Judgment Error in Lottery Play:
When the Hot-Hand Meets the Gambler’s Fallacy

Qingxia Kong*  Nicolas S. Lambert †  Chung-Piaw Teo‡

November, 2013

Abstract

We demonstrate that lottery players can be influenced to believe erroneously in the existence of “hot” numbers, where past winning numbers are perceived to have a greater probability of winning in future draws, even though past and the future events are independent. The existence of this “hot-hand” effect in lottery games is surprising, as works by Clotfelter and Cook (1993) and Terrell (1994) have documented instead the presence of the opposite effect, the “gambler’s fallacy”, in the US lottery market—which means that the amount of money bet on a particular number falls sharply after the number is drawn. We use two sets of lottery game data to show that both the gambler’s fallacy and hot-hand fallacy can prevail under different gaming environments, contingent on the design (e.g., prize structures) of the lottery games. We develop a quasi-Bayesian model that is consistent with our empirical findings to investigate the conditions in the environment that determine which fallacy dominates. Our results also provide a new explanation for the “lucky store” effect (Guryan and Kearney, 2008)—why players in the US believe that lightning will strike twice in the case of lottery vendors, but not in the case of lottery numbers.

Keywords: Perception of Randomness, Hot-Hand Fallacy, Gambler’s Fallacy, Bayesian Inference

*School of Business, Universidad Adolfo Ibáñez, Chile; q.kong@uai.cl.
†Graduate School of Business, Stanford University; nlambert@stanford.edu.
‡Department of Decision Sciences, National University of Singapore; bizteocp@nus.edu.sg.
1 Introduction

People are known to rely on heuristics or simple mental models to interpret random events (Tversky and Kahneman, 1974; Oskarsson et al., 2009). Although this ensures fast decisions that generate reasonably good outcomes under most circumstances, decision biases are inevitable. Among many others, the gambler’s and hot-hand fallacies are the most common biases in the perception of randomness. The gambler’s fallacy is an erroneous belief in the negative correlation of independent outcomes generated by a random process, while the hot-hand fallacy refers to the belief in positive correlations. Although the two fallacies seem to contradict each other, current literature posits that the two are related. They may arise from the same mechanism: belief in “the law of small numbers” or “local representativeness,” in which small samples are used to represent the characteristics of the total population (Rabin, 2002; Rabin and Vayanos, 2010).

Earlier works focus on understanding the phenomenon in which the gambler’s and hot-hand fallacies coexist in the same individual. They identify the importance of the length of streaks (sequence of repeated events) in influencing the perception of randomness—i.e., the gambler’s fallacy prevails in short streaks, but as the streaks lengthen, belief in the hot-hand fallacy dominates (Asparouhova et al., 2009; Rabin and Vayanos, 2010; Jørgensen et al., 2011). On the other hand, observations that the gambler’s fallacy is more prevalent in lottery games, and the hot-hand fallacy more prevalent in games that require skills, are often attributed to the “intentions” of the subjects involved. For example, the experimental study of Caruso et al. (2010) demonstrates that human subjects tend to predict continuation of a streak of outcomes when the agents that generate the streak are perceived to be intentional. In contrast, when the streak is generated by some kind of mechanical device (as in lottery games), people believe the streak will end, that is, the gambler’s fallacy dominates.

Boynton (2003) argues, however, that the spurious correlation in binary sequences could also be a function of the perceived response-outcome contingency, i.e., the hot-hand effect is maintained by accidental successes in prediction, and the gambler’s fallacy is reinforced by accidental failures. This is predicated on the assumption that subjects learn using a memory-based process, and have the tendency to switch responses after failure. This line of argument is supported by empirical evidence in the classic study of human superstitious responding by Hake and Hyman (1953) and
experiments by Boynton (2003). This raises an important question: Since accidental successes and failures in the lottery are determined by game design, can we identify conditions in the gaming environment that allow one form of fallacy to dominate over the other?

Green et al. (2010) develop several ingenious arguments to demonstrate that experimental cues can alter how subjects learn about the generating process (using optimal Bayesian updating with erroneous beliefs) and thus substantially affect subjects’ choices, without altering the underlying outcome probabilities. In one experiment, subjects played a series of roulette games under different conditions—one uses roulette wheel that has evenly sized colored slots, but the odds of one color’s appearing in the “random” outcome are double those of the other’s appearing. The other condition has slots proportionally sized to their chances of being selected. Subjects were to choose one of the two colors, and outcomes were generated either by a computer program or subjects’ erroneous belief that outcomes were dependent on their own motor behavior. Under different experimental cues, subjects behaved differently over time, ranging from optimal (choosing only the color with the higher probability of appearing) to probability matching behavior (choosing the colors in a 2:1 ratio).

In this paper, we provide empirical evidence using real lottery data to demonstrate the impact of experimental cues in shaping the erroneous beliefs in the hot-hand or gambler’s fallacy. The experimental cues in our case refer to the number of prizes awarded by the lottery operator. We show that the number of prizes awarded can indeed influence the way subjects learn through a Bayesian updating mechanism, and determine which fallacy will dominate in the long run. These results have important practical implications for the management of risk and decision under uncertainty, since they indicate that the perception of randomness can be manipulated, and hence behavior can be nudged with the appropriate design. For instance, several countries have attempted to influence commuters’ behavior by offering incentive schemes in which commuters earn credit for each journey taken (with triple credit for off-peak journeys) and to enter in weekly cash lotteries. The success of these schemes hinges on an insight from behavioral economics: On average people are risk-seeking when the stake are small. So a 1-in-1000 chance to win $100 is more attractive than a cash award of $0.10. Our results say that we can do better, through designing the lottery games to allow the hot-hand fallacy to dominate other(erroneous) beliefs of the commuters. Distortion in the winning probabilities is enough to make what is actually a 1-in-1000 chance of winning appear to be a much
safer bet, and thus more attractive to commuters.

2 Literature Review

Both the gambler’s and hot-hand fallacies have long been observed in the field. Among many others, Clotfelter and Cook (1993) observe that lottery players in 3-digit numbers games in the United States are subject to the gambler’s fallacy. Figure 1 shows the percentiles of betting ratios (betting volume index over average index on a particular day) on different days after the numbers have been selected as winners. Once a 3D number is drawn as the winner, the subsequent betting volumes drop immediately and then gradually pick up. The immediate drop after the winning number is drawn provides strong evidence of the gambler’s fallacy at work. In a different context, Camerer (1989) shows that betting markets for basketball games exhibit a small hot-hand bias. Guryan and Kearney (2008) examine the sales data for lottery outlets that have sold winning jackpot tickets and show that those stores experience a significant increase in game-specific ticket sales, exhibiting the “lucky store” effect. Jørgensen et al. (2011) examine a set of panel data on Lotto games and find that people usually avoid numbers that have recently been drawn (exhibiting gambler’s fallacy), but tend to bet more on winning numbers in streaks (that have been drawn several times in a row), suggesting the presence of the hot-hand fallacy.

The gambler’s and hot-hand fallacies may coexist in the same individual and appear within the same setting. One line of research posits that they both arise from representativeness bias or belief in “the law of small numbers.” Tversky and Kahneman (1971) coin this term and examine its connection with the gambler’s fallacy. Rabin (2002) and Rabin and Vayanos (2010) build theoretical models to model the law of small numbers which directly leads to the gambler’s fallacy. Asparouhova et al. (2009) run a set of lab experiments on binary choice games, and the results support the work of Rabin (2002). In recent work of Kendall (2010), the author uses a discount factor to refine the model presented in Rabin (2002) and proposes a simple but effective model to explain the two fallacies.

These works identify the importance of the length of streaks (sequence of repeated events) in influencing the perception of randomness, i.e., the gambler’s fallacy prevails in short streaks, but as the streaks lengthen, beliefs in the hot hand dominate. Different from this line of research, our
paper examines the role of game design in shaping the perception of randomness in a long run. We show that due to pattern seeking, either the gambler’s or the hot-hand fallacy can arise under appropriate conditions. We can manipulate the behaviors through careful system design. We also generalize current results in literature. In our paper, a streak follows one particular pattern (e.g., a streak of permuted numbers) that agents might recognize from previous outcomes, in particular, outcomes do not have to be perfectly identical, but they have to “look alike”. When the streak is short (i.e., the pattern has not been recognized yet), people tend to believe in the gambler’s fallacy; as the streak lengthens and become an obvious pattern, however, the hot-hand fallacy occurs. Then the gambler’s fallacy is commonly seen in single-winner lottery games, in which streaks are extremely unlikely.

In spite of excellent work that uses the law of small numbers to examine the two fallacies, another stream of research provides different explanations. These researchers argue that the type of random-event generator—whether an inanimate device or a human being—determines the occurrence of the gambler’s or hot-hand fallacy (Ayton and Fisher, 2004; Burns and Corpus, 2004). If the random process is believed to be generated by a mechanical device, people expect negative recency; if human beings generate the sequences, people expect positive recency. These results are reinforced by field
observations made in casinos and lottery stores (Sundali and Croson, 2006; Guryan and Kearney, 2008). Caruso et al. (2010) show that when outcomes are perceived to be generated by agents with intentions, people tend to predict continuation of a streak. This work has a similar flavor to ours, in the sense that an intentional mind shows that some mechanism is at work and patterns related to the intention can be perceived.

