The Gambler's Fallacy and the Hot Hand v4

Austin Bradley

October 2023

Abstract

The Hot Hand and Gambler's fallacies are both well established biases in probabilistic reasoning, referring to incorrect expectations of positive and negative auto-correlation between outcomes. If both of these opposing patterns of bias arise from the same underlying error in reasoning, what causes one to take precedence over the other?

In existing experimental and empirical work, the variety of methods and designs makes it difficult to compare results across studies. In this paper, we report the results of an experiment in which we ask subjects to predict coin tosses after observing previous outcomes, varying both the question used to elicit their predictions and the visual presentation of the past outcomes. We find the same subjects alternating between the Hot Hand and Gambler's fallacies across transparently equivalent treatments. When asked to predict individual outcomes, subjects act in accordance with the Gambler's Fallacy. When asked to state their beliefs as probabilities, they favor the Hot Hand. This presents a challenge to understanding biased reasoning about probability as it implies that people use different subjective models to answer theoretically identical questions.

1 Introduction

Many economic decisions require a person to predict outcomes of random processes. To maximize their expected utility, decision makers must infer the likelihoods of future outcomes from past information. Even under ideal circumstances, this inference process may be highly complex. As a result, either through errors or the use of heuristics, humans decision makers may hold biased beliefs about probability. Numerous experiments have demonstrated that humans differ significantly from the rational Bayesian benchmark, expecting sequences of random data to appear "more random" than the true distribution (see Bar-Hillel and Wagenaar (1991) for a review). For sequences of Independent and Identically Distributed (IID) random variables with binary outcomes, this means that people expect more reversals and thus judge streaks of consecutive outcomes to be excessively unlikely. The expectation of excessive randomness is most commonly attributed to a reliance on the representativeness heuristic to simplify complex inference about random processes — the belief that even short sequences of random data should maintain the features of the long-run distribution (Tversky and Kahneman, 1971). For sequences of IID random variables, we classify errors in judgement as manifestation of either the Gambler's or Hot Hand fallacy. The former is a false expectation of negative correlation between outcomes, and the latter is the unfounded expectation of positive correlation. For flips of a fair coin, representativeness leads to the Gambler's Fallacy because decision makers anticipate reversal of the most recent outcome as it would better approximate the long-run expected frequencies. The Hot Hand also arises from representativeness because the observer over estimates the rarity of streaks of outcomes, leading them to infer the coin must be biased (Gilovich et al., 1985). The existence of opposing patterns of bias is well-established, raising the natural question of what causes one to take precedence over the other.

Two recent studies in which subjects predict unambiguously IID binary sequences find that different biases predominate. Benjamin et al. (2017) (BMR) ask their participants to predict the next outcome in a series of coin-tosses after observing a sequence of past outcomes displayed as a string of H's and T's (for Heads and Tails). Subjects bet on the future coin-tosses with asymmetric payoffs for correctly predicting Heads or Tails. In this environment, the Gambler's Fallacy is the dominant bias — subjects believe strongly enough that streaks are more likely to end than persist to overcome the difference in potential payoffs. Frydman and Nave (2017) (FN) use a similar environment — a neutral framing and full transparency about the true IID generating process — and find the opposite: predictions, on average, conform with belief in the Hot Hand Fallacy. Examined together, their results imply a contradiction in subjects' reasoning: they simultaneously believe the next outcome is more likely to be Heads and that the probability of Heads has decreased.

If both of these opposing patterns of bias arise from the same underlying error in reasoning, what causes one to take precedence over the other? In this paper, we report the results of a laboratory experiment investigating this question. Subjects view sequences of IID coin tosses and predict the next outcome. Based on the key differences between BMR and FN, we vary both the incentive structure (predicting single outcomes versus estimating probabilities directly) and the visual presentation of previous outcomes (a random-walk graph versus a string of text). Subjects are informed unambiguously about the true data generating process and that this process is consistent across all treatments. Despite the consistent information and framing, we produce both the Hot Hand and Gambler's fallacies within the same subjects. We connect subjects' responses to the Rabin-Vayanos model of subjective probability (Rabin and Vayanos, 2010), described in detail in Section 2.

Previous authors find that streak lengths (Asparouhova et al., 2009), human versus purely mechanical data generating processes (Burns and Corpus, 2004; Rao and Hastie, 2023), and human intentions (Roney and Trick, 2009; Huber et al., 2010) influence the average direction of bias. These factors are meaningful and consistent with how people encounter probability in their day-to-day lives. However, the methods used across these studies vary greatly and present challenges when trying to develop a comprehensive understanding of subjects' decision making process. The studies which find subjects switching between the two patterns generally induce this behavior by presenting identical data framed as the result of different generating processes. Although the experimenter knows the data are IID, this is not communicated to participants and changes in framing leave ample opportunity for subjects to infer unintended information.

Previous studies establish the existence of both the Hot Hand and Gambler's fallacies across a wide variety of domains. Naturally, gamblers themselves often exhibit this bias, avoiding lottery numbers (Clotfelter and Cook, 1993) and roulette outcomes (Sundali and Croson, 2006) which have recently occurred. They continue to avoid recent winners even in parimutuel betting markets where this bias lowers their expected payoffs (Terrell, 1994, 1998). However, gamblers favor the Hot Hand Fallacy with respect to the success or failure of others' predictions. Both in laboratory coin-toss experiments (Huber et al., 2010; Powdthavee and Riyanto, 2015) and in real-world lottery markets (Bou et al., 2016), people are willing to pay for the useless recommendations of recent winners. Similarly, lottery gamblers purchase tickets more frequently from stores that have recently sold winners (Guryan and Kearney, 2008; Lien and Yuan, 2015).

