of the regression of an indicator for whether a horse won the race on its betting odds. Columns 2 and 3 however, provide compelling evidence that we can do better than the Harville formula. Column 2 adds dummy variables representing the odds of the first place horse and the odds of the second placed horse (using 100 odds groupings in each case, each grouping containing 1% of the odds distribution). Column 3 includes a full set of interactions of these fixed effects, estimating the conditional probability non-parametrically

---

<sup>12</sup>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.

from the odds of the first and second place horses; this regression is equivalent to estimating a large table showing the proportion of horses at odds of  $O_B/1$  who won the race for second, given the winner was at odds of  $O_A/1$ . In both columns 2 and 3, F-tests show that these fixed effects are jointly statistically significant.

We now use non-parametrically estimated probabilities as a robustness check of our 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 using the Lowess procedure of Cleveland, Devlin and Grosse (1988). 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 so that  $p_{B|A}$  is derived from the data.<sup>13</sup>

The results in Table 3 are almost identical to those in Table 1. For exacta, quinella and trifecta bets, the misperceptions model has greater explanatory power than the risk-love model.

## 5.1 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 bettors choice set.<sup>14</sup> 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. 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 horses  $A$ , at odds  $O_A/1$ , and  $B$ , at odds  $O_B/1$ . The exacta  $AB$  ( $AB$  represents  $A$

---

<sup>13</sup>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. Additionally, we could estimate  $p_{B|A}$  directly from column 3 of Table 2. These approaches produce qualitatively identical results.

<sup>14</sup>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.

winning and  $B$  coming second) occurs with probability  $p_A * p_{B|A}$ ; the  $BA$  exacta occurs with probability  $p_B * p_{A|B}$ . By definition the corresponding quinella pays off when the winning exacta is either  $AB$  or  $BA$  and hence occurs with probability  $p_A * p_{B|A} + p_B * p_{A|B}$ . If horse  $A$  is the favorite, the exacta  $AB$  is more likely than  $BA$  and hence less risky. This implies that under the risk-love model 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  $B$  so that the rate of return to exactas putting the favorite first will be lower than the reverse ordering.

Each model yields unique implications for the relative prices of the winning exacta and quinella bets and thus unique predictions for  $\frac{p_A p_{B|A}}{p_A p_{B|A} + p_B p_{A|B}}$ , the probability horse  $A$  wins given that  $A$  and  $B$  are the top two finishers. Specifically, consider the  $AB$  exacta at odds of  $E_{AB}/1$ , and the corresponding quinella at  $Q/1$ :

#### Risk-Love Model

(Risk-lover, Unbiased expectations)

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

$$p_{B|A} = \frac{U(O_A)}{U(E_{AB})} \quad (5)$$

*Quinella:*  $[p_A p_{B|A} + p_B p_{A|B}] U(Q) = 1$

$$p_{A|B} = U(O_B) \left( \frac{1}{U(Q)} - \frac{1}{U(E_{AB})} \right) \quad (7)$$

Hence from (1), (5) and (7):

$$\frac{p_A p_{B|A}}{p_B p_{A|B} + p_A p_{B|A}} = \frac{U(Q)}{U(E_{AB})} \quad (9)$$

#### Misperceptions Model

(Biased expectations, Risk-neutral)

*Exacta:*  $\pi(p_A) \pi(p_{B|A}) (E_{AB} + 1) = 1$

$$\pi(p_{B|A}) = \frac{O_A + 1}{E_{AB} + 1} \Rightarrow p_{B|A} = \pi^{-1} \left( \frac{O_A + 1}{E_{AB} + 1} \right) \quad (6)$$

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

$$\pi(p_{A|B}) = (O_B + 1) \left( \frac{1}{Q+1} - \frac{1}{E_{AB}+1} \right) \Rightarrow p_{A|B} = \pi^{-1} \left( \frac{(O_B+1)(E_{AB}-Q)}{(E_{AB}+1)(Q+1)} \right) \quad (8)$$

$$\frac{p_A p_{B|A}}{p_B p_{A|B} + p_A p_{B|A}} = \frac{\pi^{-1} \left( \frac{1}{O_A+1} \right) \pi^{-1} \left( \frac{O_A+1}{E_{AB}+1} \right)}{\pi^{-1} \left( \frac{1}{O_A+1} \right) \pi^{-1} \left( \frac{O_A+1}{E_{AB}+1} \right) + \pi^{-1} \left( \frac{1}{O_B+1} \right) \pi^{-1} \left( \frac{(O_B+1)(E_{AB}-Q)}{(E_{AB}+1)(Q+1)} \right)} \quad (10)$$

Equations (9) and (10) show that for any pair of horses at win odds  $O_A/1$  and  $O_B/1$  with quinella odds  $Q/1$ , each model yields different implications for how frequently we expect to observe the  $AB$  exacta winning, relative to the  $BA$  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 = 0.63; standard error = 0.014,  $n = 60,288$ ), while the misperceptions model is negatively correlated with outcomes (coefficient =  $-0.59$ ; standard error = 0.013,  $n = 60,288$ ).