3 Field Evidence

According to prevalent theory, the gambler’s fallacy would occur in lottery games, as outcomes are generally believed to be generated without human involvement. As demonstrated in Green et al., however, experimental cues in a lottery game can also play an important role in shaping players’ erroneous beliefs. In this section, we present empirical evidence to show that different fallacies and decision biases may prevail in different gaming environments, even if the games appear to be similar in design. To control for cultural biases, we collected two sets of field data from two fixed-odds lottery games to demonstrate this phenomenon. The first is a 3-digit (3D) lottery game played in China, the scheme of which resembles the “Pick 3” lottery game reported in Clotfelter and Cook (1993). It draws a single prize-winning number each day, 7 times per week. The other game, a 4-digit (4D) lottery game played in Southeast Asia, picks 23 winning numbers at a time, three times per week. The two games are of similar design and are played by two populations that are similar in both race and culture. They differ mainly in the number of winning numbers drawn. This provides an excellent basis on which to test the theory and predictions that we will develop in Section 4.

3.1 The Gambler’s fallacy in the 3D numbers game

In the 3D numbers game, a single winning number from 000 to 999 is randomly drawn with equal probability at 8:30 p.m. each day. Lottery tickets for the day’s draw are sold until 8:00 p.m. Each wager costs 2 RMB. There are three types of bets: Straight, Box 3, and Box 6. In the Straight bet, players bet on a specific 3D number and receive 1000 RMB per wager if the number wins. The Box 3 bet allows players to make one wager on the permutation of any 3D number with 2 identical digits, and in the Box 6 bet, on the permutation of any 3D number with 3 unique digits.
For example, a player who places a 2 RMB wager on 223 and chooses the Box 3 bet type actually bets 2/3 RMB on each of the 3 numbers 223, 322, and 232. The payout per wager and odds of winning are shown in Table 1.

<table>
<thead>
<tr>
<th>Bet Type</th>
<th>Match to Win</th>
<th>Payout per Wager</th>
<th>Odd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Straight</td>
<td>Match the exact order</td>
<td>1,000</td>
<td>1 in 1,000</td>
</tr>
<tr>
<td>Box 3</td>
<td>Match any order (2 identical digits)</td>
<td>320</td>
<td>1 in 333</td>
</tr>
<tr>
<td>Box 6</td>
<td>Match any order (3 unique digits)</td>
<td>160</td>
<td>1 in 167</td>
</tr>
</tbody>
</table>

Testing for the presence of the gambler’s fallacy usually requires a full dataset that includes the sales of every number in every draw. Unfortunately, such a dataset is not publicly available for the lottery of interest. Nevertheless, we were able to obtain a rich enough dataset from the official web-site of the game operator by extracting information on a daily basis for an extended period. The lottery company posts (i) the winning number drawn, (ii) number of winning wagers (of the three types), and (iii) total sales for each draw. We collected online data for 2,272 draws over a 76-month period, from May 8, 2005, to September 15, 2011. We then constructed a sub-sample of 95 draws that contains winning numbers repeating within 7 weeks. The method of data analysis we adopt is similar to that of Terrell (1994).

For example, the overall mean payout rate in our dataset was 50% and the median payout 42.7%. The lowest payout rate of 7.3% occurred on September 18, 2010, when “945” was drawn as the winning number. The total payout was only $3.869M, from a total pot of $52.63M. The same number, 945, had also been the winning number drawn on September 17, 2010, the day before. The payout was $27.85M from a total pot of $52.532M, with a payout ratio of 53%. The only other time this number had been drawn in our dataset was on April 18, 2008. The payout then was $32.1M from a total pot of $55.2M, with a payout ratio of 58.1%. This suggests that winning in the current draw will have an adverse effect on sales of the same number in subsequent draws.

More generally, we seek to estimate the impact on sales in subsequent draws after a 3D number has been drawn as a winning number. As we only have detailed information for winning numbers, we check the payout rates for winning numbers that repeat within 7 weeks. Table 2 shows the

\[^1\] There are minor payout variations by areas: in Beijing, the game pays 980 per wager for the Straight bet, and in Shanghai and Tibet, the payout is 333 for the Box 3 bet and 166 for the Box 6 bet.

payout statistics (mean, standard deviation, and median) of winning numbers that repeat within several stipulated periods. It shows that the mean and the median of payout statistics for winning numbers that repeat within 7 weeks are lower than those statistics for all winning numbers. The median payout of winning numbers that repeat within 1 week drops to around 40% and then gradually increases \(^3\). A two-sample \(t\)-test shows that the mean payout of numbers that repeat within 1 to 7 weeks is significantly smaller than 0.5, the overall mean payout rate.

<table>
<thead>
<tr>
<th>winning numbers repeating</th>
<th>Number</th>
<th>Mean</th>
<th>STD</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>within 1 week</td>
<td>12</td>
<td>43.6%</td>
<td>52%</td>
<td>25.9%</td>
</tr>
<tr>
<td>between 1 and 2 weeks</td>
<td>12</td>
<td>35.8%</td>
<td>15%</td>
<td>34.6%</td>
</tr>
<tr>
<td>between 2 and 4 weeks</td>
<td>25</td>
<td>38.4%</td>
<td>22%</td>
<td>30.3%</td>
</tr>
<tr>
<td>between 4 and 7 weeks</td>
<td>46</td>
<td>47.5%</td>
<td>23%</td>
<td>41.8%</td>
</tr>
<tr>
<td>7 weeks and beyond</td>
<td>2177</td>
<td>50.3%</td>
<td>31%</td>
<td>43.0%</td>
</tr>
<tr>
<td>Total</td>
<td>2272</td>
<td>50%</td>
<td>31%</td>
<td>42.7%</td>
</tr>
</tbody>
</table>

Let \(\log R_i\) denote the logarithmic value of the payout rate of winning number \(i\). To analyze the effects of the winning number on sales, we build a linear regression model for \(\log R_i\) as follows:

\[
\log R_i = \alpha + \beta D_i + \gamma \log \text{Last} R_i + \epsilon_i. \tag{1}
\]

We use \(D_i\) as a predictor, where \(D_i\) denotes the inverse of the number of draws it takes for a winning number \(i\) to show up again as a winning number. Note that a positive coefficient of \(D_i\) in the regression indicates that a winning draw promotes the popularity of a winning number (hot-hand fallacy), while a negative coefficient undermines the popularity of a winning number (gambler’s fallacy). We also use another independent variable, \(\log \text{Last} R_i\), in our regression, which is the logarithmic value of the payout rate when the number \(i\) last won. A positive coefficient of \(\log \text{Last} R_i\) shows that the popularity of a number persists in the lottery game, so that a high payout rate observed can attributed merely to the inherent popularity of the particular number.

Among the parameters, \(\alpha\) is a constant, \(\beta\) captures the effect of a winning draw, and \(\gamma\) represents

\(^3\)Note that such a tendency does not stand out in the mean, since the mean of numbers that repeat within one week is high due to an outlier. The 3D number 149 was drawn as the winner on Feb 13, 2008. One week later, when the same number was again the winner, the payout ratio hit 200.25%, the highest level recorded in history for this number. This noise explains the high mean and standard deviations among winners that repeat within one week. 149 is an exceptionally popular number in this game. It has been drawn 5 times, with payout ratios of 200.25%, 134.19%, 94.36%, 88.84%, 73.41% and 53.17%.
the popularity of a number. Consistent with Table 2, we only select winning numbers that repeat within 7 weeks. Regression results are shown in Table 3. The coefficient $\beta$ is estimated to be $-0.762$ and is significantly negative at the 1% level. It indicates that the expected payout value of a number drops by 17.3% ($= 10^{-0.762}$) immediately after it wins the prize, keeping other variables constant. In all, if the payout rate of a number is 50%, the regression model predicts that the mean payout of the same number will drop to 7% on the next draw, to 36.2% 10 draws after winning, and to 39.5% 20 draws after winning. Both the statistical data and regression results suggest that the gambler’s fallacy exists among the players in the 3D lottery game.

### 3.2 The hot-hand fallacy in the 4D numbers game

In the 4D lottery game, players bet on numbers 0000 to 9999. Sales for each draw start one week before and close at 6 p.m. on the draw day. There are three draws every week. The minimum cost of a bet is 1 local currency. At each draw, 23 numbers are picked as winning numbers. The 23 winning numbers are generated with replacement by rolling four boxes that each contain 10 balls. There are 5 prize categories. Table 4 shows the payout for each prize category and winning odds.

<table>
<thead>
<tr>
<th>Prize Category</th>
<th>Payout per Wager</th>
<th>Odds</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Prize</td>
<td>2000</td>
<td>1 in 10,000</td>
</tr>
<tr>
<td>2nd Prize</td>
<td>1000</td>
<td>1 in 10,000</td>
</tr>
<tr>
<td>3rd Prize</td>
<td>500</td>
<td>1 in 10,000</td>
</tr>
<tr>
<td>10 Starter</td>
<td>250</td>
<td>1 in 1,000</td>
</tr>
<tr>
<td>10 Consolation</td>
<td>60</td>
<td>1 in 1,000</td>
</tr>
</tbody>
</table>

We obtained data from a game operator in the Southeast Asia region, with information on 156 draws over a 1-year period. The dataset consists of 23 winning numbers and sales volume for each 4D number in each draw. We want to investigate how a winning draw affects sales for
winning numbers in subsequent draws. To account for daily effects, we use betting proportion as an indicator, which is defined as the ratio of the sales volume for a 4D number to total sales volume. We examine all winning numbers from draw 1 to draw 100 and calculate their betting proportions respectively, on the day they are drawn and the subsequent 56 draws. Sales increase by about 40% immediately following a winning draw (cf. Figure 2), suggesting that the hot-hand fallacy prevails at the aggregate level.

