Toward an understanding of the origins of the favourite-longshot bias: Evidence from online poker markets, a real-money natural laboratory

Leighton Vaughan Williams[a]*, Ming-Chien Sung[b], Peter A.F. Fraser-Mackenzie[b], John Peirson[c], Johnnie E.V. Johnson[b]

[a] *Nottingham Business School, Nottingham Trent University, Burton St., Nottingham, NG1 4BU; Tel: 0115 8486150*
[b] *Centre for Risk Research, School of Management, University of Southampton, Highfield, Southampton, SO17 1BJ, UK; Tel: 02380 592546*
[c] *School of Economics, Keynes College, University of Kent, Canterbury, Kent, CT2 7NP, UK; Tel: 01227 823328*

**Corresponding author: Leighton.Vaughan-Williams@ntu.ac.uk*

This is a post-print version of L. Vaughan Williams, M-C. Sung, P.A.F. Fraser-Mackenzie, J. Peirson and J.E.V. Johnson, 'Toward an understanding of the origins of the favourite-longshot bias: Evidence from online poker markets, a real-money natural laboratory', 2016, Economica, DOI: 10.1111/ecca.12200

Publisher's version can be found at:
http://onlinelibrary.wiley.com/doi/10.1111/ecca.12200/full

Toward an understanding of the origins of the favourite-longshot bias: Evidence from online poker markets, a real-money natural laboratory

Leighton Vaughan Williams[a]*, Ming-Chien Sung[b], Peter A.F. Fraser-Mackenzie[b], John Peirson[c], Johnnie E.V. Johnson[b]

[a] Nottingham Business School, Nottingham Trent University, Burton St., Nottingham, NG1 4BU; Tel: 0115 8486150
[b] Centre for Risk Research, School of Management, University of Southampton, Highfield, Southampton, SO17 1BJ, UK; Tel: 02380 592546
[c] School of Economics, Keynes College, University of Kent, Canterbury, Kent, CT2 7NP, UK; Tel: 01227 823328

**Corresponding author: Leighton.Vaughan-Williams@ntu.ac.uk*

JEL : D03, G14, L83

Key words: Risk, information, efficiency, betting, decision-making.

Toward an understanding of the origins of the favourite-longshot bias: Evidence from online poker markets, a real-money natural laboratory

Abstract

Evidence of differential returns to bets placed with different probabilities of success has revealed a broadly systematic tendency for low/high probability events to be relatively over/under-bet, a phenomenon known as the favourite-longshot bias. While most of the literature focuses on sports, especially horse racing, we report here the existence of the same phenomenon in online poker games. We find that misperception rather than risk-love offers the best explanation for the behaviour we identify. The paper contributes to the more general literature explaining betting behaviour as well as the prevalence of the favourite-longshot bias in betting markets.

I. INTRODUCTION

I.I Existence of the favourite-longshot bias

Evidence of a differential return to bets placed with different probabilities of success can be traced to laboratory experiments by Preston and Baratta (1948), Yaari (1965) and Rosett (1971). They each reported evidence of a systematic tendency by subjects to under-bet or under-value high-probability events and to over-bet or over-value low-probability events. Tests for the existence of such a bias in non-laboratory conditions can be traced to Griffith (1949), who investigated the 'pari-mutuel' betting markets characteristic of US racetracks. In these markets, winning bets share the pool of all bets. Griffith confirmed a tendency for bettors to over-value events characterized by low probability ('longshots') and to relatively under-value events characterized by higher probability ('favourites'), a tendency which is consistent with higher expected returns at lower odds than at higher odds. This behavioural tendency has become known as the 'favourite-longshot bias' (FLB), and has since been observed across a range of betting markets in a range of contexts (Vaughan Williams, 1999; Snowberg and Wolfers, 2010; Ottaviani and Sorensen, 2008, survey the literature), while Busche and Hall (1988) and Busche (1994) highlight some exceptions to the general pattern.

Evidence for the existence of the FLB tends to be derived from observation of the bias in sports betting markets, and horse race betting markets in particular. In this paper, however, we examine whether a similar bias exists in a quite distinct and different betting market, i.e. online poker. If we can identify the bias in this arena, this provides a complementary context in which we can investigate this phenomenon.

We use a very large data set of rounds of betting from an online poker website that collects this information for Texas Hold'em poker. While this setting does not appear to resemble racetrack or sports betting, we demonstrate that data drawn from online poker games can be used to test hypotheses about pricing behaviour even though no traditional 'market prices' exist. However, poker is effectively a pari-mutuel game where the pot of all player contributions is distributed (minus operator commission, called the 'rake') from losing players to the winner of each game. We use the fact that the later in a round of betting the player is required to make betting decisions, the more information is available from the bets already placed by other players. Similarly, since there are multiple stages of the game (called the preflop, flop, turn and river stage), where more cards are revealed and more bets have been placed, more information is revealed as the game proceeds. Within a round of betting, the informational advantage should result in the expected return to bets placed later in the round exceeding the expected return to bets placed earlier in the round (the last to play is the equivalent to the 'favourite' in a racetrack betting setting). 'Over-betting' by a player in a round of betting can now be interpreted, when assessed over the entire data set, as betting more than is warranted in terms of the distribution of expected returns at different points in the game. In particular, the amount bet later in the round of betting should on average be higher than earlier in the round.

If we find that the amounts bet in given positions in the round of betting do not conform to the expected returns to bets placed in those positions, this is evidence of a behavioural bias, and potentially a FLB. The FLB exists if players on average under/over-bet when acting later/earlier in the round relative to the expected return to be earned in that position.

We develop a theory of how perfect and behavioural/recreational players play poker. We use the actual behaviour of players in 'folding', 'calling' and 'betting/raising' to disaggregate the large data set into bets on outcomes that have small, moderate and high probabilities of occurring. This allows us to test a rich range of hypotheses on betting behaviour that strongly support the existence of a FLB amongst behavioural/recreational players relative to that of more perfect/informed players. These results are highly statistically significant and we demonstrate that this has a significant economic impact on the recreational players relative to the more informed group.

I.II Explanations for the favourite-longshot bias

A number of explanations have been proposed to explain the existence of the FLB. The main theoretical explanations for the bias have been reviewed and arranged by Ottaviani and Sorensen (2008) into seven convenient categories. These are misestimation of probabilities (Griffith, 1949; Snowberg and Wolfers, 2010), market power by informed bettors (Isaacs, 1953); preference for risk (Weitzman, 1965; Quandt, 1986), heterogeneous beliefs (Ali, 1977; Gandhi and Serrano-Padial, 2015), market power by uninformed bookmakers (Shin, 1991, 1992), limited arbitrage (Hurley and

McDonough, 1995), and simultaneous betting by partially informed insiders (Ottaviani and Sorensen, 2009, 2010).

In this paper, in addition to demonstrating the existence of the FLB in online poker, we seek to arbitrate between two widely cited broad classes of explanation for the bias: 'risk preferences' and 'misperceptions'. In particular, we determine which of these two explanations is most consistent with our data. To this extent, one of our research questions is similar to that asked, for example, by Jullien and Salanié (2000), Gandhi (2007) and Snowberg and Wolfers (2010), who each attempt to differentiate between preference and perception-based explanations of the FLB, but in their cases using racetrack betting data.

It is important to note here that misperceptions can refer to a number of different possibilities and in subsequent analyses we seek to arbitrate further between two different types of misperception. The first type of misperception arises from incomplete processing of information and errors in the processing of such information and leads to heterogeneity of the beliefs about the probabilities of winning hands. The more complex the information, the more heterogeneity of beliefs we would expect to observe and therefore the more variation in the estimation of the probabilities of winning hands.

Second, experimental evidence suggests that non-linearity in the probability weighting function of individuals may exist and this is captured in the general literature on cumulative prospect theory (see Tversky and Kahneman, 1992). This assumes that there are systematic biases in perception leading to under-weighting of high probability events and over-weighting of low probability events. The degree of non-linearity is not related to the amount of information and so is not expected to vary with choice complexity. Such ideas have been applied in the betting context (see, for example, Jullien and Salanié, 2000; Gandhi, 2007; Chiappori et al., 2009, 2012). It should be noted that these types of misperception are not mutually exclusive. It is possible that there may exist systematic biases in perception as well as random noise in perception as a result of information complexity.

These differences are noted by Gandhi and Serrano-Padial (2015), where the equilibrium prices predicted by a heterogeneous beliefs based model will be different for two horse races with the same fundamentals but different public information, while the predictions yielded by a probability weighting function based model would be the same. They test this at the US racetrack using different types of races characterized by different levels of information, and show that a beliefs-based model provides a significantly better explanation of what is observed than probability weighting function based theories.

These misperceptions can be distinguished in explaining the FLB from an explanation based on risk-loving behaviour (see Weitzman, 1965; Quandt, 1986). Based on this interpretation, individuals are aware of the true probabilities but act according to their preferences which, using the traditional description, we assume to follow a convex value function. This should result in a preference for risky alternatives.