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  $O_A/1$  and  $O_B/1$  with the quinella paying  $Q/1$ ). Thus, we can include a full set of fixed effects for  $O_A$ ,  $O_B$ ,  $Q$  and their interactions in our statistical tests of the predictions of each model.<sup>15</sup> 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  $O_A/1$  and  $O_B/1$  fill the quinella at odds  $Q/1$ . That is, for all races we compute the predictions of each model for the likelihood that exacta  $AB$  occurs, relative to  $BA$ , and subtract the baseline  $O_A * O_B * Q$  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.

It is clear that the misperceptions model does a much better job in predicting the winning exacta, given the horses that finish in the top two positions. It should also be clear that this test of the 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.

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 subsection shows, also consistent with the relative pricing of exacta and quinella bets. These results are robust to a range of different approaches to testing the theories.

---

<sup>15</sup>Because the odds  $O_A$ ,  $O_B$  and  $Q$  are actually continuous variables, we include fixed effects for each percentile of the distribution of each variable (and a full set of interactions of these fixed effects).

## 6 Conclusion

Employing a new dataset which is much larger than those in the existing literature, we document stylized facts about the rates of return to betting on horses. As with other authors, we note a substantial favorite-longshot bias. The term bias is somewhat misleading here. That the rate of return to betting on horses at long odds is much lower than the return to betting on favorites simply falsifies a model in which bettors maximize a function that is linear in probabilities and linear in payoffs. Thus, the pricing of win bets can be reconciled by a representative bettor with either a concave utility function, or a subjective utility function employing non-linear 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 as each category includes a wider range of competing theories.

We show that these models can be separately identified using aggregate data by requiring that they explain both choices over betting on different horses to win and choices over compound bets: exactas, quinellas and trifectas. Because the underlying risk or set of beliefs (depending on the theory) is traded in both the win and compound betting markets, we can derive unique testable implications from 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, the misperceptions model strongly dominates the risk-love model. This result is robust to a range of alternative approaches to distinguishing between the theories.

This bias likely persists in equilibrium because misperceptions are not large enough 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 their cost of 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 simply argue that our results suggest that non-expected utility theories are a promising candidate for explaining

racetrack bettor behavior. This provides cause for optimism that misperceptions may also explain anomalies in other domains of decision-making under uncertainty.

## References

- Aczél, J. 1966. *Lectures on Functional Equations and their Applications*. New York: Academic Press.
- Ali, Mukhtar M. 1977. “[Probability and Utility Estimates for Racetrack Bettors.](#)” *The Journal of Political Economy* 85(4):803–815.
- Ali, Mukhtar M. 1979. “[Some Evidence of the Efficiency of a Speculative Market.](#)” *Econometrica* 47(2):387–392.
- Allais, Maurice. 1953. “[Le Comportement de l’Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l’Ecole Americaine.](#)” *Econometrica* 21(4):503–546.
- Allais, Maurice. 1979. “The So-called Allais Paradox and Rational Decisions Under Uncertainty”. In *Expected Utility Hypotheses and the Allais Paradox*, ed. Maurice Allais and Ole Hagen. Dordrecht, Holland: Reidel Publishing Company.
- Asch, Peter, Burton G. Malkiel and Richard E. Quandt. 1982. “Racetrack Betting and Informed Behavior.” *Journal of Financial Economics* 10(2):187–194.
- Asch, Peter and Richard E. Quandt. 1987. “[Efficiency and Profitability in Exotic Bets.](#)” *Economica* 54(215):289–298.
- Busche, Kelly. 1994. “Efficient Market Results in an Asian Setting”. In *Efficiency of Race-track Betting Markets*, ed. Donald Hausch, V. Lo and William T. Ziemba. New York: Academic Press pp. 615–616.
- Busche, Kelly and Christopher D. Hall. 1988. “[An Exception to the Risk Preference Anomaly.](#)” *The Journal of Business* 61(3):337–346.
- Camerer, Colin F. and Teck Hua Ho. 1994. “Violations of the betweenness axiom and nonlinearity in probability.” *Journal of Risk and Uncertainty* 8(2):167–196.
- Cleveland, William S., Susan J. Devlin and Eric Grosse. 1988. “Regression by Local Fitting: Methods, Properties, and Computational Algorithms.” *Journal of Econometrics* 37(1):87–114.
- Conlisk, John. 1993. “The Utility of Gambling.” *Journal of Risk and Uncertainty* 6(3):255–275.
- Dolbear, F. Trenery. 1993. “[Is Racetrack Betting on Exactas Efficient?](#)” *Economica* 60(237):105–111.
- Friedman, Milton and Leonard J. Savage. 1948. “[The Utility Analysis of Choices Involving Risk.](#)” *The Journal of Political Economy* 56(4):279–304.
- Gabriel, Paul E. and James R. Marsden. 1990. “[An Examination of Market Efficiency in British Racetrack Betting.](#)” *The Journal of Political Economy* 98(4):874–885.