Decisions consistent with the Gambler's Fallacy persist in highly consequential environments. Judges, loan officers, baseball umpires (Chen et al., 2016), professional soccer goalkeepers (Misirlisoy and Haggard, 2014), and research and development managers (Criscuolo et al., 2021) all exhibit decision patterns consistent with expectations of mean reversion. Belief in the Gambler's Fallacy may explain much of the Disposition Effect, resulting in decision makers expecting assets to have negatively autocorrelated returns (Jiao, 2017; Weber and Camerer, 1998). However, investors believe in the Hot Hand Fallacy with respect to fund manager's recent returns (Loh and Warachka, 2012) despite evidence against persistent differences in skill (Chen et al., 2000).

In our experiments, we observe both Gambler's Fallacy and Hot Hand behaviors within the same subjects without changes in framing. Altering the incentive structure changes the overall pattern of bias. Subjects conform with the Gambler's Fallacy when predicting single outcomes but follow the Hot Hand Fallacy when reporting their beliefs as frequencies. Changing the visual presentation of previous outcomes has no significant impact on subjects' choices. Reproducing both patterns of bias while minimizing any potential influence of outside information allows us to more closely examine the underlying sources of the contradictory behaviors.

Subjects' choices conform with the Rabin-Vayanos model with full knowledge of the generating process only when predicting single outcomes. When reporting their beliefs as frequencies, their inference is only partially consistent with the Rabin-Vayanos model. Their reported probabilities also imply uncertainty about the true probability of the coin, using the previous outcomes to infer the Bernoulli parameter value. The opposing beliefs across different incentive structures raise important questions about the appropriate methodology for studying subjective beliefs about probability. Extremely small changes in context appear to significantly alter the model humans use to forecast future outcomes.

2 Theory: The Rabin-Vayanos Model

We examine our results in the context of the Rabin-Vayanos (RV) model of biased inference. In the full model, subjects predict the outcomes of an AR1 process with normally distributed errors based on previous observations. Agents infer the model parameters from their observations as rational Bayesians with one critical exception: they hold an explicit, unchanging belief in the Gambler's Fallacy. They expect recent 'luck' to reverse to maintain local representativeness consistent with the Law of Small Numbers (Rabin, 2002). The Gambler's Fallacy arises from this belief explicitly and the Hot Hand Fallacy arises implicitly; their belief in the Gambler's Fallacy leads them to believe that long streaks of errors in the same direction are excessively unlikely, resulting in over-inference about the mean of the distribution.

In our context, a series of coin tosses with binary outcomes, we use the Fudenberg et al. (2022) adaptation of the RV model to binary outcomes.

The agent observes a sequence of previous coin flips, revealed simultaneously, with each element of the sequence drawn from a Bernoulli distribution with constant parameter θ (the true value being $\theta = .5$, in our case) giving the probability of the coin coming up Heads.

$$s_t \sim Ber(\theta)$$

with a corresponding true error term:

$$\epsilon_t = s_t - \theta$$

In the presence of the Gambler's Fallacy bias, the agent instead believes that each flip follows the distribution:

$$s_t \sim Ber\left(\theta - \alpha \sum_{k=0}^{t} \delta^k (s_{t-k-1} - \theta)\right)$$

with α and δ both being values between 0 and 1 giving the strength of the Gambler's Fallacy bias and the rate of decay over time, respectively. These values are fixed and idiosyncratic to a given agent.

Agents use this biased model of probability to conduct inference and form their predictions about the future. Apart from their persistent belief in the Gambler's Fallacy (α and δ are fixed) they process information as rational Bayesians. With the goal of predicting the next outcome of the sequence, agents apply this reasoning in two steps. First, agents infer the value of the Bernoulli parameter θ using maximum likelihood, selecting the value of $\hat{\theta}$ which best fits the observed sequence.

$$\hat{\theta} = \underset{\theta \in [0,1]}{\arg\max}\ L(\theta|\mathbf{s}) = \prod_{t=1}^{T} \left((2s_t - 1) \cdot \sum_{n=0}^{t-1} \left((s_{t-n} - \theta) \cdot \sum_{k=1}^{n} \alpha^k \delta^n \binom{n}{k} \right) \right)$$

This estimate, $\hat{\theta}$, equates to the agent's long-run beliefs about the sequence, given their distorted, subjective model. Belief in the Gambler's Fallacy leads to *over-inference*, consistent with the Hot Hand by inflating the influence of streaks of consecutive outcomes.

In the second step, agents predict the immediate following outcome, now explicitly incorporating the Gambler's Fallacy distortion, evaluated relative to their long-run belief, $\hat{\theta}$.

$$\hat{\theta_{T+1}} = \hat{\theta} - \alpha \sum_{k=0}^{T} \delta^k (s_{T-k-1} - \hat{\theta})$$

This estimate, $\hat{\theta_{T+1}}$, is the agent's local belief and applies only to the immediate next outcome, accounting for their belief that recent, abnormal 'luck' will soon reverse.

Previous tests of this model and estimates of the α and δ parameters focus solely on the second stage of this prediction process, setting aside the first, long-run inference step under the assumption that subjects are certain of the true generating process apart from their explicit Gambler's Fallacy bias. This is justified because subjects are informed of the true parameter value ($\theta = .5$) and that the outcomes are IID, thus they know the long-run behavior with certainty and have no need to infer. Their long-run beliefs should be constant and equal to the true parameter $\hat{\theta} = \theta$. However, the second step remains and they make their prediction, accounting for the Gambler's Fallacy distortion measured against the true parameter: $\hat{\theta_{15}} = .5 - \alpha \sum_{k=1}^{14} \delta^k (s_{14-k-1} - .5)$

As the Hot Hand arises from the first step, only the Gambler's Fallacy is possible under this assumption. However, this is inconsistent with the existing literature, finding numerous examples of behavior consistent with the Hot Hand Fallacy in IID environments.