![Figure 2: Betting on previous winning numbers](image)

Next, we run a regression model with the data from 4D numbers games. The model is similar to the one defined in equation (1). Due to the presence of different prize categories, we use betting proportion instead of payout ratio to indicate the popularity of a number\(^4\). Again, let \( \log R_i \) denote the log value of the betting proportion and \( \log LastR_i \) the log value of the betting proportion when winner \( i \) last won. Similar to the analysis of 3D games, we use data that repeats within 7 weeks (21 draws). Regression results are shown in Table 5. Contrary to behavior in the 3D game, the regression model in the 4D game shows that winning in the current draw has a positive impact on subsequent sales. This compelling evidence suggests that players in 4D numbers games are subject

\(^4\)Note that in the 3D numbers game, the payout ratio is simply the betting proportion times the prize won. The payout ratio for a 4D game is more complicated because of the possibility of a number’s winning multiple prizes.
to the hot-hand fallacy.

Table 5: Results of Linear Regression on 4D Numbers Games

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>157</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ( \alpha )</td>
<td>(-0.452^{***})</td>
</tr>
<tr>
<td>Delay ( \beta )</td>
<td>(0.114^{***})</td>
</tr>
<tr>
<td>Previous Payout ( \gamma )</td>
<td>(0.886^{***})</td>
</tr>
<tr>
<td>Adjusted ( R^2 )</td>
<td>0.738</td>
</tr>
</tbody>
</table>

\(**p < 0.01; **p < 0.05; *p < 0.1\)

We use the 4D sales data to further validate the observation that the hot-hand fallacy extends beyond the winning numbers, in that the erroneous belief also affect the sales of numbers that are “similar” to the winning numbers. We say that a 4D number \( i \) is a “near miss” in the current draw if it is “similar” to a winning number but not identical. We specify two types of near misses: Permutation and Digit Replacement. If a player bets on a number that gets all 4 digits right but in a wrong order, a Permutation near miss occurs. On the other hand, if a number matches 3 digits of a winning number in the right position correctly, the bet is considered a Digit-Replacement near miss. Both empirical research and lab experiments indicate that near misses can encourage more participation in lottery games (Reid, 1986; Kassinove and Schare, 2001). These results are supported by Clark et al. (2009), who also report that game operators tend to manipulate the frequency of near misses to generate more sales. The near-miss effect is a prime suspect for a fundamental psychological factor that contributes to problem gambling.

In the following, we test whether there is a near-miss effect in the 4D sales data. We draw insights using data from the 8th draw to the 146th draw. We calculate the betting proportions of the near-miss numbers on the day the corresponding winning numbers were drawn (row 0), 7 draws before they were drawn (row -7), and the 1st, 2nd, 3rd, 7th, 14th and 20th draw after they were drawn (row 1, 2, 3, 7,14 and 20 respectively). Table 6 lists the median value of the betting proportions of the first-prize-winning numbers, the average of 23 prize-winning numbers, and near-miss numbers in specific draws. As illustrated, the proportion of bets remained largely unchanged between the current draw and the 7th prior draw (row 0 and -7, respectively). Once a number appeared as a winning number, however, its popularity surged. For first-prize-winning numbers, the surge in relative sales in the next draw is about 102% (from 0.0848 to 0.1712). The hot hand fallacy on winning numbers also affects numbers that are near miss, although the effect on sales is
less significant: For a Permutation near miss, the surge is from 0.0992 to 0.1175, and for a Digit Replacement near miss, from 0.0992 to 0.1068.

<table>
<thead>
<tr>
<th>Lag</th>
<th>23 winning numbers</th>
<th>1st Prize Winner</th>
<th>Permutation</th>
<th>Digit Replacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>-7</td>
<td>0.0838</td>
<td>0.0842</td>
<td>0.0991</td>
<td>0.099</td>
</tr>
<tr>
<td>0</td>
<td>0.0846</td>
<td>0.0848</td>
<td>0.0992</td>
<td>0.0992</td>
</tr>
<tr>
<td>1</td>
<td>0.1185</td>
<td>0.1712</td>
<td>0.1175</td>
<td>0.1068</td>
</tr>
<tr>
<td>2</td>
<td>0.1025</td>
<td>0.1434</td>
<td>0.1116</td>
<td>0.1049</td>
</tr>
<tr>
<td>3</td>
<td>0.0959</td>
<td>0.1272</td>
<td>0.1093</td>
<td>0.1041</td>
</tr>
<tr>
<td>7</td>
<td>0.0884</td>
<td>0.1135</td>
<td>0.1052</td>
<td>0.1021</td>
</tr>
<tr>
<td>14</td>
<td>0.0868</td>
<td>0.1075</td>
<td>0.1032</td>
<td>0.101</td>
</tr>
<tr>
<td>20</td>
<td>0.0852</td>
<td>0.096</td>
<td>0.1021</td>
<td>0.1004</td>
</tr>
</tbody>
</table>

4 Model and Theoretical Analysis

The empirical findings of the previous section suggest that lottery players can exhibit either the hot-hand fallacy or the gambler’s fallacy, depending on the gaming environment. In this section, we develop a model consistent with the empirical evidence that sheds light on which aspects of the environment can trigger one fallacy or the other.

Our model features a representative individual, the decision maker, who observes a stream of data generated at random by a fixed source. For example, a lottery player observes the past winning numbers generated by some lottery mechanism. He faces uncertainty and forms beliefs about the probabilities of future events. Our model takes a quasi-Bayesian approach: The decision maker does not know the data-generating process, but learns progressively from past observations. He is rational in updating his beliefs (following the standard Bayesian updating process), but holds a wrong prior belief that rejects the true data-generating process. As a result, the decision maker ends up forming biased views of the uncertain environment, which may take the form of the hot-hand or gambler’s fallacy and may lead to suboptimal decisions.

We present the model in its full generality. As it only requires a Markovian structure, it applies to a broad range of situations that extend beyond lottery games. After introducing the model, we derive the conditions that drive belief convergence. We then use these results to explain when and how the hot-hand or gambler’s fallacy arises. Finally, we apply the results to the case of a player of 3D and 4D lottery games.
The two most prominent quasi-Bayesian models are those of Barberis et al. (1998) and Rabin (2002). The model we develop in this paper includes elements present in both. Note that the models of Barberis et al. and Rabin provide explanations for the existence of the gambler’s fallacy and the hot-hand fallacy in different environments. Neither model applied, however, to the environment considered in this paper.

4.1 Model

At every period $t = 1, 2, \ldots$, the source generates a random outcome from a finite set $\mathcal{X}$. For example, if the source models a lottery mechanism, outcomes are the daily draws of winning numbers. We denote by $\{X_t\}_{t=1}^{\infty}$ the process associated with the sequence of outcomes.

The source distributes the outcomes according to the law $p^*$. For lottery mechanisms, $p^*$ distributes the outcomes of each period uniformly, and identically and independently across periods. In the general case, we simply assume that $p^*$ belongs to the collection $\mathcal{L}$ of the probability laws $p$, where $\{X_t\}_{t=1}^{\infty}$ is a stationary\(^5\) time-homogeneous Markov process whose transition matrix has only positive entries.\(^6\) In particular, every law of $\mathcal{L}$ is fully specified by its transition matrix. In the sequel we denote by $p(x|y)$ the probability that $X_{t+1} = x$ given $X_t = y$, and by $p(x)$ the unconditional probability that $X_t = x$.

We assume a population of players, each holding prior beliefs on different theories on the random outcomes. We assume also $\theta$ fraction of the population holds the true law $p^*$ in its set of priors, whereas a fraction $1 - \theta$ has ruled out $p^*$ completely. The latter corresponds to the group of conspiracy theorists who do not believe that the winning numbers are drawn randomly. The existence of the latter group of players explains the market for guide books on lottery number picking and also proliferation of web sites purporting to have spotted patterns of the winning numbers.

We examine next the action of each player in this population, in response to the stream of random outcomes (winning numbers drawn). Each player observes the sequence of outcomes unravel over time. He is unable to form an accurate cognitive representation of the true law $p^*$ and ends

\(^5\)The stationarity of $p$ is without loss of generality. The other conditions ensure that $p$ is irreducible and aperiodic, and so admits a unique stationary distribution. Because we are interested in its long-run properties, we can assume without loss of generality that we are already in steady state.

\(^6\)Saying that the transition matrix only has positive entries is equivalent to saying that every sequence of outcomes of finite length occurs with positive probability.
up forming approximations of it. He holds a prior belief on different models, or theories, that can describe the law of the outcome process. We denote these theories by $p^1, \ldots, p^K$, where $p^k \neq p^l$ if $k \neq l$. Note that a theory is also expressed as a probability law, but the words theory or model are used to capture the idea that theories reflect approximate and imperfect representations of the true law. For simplicity, we assume that the theories also belong to the set $\mathcal{L}$, that is, the theories are Markov processes. In effect, this limits the memory of the decision maker to one period; however, this is not essential to our results: The analysis extends directly to Markov theories of arbitrary order. Let $\mu_0^1, \ldots, \mu_0^K > 0$ be the prior likelihood that the decision maker assigns to the theories $p^1, \ldots, p^K$, respectively. Therefore, a prior belief can be summarized as a pair $((p^1, \ldots, p^K), (\mu_0^1, \ldots, \mu_0^K))$. As more data become available, the decision maker updates his prior beliefs according to Bayes’ Law. We denote by $\mu_t^k(X_t, \ldots, X_1)$ the posterior probability assigned to theory $p^k$ after observing the outcomes up to time $t$. Denote $\mu_0^k$ as the prior belief assigned to theory $p^k$. Bayes’ Law gives the posterior beliefs as

$$
\mu_t^k(X_t, \ldots, X_1) = \mu_{t-1}^k(X_{t-1}, \ldots, X_1) \cdot \frac{p^k(X_t \mid X_{t-1})}{\sum_l \mu_{t-1}^l(X_{t-1}, \ldots, X_1)p^l(X_t \mid X_{t-1})}.
$$

We stress that a key element of our model, common among quasi-Bayesian models, is that the decision maker can assign a zero prior to the true law of the outcome process. This captures the idea that the decision maker has a biased view of randomness. Indeed, if the decision maker were to put a positive weight on the true law, learning would eventually lead him to uncover the true distribution.