- Gandhi, Amit. 2007. “Rational Expectations at the Racetrack: Testing Expected Utility Using Prediction Market Prices.” University of Wisconsin-Madison, *mimeo*.
- Golec, Joseph and Maury Tamarkin. 1998. “[Bettors Love Skewness, Not Risk, at the Horse Track.](#)” *Journal of Political Economy* 106(1):205.
- Griffith, R.M. 1949. “[Odds Adjustments by American Horse-Race Bettors.](#)” *The American Journal of Psychology* 62(2):290–294.
- Harville, David A. 1973. “[Assigning Probabilities to the Outcomes of Multi-Entry Competitions.](#)” *Journal of the American Statistical Association* 68(342):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.](#)” *The 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.
- Kahneman, Daniel and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica* 47(2):263–292.
- Levitt, Steven D. 2004. “Why are Gambling Markets Organised so Differently from Financial Markets?” *The Economic Journal* 114(495):223–246.
- Manski, Charles F. 2006. “Interpreting the Predictions of Prediction Markets.” *Economics Letters* 91(3):425–429.
- McGlothlin, William H. 1956. “[Stability of Choices among Uncertain Alternatives.](#)” *The 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.](#)” *The Quarterly Journal of Economics* 101(1):201–208.
- Rosett, Richard N. 1965. “[Gambling and Rationality.](#)” *The Journal of Political Economy* 73(6):595–607.
- Samuelson, Paul A. 1952. “[Probability, Utility, and the Independence Axiom.](#)” *Econometrica* 20(4):670–678.

- Sauer, Raymond D. 1998. “The Economics of Wagering Markets.” *Journal of Economic Literature* 36(4):2021–2064.
- Segal, Uzi. 1990. “Two-Stage Lotteries without the Reduction Axiom.” *Econometrica* 58(2):349–377.
- Shin, Hyun Song. 1992. “Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias.” *The Economic Journal* 102(411):426–435.
- Snowberg, Erik and Justin Wolfers. 2007. “The Favorite-Longshot Bias: Understanding a Market Anomaly”. In *Efficiency of Sports and Lottery Markets*, ed. Donald Hausch and William Ziemba. Elsevier: Handbooks in Finance series.
- Thaler, Richard H. and William T. Ziemba. 1988. “Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries.” *The Journal of Economic Perspectives* 2(2):161–174.
- Weitzman, Martin. 1965. “Utility Analysis and Group Behavior: An Empirical Study.” *The Journal of Political Economy* 73(1):18–26.
- Williams, Leighton Vaughn and David Paton. 1997. “Why is There a Favourite-Longshot Bias in British Racetrack Betting Markets?” *The Economic Journal* 107(440):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 6,403,712 observations in 865,934 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 and 5. 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 5,606,336 valid observations on win bets from 678,729 races and 1,485,112 observations from 206,808 races include both valid win odds and payoffs for the winning exotic bets.