The poker data provides us with the rare opportunity of examining the existence of FLB and undertaking tests to arbitrate between the risk-love and different types of misperception as possible explanations in a real world context. In attempting this, we are seeking to address a concern highlighted by Pope and Schweitzer (2011, p.129): "Critics of the decision bias literature believe that biases are likely to be extinguished by competition, large stakes and experience." This idea echoes Barberis (2013) and Levitt and List (2008, p.909),: "Perhaps the greatest challenge facing behavioral economics is demonstrating its applicability in the real world. In nearly every instance, the strongest empirical evidence in favour of behavioural anomalies emerges from the lab. Yet, there are many reasons to suspect that these laboratory findings might fail to generalize to real markets." The remainder of the paper is structured as follows: We introduce the nature of 'Texas Hold'em' and the theory behind our hypotheses in section II. The data and analysis is described in section III. Results are presented and discussed in section IV, and some concluding remarks follow in section V.

II. IDENTIFYING FAVOURITE-LONGSHOT BIAS IN POKER MARKETS

II.I "Texas Hold'em" – the nature of the game.

Texas Hold'em can be played by different numbers of players but our dataset consists only of six player games. Consequently, throughout our paper, the analysis is based on six player Texas Hold'em games.

Before any cards are dealt in a Hold'em hand, two players post forced bets of a predetermined amount, called 'blinds'—a 'small blind' by the player to the dealer's immediate left and a 'big blind' by the next player to the left . The small blind is equal to one-half the amount of the big blind in our dataset. The deal rotates after each hand so that all players participate equally in the posting of any forced blind bets.

Once the blinds have been contributed to the pot, each player is dealt two cards face down, followed by a round[1] of betting (the *preflop*). After this first round of betting, three (community) cards available to all the players (the *flop*) are dealt face upright, followed by a second round of betting. After this round of betting a fourth card (the *turn*) is made available to all players by being dealt upright on the table, followed by another round of betting. A final community card (the *river*) is then dealt upright, and this is followed by a final round of betting.

[1] This and any other betting stage can involve multiple rounds of betting if players keep raising instead of simply calling. For simplicity, we just refer to such sequences of betting during a given stage as a "round" of betting.

In the first round of betting, the player immediately to the left of the big blind is the first to act, and so on leftward around the table. In subsequent rounds, the small blind plays first, followed by the other players in order, with the 'Button' playing last. When it is a player's turn to act, they may choose to *bet* (choose an amount to bet on the hand if there has been no previous bet made in the round—an integer multiple of the big blind), or *call* (stake an amount equal to the current outstanding bet) or *raise* (match the current outstanding bet and stake additional money) or *check* (pass without betting, which is allowed if there is no current outstanding bet) or *fold* (withdraw from the hand, leaving all previously wagered money in the pot). Betting moves around clockwise until each player who is still in the round contributes the same amount into the pot, or until there is only one player left active. In that case, the remaining player wins the pot, and has the right to keep their hidden cards secret. If more than one player remains in the hand after the final round of betting (river stage), there is a *showdown* and the player with the best poker hand—formed by using any combination of the five community cards and the player's two hidden cards—wins the pot. In the event of a tie, the pot is divided.

II.II Identifying the existence of FLB

In the context of poker, we argue that a FLB exists if we find evidence of over-betting on hands with a low probability of winning and under-betting on hands with a high probability of wining, to a degree that does not maximise expected value. Consequently, in order to demonstrate the existence of the FLB, we need to show that: (1) some hands have a lower probability of winning than others; (2) those hands with a higher probability of winning are under-bet and those with a lower probability are over-bet; and (3) players demonstrating element 2 are not maximising expected value. These components are developed below into three hypotheses (H1.1- H.1.3) which are tested to confirm that these elements are present in online poker.

II.II.I Seating Position-Based Favourite-Longshot Bias

In poker, new information is presented to each player at each stage of the game (Preflop, Flop, Turn, River) and within the betting at each stage. This occurs as each player in each seating position (Early, Middle, Cutoff and Button[2]) sequentially either raises, calls or folds. The sequential nature of the game means that players who make betting decisions later in a given round (indicated by seating position) or at later stages of the game (i.e. the river and turn compared to the flop and preflop stages) have more information available. For example, they know whether earlier players have folded or if earlier players have contributed a particularly large amount to the pot.

We first explore whether some seating positions do indeed have an informational advantage. In particular, we examine whether the earliest position (i.e. Early) has lower win rates (i.e. a longshot) compared to the latest seating position (i.e. Button), which should have the greatest win rate (i.e. the favourite). Consequently, we test **Hypothesis 1.1**: ***There is a significant increase in win rate from the Early to the Button seating position*.**

II.II.II Recreational vs More Informed Players

In order to determine whether the players in a poker game are betting in a manner which consistently under/over-bets in favourable/unfavourable seating positions (i.e. demonstrating a FLB), we need a benchmark against which to assess optimal staking. To achieve this, we assume that informed players play perfectly. Consequently, we assume that the informed player has, at each stage of the game, a subjective probability

[2] Given that the small blind and big blind are forced to contribute to the pot, we exclude these seating positions from the analysis. These blind positions provide the initial liquidity for each hand.

density function over three dimensions: new and future contributions that they will make to the pot (c), size of final pot (p) and probability of winning (q). This subjective probability density function $\phi(c, p, q)$ which refers to the probabilities and density associated with one particular hand of cards, is revised with each piece of new information. Their expected profit is thus given by:

$$E(\pi) = \int_c \int_p \int_q \phi(c, p, q) \left[q * p - c \right] dq \, dp \, dc \qquad (0)$$

Each piece of information is valuable as it changes the subjective probability density function. Typically, but not always, the change in $\phi(c, p, q)$ reduces the range of possibilities that have to be considered in constructing the new $\phi(c, p, q)$. For example, if the previous player folds, this reduces the range of future possibilities. The decision that a player has to make at each stage is whether to fold, call or raise and, in the latter case, how much to raise. The action that gives the greatest new expected profit determines the preferred decision.

We assume that the informed player has a subjective probability density function that uses the available information to the greatest possible extent. By comparison, the recreational player's subjective probability density function suffers from biases caused by poor information processing skills, and cognitive bias as predicted by Prospect Theory, many other behavioural theories and experimental evidence (e.g., see Tversky and Kahneman, 1992). These two possible causes of biases relate to the two possible types of misperceptions discussed earlier. The informed player is better at processing and using new information and better at making decisions. Thus, they are assumed to maximise expected profit. As such, we expect that their contributions to the pot should reflect the informational advantage based on the seating position.

We assume, in line with the evidence presented in Fiedler (2011) and Fiedler and Rock (2009), that most online poker players are recreational players and these generally play in small blind games, whereas those involved in higher stake games (i.e., in larger blind games) tend to be more sophisticated players. We expect these larger blind players to be less prone to decision biases in the face of the increased information (see, for example, Sonsino, Benzion and Mador, 2002), which occurs through sequential rounds of betting.

Clearly, it is possible to argue that low stakes games may be populated by lower income individuals, and that income effects may play a role, but an examination of the gambling habits of online poker players by Fiedler (2011) suggests instead that a more useful way of framing behavioural differences between higher and lower-blinds players is in terms of player experience and sophistication. Based on this view, we aim to determine the degree of under/over-betting by recreational, lower stake, players by comparing the contributions in low stake games in each seating position against that of the higher stake games containing relatively more informed players. Consequently, we test **Hypothesis 1.2**: ***Recreational players have a lower/higher contribution rate when in the Button/Early seating positions relative to the contribution rates of the more informed players.***

Assuming that the evidence supports H1.1 and H1.2, the final piece of evidence required to demonstrate that this is indicative of a FLB is that the recreational players should attain a lower rate of return than the more informed players in the initial seating positions[3] and recreational players should obtain higher rates of return in the later

[3] Note that we do not expect all the higher-stake players to exhibit perfect betting behaviour. However, we still expect them to more closely reflect the true informational advantage of seating position and therefore attain higher rates of return compared to recreational players.

seating positions. We, therefore, test **Hypothesis 1.3:** ***Recreational players have a lower rate of return in the earlier seating positions relative to that of more informed players and a lower rate of return compared to the returns for recreational players in the later seating positions.***

Our results, as indicated below, confirm hypotheses 1.1-1.3, so we next explore the observed FLB in more detail.

II.II.III Explaining the Favourite-Longshot Bias: Risk Love vs. Misperception

To examine the degree to which the FLB is displayed in the actions of poker players and to examine its causes, we undertake a more disaggregated analysis that examines betting on different types of poker hands. In particular, we first explore, for preflop card hands, whether there is under/over-betting on higher/lower probability hands when the return is higher (and positive)/lower (and negative).