We are interested in the long-run properties of the decision maker’s posterior belief. For a given realization of outcomes $x_1, x_2, \ldots$, we say that the posterior belief converges when, for every $k$, the posterior $\mu_t^k(x_t, \ldots, x_1)$ converges as $t$ grows to infinity. As outcomes are randomly generated, it is natural to consider convergence properties that must be satisfied not only for one but for most realizations. We say that the posterior belief converges almost surely when, for every $k$, the belief process $\{\mu_t^k(x_t, \ldots, x_1)\}$ converges with probability 1 (under the true law of the outcome process).
4.2 Convergence of beliefs

We first establish the general results of the decision maker’s beliefs in the long run. Knowing when and where beliefs converge is important for understanding what leads to the hot-hand or the gambler’s fallacy, which is the aim of the next section.

In our simple setting, establishing the conditions for convergence turns out to be relatively straightforward. The conditions for almost-sure convergence are expressed in terms of the Kullback-Leibler divergence rate. The K-L divergence rate is used as a measure of the difference between a theory that approximates the true law and the actual, true law of the outcome process. Given two laws \( p, q \in \mathcal{L} \) for the Markov process \( \{X_t\}_1^\infty \), the K-L divergence rate between \( p \) and \( q \) is defined by

\[
\mathcal{D}_{KL}(p \parallel q) = \lim_{t \to \infty} \frac{1}{t} \mathcal{D}_{KL}(p(X_t, \ldots, X_1) \| q(X_t, \ldots, X_1)),
\]

where \( \mathcal{D}_{KL}(p(X_t, \ldots, X_1) \| q(X_t, \ldots, X_1)) \) is the K-L divergence between the distributions \( p \) and \( q \) restricted to the first \( t \) periods. We refer to Gray (1990) or Cover and Thomas (1991) for general properties of the K-L divergence rate. Given the Markov assumptions made on \( \mathcal{L} \), in our model the K-L divergence rate is easily written as

\[
\mathcal{D}_{KL}(p \parallel q) = \mathbb{E}_p \left[ \log \frac{p(X_{t+1} \mid X_t)}{q(X_{t+1} \mid X_t)} \right] = \sum_{x,y \in \mathcal{S}} p(y)p(x \mid y) \log \frac{p(x \mid y)}{q(x \mid y)},
\]

where \( \mathbb{E}_p \) denotes the expectation over the process \( \{X\}_1^\infty \) under the law \( p \). Recall that \( \mathcal{D}_{KL}(p \parallel q) \geq 0 \), with equality if and only if \( p = q \). The K-L divergence rate gives a sense of how distinct the laws \( p \) and \( q \) are from each other.

The first proposition below is from Shalizi (2009).\(^7\) If there is a single theory \( p^k \) that is closest to the true law of the outcome process, the posterior belief converges to \( p^k \) almost surely, discarding every other theory \( p^l, l \neq k \), after observing a sufficient number of periods.

**Proposition 1.** Consider a law \( p^* \) for \( \{X_t\}_1^\infty \) and a prior belief \( (p^1, \ldots, p^K, \mu^0_1, \ldots, \mu^0_K) \). If, for some \( k \neq l \), \( \mathcal{D}_{KL}(p^* \| p^k) < \mathcal{D}_{KL}(p^* \| p^l) \), then almost surely the posterior on theory \( p^l \) converges

---

\(^7\)The setting of Shalizi (2009) is considerably more general. For our simpler setting, there exists an elementary proof of Proposition 1, presented Appendix A, which is used for our next proposition.
toward zero:

$$\lim_{t \to \infty} \mu_t^l(X_t, \ldots, X_1) = 0 \quad \text{almost surely.}$$

In particular, if there exists $k$ such that for every $l \neq k$, $\mathcal{D}_{KL}(p^* \parallel p^k) < \mathcal{D}_{KL}(p^* \parallel p^l)$, then almost surely the posterior belief converges to the theory $p^k$:

$$\lim_{t \to \infty} \mu_t^k(X_t, \ldots, X_1) = 1 \quad \text{almost surely.}$$

Proof: See Shalizi (2009) and Appendix A.

Proposition 1 states that no theory can be deemed consistent with the data in the long run if it does not minimize the K-L divergence rate. In particular, that the K-L divergence rate is minimized for a single theory is a sufficient condition for almost sure convergence. Our next result states that this condition is also necessary. If there are two or more theories that minimize the K-L divergence rate, the posterior belief is guaranteed not to converge for almost every sequence of outcomes. The posterior oscillates between those minimizers, each theory being nearly completely discarded at different times and infinitely often. In such circumstances the decision maker switches between different explanations without ever making up his mind. One implication of the two propositions is that for any given law that governs the outcome process, the behavior of the posterior belief is always extreme. If the posterior converges with positive probability, then it converges with probability 1, and discards every theory except one. If the posterior is non-convergent with positive probability, then it is non-convergent with probability 1.

**Proposition 2.** Consider a law $p^*$ for $\{X_t\}_{1}^{\infty}$ and a prior belief $\langle (p^1, \ldots, p^K), (\mu^1_0, \ldots, \mu^K_0) \rangle$, and let $\mathcal{M}$ be the set of indices $k$ that minimize the K-L divergence rate $\mathcal{D}_{KL}(p^* \parallel p^k)$. If $\mathcal{M}$ contains two or more minimizers, then almost surely the posterior belief does not converge:

$$\lim_{t \to \infty} \mu_t^l(X_t, \ldots, X_1) = 0 \quad \text{for every } l \not\in \mathcal{M}, \text{ almost surely.}$$

$$\liminf_{t \to \infty} \mu_t^k(X_t, \ldots, X_1) = 0 \quad \text{for every } k \in \mathcal{M}, \text{ almost surely.}$$

Proof: See Appendix B.

Note that if the decision maker were to put a positive prior on the true law $p^*$, then he would
gradually learn the truth and discard every other theory. This result is well known, starting from Doob (1949) in the i.i.d. case. In our setting it is a direct consequence of Proposition 1: Since $\mathcal{D}_{KL}(p^*||p^k) = 0$ if and only if $p^k = p^*$, the K-L divergence rate is always positive otherwise.

4.3 The hot-hand and gambler’s fallacies

We now turn to modeling the hot-hand and gambler’s fallacies in the lottery game (with $N$ numbers to be played). These biases occur when the decision maker overestimates or underestimates, respectively, the chance that an event will occur when an identical or similar event has occurred in the past. To model these phenomena, we complement our setting with $N$ binary functions $\mathcal{E}^i : \mathcal{S} \rightarrow \{0, 1\}$, for $i \in \{1, \ldots, N\}$. These functions can be interpreted as the indicator variables of dichotomous events that can be observed at every period. In the lottery context, we use $\mathcal{E}^i$ to represent the event that $i$ is a winning number.

We formalize the notion of “similarity” between events in an abstract form with a collection $\mathcal{R} \subset \{1, \ldots, N\}^2$. If $(i, j) \in \mathcal{R}$, then it means the events $\mathcal{E}^i$ and $\mathcal{E}^j$ are similar. We call $\mathcal{R}$ the similarity set.

We define two classes of theories, $\mathcal{C}^+$ and $\mathcal{C}^-$. The first class includes all the theories that are associated with the hot-hand fallacy, while the second class includes the those corresponding to the gambling fallacy. The set $\mathcal{C}^+$ contains all the theories $p$ such that

\[
p(x|y) > p^*(x|y) \quad \text{if } (x, y) \in \mathcal{R},
\]

\[
p(x|y) < p^*(x|y) \quad \text{if } (x, y) \notin \mathcal{R}.
\]

Similarly, the set $\mathcal{C}^-$ contains all the theories $p$ that satisfy

\[
p(x|y) < p^*(x|y) \quad \text{if } (x, y) \in \mathcal{R},
\]

\[
p(x|y) > p^*(x|y) \quad \text{if } (x, y) \notin \mathcal{R}.
\]

In this way, we use $\mathcal{R}$ to model the biased beliefs that the transition probability from $\mathcal{E}^j$ to $\mathcal{E}^i$ does not follow the true law $p^*(\cdot|\cdot)$, whenever $(i, j) \in \mathcal{R}$.

An obvious choice for $\mathcal{R}$ is the set $\{(i, i); i = 1, \ldots, N\}$ of pairs of identical events. The theories
in $C^+$ are associated with the hot hand beliefs that once a number $i$ has won in a draw, its chances of winning in the next draw will be higher than the true probability $p^*(i|i)$. Similarly, the theories in $C^-$ are associated with the gambler’s fallacy that once a number $i$ has won in a draw, its chances of winning in the next draw will be lower than the true probability $p^*(i|i)$.

The use of more general similarity sets allows us to consider phenomena such as near-miss effects in lottery games, which are also used to characterize hot-hand and gambler’s fallacies. For example, in lottery games, it is common to consider one-digit replacement and permutations of a number as a related number. In this way, we say two numbers $x$ and $y$ are similar if one can be obtained from another using a single digit replacement operator, or a permutation operator.