## Appendix B Pricing of Compound Lotteries using Conditional Independence

In the text we derived the pricing formulae for exacta bets explicitly; this appendix extends that analysis to also include the pricing of quinella and trifecta bets. The derivations of these pricing formulae depend on the following two formulae originally derived in Section 3:

### Risk-Love Model

(Risk-lover, Unbiased expectations)

$$U(O) = \frac{1}{p} \quad (\text{B.1})$$

### Misperceptions Model

(Biased expectations, Risk-neutral)

$$\pi(p) = \frac{1}{O + 1} \quad (\text{B.2})$$

As in the text, we derive pricing formulae by imposing that the expected utility of all bets is equal. Consider a horse race which includes horses  $A$ ,  $B$  and  $C$ . 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 quinella bet on horses  $A$  and  $B$  gives odds  $Q_{AB}$ . A trifecta is a bet on three horses to finish first, second and third, and the bettor must correctly specify their order. A trifecta bet on horses  $A$ ,  $B$  and  $C$ , in that order, gives odds  $T_{ABC}$ . Thus the quinella and trifecta analogues to equations (3)

and (4) in the main text are:

### Risk-Love Model

(Risk-lover, Unbiased expectations)

*Quinella:*

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

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

*Trifecta:*

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

$$T_{ABC} = U^{-1}(U(O_A)U(O_{B|A})U(O_{C|A,B})) \quad (3t) \quad T_{ABC} = (O_A+1)(O_{B|A}+1)(O_{C|A,B}+1)-1 \quad (4t)$$

### Risk-Love Model

(Risk-lover, Unbiased expectations)

*Quinella:*

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

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

*Trifecta:*

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

The odds data,  $O_A$ ,  $O_B$  and  $O_C$  are directly observable. The utility  $U(\cdot)$  and probability weighting  $\pi(\cdot)$  functions that we use are shown in Figure 3. In order to price these compound bets we also need the conditional probabilities  $O_{B|A}$ ,  $O_{A|B}$  and  $O_{C|A,B}$ .

As noted in Section 3.3, we provide two approaches to recovering these unobservables. First, 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  $p_{B|A}$ ,  $p_{A|B}$ , and  $p_{C|A,B}$  using Lowess smoothing as described in Cleveland, Devlin and Grosse (1988). Under both the Harville and Lowess approach these probability estimates and (B.1) and (B.2) are used to recover the relevant odds  $O_{B|A}$ ,  $O_{A|B}$  and  $O_{C|A,B}$ .

Colin Camerer suggested several additional analysis of our data. Although unrelated to the analysis at hand, we include two of these suggested analyses in Figures B.1 and B.2 in the hope that they will be useful to other scholars.

Table 1: Mean Error Based Tests of Risk-Love versus Misperceptions Model

<i>Test:</i>	(1)	(2)	(3)
	Absolute Error:  Prediction – Actual	Absolute % Error: $\frac{ \text{Prediction} - \text{Actual} }{\text{Actual}}$	Which Prediction is Closer to Actual? (%)
Panel A: Exacta Bets (n=197,551)			
Risk-Love Model	0.0139	34.3%	42.1%
Misperceptions Model	0.0125	28.0%	57.9%
Risk-Love Error – Misperceptions Error	0.00137 (.00002)	6.3% (.1%)	
Panel B: Quinella Bets (n=70,169)			
Risk-Love Model	0.0274	39.0%	46.0%
Misperceptions Model	0.0258	36.3%	54.0%
Risk-Love Error – Misperceptions Error	0.00155 (.00003)	2.7% (.2%)	
Panel C: Trifecta Bets (n=137,756)			
Risk-Love Model	0.00796	100%	28.9%
Misperceptions Model	0.00532	57.4%	71.1%
Risk-Love Error – Misperceptions Error	0.00264 (.00001)	42.9% (.2%)	

Notes: Standard errors in parenthesis. Predictions and actual outcomes are measured in the price of a contract that pays \$1 if the event occurs, zero otherwise.

Table 2: Predicting the Conditional Probability of a Second Place Finish

<i>Dependent Variable:</i> Indicator for whether a horse came in second (Conditional on not winning)			
<i>Specification:</i>	(1)	(2)	(3)
Prediction from Conditional Independence (Harville Formula)	0.793 (.0012)	0.908 (.0077)	
Odds of this horse and odds of first horse (100 dummy variables for each)		F=32.3 p=0.00	
Full set of interactions: (this horse * first horse) (10,000 dummy variables)			F=43.5 p=0.00
$R^2$	0.0782	0.0794	0.0813