It is important to examine the preflop stage when testing our hypotheses for a number of reasons. In particular, at this stage, all players are present, only the blind amounts have been contributed to the pot, and each player has only two cards. As such, the probability mix is identical for both the higher stake games and the lower stake games. At later stages, some players may have folded and so the number of players varies between games, the players will have contributed different amounts relative to their stack, more community cards will have been turned over, etc. Therefore, from an experimental perspective, the preflop stage is naturally more controlled than later stages.

For this reason, players' decisions at the preflop stage are likely to be closely related to the quality of the cards they are dealt and seating position, and we can examine their responses (i.e. fold, call and raise) to determine whether the hands are of poor, medium or high quality. Therefore, the preflop stage provides us with the best chance of differentiating those with high vs. low quality hands on the basis of their responses. The reason being that the decision to fold, call or raise at this stage of the game depends on the player's probability of winning with the hand they are dealt and the informational advantage afforded by their seating position. This is also the only stage in which all players are guaranteed to be present, as subsequently any player may have folded. Furthermore, the decision to fold, call or raise in later stages of the game (i.e. flop, turn and river) will depend increasingly on the additional information gained from observing the behaviour of other players.

It is assumed that informed players attempt to maximise expected profit using their subjective probability density functions. In order to represent this behaviour schematically, Figure 2 illustrates a stylized example of the probability density function for different hands in ascending order of their probabilities of winning. In Figure 3, we contrast stylised subjective relative frequency distributions of the recreational[4] and more informed players that ought to result in a FLB for recreational players. For a FLB to occur, the recreational players must over-bet low probability outcomes. Consequently, the recreational players' relative frequency function will display a greater probability mass to the right for low probability events (cf. the more informed players' true probability density function). In addition, the recreational players are likely to under-bet high probability outcomes (cf. the more informed players) and consequently there is a lesser probability mass to the right for high probability events (cf. the more informed player's true probability density function).

An explanation for this shape of the probability density function of the recreational player can be predicted by cognitive bias in the form of a non-linear probability weighting function (i.e. Prospect Theory). The critical probabilities of

[4] The subjective probability density function for the recreational player includes a non-linear probability weighting function.

winning that ensure positive expected profit from calling and greater expected profit from raising are represented by the two vertical lines (q_{FC} and q_{CR}), and the associated probability masses are denoted by the letters F, C and R. These lines and areas are represented schematically in Figure 3[5]. The vertical lines are the same for the recreational and more informed players on the assumption that they each attempt to maximise expected profit using subjective probabilities.[6] It is clear from the distribution for the recreational players that, because high probabilities ('raise situations') are underweighted, they will tend to call when they should raise. Equally, and because they overweight low probability events ('fold situations'), they will tend to call when they should fold. Overall, therefore, we should expect higher call rates and lower fold and raise rates amongst the recreational (cf. more informed) players.

The heterogeneity explanation of a FLB considers the information processing abilities of recreational and informed players. With no information at all, the probability of winning in a six person game is simply one sixth. Information on seating position, quality of the preflop cards and the observed behaviour of rivals allows the informed player to generate a more informed opinion of the likelihood of winning. By comparison, a less informed, recreational player with the same information would move less far from the original estimate of one sixth. In addition, the recreational player may exhibit cognitive biases in their information processing and a degree of random error. Heterogeneity exists in such players' beliefs from three possible sources: the degree of use of new information, the degree of bias in judgements based on new information and sheer random error in judgements. The incomplete use of information by recreational players and cognitive bias (as described in the previous paragraph)

[5] Note that both q_{FC} and q_{CR} do not necessarily have to be above the cross-over points of the density functions, as shown. It is equally possible that one or both could be below and the conclusions would still hold.
[6] It should be noted that moving from calling to raising cannot be exactly represented on the two dimensional figure as the pot and additional contributions are likely to alter with the action of raising.

are likely to lead to less folding and raising by such players. An independent random error in the use of information by less informed players will lead to a more widely dispersed frequency distribution with more probability mass in the tails. We believe that such a random effect is likely to be less important than the incomplete use of information and cognitive bias effects. We come to this view as the heuristics and bias literature demonstrates the importance of incomplete use of information and cognitive bias effects as explanations of behaviour, with less emphasis being placed on random errors (Kahneman, Slovic and Tversky, 1982). On the assumption that random errors are less important, we predict less folding and raising by recreational players.

Importantly, these patterns are not what we ought to expect given greater risk love by recreational players. Risk love predicts an aversion to folding and a preference to call but also a preference to raise since we assume that risk love simply pushes the critical probabilities (q_{FC} and q_{CR}) to the left. Therefore, we can directly test risk-love vs non-linear weighting function explanations by comparing two hypotheses.

The risk-love prediction is as follows, **Hypothesis 2.1:** ***At the preflop stage, there is less folding and more calling and raising by recreational (cf. more informed) players.***

The misperception prediction is as follows, **Hypothesis 2.2:** ***At the preflop stage, there is less folding and less raising but more calling by recreational (cf. more informed) players.***

Since both Hypothesis 2.1 and 2.2 have identical predictions for calls and folds, our main evidence for risk-love vs. misperception is whether or not recreational players raise more (Hypothesis 2.1) or raise less (Hypothesis 2.2) than the more informed players.

Our results actually indicate that Hypothesis 2.2 is true and based on this we can go further to differentiate the two types of misperception explanation. It would be expected that, for each type of game, the rate of return would be higher for more informationally favourable positions. In addition, since we expect recreational players to call when they should fold, the rate of return from calling may be negative in the Early position and then increase across superior positions. This effect may exist in both types of games but more weakly in the higher stakes games as it is populated by relatively fewer recreational players.

Importantly, however, if recreational players call when they should raise because of a non-linearity in the weighting function, then average rates of return in lower stake games from raising may actually be higher than otherwise and, in particular, in comparison with higher stake games. If, on the other hand, heterogeneity of beliefs is the explanation then recreational players ought to process information less accurately in general than informed players and make random mistakes. As such, they may raise inappropriately on some card hands and this would reduce the average rate of return. Therefore, only if recreational players' average rates of return, on those few favourites that are raised, are less than those of informed players then this would support the view that the recreational players make random heterogeneous probability mis-estimations that cause these implied losses.

We test the heterogeneity of beliefs explanation in **Hypothesis 2.3:** ***Recreational players who have raised at the preflop stage will achieve lower rates of return (cf. more informed players) on those raises.***

In summary, risk love is a parsimonious explanation only if Hypothesis 2.1 is true whereas misperception is a better explanation if Hypothesis 2.2 is true. If we also find that Hypothesis 2.3 is true, then we find evidence in favour of heterogeneity of beliefs occurring.

II.III Origins of the Favourite-Longshot Bias: Further examination of the impact of new information on misperception.

To arbitrate further between different misperception explanations for the FLB, we compare a key measure of voluntary contribution (bets and raises) to the pot at various points in the cycle of betting rounds. In doing so, we examine how the impact of additional information affects the risk-taking since as the game progresses from the preflop to the river stage, more complex information is revealed and the uncertainty is converted into more clearly defined risks for the informed players that are able to process this complex information.

It is important, here, to make a distinction between risk and uncertainty/ambiguity and how these elements change during the course of a game of poker. At the preflop stage, little information has been revealed beyond seating position and hand quality. Hence, the final risks are not so clearly defined and uncertainty/ambiguity is at its highest, but the new information load for players is low. At later stages, more information is revealed which increases the complexity of the decision for players, but much of the uncertainty is resolved by informed players, revealing more clearly defined risks.

The change in risk over the course of the game is difficult to determine. On the one hand, the number of players will tend to decline, and correspondingly, the win rate at each position will tend to increase, as the rounds of betting progress (i.e. as some players will fold). Consequently, risk, measured by the probability of losing, will tend to be higher in the earlier rounds. However, the potential pot size increases throughout the game and so the potential monetary loss also increases.

What is certain, however, is that the advantage of the Button seating position compared to the Early Position should continue to exist. We can determine this by examining the ratio of voluntary pot contributions in the Button compared to the Early position as the game progresses.

We discuss above that the non-linear weighting function explanation is not expected to depend on new information (Gandhi & Serrano-Padial, 2015). Therefore, if non-linear weighting functions explains behaviour, then the degree of bias exhibited by recreational players at the preflop stage compared to more informed players should be similar to the degree of bias compared to more informed players at the river stage. In essence, the degree of FLB between the Button and the Early position should be the same throughout the game. This is tested in **Hypothesis 3.1** ***The degree of under-betting the Button and over-betting the Early position by recreational players compared to the more informed players will remain the same throughout the game.***

However, unlike non-linear weighting functions, heterogeneity is expected to increase with the amount/complexity of information. This can be the result of increased complexity which, through the need to deal with more information (as experienced in the latter stages if the game), has been shown to lead to proportionally less exhaustive analysis and a propensity to use heuristics and biases in forming judgements (see, for example, Katsikopoulos, 2011; Kahneman, Slovic and Tversky, 1982).