**Proposition 3.** Assume that every theory in the support of the player’s prior belief belongs to one of the two classes $C^+$ or $C^-$, but that the support is not entirely contained in one class. Then there exists two thresholds, $\beta_L, \beta_H \in (0,1)$, that depend solely on the support of the player’s prior, such that:

- If $p^*(R) > \beta_H$, then in the long run, the player believes in the hot-hand fallacy.
- If $p^*(R) < \beta_L$, then in the long run, the player believes in the gambler’s fallacy.

**Proof:** By a direct application of Proposition 1, for the gambler’s fallacy to dominate in the long run, it is enough to show that

$$D_{KL}(p^*||p^+) > D_{KL}(p^*||p^-)$$

for every $p^+ \in C^+ \cap \mathcal{S}, p^- \in C^- \cap \mathcal{S}$, where $\mathcal{S}$ is the support of the individual’s prior belief. Define

$$a = \min_{p^+ \in C^+ \cap \mathcal{S}, \ p^- \in C^- \cap \mathcal{S}, (x,y) \in \mathcal{R}} \log \frac{p^-(x|y)}{p^+(x|y)}$$

$$b = \min_{p^+ \in C^+ \cap \mathcal{S}, \ p^- \in C^- \cap \mathcal{S}, (x,y) \notin \mathcal{R}} \log \frac{p^-(x|y)}{p^+(x|y)}$$
Note that \(a < 0\) and \(b > 0\). Let \(\beta^L = b/(b-a) \in (0,1)\), and suppose that \(p^*(\mathcal{R}) < \beta^L\). Then,

\[
\mathcal{D}_{KL}(p^* || p^+) - \mathcal{D}_{KL}(p^* || p^-) = \sum_{x,y \in \mathcal{X}} p^*(x,y) \log \frac{p^-_y(x)}{p^+_y(x)} - \sum_{(x,y) \notin \mathcal{R}} p^*(x,y) \log \frac{p^-_y(x)}{p^+_y(x)} = p^*(\mathcal{R})a + (1 - p^*(\mathcal{R}))b > 0 .
\]

Therefore, if \(p^*(\mathcal{R}) < \beta^L\), in the long run the player believes in the gambler’s fallacy. By a symmetric argument, there exists \(\beta^H \in (0,1)\) such that if \(p^*(\mathcal{R}) > \beta^H\), in the long run the player believes in the hot-hand fallacy.

In general the thresholds \(\beta^L\) and \(\beta^H\) differ, but they match under some additional conditions. In particular, when the player believes that the true law can be only one of two possible theories, \(p^+\) and \(p^-\). The theory \(p^+\) is associated with positive serial correlations, while the theory \(p^-\) is associated with negative serial correlations. Initially, the player puts positive weight on both theories. Both theories are fully specified by a parameter that controls the amount of serial correlation between consecutive states. The theory \(p^+\) has a parameter \(\alpha^+ > 1\) and takes the form

\[
p^+_y(x) = \begin{cases} 
\frac{\alpha^+}{N_y \alpha^+ + \alpha^- N_y} & \text{if } (x,y) \in \mathcal{R} , \\
\frac{1}{N_y \alpha^+ + \alpha^- N_y} & \text{if } (x,y) \notin \mathcal{R} ,
\end{cases}
\]

where \(N_y = |\{x|(x,y) \in \mathcal{R}\}|\) is the number of states similar to the state \(y\), and \(N_y = |\{x|(x,y) \notin \mathcal{R}\}|\) is the number of states that are not. The theory \(p^-\) is also defined by a parameter \(\alpha^- < 1\) and takes the form

\[
p^-_y(x) = \begin{cases} 
\frac{\alpha^-}{N_y \alpha^- + \alpha^+ N_y} & \text{if } (x,y) \in \mathcal{R} , \\
\frac{1}{N_y \alpha^- + \alpha^+ N_y} & \text{if } (x,y) \notin \mathcal{R} ,
\end{cases}
\]

We make the following assumptions about the parameters: (i) \(0 < \alpha^+ + \alpha^- - 2 < 2(\alpha^+ - 1)(1 - \alpha^-)\), and (ii) \(\log \frac{\alpha^+}{\alpha^-} < \alpha^+ - \alpha^-\) and \(\log \frac{\alpha^+}{\alpha^-} < \frac{\alpha^+-\alpha^-}{\alpha^+\alpha^-} \).

The implications of assumption (i) are two folds. The left-hand sided inequality guarantees that
the sum of $\alpha^+$ and $\alpha^-$ are not too small. A small $\alpha^+$ indicates that the theory $p^+$ is close the true theory; while a small $\alpha^-$ implies that the theory $p^-$ is far from the true theory. If this constraint is violated, theory $p^+$ systematically dominates in the long run. Similarly, the right-sided inequality ensures that $p^-$ does not always dominate. The assumption (ii) requires that $\alpha^+$ is large enough while $\alpha^-$ is small enough, so that both theories are sufficiently different from the true theory.

In this special case, there is a threshold parameter on the probability $p^*(R)$ that two consecutive states are similar, such that, above the threshold, the player is prone to the gambler’s fallacy after observing many periods, while below the threshold, the player is biased towards the hot-hand fallacy. If the probability exactly equals the threshold, the player does not make up his mind, and switches permanently between a hot-hand behavior and a gambler’s behavior.

**Proposition 4.** Assume that the player holds a prior belief composed of the parameterized theories $p^+$ and $p^-$ as defined above. Then, there exists a threshold $\beta$ such that:

- If $p^*(R) > \beta$, the player falls into the hot-hand fallacy in the long run.
- If $p^*(R) < \beta$, the player falls into the gambler’s fallacy in the long run.
- If $p^*(R) = \beta$, the player oscillates between the hot-hand and the gambler’s fallacy.

**Proof:** We consider the difference of the two K-L divergence rates

$$D_{KL}(p^*\|p^+) - D_{KL}(p^*\|p^-).$$  \hspace{1cm} (3)

Proposition 1 and 2 dictate that the theory $p^-$ eventually dominates if (3) is strictly positive, that $p^+$ dominates if it is strictly negative, and that neither theory always dominates the other in the long run if (3) is zero, instead, they both dominate infinitely often at different time periods. We
re-write (3) as

\[
\sum_{x,y \in \mathcal{X}} p^*(y)p^*(x|y) \log \frac{p^-(x|y)}{p^+(x|y)},
\]

\[
= \sum_{(x,y) \in \mathcal{R}} p^*(y)p^*(x|y) \log \frac{p^-(x|y)}{p^+(x|y)} + \sum_{(x,y) \not\in \mathcal{R}} p^*(y)p^*(x|y) \log \frac{p^-(x|y)}{p^+(x|y)},
\]

\[
= \sum_{(x,y) \in \mathcal{R}} p^*(y)p^*(x|y) \log \frac{\alpha^- (N_R + \alpha^+ N_R)}{\alpha^+(N_R + \alpha^- N_R)} + \sum_{(x,y) \not\in \mathcal{R}} p^*(y)p^*(x|y) \log \frac{N_R + \alpha^+ N_R}{N_R + \alpha^- N_R},
\]

\[
= p^*(\mathcal{R}) \log \frac{\alpha^-}{\alpha^+} + \log \frac{1 + p^*(\mathcal{R})(\alpha^+ - 1)}{1 + p^*(\mathcal{R})(\alpha^- - 1)}.
\]

where we note that \(N_R/N_R\) is the odd ratio \(p^*(\mathcal{R})/(1 - p^*(\mathcal{R}))\). The conclusion follows from observing that under the assumptions on the parameters \(\alpha^+\) and \(\alpha^-\), the function \(f(x) = x \log \frac{\alpha^-}{\alpha^+} + \log \frac{1 + (\alpha^+ - 1)x}{1 + (\alpha^- - 1)x}\) is not monotone, but has with a single zero in \((0, 1)\), it is positive to the left and negative to the right (see Appendix).

\[\Box\]

4.4 Application to the 3D and 4D numbers lottery games

We now apply the results to the 3D and 4D lottery games considered in this paper. For concreteness, let us consider the case of a lottery mechanism that, at every period, draws \(M\) balls at random from an urn that contains \(N\) balls numbered 1 through \(N\) without replacement.\(^8\) The number of lottery outcomes, each consisting of \(M\) draws, is \(\binom{N}{M}\). Hence, for every outcome \(x\),

\[
p^*(x) = 1/\binom{N}{M}.
\]

For simplicity, we assume that two numbers are related only when they are identical, and two outcomes (each with \(M\) numbers) are “similar” only when they share at least one winning number. Given an arbitrary reference lottery outcome, the number of lottery outcomes that do not have any winning number in common is

\[
\binom{N - M}{M},
\]

\(^8\)In the actual 3D and 4D lottery games, the winning numbers are drawn with replacement; however, this difference is irrelevant for our analysis.
so that the number of outcomes that do share a common winning number is
\[
\binom{N}{M} - \binom{N-M}{M}.
\]

Therefore, the probability that two consecutive outcomes are similar, i.e., share at least one winning number, is
\[
p^*(R) = 1 - \frac{\binom{N-M}{M}}{\binom{N}{M}}.
\]