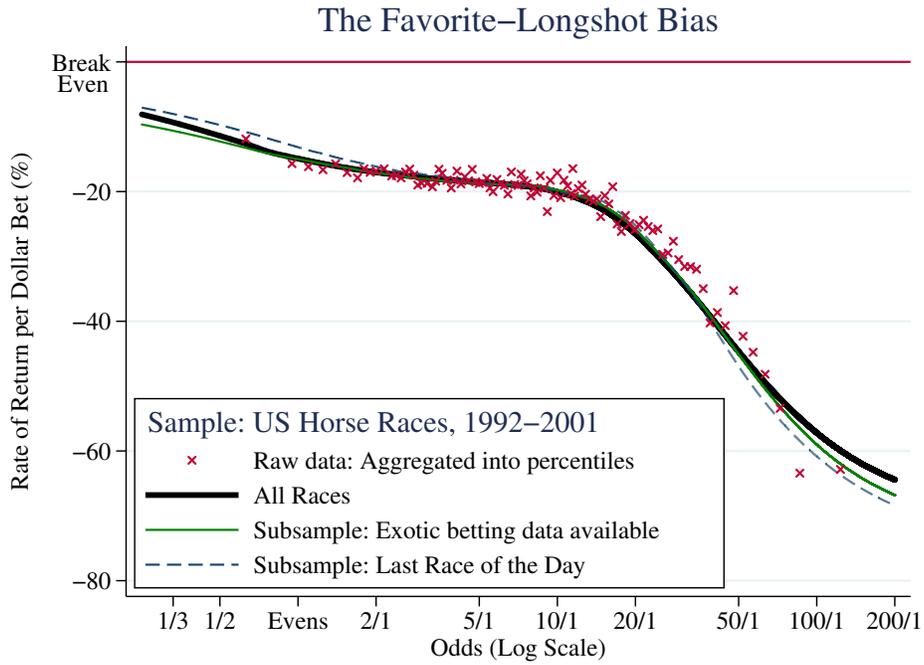
Notes: Standard errors in parenthesis. Predictions and actual outcomes are measured in the price of a contract that pays \$1 if the event occurs, zero otherwise.

Table 3: Robustness to Relaxing Conditional Independence Assumption

<i>Test:</i>	(1)	(2)	(3)
	Absolute Error:  Prediction – Actual	Absolute % Error: $\frac{ \text{Prediction} - \text{Actual} }{\text{Actual}}$	Which Prediction is Closer to Actual? (%)
Panel A: Exacta Bets (n=197,551)			
Risk-Love Model	0.0117	33.7%	42.9%
Misperceptions Model	0.0109	24.4%	57.1%
Risk-Love Error – Misperceptions Error	0.00082 (.00001)	9.3% (.1%)	
Panel B: Quinella Bets (n=70,169)			
Risk-Love Model	0.0240	37.7%	48.7%
Misperceptions Model	0.0235	33.8%	51.3%
Risk-Love Error – Misperceptions Error	0.00046 (.00002)	3.9% (.2%)	
Panel C: Trifecta Bets (n=137,756)			
Risk-Love Model	0.00650	98.0%	30.4%
Misperceptions Model	0.00464	49.0%	69.6%
Risk-Love Error – Misperceptions Error	0.00186 (.00001)	49.0% (.1%)	

Notes: Standard errors in parenthesis. Predictions and actual outcomes are measured in the price of a contract that pays \$1 if the event occurs, zero otherwise.

Figure 1: The rate of return on win bets declines as risk increases.



Notes: Sample includes 5,610,580 horse race starts in the United States from 1992–2001. Lines reflect Lowess smoothing (bandwidth=0.4).

Figure 2: The favorite longshot bias has persisted for over 50 years.