Therefore, we should expect greater heterogeneity with regard to incomplete information processing at the river stage compared to the preflop stage. This element of heterogeneity for recreational players should lead to greater perceived uniformity in information and assessment of probabilities and, thus, pot-contributions between the Button and Early positions, as the game progresses. This could, therefore, cause an increase in the FLB since it would result in under-betting seating positions that should be raised more (notably the favourite, Button position) and over-betting positions that should be raised less (notably the longshot, Early position).

This is tested in **Hypothesis 3.2** ***The degree of under-betting the Button and over-betting the Early position by recreational players compared to the more informed players will increase through the stages of the game as recreational players adjust less to the information advantage afforded by seating position.***

In summary, if non-linear weighting functions explain the FLB, then the degree of bias relative to more informed players should be the same throughout the game (Hypothesis 3.1). However, if heterogeneity plays a role in the FLB, then as more potential information is revealed at each stage of the game, then the degree of bias should increase (Hypothesis 3.2).

II.IV Economic Significance

To establish the importance of any observed FLB amongst recreational players, it is important to establish the economic impact of this bias. Consequently, we determine the differential in the returns of recreational and more informed players in each seating position. If the FLB causes differences in these returns, then we would expect that those players who bet more/less on the longshot (Early position) than the favourite (Button position) should have a worse/better return. Consequently, we determine the relative returns of players who bet more and less in accordance with the FLB in order to assess the penalty imposed by any observed FLB.

III. DATA AND PROCEDURES

III.I Datasets

We employ a data set of 19,171,284 hands from a real-world natural laboratory, namely online real-money 'Texas Hold'em' poker games. The data was purchased from www.hhdealer.com and comprises 12,466,677 hands at the $0.5/$1 blind level ('more recreational' players), 6,704,607 hands at the $5/$10 blind level ('more informed' players), from the period 10th August 2007 to 29th May 2011.

The numbers of players in poker games may vary but we employ data from the popular variant involving only six players. Those players seated in the 'Blind' positions are forced to make fixed contributions to the pot. Consequently, they are unable to regulate their initial bet based on the information advantage they may possess. In addition, unlike the 'non-Blind' positions, they are required to act at different points in the betting cycle depending on the phase of the betting (for example, the small blind acts next to last before the 'flop', but first after the 'flop'). As a result, we exclude these players from our analysis, but note that, as poker is a nearly zero sum game, our results below indicate that most non-Blind positions win money and the Blind seating positions are those that tend to lose money.

In order to test our hypotheses, we largely employ data from the preflop stage. We focus on this stage because it is naturally more controlled, from an experimental perspective, than later stages (see the reasons outlined in Section II.II.III).

III.II Variables

III.II.I Win Rate

The win rate in a given position is defined as the total number of games won in that position divided by the total number of games played, e.g. for the Early position 1,499,063 games are won from 12,466,677 games played, leading to a win rate of 12.02%. In addition, we define the 'win rate ratio' as the win rate in a position divided by the win rate in the Early position (the position with the least information advantage). For example, the win rate ratio for the Middle position is the win rate for that position (13.40%) divided by the win rate in the Early position (12.02%) giving a measure of 1.11.

III.II.II Contribution Rate

We define contribution rate as the percentage contributed in a given position divided by the total contributed in all positions. For example, the total contributed to the pot by all players in the 'Early' position, in games involving recreational players ($0.5/$1 blind games), was $33,038,014.10, and this was compared to the total amount contributed to the pot in all positions ($258,663,676.75); thus the market pot contribution for the 'Early' position was determined as 12.77%. We also assess the extent to which the market pot contribution in a given position exceeds the market pot contribution in the position with the least information advantage (i.e. the Early position), by calculating the ratio of the market pot contributions in these two positions: the 'pot contribution ratio'. For example, the market pot contribution in the Middle position was 13.90% and this divided by the market pot contribution in the Early position (12.77%) yields a pot contribution ratio of 1.09. Finally, we define the 'ratio difference' as the difference between the pot contribution ratio and the win rate ratio.

III.II.III Rate of Return

We define the 'return rate' for a position as the total net return for all players in that position divided by the total contribution to the pot by all players in that position. A player's 'net return' in a given game is defined as the total returned to the player from the pot in a game, minus that player's total contribution in that game (including 'the rake'). The rake is a transaction cost whereby the game host (casino/website) takes a specified percentage of the pot. For example, in a game with a rake of 5 per cent, where two players contribute $10, $10, and the rest $0, the player who wins the hand wins the pot and receives a net return of $20 minus his/her contribution to the pot and minus the $1 he paid for the rake. The losers make a negative net return equal to their contribution to the pot. By combining the results of all games in the sample we are able to assess the net return across all players in each position.

III.II.IV Preflop Actions

The dataset enables us to determine the actions taken by players at each stage of the game (preflop, flop, turn, river). Since there may be multiple actions in a given round, we define a 'preflop fold' as any string of actions at the preflop stage which ended in a fold, a 'preflop raise' as any string of preflop actions that did not end in a fold (i.e. continued to at least the flop stage), and had at least one raise action, and we define a 'preflop call' as any string of preflop actions that did not end in a fold and did not include any raises.

III.II.V Percent Voluntary Put in Pot (%VPIP)

The %VPIP is the percentage of games which reached a particular stage (i.e., the preflop, flop, turn and river) in which a voluntary contribution is made to the pot from a given position (excluding, where someone is *required* to bet a given amount when they are sitting in a Blind position or *reactive* bets, where someone calls to remain in the game). We use %VPIP to measure the degree of proactive risk taking, as it only accounts for bets, raises and re-raises.

III.II.VI. Blind Adjusted Net Return

In the larger and smaller blind games, players in the large blind position pay $10 or $1, respectively. Consequently, in order to make comparisons between the average net return per game in these two types of game, we divide the average net return per game by the blind levels ($10 and $1 respectively).

III.3 Bootstrapping

Our analysis is undertaken at an aggregate level and it was therefore necessary to estimate the statistical significance of aggregated contribution rates, the ratio of contributions in seating positions relative to the Early position and aggregate return rates. To achieve this, we employed bootstrapping to estimate the standard errors of these aggregate statistics. Due to the large dataset (19,171,284 hands each of which has 6 seating positions resulting in 115,027,704 database rows), in the interests of processing efficiency, we adopted *m* of *n* bootstrapping to test hypotheses 1.2 and 1.3 (Santana, 2009).

Standard bootstrap procedures involve: (1) Taking a bootstrap sample of *n* independent observations, $X_1, X_2, X_3, \ldots, X_n$, with replacement, from the population, where n is the size of the population; (2) calculating the relevant statistic $\theta_n = \theta_n(X_1, X_2, X_3, \ldots, X_n)$ for the generated sample; (3) independently repeating steps 1 and 2, B times. The resulting B replications of the statistic ($\theta_{n1}, \theta_{n2}, \theta_{n3}, \ldots, \theta_{nB}$) are then used to estimate the statistic's sampling distribution. The mean of the replications is the estimate for the statistic $\hat{\theta}_n = \theta_{n1}, \theta_{n2}, \theta_{n3}, \ldots, \theta_{nB}$ and the standard error of the statistic $\hat{\theta}_n$ can be expressed as (Santana, 2009);

$$SE(\hat{\theta}_n) = \sqrt{Var(\hat{\theta}_n)} \qquad (2)$$

For the *m* of *n* bootstrap, rather than take a bootstrap sample of size *n* we take a smaller sample, with replacement, of size *m* based on a scaling parameter, *k*, where $m = n^k$ (in our case we used k = .7). For this the standard error is calculated as follows;

$$SE(\hat{\theta}_n) = \sqrt{\frac{m}{n} Var(\hat{\theta}_n)} \qquad (3)$$

The disaggregated analysis we performed to test hypotheses 1.1, 2.1, 2.2, 3.1 and 3.2 involved smaller subsets of data. This enabled us to employ full *n* bootstrapping for the relevant statistics.

IV. RESULTS

IV.I Seating position favourite-longshot bias

The results displayed in Table 1 indicate that the win rate increases monotonically from the Early position (around 12%) to the Button position (around 18%) for games involving both recreational and informed players. These results confirm Hypothesis 1.1 and suggest that later seating positions have an information advantage.

The results also demonstrate that pot contributions of recreational and more informed players echo the monotonic win rate pattern over seating positions, with monotonically increasing pot contributions (significant at .01% level) for every position from the Early to the Button. However, the recreational players contribute significantly less (cf. more informed players) in the 'favourite' position (Button) and significantly more in the longshot position (Early) ($p < .001$). These differences in the behaviour of recreational and informed players are highlighted by the contribution ratio statistics in Table 1. These reveal that informed players contributed 1.62 times more to the pot in the Button compared to the Early position, whereas this ratio for recreational players is significantly less (1.39, $p < .001$). Taken together, these results support Hypothesis 1.2 that recreational players have a lower/higher contribution rate when in the Button/Early seating positions relative to the contribution rates of the more informed players.