Without the parameter certainty assumption, the model has no closed-form solution for the impact of a particular piece of information on the overall beliefs and behavior of the agent. Because of this, we employ numerical methods to simulate beliefs formed through this model and establish expected results if subjects in our experiments behave consistently with the RV model.

For the case where agents are uncertain about the true Bernoulli parameter, we use numerical methods to estimate beliefs in response to observing each possible sequence of 14 coin tosses for the parameter certainty case, values are calculated directly. We use $\alpha = .2$ and $\delta = .7$ for Gambler's Fallacy parameters based on previous estimates (Benjamin et al., 2017). Using these estimates, we examine the average effect of Heads occurring in each of the 14 possible positions on both the agent's long-run and local beliefs: $\hat{\theta}$ and $\hat{\theta_{15}}$. To do so, we regress the simulated beliefs on dummy variables, each representing Heads in the 14 possible positions.

$$\hat{\theta}(\mathbf{s}) = \beta_0 + \sum_{i=1}^{14} s_i \cdot \beta_i$$

The point estimates for the coefficients on each of the 14 positions, both of the overall belief and the local prediction $\hat{\theta}_{15}$, are presented in figure 1.

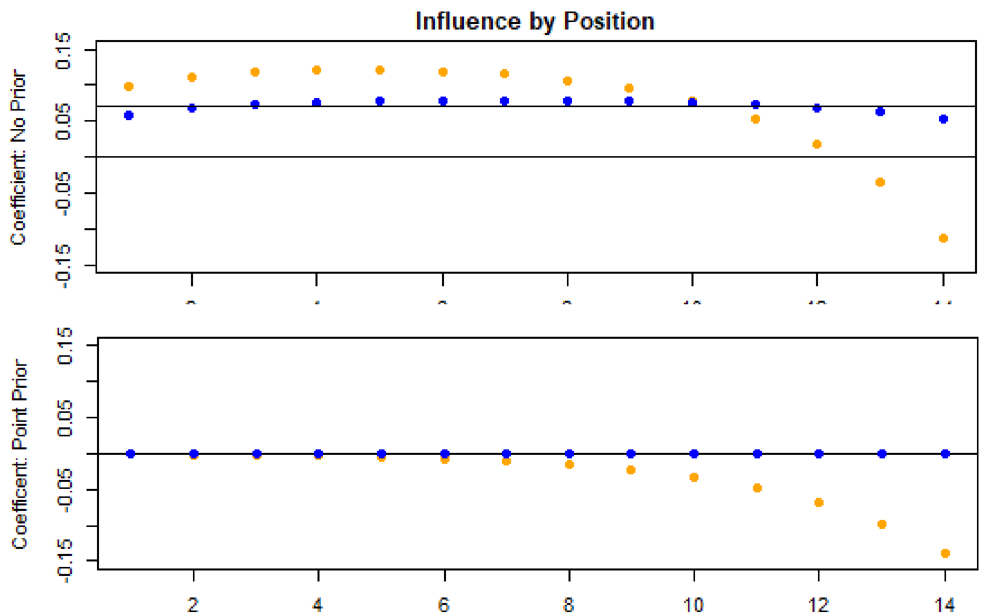

Figure 1: Regression model estimates for average effect of a Heads in each position on the decision maker's estimate of the likelihood of Heads. Long-run belief, $\hat{\theta}$ is plotted in blue and the local belief about the immediate following outcome, $\hat{\theta}_{15}$ is plotted in orange. The black reference line plots the weight placed by a rational Bayesian for comparison.

For a rational Bayesian decision maker, unaffected by the Gambler's Fallacy beliefs, the weight assigned to each of the 14 positions would be equal and positive (or zero in the case of parameter certainty). A few key features emerge in the expected responses between the two cases. First, as stated previously, there is no possibility of the Hot Hand Fallacy under parameter certainty. As such, all coefficient estimates in the local belief model are negative, with the magnitude steadily decaying for less recent outcomes. The long-run belief coefficient is trivially zero for all positions as the agent is assumed to conduct no long-run inference.

When uncertain about the Bernoulli parameter, the agent's pattern of inference differs in several key aspects. First, while the influence on the local beliefs remains negative in the positions in the most recent past, they become positive in earlier positions, exceeding the magnitude of the rational Bayesian baseline. The Gambler's Fallacy belief inflates the influence of consecutive outcomes and reduces the influence of reversing outcomes, leading to this over-inference. Further, the influence on long-run beliefs is always positive, in the direction of the Hot Hand. This influence is strongest near the middle of the sequence, exceeding the magnitude of a rational Bayesian, conducting inference without the Gambler's Fallacy distortion. Counter-intuitively, the coefficient estimates for local beliefs are positive and greater in magnitude than for long-run beliefs. The positive influence of the early outcomes on long-run beliefs also decreases the magnitude of the explicit Gambler's Fallacy effect when agents form their local beliefs.

3 Hypotheses

In the context of this model of subjective probability, we conduct an experiment to examine the causes of the opposing biases observed in BMR and FN. To do so, we ask subjects to predict the outcome of a coin-toss after observing a sequence of 14 prior results. We employ a 2 by 2 design, with each factor corresponding to one of the notable design differences between BMR and FN. First, we vary the incentive structure used to elicit subjects' predictions, either asking subjects to either bet on the specific, binary outcome of the next flip as in BMR or asking for an explicit frequency estimate as in FN. Additionally, we vary the visual presentation of the previous outcomes, presenting them as strings of text (ex. HTHT...) as in BMR or as a random walk graph as in FN. Using subjects responses from this game, we test three primary hypotheses.