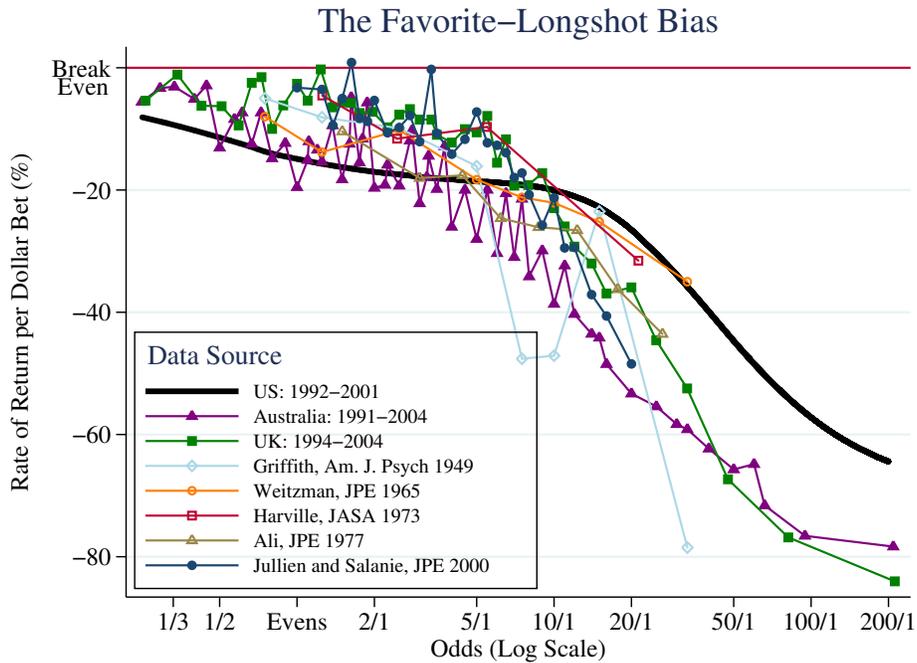


Figure 3: The win data is completely rationalized by both models.