We consider a lottery player with a prior belief composed of theories associated with conservatism and theories associated with the representativeness heuristic. Every theory predicts that every outcome is unconditionally equally likely, and so there is no a priori bias toward certain winning numbers. However, the theory associated with conservativeness tends to predict that the next outcome is more likely to contain a winning number that won in the previous period, while the theory associated with representativeness tends to predict the opposite. To keep matters simple, suppose that theories distribute similar and dissimilar outcomes uniformly. Thus, we can define a theory \(p^+\) that corresponds to the hot-hand fallacy. Under \(p^+\), an outcome similar to the previous winning numbers is \(\alpha\) \((\alpha > 1)\) times as likely as an unrelated outcome.

\[
p^+(x|y) = \begin{cases} 
\frac{\alpha}{1 + p^*(R)(\alpha - 1)} \cdot p^*(x) & \text{if } x \text{ and } y \text{ share a winning number}, \\
\frac{1}{1 + p^*(R)(\alpha - 1)} \cdot p^*(x) & \text{if } x \text{ and } y \text{ do not share a winning number}. 
\end{cases}
\] (4)

Similarly, we can define another theory \(p^-\) that gives rise to the gambler’s fallacy. Under \(p^-\), an outcome is \(\beta\) \((0 < \beta < 1)\) times as likely to occur when it contains a number that won during the previous period than when it does not.

\[
p^-(x|y) = \begin{cases} 
\frac{\beta}{1 + p^*(R)(\beta - 1)} \cdot p^*(x) & \text{if } x \text{ and } y \text{ share a winning number}, \\
\frac{1}{1 + p^*(R)(\beta - 1)} \cdot p^*(x) & \text{if } x \text{ and } y \text{ do not share a winning number}. 
\end{cases}
\] (5)

We define the function \(f(x) = x \log \frac{\alpha}{\beta} + \log \frac{1+(\beta-1)x}{1+(\alpha-1)x}, x \in [0, 1]\), which will be useful for characterizing the behavioral biases of the lottery player. The following lemma explores some of the properties of function \(f(x)\) within \([0, 1]\).
Lemma 1. Assume that the following conditions are satisfied:

1. $0 < \alpha + \beta - 2 < 2(\alpha - 1)(1 - \beta)$.
2. $\log \alpha - \log \beta < \alpha - \beta \land \log \alpha - \log \beta < \frac{\alpha - \beta}{\alpha \beta}$.

Then there exists a unique solution $0 < x^* < 1$ that solves $f(x^*) = 0$; moreover, $f(x) < 0$ when $x \in (0, x^*)$ and $f(x) > 0$ when $x \in (x^*, 1)$.

The implications of condition 1 are twofold. The left-hand side of the inequality guarantees that the sum of $\alpha$ and $\beta$ is not too small. A smaller $\alpha$ indicates that theory $p^+$ is closer to the true theory, while a smaller $\beta$ implies that theory $p^-$ is more distinct from the true theory. If this constraint is violated, theory $p^+$ dominates in the long run. The second constraint regulates the two parameters to ensure that $p^-$ does not dominate. Condition 2 requires that $\alpha$ is big enough and $\beta$ is small enough. This assumption ensures that both theories are sufficiently different from the true theory. Lemma 1 leads to the following result in Proposition 5.

Proposition 5. Fix number of balls $N$ and a number of balls drawn $M$. Suppose the lottery player’s prior belief satisfies the conditions in Lemma 1. If

$$\binom{N-M}{M} < (1-x^*)\binom{N}{M},$$

then in the long run the lottery player exhibits the hot-hand fallacy. If

$$\binom{N-M}{M} > (1-x^*)\binom{N}{M},$$

then in the long run the lottery player exhibits the gambler’s fallacy. Finally, if there is equality, then the beliefs do not converge, and in the long run the lottery player alternates between the hot-hand and gambler’s fallacies.
**Proof:** We write \( x = y \) if \( x \) and \( y \) share a winning number. Let us consider the difference

\[
\mathcal{D}_{KL}(\mathbf{p}^*\|\mathbf{p}^-) - \mathcal{D}_{KL}(\mathbf{p}^*\|\mathbf{p}^+)
\]

\[
= \frac{1}{(N/M)^2} \sum_{x,y \in \mathcal{X}} \log \frac{\mathbf{p}^+(x|y)}{\mathbf{p}^-(x|y)}
\]

\[
= \left\{ \sum_{x=y} \log \left( \frac{\alpha}{\beta} \cdot \frac{1 + \mathbf{p}^*(\mathcal{R})(\beta - 1)}{1 + \mathbf{p}^*(\mathcal{R})(\alpha - 1)} \right) + \sum_{x \neq y} \log \frac{1 + \mathbf{p}^*(\mathcal{R})(\beta - 1)}{1 + \mathbf{p}^*(\mathcal{R})(\alpha - 1)} \right\}/(N)^2
\]

\[
= f(\mathbf{p}^*(\mathcal{R}))
\]

where \( \sum_{x=y} 1 = \mathbf{p}^*(\mathcal{R})(N/M)^2 \) and \( f \) is the function

\[
f(x) = \left( \log \frac{\alpha}{\beta} \right) x + \log \frac{1 + x(\beta - 1)}{1 + x(\alpha - 1)}.
\]

Proposition 5 lists conditions under which function \( f(x) \) is negative or positive (see Appendix C for details). According to Propositions 1 and 2, if \( \mathbf{p}^*(\mathcal{R}) > x^* \), every theory associated with the gambler’s fallacy vanishes a posteriori in the long run, and the theories that dominate are associated with the hot-hand fallacy. If \( \mathbf{p}^*(\mathcal{R}) < x^* \), the effect is the opposite: The posterior belief for the theories associated with the hot-hand fallacy converges toward zero. Finally, if both are equal, then the posterior belief oscillates between the pair of theories closest to \( \mathbf{p}^* \) in the sense of the K-L divergence rate.

In the 3D and 4D games, we can easily infer in which direction the agent’s belief converges by calculating the value of \( \mathbf{p}^*(\mathcal{R}) \). Assume that \( N = 1,000 \) and change \( M \) from 1 to 100. Consider the case \( \alpha = 2 \) and \( \beta = 1/2 \). Note that \( x^* = 1/2 \) in this case. Figure 3 shows that the value of \( \mathbf{p}^*(\mathcal{R}) \) as \( M \) increases. Below the line \( p^* = 1/2 (M < 26) \), the gambler’s fallacy occurs while the hot-hand fallacy appears above this line (when \( M \geq 26 \)).

### 4.4.1 Simulation Results

In the following, we simulate the Bayesian updating process in the context of the 4D numbers game, choosing \( N = 10,000 \) and \( M = 23 \). In the simulation, we assume that the similarity set contains identity, permutation, and one-digit replacement. We set \( \alpha = 3 \), \( \beta = 1/3 \), and a prior belief of the agent unbiased toward \( p^+, p^- \). Figure 4 illustrates the updating process of the weight
assigned to theory $p^+$ over 1,000 periods. Results show that on average $p^*(ℛ) = 92.32\%$ and $E[f(p^*(ℛ))] = 0.0265 > 0$, hence the hot-hand fallacy dominates in this case.

**Remark:** It may be expected that as $M$ grows larger, the faster the belief will converge to the hot-hand theory $p^+$. This is, however, not the case: The convergence rate first increases, then decreases in $M$. Indeed, the convergence rate is determined by $|E[f(p^*(ℛ))]|$: The higher the value, the faster the convergence. When $M = 39$, for example, the probability that two consecutive outcomes will be related is larger than 99%, yet $E[f(p^*(ℛ))] < 3 \cdot 10^{-4}$, and so it takes about 10,000 rounds (roughly 30 years of gambling, assuming one draw per day) for the player to converge to the hot-hand belief. In this case, when a related outcome occurs, the weight assigned to $p^+$ increases slightly, but whenever an unrelated outcome appears, the weight drops significantly. Even though the weight rarely drops, if the drop is significant, it can easily cancel the benefit of many periods of positive reinforcement of $p^+$ via a single unrelated outcome.

This observation leads to interesting implications. In the following simulations, we set $\alpha = 3$, $\beta = 1/3$, $N = 10,000$, and vary the values of $M$. Instead of an unbiased prior, we assume that
Figure 4: The Bayesian updating process given $N = 10000$, $M = 23$ and $\alpha = 3$, $\beta = 1/3$.

the player is biased toward the gambler’s fallacy with an initial weight of 99%. In Figure 5 we plot the weight for $p^+$, taking $M = 1$, $M = 23$, and $M = 39$, respectively, over 10,000 rounds. When $M = 39$, the belief remains heavily biased toward the gambler’s fallacy, even though two consecutive outcomes will be extremely likely to occur. As $M$ increases, we observe first the gambler’s fallacy, then the hot-hand fallacy, and return to the gambler’s fallacy as the player does not have enough time to converge to the hot-hand belief within the time frame of the play. This result is reminiscent of Rabin and Vayanos (2010), albeit in a quite different context.9

5 Application: Proportion of Players with Fallacious Beliefs

In this section, we exploit the model built in Section 4.4 and develop simple models to predict that proportion of sales 10 in 4D numbers game that are affected by fallacious beliefs that the numbers are not drawn independently.

9In Rabin and Vayanos (2010), they demonstrate that when an agent who is subject to the gambler’s fallacy observes a series of related binary outcomes, he tends to underreact to short streaks (gambler’s fallacy), overreact to longer streaks (hot-hand fallacy) and underreact to very long streaks (returning to the gambler’s fallacy).