- Hypothesis 1: The incentive structure used to elicit beliefs matters. Subjects' predictions, on average, imply belief in the Gambler's Fallacy when making binary predictions and imply belief in the Hot Hand Fallacy when reporting probabilities directly.
- Hypothesis 2: Subjects process information differently depending on the visual presentation of previous outcomes. They favor the Hot Hand when shown past flips graphically and the Gambler's Fallacy when shown a string of text. We test this against the null hypothesis that regression coefficients for models fit on data from each presentation type are equal.
- Hypothesis 3: The subjects do not fully believe the stated Bernoulli parameter, leading to the possibility of the Hot Hand Fallacy. We test this against the null hypothesis that the regression coefficients for all 14 positions are less than or equal to zero; none of the positions are positively correlated with the agent's predictions.

4 Methods

Our experiment consists of an online task conducted on Amazon MTurk. This online labor market allows requesters to post small jobs referred to as Human Intelligence Tasks (HITs) which individuals then perform. Typical HITs are extremely brief with many taking only a few seconds to complete. Using this system offers us a large, diverse pool of potential participants. Results from experiments conducted on this platform are similar to those conducted in a laboratory setting (Horton et al., 2011).

Our task asks individuals to make predictions about random binary events framed as coin tosses. We use four treatments involving two different incentive structures and two different visual presentations of sequences of flips. Participants are informed that they will be asked questions about random events based about a data-set generated prior to the start of the task. This data-set consists of one million realizations of sequences of 15 simulated tosses of a fair coin. From this pool of sequences, we select one and show the participant the first 14 flips. Participants are then asked to make predictions about the result of the 15th flip. Subjects encounter two incentive structures across Bet and Frequency treatments. Frequency questions ask participants to predict the frequency that they expect Heads to occur and others ask them to bet on the next outcome. This method of sequence generation, the questions used to elicit beliefs, and the instructions are adapted from Benjamin et al. (2017).

When participants are asked to predict the frequency, they are paid based on how close their estimate is to the true frequency across our pool of sequences. Their payoff is calculated as:

$$ECU = 100 - |\text{Estimated frequency} - \text{True frequency}|$$

with ECU being converted to final payment at a rate of 3000 ECU = $1.00.

When participants bet on the next outcome, the payoffs for correctly predicting heads or correctly predicting tails are asymmetric to reveal strictly biased beliefs. The payoff for correctly predicting heads is always 50+x ECU and 50-x ECU for tails, with x being randomly drawn from -5 to 5. This incentive structure is also adapted from BMR.

We use two methods to present participants the 14 flip sequences. In 'Text' treatments, participants are shown a string of H's and T's. In 'Graph' treatments, the previous flips are converted into a graphical representation where heads is shown as an upward move of one unit on a time series graph and tails as a downward movement. This presentation method is designed to closely follow the environment found in the economic decision task of FN.

In designing the graphical presentation, we take care to avoid features that may impact subject's beliefs about the generating mechanism. Specifically, we avoid labeling the axis and we ensure that there is ample space left between the maximum and minimum values observed. This is done to avoid giving the impression that the sequence is approaching an upper or lower bound of possible values which may push participants towards the Gambler's Fallacy. This also avoids the risk that participants may assign special significance to particular numerical values e.g. believing that a sequence is unlikely to cross from positive to negative values.

5 Results

5.1 Descriptive Statistics

Our data include observations from 201 individuals. Of those 201 subjects, 18 of them gave the unique expected value maximizing prediction for every question. Participants in our study earned $1.00 for participation as well as a bonus based on the accuracy of their predictions during the experiment. A typical participant earned $1.25 in bonus payment. The task included 60 questions and the average total participation time was 11.3 minutes.

Each subject predicts the 15th outcome through four blocks of 15 questions. Their participation time during the prediction task is summarized in Table 1. The mean overall time spent during each 15 question block is 88 seconds for a total average participation time of approximately 6 minutes, not including consent forms and initial instructions.

Subjects progress through each block in less time as they progress through the study, having gained familiarity with the task. On average, subjects spend approximately 50% longer on Graph treatment blocks relative to Bet treatments and nearly twice as long answering each frequency question as they do in each betting question. Much of this is likely

Table 1: Mean duration by treatment and order

	Order of treatment				
treatment	1	2	3	4	Total
Graph Bet	150.3	50.1	44.1	55.3	72.8
Graph Freq	151.7	131.0	127.8	147.3	138.4
Text Bet	80.8	56.7	39.9	21.2	50.7
Text Freq	117.6	88.0	89.5	64.9	91.3
Total	121.9	85.4	73.6	72.2	88.3

due to the differences in interfaces with frequency questions requiring subject to type their responses and graph questions requiring scrolling past large images.

Because there is a unique, payoff maximizing choice for all questions, we also briefly verify that our subjects are sufficiently susceptible to bias to warrant analysis. To do so, we examine responses to all frequency questions, grouped at the subject level, and calculate the standard deviation of each participant's reported estimates. Figure 2 plots a histogram of these standard deviations.

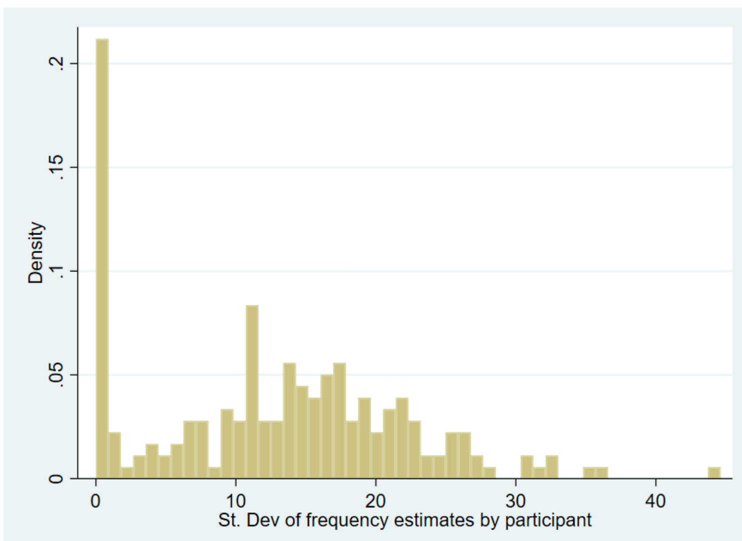

Figure 2: Distribution of the standard deviation of frequency estimates across all prompts for each participant.