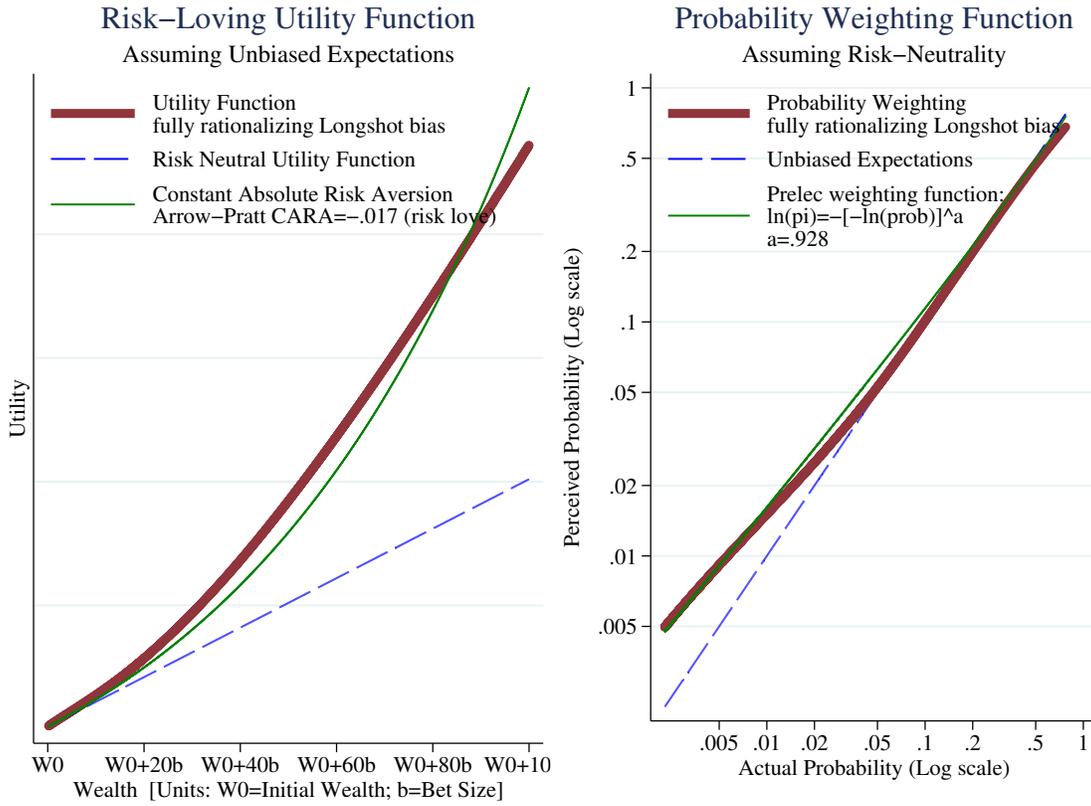


Figure 4: Predicted Exacta Pricing—Risk-Love and Misperception Models

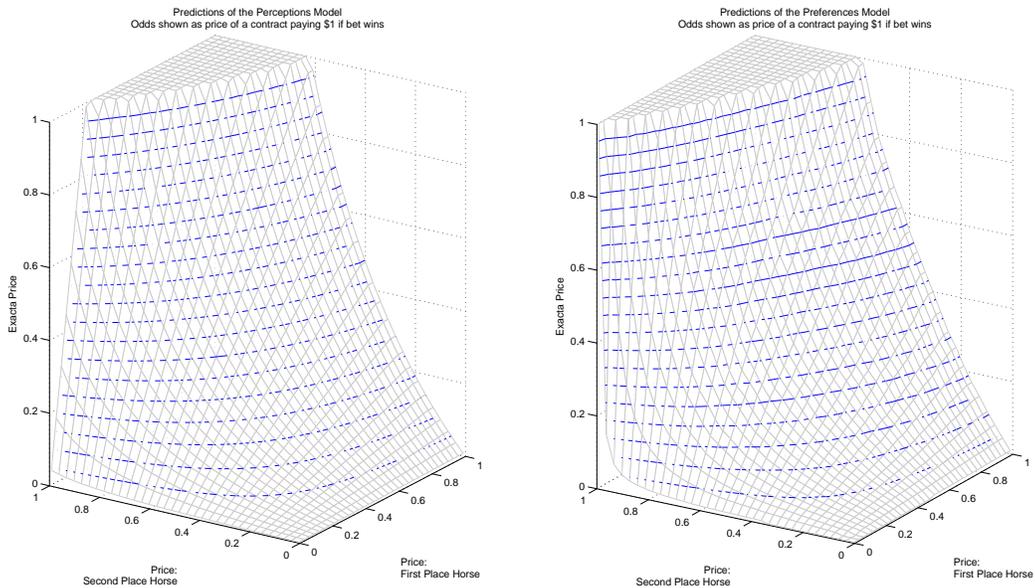
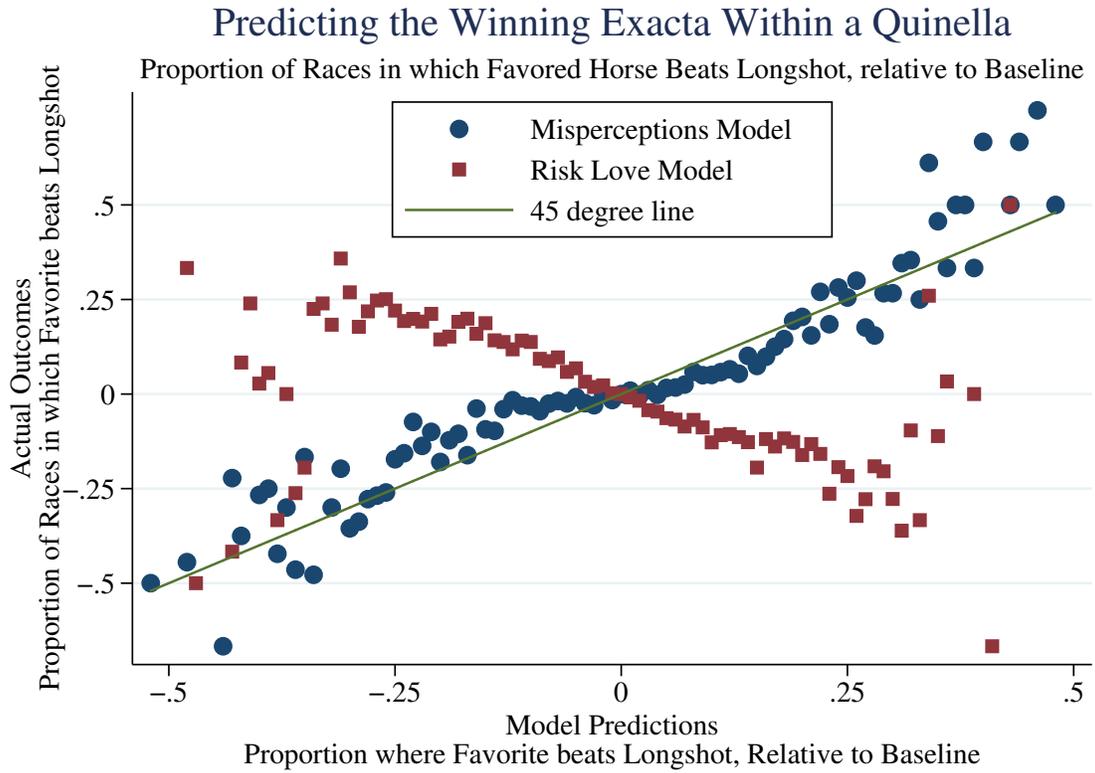
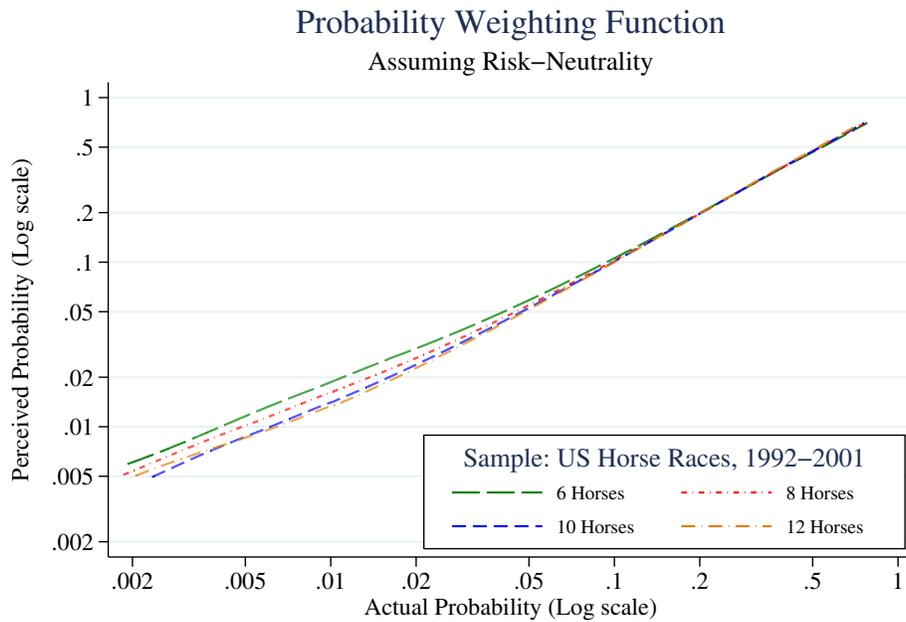