10Since the sales volume vary a lot in the course of one year, we normalize the sales volume in each period and predict instead sales proportions for each of the 10000 4D numbers in the next period.
In 4D numbers game, we have observed significant increase in the sales of 4D numbers that were either drawn as or related to previous winning numbers, indicating the existence of Hot-hand believers. We build a “Fallacious Model” to incorporate the Hot-hand behaviors among the lottery players. In the model, we assume that \( \theta \) proportion of players are “fallacious believer” who follow the Hot-hand fallacy when they bet. The fallacious believers form similarity sets that contain identity, permutation and one-digit replacement of the winning numbers in the current period and they predicted the sales proportion in the next period according to theory specified in Equation 4. The rest \( 1 - \theta \) fraction of the players are not affected by any fallacious beliefs so their predictions of outcome of the next period will be independent by the previous winning numbers. We call those players as “naive players”, for whom the betting proportion for the next period is basically the same as in current period. In the following, we will estimate the parameter \( \theta \) in the Fallacious Model and test the performance of this model.

Let \( p_{i,t} \) denote the (actual) sales proportion of 4D number \( i \) at period \( t \) and \( \hat{p}_{i,t+1} \) the predicted sales proportion in period \( t + 1 \). In the base model where all the players are naive, \( \hat{p}_{i,t+1} = p_{i,t} \). In
the Fallacious Model,

\[ \hat{p}_{i,t+1} = \theta \times p_{i,t} + (1 - \theta) \times \hat{p}_{i,t+1}^{\pm} \]  

(6)

where \( \hat{p}_{i,t+1}^{\pm} \) is the predicted sales proportion predicted by the fallacious believers.

Besides, we also introduce a simple “Linear Model” where we use both the sales proportions and the winning number indicator variable (of current period) as inputs to predict the sales in next period. Let \( w_{i,t} = 1 \) if 4D number \( i \) is a winning number at period \( t \); \( w_{i,t} = 0 \) otherwise. We assume that each player has certain chances to be attracted to play on the winning numbers drawn in the current period. The predicted sales in the next period is thus

\[ \hat{p}_{i,t+1} = a + b \times p_{i,t} + c \times w_{i,t}. \]  

(7)

Note that similar as the Fallacious Model, the Linear Model also assume that certain players are affected by fallacious beliefs. We use this model to benchmark the performance of the Fallacious Model.

Forecast error is given by the difference between the predicted sales \( \hat{p}_{i,t+1} \) and the actual sales proportion \( p_{i,t+1} \) in period \( t + 1 \). In the Forecast chapter (Chapter 3), Stevenson (2012) introduces three commonly used measures for forecast accuracy. In this study, we apply the Mean Squared Error (MSE) to measure and compare performances of different models. If we use data of \( m \) periods, then

\[ MSE(m) = \frac{1}{m-1} \sum_{t=1}^{m} \sum_{i=1}^{N} (\hat{p}_{i,t+1} - p_{i,t+1})^2. \]

We use the first 100 draws of 4D sales data described in Section 3 to calibrate the models (so as the MSE is minimized), and the remaining 56 draws to evaluate out-of-sample predictions.

- In the Naive model, \( MSE = 2.42 \times 10^{-6} \).

- In the linear model, estimation results show that \( b = 0.9552 \) and \( a, c \) are close to zero \((a = 3.3 \times 10^{-6} \text{ and } c = 1.11 \times 10^{-5})\). The MSE is \( 2.31 \times 10^{-6} \). The in-sample improvement is 4.55%.

- In the fallacious model, the parameter set \( \{\alpha, \theta\} \) is estimated to be \( \{4.37, 0.045\} \). It shows that around 4.5% of the players follow fallacious beliefs. When a number is deemed to be related with the winning numbers, it is around 3.4 times more likely to be the winner in the
next draw. The MSE is $2.22 \times 10^{-6}$ and the in-sample improvement is 8.2% compared to the naive model.

Finally we apply the models with the calibrated parameters to forecast the proportion of sales in the next 56 draws and compare the performances of the three models. Our results indicate that

- the Fallacious model shows an out-sample improvement by 7.82% while the linear model improves the MSE by 2.9% compared to the Naive Model.

We calculate the mean absolute forecast deviation for each 4D number in the out-of-sample tests. Figure 6 shows the histograms of the forecast deviation of each 4D numbers under the three different models.

![Figure 6: Histograms of absolute forecast error in the three models](image)

In the fallacious model, the forecast error (measured by mean absolution deviation) of around 50% of the 4D number falls within $0.75 \times 10^{-5}$, while only around 40% in the Naive model falls within this range. The fallacious model, after incorporating about 4.5% of the sales with fallacious beliefs,
are able to exhibit significantly better forecast accuracy on the proportion of bets on individual 4D number, than both the Naive model and the linear model.

6 Concluding Remarks

In this paper, we collect data from two different lottery games played in the region and provide empirical evidence that demonstrates the impact of experimental cues in shaping erroneous beliefs in the hot-hand or gambler’s fallacy. Field data show that once people recognize a sufficient number of patterns among the outcomes, they tend to believe that the same outcome or related outcomes will repeat in the future. We then develop a Bayesian updating model to examine how beliefs evolve through time. Our model predicts that when the set of outcomes is relatively large enough for agents to infer sufficient relationships (i.e., higher $p^*(R)$), the hot-hand belief may dominate.

The notion that humans can be manipulated to believe in the hot-hand fallacy through appropriate game design has important ramifications in various fields. First, it provides a behavioral explanation for the “medium prizes puzzle” (Haruvy et al., 2001)—why lottery-game operators typically offer prize distributions of a few large prizes and a large number of medium ones. This is surprising, especially if we assume that gamblers are typically risk-seeking. There are two common explanations for this observation. One relies on prospect theory: A large number of medium prizes reduces the probability of losing from near certainty to some smaller probability. Another explanation follows the line of adaptive learning, in that human behavior is best captured by simple adaptive learning models, and actions that did better in the past will tend to be adopted more frequently compared to actions that did worse. Thus the presence of medium prizes slows down the agent’s inclination to gamble less. Both explanations, however, fail to account for the decision biases in the gambler’s and hot-hand fallacies. They do not explain, for instance, why players in these games normally prefer to choose their own numbers to bet on. Our paper provides another explanation: That a large number of medium prizes can induce more players to believe in the hot-hand fallacy. This reduces players’ inclination to quit the game and also increase to bet on those numbers they believe to have a larger probability of winning.

Second, it provides guidelines for designing lottery games that will induce desirable behavior. There is a recent trend for governments to encourage good civic behavior through the use of lottery
games. Richard Thaler has written about the merits of this approach.\textsuperscript{11} New Taipei City in Taiwan recently initiated a lottery as an inducement for dog owners (and other citizens) to clean up after their pets in order to win gold ingots worth as much as $2000. And, as previously noted, the Singapore government is experimenting with an incentive scheme in which commuters earn credit for each journey taken (with triple credit for off-peak journeys) for a chance to win cash prizes in weekly lotteries. Our study highlights two features that will make these lottery games more effective at influencing behavior: (1) there should be a sufficiently large number of medium prizes (to induce more hot-hand believers), and (2) some mechanism that allows the players to bet on numbers they believe to have a higher probability of winning—and thus increase their incentive to participate in lottery games—is desirable. One way to do this is to use personalized numbers (instead of random numbers) that players can easily relate. The Dutch government uses this principle very effectively; one of its state lotteries is based on postal codes. The idea is to make use of the near-miss effect, and also to exploit the lucky-store effect that has been shown to exist in various lottery games. The chances of winning may be falsely believed to be higher if the postal code has been drawn before (or is a near miss) in the previous draws.

Our results also shed light on a question raised by Guryan and Kearney (2008): Why do lottery players believe that lightning will strike twice with lottery vendors, but not with numbers? Guryan and Kearney have documented a sharp increase in sales for stores after selling a winning lottery ticket in the Lotto Texas game, while Clotfelter and Cook (1993) have documented that sales in a pick-3 game falls sharply after the number is drawn. One possible explanation offered by Guryan and Kearney is along the lines of an animate/inanimate versus intended/non-intended distinction: The winning number is drawn from a mechanical device, but the location of the winning outlet is chosen deliberately by the person buying the winning ticket. Another possible explanation is related to the pattern seeking introduced in this paper: Instead of millions of lottery numbers, there were only 669 retail outlets in Texas, and 68 had sold winning jackpots over the 2.5 years analyzed in the study. Furthermore, the data indicate that the lucky-store effect can linger for as long as 40 weeks after a store sells a the winning ticket. This means that the memory of a store's selling a winning ticket remains for a long time, and thus has a higher chance of being associated with pattern-seeking behavior. This is enough to induce some players to believe in the hot-hand

phenomenon, and lead to an increase in sales for the lucky store.

References


A Proof of Proposition 1

Shalizi (2009) proves Proposition 1, subject to certain conditions, in a very general form—in particular, for probability laws well beyond the Markovian case, and for a continuum of theories, parameterized in an infinite dimensional space. Note that in our restricted setting, the proof of the proposition can be obtained directly via a simple method, which we describe below. This method will be used for the proof of Proposition 2.