The modal subject reports the expected-payoff maximizing response of 50% for every frequency question. However, this account for only a fifth of the total subject pool. The remaining subjects all show some degree of bias, altering their frequency estimates in response to the different sequences of past coin flips.

5.2 BMR and FN Comparison

Our experiment is primarily motivated by the contrasting results from BMR and FN, given their apparent similarity. Although both ask subjects to predict outcomes of a transparently random process with incentive structures penalizing biased beliefs, BMR finds that most subjects follow the Gambler's Fallacy whereas the Hot Hand Fallacy predominates in FN. With the goal of investigate the source of this discrepancy, we first confirm that the results from our equivalent treatments produce the same patterns of bias. In our design, Bet Text and Graph Frequency approximate the methods in BMR and FN respectively.

Table 2: Replication of Frydman-Nave Econ. Decision Task

	(1) Frequency Estimate
ending.streak	0.365**
	(2.65)
Constant	48.12***
	(119.02)
Observations	2998

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Replication of BMR result 2

	Predicts Heads
bet.payoff.heads	0.0474***
	(11.52)
heads.shown	-0.0134**
	(-3.23)
Constant	-1.744***
	(-8.37)
Observations	3011

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Fitting the same regression specifications used by the authors of each of the previous studies, our data agree with both of their respective conclusions. Subjects in our Bet Text treatments, most similar in design to BMR, follow the Gambler's Fallacy with respect to streaks of outcomes at the end of the sequence. There is a negative correlation between the length and direction of the final streak of outcomes and subjects' predictions. This relationship is significant at the 5% level, after controlling for the differences in incentives. We also note, at this point, the strength of the incentive effects. Despite the small size — biased responses lower expected payoffs by a fraction of a cent — subjects respond significantly.

Our Graph-Frequency treatment is most similar to the FN environment, where the authors found that the Hot Hand Fallacy predominates. We fit the same regression specification used by FN and present the results in Table 2. Our results agree with the FN findings, consistent with the Hot Hand Fallacy. When evaluating the average direction of bias with respect to the concluding streak of the sequence, we find that streaks are positively and significantly correlated with the length and direction of the final streak of outcomes. Subjects' frequency estimates imply that streaks are more likely to continue than they are to reverse.

These initial tests confirm that our game designs are similar enough to the previous experiments to reproduce the opposing patterns of bias. Further, variation between our treatments results in subjects switching the direction of their bias. This occurs with identical participants observing identical sequences across the four treatments. The remainder of the results examine these changes in detail.

5.3 Indicator Variable Regressions

We next examine the overall pattern of inference by position with our subjects' choices as we did in our simulations. We use a mixed effects model with standard errors clustered at the subject level. For each of the types of questions, we fit a separate regression model regressing either the subjects' frequency estimates or outcome prediction on indicator variables. For Frequency treatments, we fit:

$$\hat{\theta}(\mathbf{s}) = \beta_0 + \sum_{i=1}^{14} s_i \cdot \beta_i$$

For betting questions, we also include the weight of the payoffs in favor of Heads:

$$\hat{\theta}(\mathbf{s}) = \beta_0 + \sum_{i=1}^{14} s_i \cdot \beta_i + \beta_{bw} \cdot \text{Bet Weight}$$

The results of these regressions are presented in Table 4

Table 4: Predictions vs. Indicator variables by position

	Freq. Graph t-stat	Freq. Text t-stat	Bet Graph t-stat	Bet Text t-stat
flip1	1.81*	0.079	0.43	1.59
flip2	0.94	2.22**	1.46	1.66*
flip3	2.12**	1.84*	1.61	0.050
flip4	-0.94	0.51	-2.61***	-0.81
flip5	1.06	0.56	-2.09**	0.62
flip6	2.80***	2.56**	0.12	0.28
flip7	2.86***	1.98**	-0.22	0.10
flip8	3.81***	1.79*	-1.63	0.40
flip9	3.89***	2.91***	-0.35	-0.23
flip10	4.81***	2.37**	0.83	0.61
flip11	1.21	2.28**	-1.01	-1.18
flip12	1.65*	1.68*	-2.00**	-2.02**
flip13	0.47	-1.02	-1.47	-3.25***
flip14	1.43	-0.56	-0.79	-3.01***
Bet weight			3.84***	4.59***
Observations	2624	2622	2627	2637

Mixed effect estimator; clustered by participant

$*p < 0.10, **p < 0.05, ***p < 0.01$

Examining the coefficient estimates for each model, we first note a stark difference between Frequency and Bet treatments. With Frequency treatments, all significant coefficient estimates ($p < 0.05$) are positive, with the strongest estimates located near the middle of the sequence. This is inconsistent with the Rabin-Vayanos model's prediction of agent's predictions for the immediately following flip. Regardless of whether subjects believe the coin is fair with certainty, decision makers are expected to have their decisions correlate negatively with flips in the recent past.

The regression specification above assumes that the information in each position has a linear effect on the overall decision. However, this is inconsistent with many previous interpretations of the Hot Hand and Gambler's fallacies which place special significance on streaks of outcomes. We estimate the same regression models as above, with the addition of a variable for the length and direction of the streak of consecutive outcomes at the tail end of the sequence. Sequences ending in Heads are encoded as positive values and streaks of Tails are encoded as negative. Table 5 presents the results.