The contribution ratios are compared to the win rate ratios used to capture informational advantage (see Table 1). For example, the win rate ratio for informed players in the Button position (1.52) is significantly greater than that for recreational players (1.47, p<.001). Comparing the ratio difference between the win rate ratio (WR) and the contribution rate ratio (CR), we observe that the Middle, Cutoff and Button positions, have ratio differences (CR-WR) which are negative for the recreational players and positive for the informed players. Bootstrapping estimates reveal that the CR-WR statistics are significantly lower for recreational players than informed players ($p < .001$ at each position). Further evidence that this behaviour is indeed a FLB is indicated by the fact that the recreational players have lower rates of return in every position compared to the informed players ($p < .001$) and have lower rates of return in the initial seating positions compared to the later positions for recreational players ($p < .001$) . This result confirms Hypothesis 1.3. It is interesting to note that the rates of return for the informed players are higher despite playing against other similarly informed players.

Additional evidence indicative of a FLB is provided by the results displayed in Table 2. In particular, these results demonstrate that, by comparison with informed players, recreational players over-bet weak hands (i.e. those likely to have a low probability of winning) by over-calling and under-bet strong hands (i.e. those likely to have a high probability of winning) by under-raising. For example, informed players in the Early seating position fold 81.8% of the time, raise 15.6% of the time and call 2.6% of the time whereas recreational players in the Early seating position fold 80.2% of the time, raise 12.9% of the time and call 6.8% of the time. The same pattern of differences between recreational and informed players is observed in all seating positions and these differences are statistically significant ($p < .001$) in all cases.

IV.II Risk-Love vs. Misperception

The results displayed in Table 2 support Hypothesis 2.2 rather than Hypothesis 2.1. In particular, recreational players fold and raise significantly less often across all seating positions than more informed players ($p < .001$ for all seating positions). This indicates that recreational players call more often than more informed players across all seating positions. Taken together the evidence suggests that misperception rather than risk love is the explanation for the FLB.

The results displayed in Table 2 also indicate that the percentage of games called or raised at the preflop stage increases monotonically with seating position for both recreational and informed players ($p < .001$). In addition, the results displayed in Table 3, indicate that the recreational players' rates of return are significantly lower (cf. more informed players; $p < .001$) in all seating positions for games called and raised preflop. Notably, the rates of return in all seating positions for recreational players in games which they call in the preflop are all negative. By contrast, the more informed players are profitable when calling in all seating positions, other than the Early which has the greatest information disadvantage. These results suggest that the lower rates of return achieved by recreational players result from their sub-optimal betting decisions.

It is of interest to disaggregate players' hands into call and raise hands. This allows investigation of high probability events (e.g. when a player raises) and the associated rates of raise/call frequencies, win and return and compares these results between informed and recreational player games. Table 4 shows the results of this disaggregation for raise hands at the preflop stage and shows lower raise frequencies but also lower rates of win and return for recreational (cf. more informed) games. These results suggest that the misperception arises from poor information processing and can be considered as evidence in favour of heterogeneity in beliefs causing the resulting bias. These results support Hypothesis 2.3 and suggest recreational players do lose money despite raising less often, suggesting misperception causes errors in those raising decisions rather than simply a systematic under weighting of high probability events. These differences are all statistically significant at $p<0.001$.

Overall, we observe a systematic over-betting on low-probability events (Early seating positions with a poor hand) and under-betting on high-probability events (Button seating positions with good hands) by recreational players relative to more informed players. These differences in betting patterns are economically significant given that the recreational players attain a lower average rate of return than the more informed players, despite the fact that the more informed players are playing against better competition.

Taken together, the results presented above suggest that recreational players are subject to the FLB. Ultimately, the tendency to under-raise in the lower stake games suggests that the FLB they display may be caused by their misperception of probabilities rather than risk-love. Furthermore, the fact that these player attain lower rates of return on those raises compared to more informed players is suggestive of heterogeneity of beliefs.

IV.III Misperception under increasing information complexity

The non-linear weighting function explanation and the heterogeneity of beliefs explanation differ in their predictions regarding the impact of new information (Hypothesis 3.1 vs. 3.2). We test which prediction is true in this section. The results presented in Table 5 show that in the preflop betting round recreational players are less likely to bet than informed players in all seat positions and these differences are statistically significant at the 0.1% level. However, moving through the betting rounds these differences are mostly reversed. Similarly, Table 5 indicates that the proportion of games in which a voluntary contribution is made to the pot at various stages of the game increases the later in the betting round (from 'Early' to 'Button') that the player acts in 45 out of the 48 possible comparisons and these are all statistically significant at the 1% level.

This result holds for both the recreational and more informed players. In fact, differences in proportions tests, comparing differences in proportions of voluntary contributions in different positions, confirm that these are all significant at the 0.1% level for both recreational and informed players for all stages of play (preflop, flop etc.). This shows a general tendency for both the recreational and more informed players to bet more when in stronger seating positions.

As the game progresses from the preflop to the river stage, the pro-active risk ratio (indicating the degree of risk taken in terms of bets and raises) remains above 1.2 for the more informed players. However, it reduces for recreational players. The pro-active risk ratio for recreational players significantly exceeds that of more informed players in the preflop ($p<.001$) by a small amount but this ratio for more informed players exceeds that of the recreational players in later stages of the game. In fact, the effect size of the differences in these ratios between games involving more recreational and more informed players is lowest at the preflop stage and tends to increase through the stages of the game (Cohen's *d* statistics[7]: $d_{preflop} = 15.98$, $d_{flop} = 42.29$, $d_{turn} = 405.27$, and $d_{river} = 258.26$). This is important, since it reveals that there is an increase in the difference between the more informed and the recreational players as the game progresses. This is initial evidence in support of Hypothesis 3.2 (heterogeneity) rather than Hypothesis 3.1 (non-linear weighting functions).

[7] Where $d = (X_1 - X_2) / s_p$, where X1 and X2 are the means s_p is the square root of the pooled standard deviation (Hedges, 1982).

Importantly, at the preflop stage, t-tests also reveal that the recreational players have a higher Button/Early pro-active risk ratio (i.e. betting more in the Button relative to the Early position) compared to the more informed group (p < .001). However, at later stages, the recreational players have lower ratios compared to the more informed group (p < .001 for all) confirming Hypothesis 3.2 rather than 3.1.

The more informed players continue to use the seating position information advantage throughout the game, whereas the recreational players appear to make less use of this information advantage as the game advances. The full use of seating position advantage may be regarded as requiring good information processing skills that most recreational players lack. However, there are advantages of seat position that come automatically, such as players folding before others players have to make a decision about folding, calling or raising.

These findings are consistent with the recreational players being subject to a specific type of misperception that increases as more information is revealed. This finding, as expected, fails to confirm Hypothesis 3.1 but does confirm Hypothesis 3.2, suggesting that misperceptions (leading to heterogeneity)rather than non-linear value functions is the underlying cause of this result.

IV.IV Demonstrating Economic Significance

The results presented in Table 6 demonstrate that recreational players make significantly less money than informed players in the advantageous positions. The 'game cost' indicated in Table 6 is the difference in the blind adjusted net return per game between the recreational and the more informed players. When multiplied by the average number of games played by a recreational player (437.3) this provides the expected cost to recreational (cf. informed) players over their gaming lifespan. The final column in Table 6 indicates the 'market cost', that is the game cost multiplied by the number of games in the recreational dataset (12,466,467). This final column indicates that the total expected cost to recreational players (cf. informed players) in our data set is over $3.1 million.

Furthermore, we find that those recreational/more informed players that clearly displayed FLB (i.e. bet more in the Early than the Button position) earned on average $0.64/$0.66 ($p < .001$) blind adjusted dollars per game less than those that bet more in the Button position than the Early Position. This demonstrates the cost per game when acting against the informational advantage of the seating positions. However, it is possible that over-betting on the Button relative to the Early (i.e. an inverse FLB) could also be detrimental to performance. The relationship between the average blind adjusted net return per trade for a player should then be a second order polynomial function of the ratio of Button/Early contributions, where the peak of the function should indicate the optimal distribution of Button to Early contributions.

Further evidence in favour of the heterogeneous beliefs explanation of the FLB in poker data is explored through estimating second order polynomial functions of the rate of return against Button/Early contribution ratios for the lower and higher stake games, as shown in Figure 4. The highest rate of return is given by the Button/Early contribution ratios in the region of 1.2. There is wide variation in the distribution of players' contribution ratios and this suggests a strong heterogeneity in beliefs of players. Despite this heterogeneity, the representative/median player gets the decision very roughly right in assessing the informational advantage afforded by the Button seating position over the Early position. This is shown by the fact that the highest rate of return is approximately in the middle of the range of the different Button/Early contribution ratios.