Figure 5: Dropping Conditional Independence



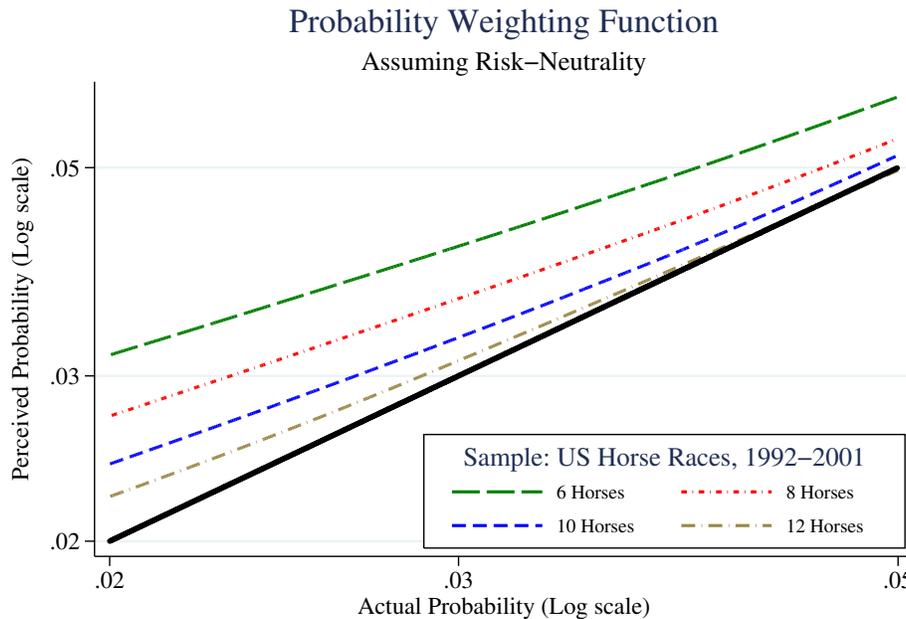
Notes: Chart shows model predictions from (3) and (4) and actual outcomes relative to a fixed-effect region baseline. For the first-two finishing horses the baseline controls for: (a) the odds of the favored horse, (b) the odds of the longshot, (c) the odds of the quinella, and (d) all interactions of (a), (b) and (c). The plot shows model predictions after removing fixed effects, rounded to the nearest percentage point, on the  $x$ - axis and actual outcomes, relative to the fixed effects, on the  $y$ -axis.

Figure B.1: The relation between actual and perceived probability for races with different numbers of horses.



Notes: All lines are generated from data using Lowess smoothing.

Figure B.2: Perceived probabilities are higher for races with fewer horses.



Notes: Thick black line indicates the 45-degree line. All other lines are generated from data using Lowess smoothing.