Table 5: Predictions vs. Indicator variables by position with streak control

	Freq. Graph t-stat	Freq. Text t-stat	Bet Graph t-stat	Bet Text t-stat
flip1	0.29	-0.62	0.61	1.63
flip2	0.91	2.16**	1.42	1.62
flip3	2.31**	1.94*	1.61	0.056
flip4	-1.33	0.24	-2.60***	-0.97
flip5	0.58	0.30	-2.08**	0.61
flip6	3.30***	2.77***	0.041	0.18
flip7	2.94***	2.01**	-0.019	0.35
flip8	2.49**	1.08	-1.68*	0.28
flip9	3.63***	2.76***	-0.38	-0.29
flip10	3.70***	1.73*	0.93	0.80
flip11	0.41	1.66*	-1.10	-1.32
flip12	1.11	1.35	-1.99**	-1.99**
flip13	0.43	-1.06	-1.53	-3.28***
flip14	0.54	-1.00	-0.83	-3.06***
Streak	1.76*	1.08	0.44	0.60
Bet weight			3.03***	3.53***
Observations	2624	2622	2627	2637

Mixed effect estimator; clustered by participant

$*p < 0.10, **p < 0.05, ***p < 0.01$

Including the streak control does not alter the overall results in the regressions model. In the frequency treatments, regardless of the visual presentation method, all significant coefficient estimates are positive, consistent with the Hot Hand Fallacy. As in the previous regression specifications, the significant effects are concentrated around the middle of the sequence. Flips near the beginning and end of the sequence have no statistically significant impact on subjects' frequency estimates.

In betting treatments, the added streak variable does not alter the tendency towards the Gambler's Fallacy observed in the previous specification. All significant coefficient estimates are negative across both visual presentations. In text treatments, these coefficients are concentrated near the end of the sequence. There is no clear pattern in the location of the significant coefficients in the graphical treatments.

With and without the controls for streaks, the results have the same implications for our hypotheses. First, the evidence supports Hypothesis 1: the incentive structure used to elicit beliefs matters. When subjects bet on individual outcomes, all significant coefficient estimates are negative, consistent with the Gambler's Fallacy. In treatments where subjects predict frequencies, all significant coefficient estimates are positive, consistent with the Hot Hand. Second, we find no evidence supporting Hypothesis 2. There is no evidence of differences in responses depending on whether Text or Graph presentations are used. We test this further in the following subsection. Finally, the differing responses to the changing incentive structures provide evidence regarding Hypothesis 3: the subjects' choices are consistent with parameter certainty under the RV model. In Frequency questions, all significant coefficient estimates are positive. This is only possible in the RV model if decision makers attempt to infer the model parameters given the observed information. In Betting questions, we find no evidence to reject the assumption of parameter certainty. All significant coefficient estimates are negative, consistent with the RV model's agents' belief in reversion to the mean without uncertainty about the model parameters.

5.4 Grouped flips

The results from the indicator variable regressions reveal diverging responses to the same sequences depending on incentive structure. In frequency questions, predictions are positively correlated with flips near the middle of the sequence. The patterns are less clear in betting questions. In both Bet-Graph and Bet-Text treatments, all significant coefficient estimates are negative with negative estimates primarily near the end of the sequence. However, we do not have the statistical power to confidently establish this pattern in the Bet-Graph treatment.

To address this, we consider an alternative specification, combining positions in the sequence into groups of four, reducing the number of independent variables in our regression model. In this new regression specification, we use the number of Heads in the first four positions, the middle four positions, (6 through 9,) and final four positions. By aggregating the flips together, we increase the power we have to detect the patterns observed in the indicator variable regressions: negative significant coefficients near the end of the sequence in betting questions and positive, significant coefficients in the middle in frequency treatments. The results are presented in Table 6

Table 6: Predictions vs. Grouped flips

	Freq. Graph t-stat	Freq. Text t-stat	Bet Graph t-stat	Bet Text t-stat
First 4	1.82*	3.04***	-0.71	1.04
Middle 4	6.45***	5.20***	-0.55	1.32
Last 4	2.43**	1.66*	-3.73***	-6.84***
Bet weight			6.48***	7.01***
Observations	2624	2622	2627	2637

Mixed effect estimator; clustered by participant

$*p < 0.10, **p < 0.05, ***p < 0.01$

The alternative specification agrees with the results of the indicator variable regressions. In frequency question treatments, regardless of visual presentation, the significant coefficient, ($p < 0.05$,) estimates are all positive, consistent with belief in the Hot Hand. For betting questions, the significant coefficient estimates are all negative, excluding the significant positive estimate for the bet weight control. One notable difference in this model is that we get a much clearer picture of the main source of the Gambler's Fallacy tendencies in the Bet-Graph treatments compared to the individual position regressions. Bet-Graph treatments follow the same patterns as Bet-Text treatments, with significant coefficient estimates consistent with the Gambler's Fallacy in the final four flips but no significant response to earlier flips in the sequence.

As in the indicator variable regression model, we estimate our grouped model again with a control variable for the length and direction of the ending streak. The results for each treatment are presented in Table 7

Table 7: Predictions vs. Grouped flips with streak control

	Freq. Graph t-stat	Freq. Text t-stat	Bet Graph t-stat	Bet Text t-stat
First 4	1.38	3.02***	-0.72	1.03
Middle 4	6.84***	5.17***	-0.57	1.29
Last 4	1.34	1.62	-3.71***	-6.79***
Streak	3.11***	-0.16	-0.23	-0.23
Bet weight			6.34***	6.85***
Observations	2624	2622	2627	2637

Mixed effect estimator; clustered by participant

$*p < 0.10, **p < 0.05, ***p < 0.01$

Including a control variable for streak length does not alter the primary results from the previous regression specifications. In frequency questions, all significant coefficient estimates are positive with the strongest effect near the middle of the sequence. In betting questions, all significant coefficient estimates are negative with only flips in the final positions affecting subjects' decisions.