V. CONCLUSIONS

Evidence of differential returns to bets placed with different probabilities of success has been reported in many different environments over many decades. In general, the evidence is of a higher expected return to bets placed on higher-probability outcomes, known as the favourite-longshot bias (FLB). There are many potential explanations for the persistence of this bias. The two we seek to arbitrate between here are broadly categorized as preference-based (e.g. risk-love) and misperception-based (notably belief heterogeneity and non-linear weighting function) theories. Using very large data sets of real-money online poker games, we identify the existence of the bias in this real-world natural laboratory. In fact, we demonstrate consistent over-betting on low probability hands and under-betting on high probability hands by recreational compared to more informed poker players, to the extent that they attain a significantly lower rate of return. Snowberg and Wolfers (2010), demonstrated that the FLB, which they found in racetrack odds, imposes a significant cost on bettors. Equally, we find that amongst recreational poker players (i.e. in games involving lower minimum stake levels), the cost of the FLB is substantial, amounting to a penalty (cf. more informed players) of more than $3.1M over the games we examine.

Having found evidence in favour of misperception rather than risk love in the raising behaviour of recreational players at the preflop stage, we then looked for evidence to distinguish between two different types of misperception at all stages of poker games. Our results revealed increasing bias by recreational players (cf. more informed players) as information complexity increased throughout the game, evidence consistent, in a completely different domain, with Gandhi and Serrano-Padial's (2014) recent findings using US racetrack data. In particular, our results suggest that heterogeneity of beliefs, rather than a non-linear weighting function, best explains the FLB displayed by recreational players.

REFERENCES

ALI, M.M., (1977). Probability and Utility Estimates for Racetrack Bettors. *Journal of Political Economy*, **85**(4), pp. 803-815.

BARBERIS, N.C. (2013). Thirty Years of Prospect Theory in Economics: A Review and Assessment. *Journal of Economic Perspectives*, **27**(1), 173-96.

BUSCHE, K. and HALL, C.D., (1988), An Exception to the Risk Preference Anomaly. *Journal of Business*, **61**, pp. 337-346.

BUSCHE, K. (1994), Efficient Market results in an Asian Setting, in Hausch, D.B., Lo, V.S.Y and Ziemba, W.T. (eds.), Efficiency of Racetrack Betting Markets, London: Academic Press, 615-616.

CHIAPPORI, P-A., A. GANDHI, B. SALANIE and F. SALANIE (2009), From Aggregate Betting Data to Individual Risk Preferences, Columbia University, Department of Economics, Discussion Paper Series, Discussion Paper No. 1213-08, November.

CHIAPPORI, P-A., A. GANDHI, B. SALANIE and F. SALANIE (2012), Identifying Preferences Under Risk from Discrete Choices, *American Economic Review: Papers and Proceedings*, **99** (2), 356-362.

FIEDLER, I. (2011), The Gambling Habits of Online Poker Players. *The Journal of Gambling Business and Economics*, **6** (1), 1-24.

FIEDLER, I. and J-P ROCK (2009), Quantifying Skill in Games – Theory and Empirical Evidence for Poker. *Gaming Law Review and Economics*, **13**, 50-57.

GANDHI, A.K. (2007), Rational Expectations at the Racetrack: Testing Expected Utility Using Prediction Market Prices, Mimeo, University of Wisconsin-Madison.

GANDHI, A. and R. SERRANO-PADIAL (2015), Does Beliefs Heterogeneity Explain Asset Prices: The Case of the Longshot Bias, *Review of Economic Studies*, 82 (1), **290**, 156-186.

GRIFFITH, R.M., (1949), Odds Adjustments by American Horse-Race Bettors. *American Journal of Psychology*, **62**, 290-294.

HEDGES, L. V., (1982), Estimation of effect size from a series of independent experiments. *Psychological Bulletin,* **92**, 490-499.

HURLEY, W. and MCDONOUGH, L., (1995), A Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Parimutuel Betting. *American Economic Review*, **85**(4), 949-955.

ISAACS, R., (1953), Optimal Horse Race Bets. *American Mathematical Monthly*, **60**, pp. 310-315.

JULLIEN, B., and SALANIÉ, B., (2000), Estimating Preferences under Risk: the Case of Racetrack Bettors. *Journal of Political Economy,* **108**, 503-530.

KAHNEMAN, D., SLOVIC, P. and TVERSKY, A. (1982), Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.

KATSIKOPOULOS, K.A. (2011), Psychological heuristics for Making Inferences: Definition, Performance, and the Emerging Theory and Practice, *Decision Analysis*, **8** (1), 10-29.

LEVITT, S.D. and LIST, J.A. (2008), Homo Economicus Evolves, *Science*, **319** (5865): 909-10.

OTTAVIANI, M., and SØRENSEN, P.N., (2008), The Favorite-Longshot Bias: An Overview of the Main Explanations. In Hausch, D.B., Ziemba, W.T. (Eds.). Handbook of Sports and Lottery Markets. Amsterdam: North Holland, 83-101.

OTTAVIANI, M. and SØRENSEN, P.N. (2009), Surprised by the Parimutuel Odds? *American Economic Review*, **99** (5), 2129-34.

OTTAVIANI, M. and SØRENSEN, P.N. (2010), Noise, Information and the Favorite-Longshot Bias in Parimutuel Predictions, *American Economic Journal*, **2**(1), 58-85.

POPE, D.G. and SCHWEITZER, M.E. (2011), Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes. *American Economic Review*, **101** (1), 129-157.

PRESTON, M.G. and BARATTA, P., (1948), An Experimental Study of the Auction-Value of an Uncertain Outcome. *American Journal of Psychology*, **61**, 183-193.

QUANDT, R.E., (1986), Betting and equilibrium. *Quarterly Journal of Economics* **101**, 201-208.

ROSETT, R.N., (1971), Weak Experimental Verification of the Expected Utility Hypothesis. *Review of Economic Studies*, **38**(116), 481-492.

SHIN, H.S., (1991), Optimal Betting Odds against Insider Traders. *Economic Journal*, **101**, 1179-1185.

SHIN, H.S., (1992), Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias. *Economic Journal*, **102**, 426-435.

SNOWBERG, E. and WOLFERS, J. (2008), Examining Explanations of a Market Anomaly: Preferences of Perceptions? In: D.B. Hausch and W. T. Ziemba (eds.), Handbook of Sports and Lottery Markets. North-Holland, Amsterdam, pp. 103-133.

SNOWBERG, E. and WOLFERS, J. (2010), Explaining the Favorite-Longshot Bias. Is it Risk-Love or Misperceptions? *Journal of Political Economy*, **118** (4), 723-746.

SONSINO, B., BENZION, U. and MADOR, G. (2002), The Complexity Effects of Choice with Uncertainty – Experimental Evidence, *Economic Journal*, **112** (482), 936-965.

TVERSKY, A., and KAHNEMAN, D., (1992), Advances in Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty*, **5**, 297-323.

VAUGHAN WILLIAMS, L., (1999), Information Efficiency in Betting Markets: A Survey. *Bulletin of Economic Research*, **51**(1), 1-30.

WEITZMAN, M., (1965), Utility Analysis and Group Behaviour: An Empirical Study.

Journal of Political Economy, **73**(1), 18-26.

YAARI, M.E., (1965), Convexity in the Theory of Choice under Risk. *Quarterly Journal of Economics*, **79**, 278-290.