The posterior probabilities are obtained through Bayes’ Law.

\[
\mu^l_t(X_t, \ldots, X_1) = \frac{\mu^l \cdot p^l(X_t, \ldots, X_1)}{\sum_k \mu^k \cdot p^k(X_t, \ldots, X_1)}
\]

\[
= \left[ 1 + \sum_{k \neq l} \frac{\mu^k}{\mu^l} \cdot \frac{p^k(X_1)}{p^l(X_1)} \times \prod_{r=2}^{t} \frac{p^k(X_r|X_{r-1})}{p^l(X_r|X_{r-1})} \right]^{-1}.
\]

Since

\[
\frac{\mu^k}{\mu^l} \cdot \frac{p^k(X_1)}{p^l(X_1)} > 0,
\]

if, for some \( k \neq l \),

\[
\prod_{r=2}^{t} \frac{p^k(X_r|X_{r-1})}{p^l(X_r|X_{r-1})} \to \infty
\]

almost surely, then \( \mu^l_t(X_t, \ldots, X_1) \to 0 \) almost surely.
Let $Y_t = (X_{t+1}, X_t)$. Under $p^*$ the process $\{Y_t\}_{t=1}^{\infty}$ is a time-homogeneous stationary Markov process, with stationary distribution $\pi(x, y) = p^*(y)p^*(x|y)$. Let $f(x, y) = \log[p^k(x|y)/p^l(x|y)]$, and consider the sum

$$S_t = \sum_{r=1}^{t} f(Y_r).$$

Clearly, if $S_t \to +\infty$, then condition (8) is satisfied. First, suppose that for some $k \neq l$, $D_{KL}(p^*||p^l) > D_{KL}(p^*||p^k)$. By the ergodic theorem of Markov processes, $S_t/t$ converges almost surely to

$$\sum_{x,y \in S} \pi(x, y)f(x, y) = D_{KL}(p^*||p^l) - D_{KL}(p^*||p^k) > 0.$$

Therefore, $S_t \to +\infty$ and condition (8) is satisfied almost surely, and so $\mu_l^t(X_t, \ldots, X_1) \to 0$ almost surely.

**B Proof of Proposition 2**

From Proposition 1, we already know that, for every $l \not\in \mathcal{M}$, $\mu_l^t(X_t, \ldots, X_1) \to 0$ almost surely.

Consider the function $f$ and processes $\{Y\}_{t=1}^{\infty}$ and $\{S\}_{t=1}^{\infty}$ as defined in the proof of Proposition 1. Assume there are $l, k \in \mathcal{M}$, $l \neq k$. Then $E[f(Y_t)] = D_{KL}(p^*||p^l) - D_{KL}(p^*||p^k) = 0$ for every $t$.

Fix a pair of states $(x, y)$, and define $T^{(n)}$ to be the $n$-th passage time for $(x, y)$: $T^{(0)} = 0$ and

$$T^{(n+1)} = \inf\{t > T^{(n)} \mid Y_t = (x, y)\}.$$

with $\inf\emptyset = \infty$. Every pair of states is recurrent and therefore, almost surely, $T^{(n)} < \infty$ for every $n$. Define the random process $\{Z_t\}_{t=1}^{\infty}$ by

$$Z_t = \sum_{r=T^{(t-1)}+1}^{T^{(t)}} f(Y_r),$$

if both $T^{(t-1)}$ and $T^{(t)}$ are finite, and $Z_t = 0$ otherwise. The process $\{Z_t\}_{t=1}^{\infty}$ is identically and independently distributed at every period $t \geq 2$. Moreover $E[Z_t] = 0$, while $Z_t$ is bounded with finite but positive variance, since $p^k \neq p^l$. Therefore, using, for example, the law of the iterated
logarithm, we get that almost surely,

$$\limsup_{t} \sum_{r=1}^{t} Z_r = +\infty.$$  

Since \( S_T(t) = \sum_{r=1}^{t} Z_r \) whenever \( T(t) < \infty \), we conclude that with probability one,

$$\limsup_{t \to \infty} S_t = +\infty,$$

and therefore, by the same argument that concludes the proof of Proposition 1, with probability one,

$$\liminf_{t \to \infty} \mu_l(t, X_t, \ldots, X_1) = 0.$$  

### C Complements to the Proofs of Propositions 5

Consider the function defined on \([0, 1]\) by

$$f(x) = (\log \frac{\alpha}{\beta}) x + \log(1 + (\beta - 1)x) - \log(1 + (\alpha - 1)x),$$

where \( \alpha > 1 \) and \( 0 < \beta < 1 \). Thus, we can easily calculate the value of \( f(x) \) at \( x = 0, 1 \), respectively, as: \( f(0) = f(1) = 0 \). Take the first-order derivatives of \( f(x) \):

$$f'(x) = (\log \frac{\alpha}{\beta}) x + \frac{\beta - 1}{1 + (\beta - 1)x} - \frac{\alpha - 1}{1 + (\alpha - 1)x}.$$  

Thus the first-order derivative at \( x = 0 \) and \( x = 1 \) are: \( f'(0) = \log \frac{\alpha}{\beta} - (\alpha - \beta) \); \( f'(1) = \log \frac{\alpha}{\beta} - \frac{\alpha - \beta}{\alpha \beta} \).

Take the second-order derivative of \( f(x) \):

$$f''(x) = \frac{\alpha - \beta}{[1 + (\alpha - 1)x]^2[1 + (\beta - 1)x]^2}[(\alpha - \beta - 2) - 2(\alpha - 1)(1 - \beta)x].$$

Under condition that \( \alpha + \beta - 2 \leq 0 \), \( f''(x) \leq 0 \) for any \( x \in [0, 1] \), i.e., \( f(x) \) is concave in \([0, 1]\). Thus, \( \forall t, t \in (0, 1), f(t) > f(0) + tf(1) = 0 \). Similarly, under the condition \( \alpha + \beta - 2 \geq 2(\alpha - 1)(1 - \beta) \), \( f(x) \) is convex in \([0, 1]\). Thus, \( \forall t, t \in (0, 1), f(t) < 0 \).

Under the conditions \( 0 \leq \alpha + \beta - 2 \leq 2(\alpha - 1)(1 - \beta) \), define \( x^* = \frac{\alpha + \beta - 2}{2(\alpha - 1)(1 - \beta)} \). Thus, \( f(x) \) is
that we prove that there exists a unique \( x_0, x_0 \in (0,1) \) that solves \( f(x) = 0 \) such that \( f(x) < 0 \) when \( x \in (0,x_0) \) and \( f(x) > 0 \) when \( x \in (x_0,1) \).

Since \( f(0) = 0 \) and \( f'(0+) < 0 \), we can easily get that \( f(0+) < f(0)0 \). Similarly, \( f(1) = 0 \) and \( f'(1-) < 0 \) imply that \( f(1-) > 0 \). The continuity of function \( f(x) \) on \([0,1]\) implies that there exists such a \( x_0 \) that \( f(x_0) = 0 \). Therefore, there exist \( x_1, x_1 \in (0,x_0) \) and \( x_2, x_2 \in (x_0,1) \) such that \( f'(x_1) = 0 \) and \( f'(x_2) = 0 \). Next, we will prove that \( f(x) \) is strictly increasing in \([x_1, x_2]\). First, it is obvious that \( x^* \) is in \([x_1, x_2]\). Then \( f'(x) \) is increasing in \([x_1, x^*]\) and decreasing in \([x^*, x_2]\). Given that \( f'(x_1) = 0 \) and \( f'(x_2) = 0 \), \( f'(x) > 0 \) for all \( x \in (x_1, x_2) \). There exists one and only one \( x_0, x_0 \in (x_1, x_2) \) such that \( f(x_0) = 0 \). Due to the property of \( f(x) \), we know that for all \( x \in (0, x_1) \), \( f'(x) \) is increasing, thus \( f'(x) < 0 \), which implies that \( f(x) < f(0) = 0 \). So we can prove that \( f(x) < 0 \) for all \( x \in (0, x_0) \). Similarly, we can prove that \( f(x) > 0 \) for all \( x \in (x_0, 1) \).

This completes the proof of Proposition 5.

Conditions \( \alpha - \beta < \log \frac{a}{\beta} < \frac{\alpha - \beta}{\alpha^3} \) imply that \( f'(0+) > 0 \) and \( f'(1-) < 0 \). We are going to prove that \( f(x) > 0, \forall x \in (0,1) \).

Since \( f''(x) \geq 0 \) for all \( x \in (0, x^*] \), take Taylor expansion with regard to \( x, x \in (0, x^*] \) at \( x = 0 \):

\[
f(x) = f(0) + f'(0)x + f''(\theta x^*) \frac{(\theta x^*)^2}{2!} > 0.
\]

Since \( f(x) \) is concave \( \forall x \in (x^*, 1) \), then for all \( x \in (x^*, 1) \),

\[
f(x) = f(t \cdot x^* + (1-t) \cdot 1) > t \cdot f(x^*) + (1-t) \cdot f(1) > 0.
\]

So \( f(x) > 0 \) for all \( x \in (0, 1) \).

Under the conditions \( f'(0+) < 0 \) and \( f'(1-) > 0 \), i.e., \( \frac{\alpha - \beta}{\alpha^3} < \log \frac{a}{\beta} < \alpha - \beta \), we are going to prove that \( f(x) < 0, \forall x \in (0,1) \). Similarly, we can first prove that \( f(x) < 0 \) for all \( x \in [x^*, 1) \). Then, due to the convexity of \( f(x), x \in (0, x^*) \), we can prove that \( f(x) < 0, \forall x \in (0, x^*) \).

Finally, we will prove that \( f'(0+) > 0 \) and \( f'(1-) > 0 \) does not occur under the current conditions. Since \( f(x) \) is convex when \( x \in (0, x^*) \), i.e., \( f'(x) \) is increasing, so \( \forall x \in (0, x^*), f'(x) > 0 \).
\( f'(0+) > 0 \). Similarly, since \( f'(x) \) is decreasing when \( x \in (x^*, 1), f'(x) > f'(1-) > 0 \). So \( f(x) \) is increasing in \([0, 1]\), which contradicts \( f(0) = f(1) = 0 \).