5.5 Visual Presentation Comparison

Using the grouped flip specification, we also test explicitly for differences across graphical treatments. We include a dummy variable for the visual treatment, interacted with the other variables to test whether certain features in the model have unequal impacts depending on the visual presentation.

Table 8: Predictions vs. Grouped flips with graph interactions

	(1) Freq. Questions	(2) Betting Questions
Graph	-1.702	0.00852
	(-0.73)	(0.24)
First 4	1.023***	0.00173
	(2.64)	(0.24)
Middle 4	1.602***	0.0101
	(4.17)	(1.19)
Last 4	0.497	-0.0534***
	(1.06)	(-4.61)
Graph × First 4	-0.304	-0.0129
	(-0.59)	(-1.37)
Graph × Middle 4	0.239	-0.0174*
	(0.51)	(-1.68)
Graph × Last 4	0.477	0.0202*
	(0.95)	(1.82)
Bet weight		0.0505***
		(7.73)
Observations	5995	6013

t statistics in parentheses

Mixed effect estimator; clustered by participant

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Including the interaction between the graphical treatment and the groups of four positions, we find no evidence that visual presentation affects choices. The previous results hold with regard to the treatment types — significant, negative coefficients for the last four flips in betting questions and positive, significant coefficients throughout in frequency questions.

There is no significant impact from the interactions of a dummy variable indicating graphical presentations and the variables for total heads in the first, middle, and final four positions.

As in the indicator variable regression specifications, we estimate the grouped flip model with a dummy variable indicating graphical treatments, including a control variable for streaks. The results are presented in Table 9.

Table 9: Predictions vs. Grouped flips with graph interactions with streak control

	(1) Freq. Questions	(2) Betting Questions
Graph	-1.204	0.00922
	(-0.52)	(0.26)
First 4	1.041***	0.000846
	(2.61)	(0.12)
Middle 4	1.656***	0.00960
	(4.28)	(1.14)
Last 4	0.469	-0.0545***
	(0.99)	(-4.68)
Streak	0.0442	-0.00150
	(0.26)	(-0.59)
Graph × First 4	-0.497	-0.0129
	(-0.94)	(-1.37)
Graph × Middle 4	0.236	-0.0176*
	(0.50)	(-1.70)
Graph × Last 4	0.113	0.0212*
	(0.20)	(1.85)
Graph × Streak	0.360	-0.00128
	(1.31)	(-0.46)
Bet weight		0.0514***
		(7.52)
Observations	5995	6013

t statistics in parentheses

Mixed effect estimator; clustered by participant

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Including streaks in the model does not reveal any significant differences between the visual presentation methods. There are no significant coefficient estimates at the 5% level for any of the interaction terms which include the Graph variable.

The results from the regressions involving grouped flips agree with the results from the indicator variable regressions. First, the evidence from the grouped regressions supports Hypothesis 1, the inference patterns are substantially different in Frequency treatments compared with Bet treatments. We cannot directly compare the coefficients in the two models as the dependent variables differ, but the significant effects differ in direction, not just in magnitude. With betting questions, all significant coefficient estimates are negative and consistent with the Gambler's Fallacy while all significant coefficients are positive, consistent with the Hot Hand with frequency questions. Second, we continue to find no evidence of differences between the visual treatments. We fail to find any statistically significant difference in coefficients fit using Text treatments relative to Graphical treatments. Finally, regarding Hypothesis 3, we find the same mixed results found in the indicator variable regressions. In betting questions, we find no evidence inconsistent with the RV model under parameter certainty. All significant coefficients are negative the GB and TB treatments, consistent with decision makers reacting only to the expectation that "luck" reverses without conducting any parameter inference. However, the evidence in frequency treatments is inconsistent with the RV model under parameter certainty. All significant coefficient estimates are positive, which requires that decision makers are inferring the model parameter, implying that they are not certain about the true values.

5.6 Alternative Features

Our regression specifications thus far have deviated from the previous literature by considering each flip instead of summary statistics describing potentially relevant features of the sequence. The indicator variable regression models reveal a clear discrepancy between beliefs depending on the question used to elicit subject beliefs. However, this assumes an independent, linear effect by each flip in the sequence. This is inconsistent with much of the previous literature which directs attention towards streaks of outcomes and number of reversals in a sequence in addition to the total number or Heads or Tails. To address this, we estimate an alternative regression model, replacing the indicator variables with these common features and their interactions. As before, we also include the weight advantaging heads in betting questions as a control variable.

Table 10: Predictions vs. Common features

	(1) Freq. Graph	(2) Freq. Text	(3) Bet Graph	(4) Bet Text
Reversals	0.339	-0.940	-0.00214	-0.0433**
	(0.38)	(-1.09)	(-0.10)	(-2.06)
Heads	1.100	0.0975	-0.0189	-0.0529***
	(1.48)	(0.14)	(-1.10)	(-3.06)
Streak	0.522	-0.0195	0.0201	0.0433***
	(1.25)	(-0.05)	(1.62)	(3.45)
Reversals × Heads	0.0176	0.184	0.000317	0.00820**
	(0.14)	(1.49)	(0.09)	(2.32)
Reversals × Streak	-0.0334	-0.0237	-0.00571**	-0.0109***
	(-0.47)	(-0.34)	(-2.27)	(-4.30)
Bet weight	0	0	0.0518***	0.0466***
	(.)	(.)	(10.83)	(9.68)
Observations	2998	2997	3002	3011

t statistics in parentheses

Mixed effect estimator; clustered by participant

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Overall, where we find significant coefficient estimates in these new regression specifications, they agree with the findings from the indicator variable model — Frequency treatments tend towards the Hot Hand Fallacy while Bet treatments favor the Gambler's Fallacy. For Frequency treatments, the total number of Heads is significantly, positively correlated with the reported frequency estimates for Heads in the Frequency-Graph treatment at the 5% level. There are no significant coefficient estimates in the Frequency-Text treatment.