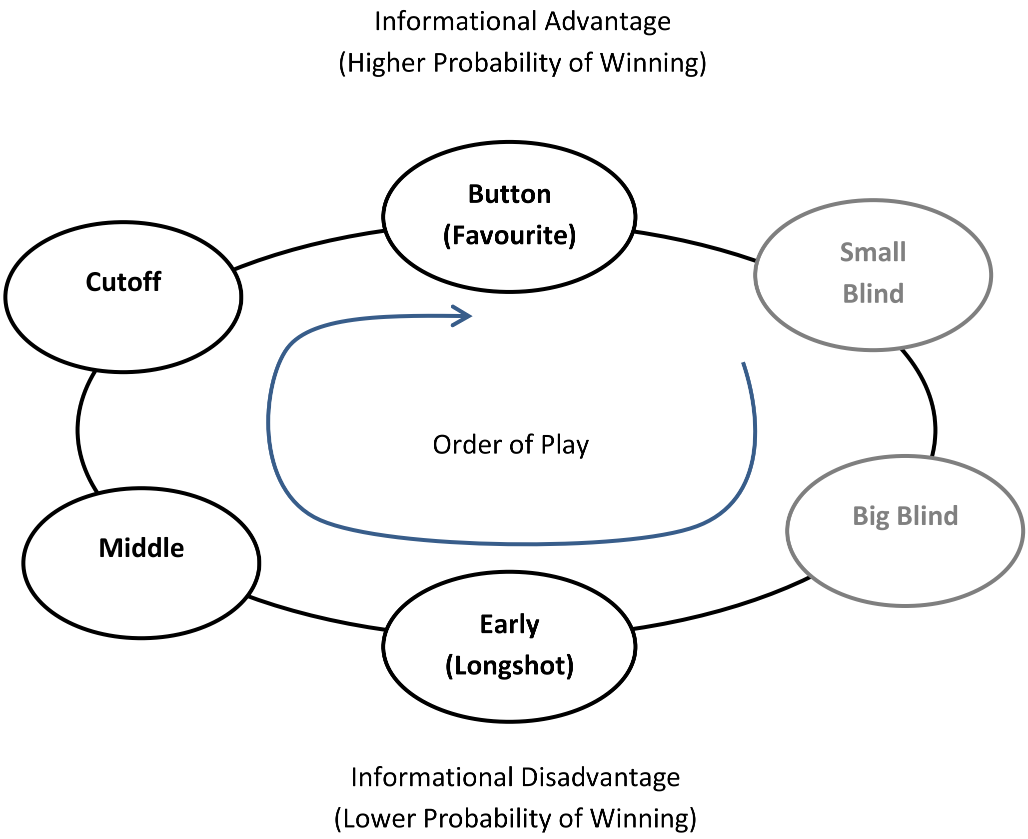
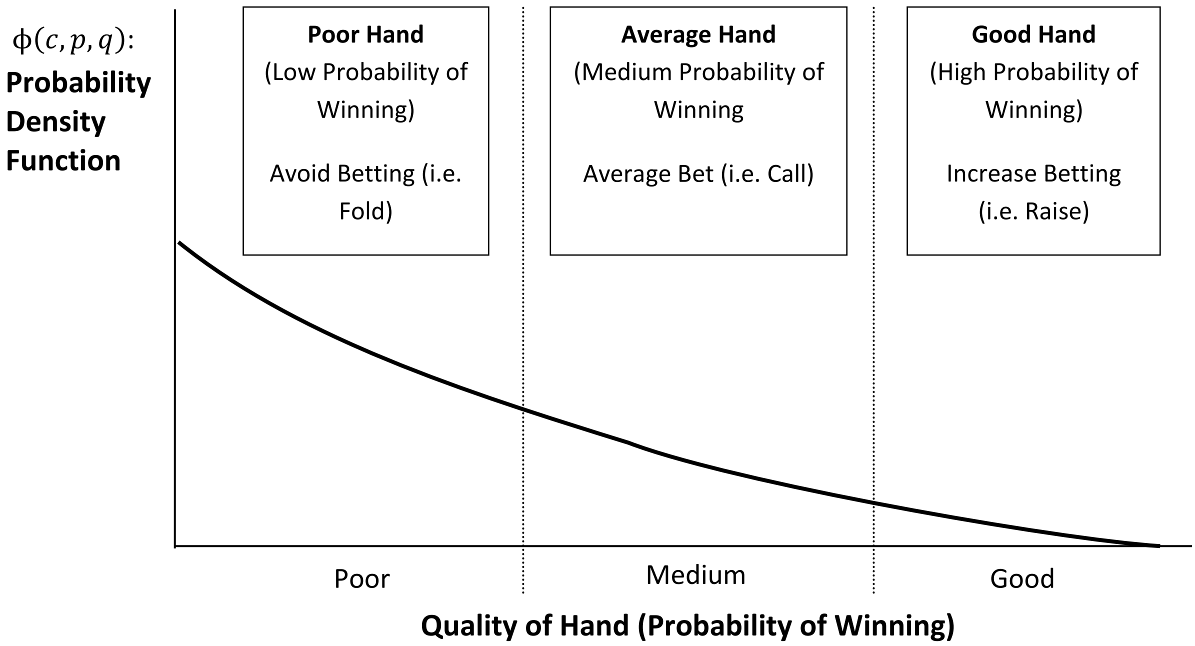


Figure 1. Informational Advantage of Different Seating Positions.

The order of play progresses from the Early through to the Button seating position and hence the informational advantage increases the later the player has to act in each round of betting.



		Preflop Hand	
Seating Position	Poor Hand (Longshot)	Medium Hand	Good Hand (Favourite)
Early (Longshot)	Very Low Probability (Fold)	Low Probability (Fold/Call)	High Probability (Call/Raise)
Middle	..	..	..
Cutoff	...	...	...
Button (Favourite)	Low Probability (Fold/Call)	Medium Probability (Call/Raise)	Very High Probability (Raise)

Figure 2. The disaggregation of preflop hands in different seating positions into low, medium and high probability of winning, with illustration of the hand probability density function.

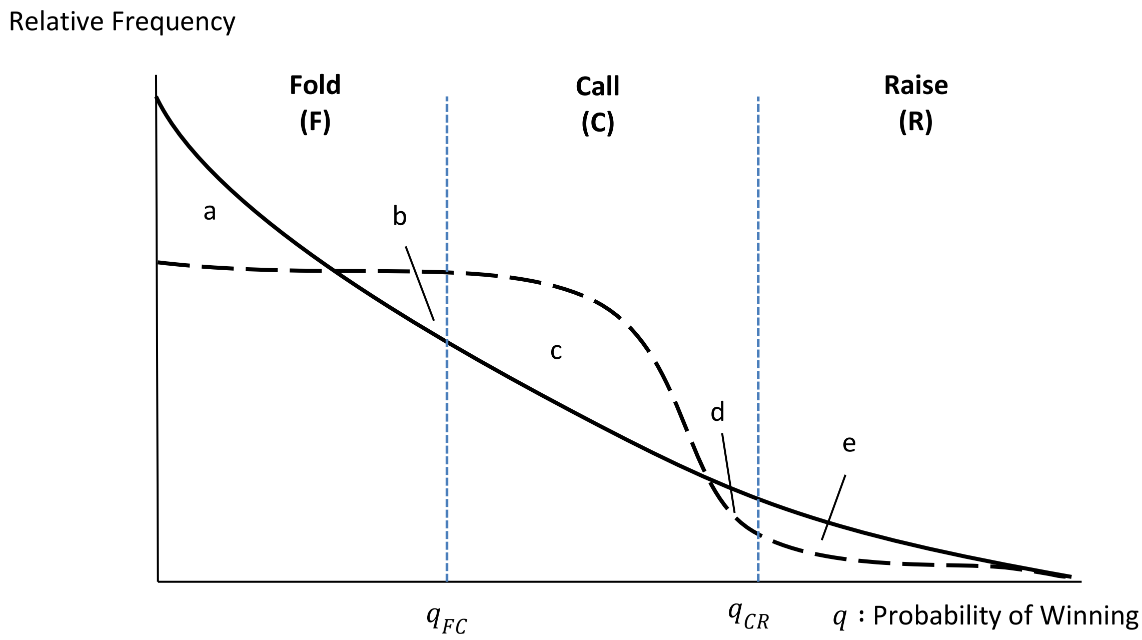
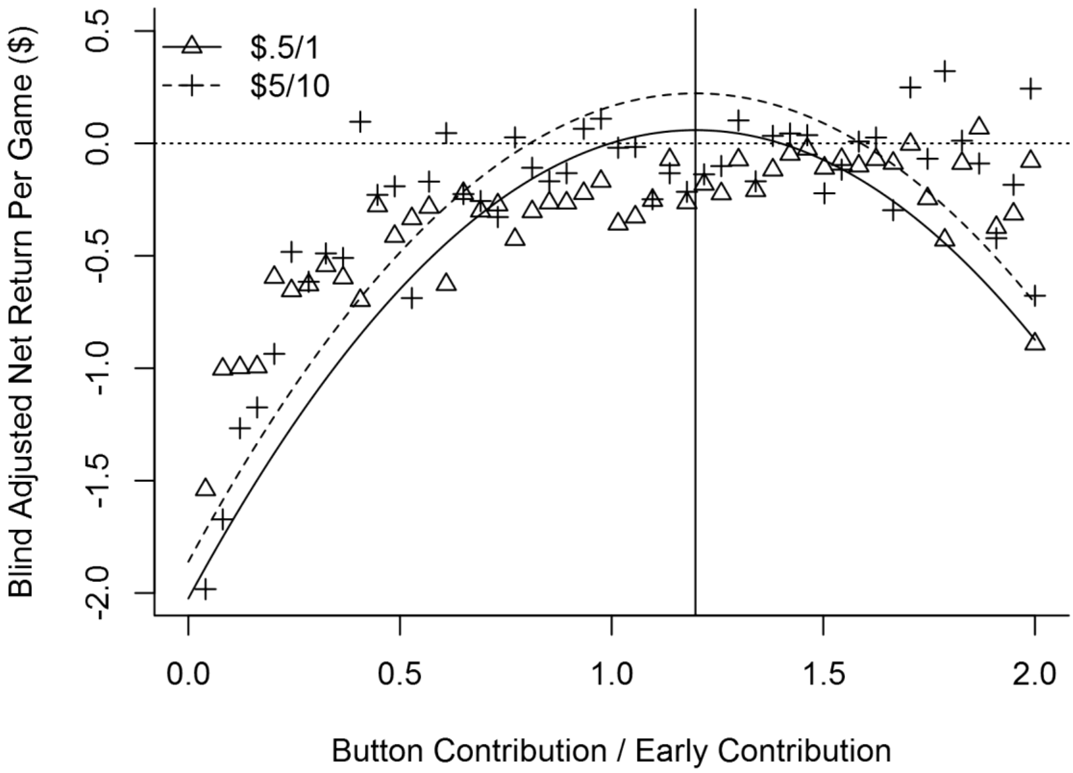


Figure 3: Relative Frequency of Probability of Winning

Illustrative example of the subjective relative frequency distributions of probability of winning for recreational and informed players. The non-linear weighting misperception can be seen in the differences between the subjective relative frequency distributions for informed players (solid black line) and recreational players (dashed line). As shown, we expect recreational players to exhibit fold errors by under folding poor cards (a) and perhaps fold some cards they ought to call (b). We also expect raise decision errors by failing to raise some cards they ought to raise (e). In general, we expect less raising (e), less folding (a < b), but more calling (c > d) overall by recreational players compared to more informed players. Rational requires that both functions integrate to one and that a+d+e=b+c.



(See attached "Figure 4.eps" file for source file)

Figure 4. Relationship between Players' Button/Early contribution ratios and the Blind Adjusted Net Returns per Game.

The local average blind adjusted net returns per game ($) at 0.04 intervals of Button/Early contributions ratios for recreational players ($.5/$1 blind games) are shown as triangles with a solid 2nd order polynomial line fit and for more informed players ($5/$10) are shown as crosses with a dashed 2nd order polynomial line fit.

Table 1. The win rate measure of relative table position advantage, pot contribution and return rates in each seating position. *Results for the lower blind ($0.5/$1), recreational, games are shown in normal text and for higher blind ($5/$10), more informed, games are shown in italics.*

Position	Win Rate (%)	Win Rate Ratio (WR)	Total Pot Contribution (%)	Contribution Ratio (CR)	Ratio Difference (CR-WR)	Return Rate (%) ~
Early	12.02^	1.00^	12.77^	1.00^	0.00^	0.52^
	*12.53^**	*1.00^**	*11.37^**	*1.00^**	*0.00^**	*3.63^**
Middle	13.40	1.11	13.93	1.09	-.02	1.60
	*13.33**	*1.06**	*12.70**	*1.12**	*0.05**	*4.64**
Cutoff	15.61	1.30	15.83	1.24	-.06	2.65
	*15.70**	*1.25**	*15.29**	*1.34**	*0.09**	*5.04**
Button	17.72	1.47	17.71	1.39	-0.09	4.84
	*19.05**	*1.52**	*18.43**	*1.62**	*0.10**	*6.41**

^ indicates a main effect of seating position on the size of the statistic shown in the column, such that the variable in the column is found to vary according to the seating position at $p < .001$ level.

* indicates a (Bonferroni adjusted) significant effect of a pairwise comparison between the statistic for the recreational vs the more informed players ($p < .001$).

~Note that since Poker is a negative sum game and the Early, Middle, Cutoff and Button positions have positive returns, it is the small blind and big blind positions that are in net loss: The total return rate (%) of both the Blind Positions for the ($0.5/$1) games was -14.7% whereas for the ($5/$10) games the total return rate was -15.4%.

Table 2. The fold, call and raise rates in the preflop stage by players in each seating position. *Results for lower blind ($0.5/$1) recreational, games are shown in normal text and for higher blind($5/$10), more informed, games are shown in italics*

Seating Position	Games Folded Preflop	Games Called Preflop	Games Raised Preflop
Early	80.2%^	6.8%^	12.9%^
	*81.8%^**	*2.6%^**	*15.6%^**
Middle	78.3%	7.5%	14.0%
	*80.7%**	*3.6%**	*15.7%**
Cutoff	75.1%	8.7%	15.6%
	*77.3%**	*5.1%**	*17.5%**
Button	72.2%	10.9%	16.9%
	*72.6%**	*7.5%**	*19.8%**

^ indicates a main effect of seating position on the size of the statistic shown in the column, such that the variable in the column is found to vary according to the seating position at $p < .001$ level.

* indicates a (Bonferroni adjusted) significant effect of a pairwise comparison between the statistic for the recreational vs the more informed players ($p < .001$).

Table 3. Rates of return for games that are called or raised preflop by players in each seating position. *Results for lower stake ($0.5/$1), recreational, games are shown in normal text and for higher stake ($5/$10), more informed games are shown in italics (note that the rate of return for folded games is -100%).*

Seating Position	Games Called Preflop	Games Raised Preflop
Early	-4.7%^	5.2%^
	*-3.0%^**	*8.0%^**
Middle	-3.4%	7.0%
	*0.9%**	*9.7%**
Cutoff	-2.1%	9.0%
	*1.9%**	*11.1%**
Button	-0.1%	11.6%
	*2.9%**	*13.3%**

^ indicates a main effect of seating position on the size of the statistic shown in the column, such that the variable in the column is found to vary according to the seating position at $p < .001$ level.

* indicates a (Bonferroni adjusted) significant effect of a pairwise comparison between the statistic for the recreational vs the more informed players ($p < .001$).

Table 4 – Percentage of hands including a raise with win and return rates at the preflop betting round

Position	Blinds	Games %	Win Rate %	Return Rate %
Early	**.5/1**	12.88*	69.95*	5.23*
	5/10	15.61*	72.58*	8.02*
Middle	**.5/1**	14.00*	72.48*	6.99*
	5/10	15.72*	74.88*	9.66*
Cutoff	**.5/1**	15.64*	75.41*	9.04*
	5/10	17.45*	77.57*	11.15*
Button	**.5/1**	16.88*	78.82*	11.64*
	5/10	19.84*	80.59*	13.30*

*Indicates significance at the .1% level.

Table 5: %VPIP Bets or Raises for $0.5/1 and 5/10 dollar games: *The percentage games reaching each stage in which a proactive voluntary contribution is made to the pot (i.e. only bets and raises) from a given position in the preflop, flop, turn and river stages of the game. Lower blind, recreational, ($0.5/$1) games shown in normal text and for higher blind, more informed) ($5/$10) games shown in italics*

Position	Blinds	Preflop	Flop	Turn	River
Games	.5/1	12,466,677	6,497,363	3,973,850	2,823,475
	5/10	*6,704,607*	*2,794,629*	*1,755,462*	*1,266,035*
Small blind	.5/1	10.76%	13.76%	13.27%	11.89%
	5/10	*13.72%*	*13.50%*	*11.64%*	*9.64%*
Big Blind	.5/1	5.99%	15.26%	16.37%	14.56%
	5/10	*7.26%*	*13.35%*	*15.06%*	*13.34%*
Early	.5/1	13.93%	12.29%	8.64%	6.95%
	5/10	*17.45%*	*12.63%*	*7.75%*	*5.82%*
Middle	.5/1	15.23%	12.67%	8.83%	6.91%
	5/10	*17.73%*	*12.63%*	*7.78%*	*5.88%*
Cutoff	.5/1	17.15%	13.57%	9.35%	7.03%
	5/10	*20.00%*	*13.78%*	*8.63%*	*6.40%*
Button	.5/1	18.33%	13.81%	9.58%	7.02%
	5/10	*22.62%*	*15.24%*	*9.73%*	*7.08%*
Button/Early					
Pro-active	.5/1	1.32	1.12	1.11	1.01
Risk Ratio	*5/10*	*1.30**	*1.21**	*1.25**	*1.22**

* indicates a significant difference at the .01% level.

Table 6: The average (blind adjusted) net return and costs per game ($) for each seating position.

Lower stake ($0.5/$1), recreational, games are shown in normal text and higher stake ($5/$10), more informed games are shown in italics.

	Avg. Net Return ($) per Game	Average Cost ($) per Game	Average Cost per Player ($)	Total Cost ($)
Early	0.0140 (.0024) *0.0813 (.0031)*	-0.0673 *	-29.43	-838,993.23
Middle	0.0461 (.0026) *0.1159 (.0033)*	-0.0697 *	-30.50	-869,421.04
Cutoff	0.0877 (.0027) *0.1528 (.0036)*	-0.0651 *	-28.47	-811,567.00
Button	0.1773 (.0028) *0.2307 (.0037)*	-0.0534 *	-23.35	-665,709.34
			-111.75	-3,185,690.61

Standard Deviations shown in brackets.

Average Cost per Game calculated as the difference between the average (blind adjusted) net return attained in the higher stake games minus the average net return attained in the lower stake games. Average cost per player equals the average cost per game multiplied by average number of games played by a recreational player in the dataset (437.3). Total cost is the average cost per game multiplied by the number of games played by recreational players (*12,466,467)*.

* Indicates significance at the .01% level.