Bet treatments are more complex. The Gambler's Fallacy is still the predominating bias, but is the result of several factors interacting. In the Bet-Graph treatment, the only feature, other than the control variable for the weight of incentives, significant at the 5% level is the interaction between the number of reversals and the length of the final streak. The estimated effect is negative, consistent with previous findings about the Gambler's Fallacy. The longer the streak, and the more varied the sequence before the beginning of the streak, the more likely the decision maker is to expect the streak to end.

In Bet-Text treatments, we find more features with significant coefficient estimates, several of which appear in contrast to the Gambler's Fallacy pattern previously observed. The length of the concluding streak is significant at the 5% level and positive. However, this is counteracted by the interaction between streak length and number of reversals, which is significant at the 1% level and negative. In our sample, the mean number of reversals in a sequence is 6.6 with a minimum of 4. Even at the minimum, the magnitude of the interaction coefficient exceeds the streak alone coefficient. The Gambler's Fallacy remains the dominant direction of bias. The number of heads has a significant, negative effect which is mitigated by the significant positive impact of the interaction between number of heads and total reversals. Given their respective magnitudes, the hot hand only dominates with respect to the number of heads, when the number of reversals is high.

This alternative regression specification allows us to examine the non-linear features previously found to influence biased probability estimates. Overall, the estimates, when significant conform with the previous, indicator variable regression model but do little to reveal a meaningful alternative understanding. Significant estimates in frequency questions favor the Hot Hand while the model conforms with the Gambler's Fallacy for betting questions for the typical sequence. However, beyond this most general agreement, these estimates do not suggest a coherent alternative model. What we can say with relative confidence is that these regressions provide further support for the key finding in our indicator variable model — subjects approach frequency and betting questions using different methods for each environment.

6 Conclusion

In this paper, we investigate the formation of subjective beliefs about probability. Specifically, we examine the factors which influence the direction of this bias towards either the Gambler's Fallacy or the Hot Hand Fallacy. To do so, we conduct an incentivized experiment where subjects predict future outcomes of an IID random process. These subjects are transparently and unambiguously informed of the true generating process which allows us to examine subjects' beliefs about the random process free from any ambiguity introduced by changing framings. Based on the opposing results of BMR and FN, we vary the incentives used to elicit beliefs as well as the visual presentations of past outcomes. In doing so, we successfully produce behavior consistent with both the Gambler's and Hot Hand fallacies within the same individuals, responding to the same sequences of data.

We find consistent evidence that the incentive structure used to elicit subjects' beliefs changes the overall tendency from the Gambler's Fallacy to the Hot Hand Fallacy. When predicting individual outcomes, (betting whether the next outcome will be Heads or Tails,) decision makers act in accordance with the Gambler's Fallacy. When giving their beliefs directly as probabilities, they favor the Hot Hand. Subjects switch between the Hot Hand and Gambler's Fallacies even though subjects are reacting to identical information across both all treatments, implying directly contradictory beliefs. They appear to simultaneously hold the beliefs that Heads is more likely than Tails and that Tails is more likely than Heads. We find no substantial evidence that the visual presentation method affects subjects' choices.

Overall, the differences between the two elicitation questions used suggest a serious challenge to the current understanding of probability biases. In response to identical information, subjects in our experiments react in discordant ways depending on the way they are asked to report their beliefs. The difference between these questions is, purposefully, extremely subtle. In Bet treatments, we ask subjects to predict the immediate next outcome. In frequency treatments, we ask subjects to predict the frequency of the immediate next flip being heads out of all of the cases where the first 14 flips are the same. We repeat this to emphasize that the two questions are as close to identical as possible. We ask subjects to make their predictions for the same future event given identical information and histories. We do not ask subjects to estimate the long-run frequency of heads if a single sequence continues indefinitely, which would make our results consistent with the existing understanding of the Gambler's Fallacy where the influence of the distortion decreases over time.

Despite the questions' similarity, the results suggest that two discrepancies arise in subjects' implied beliefs. First, betting questions lead respondents to report beliefs consistent with the Rabin-Vayanos model, reporting the belief for the single, immediately following flip. The Gambler's Fallacy predominates and there is no evidence that they doubt the fairness of the coin — they show no tendency to positively correlate their beliefs with previous flips. In contrast, subjects respond to frequency questions in a way consistent with the overall, long-run beliefs for agents in the Rabin-Vayanos model. Their beliefs are correlated positively with previous flips and these effects are strongest near the middle of the sequence — the Hot Hand is the prevailing bias. Further, this behavior implies that subjects doubt the fairness of the coin.

These results reveal a significant challenge towards understanding biased beliefs about probability. The features which are significant to decision makers, and the ways which they respond to them, are difficult to predict ex-ante. Fully understanding these common distortions requires further, detailed investigation.

References

Asparouhova, E., Hertzel, M., and Lemmon, M. (2009). Inference from streaks in random outcomes: Experimental evidence on beliefs in regime shifting and the law of small numbers. *Management Science*, 55(11):1766–1782.

Bar-Hillel, M. and Wagenaar, W. A. (1991). The perception of randomness. *Advances in applied mathematics*, 12(4):428–454.

Benjamin, D. J., Moore, D. A., and Rabin, M. (2017). Biased beliefs about random samples: Evidence from two integrated experiments. Technical report, National Bureau of Economic Research.

Bou, S., Brandts, J., Cayón, M., and Guillén, P. (2016). The price of luck: paying for the hot hand of others. *Journal of the Economic Science Association*, 2:60–72.

Burns, B. D. and Corpus, B. (2004). Randomness and inductions from streaks:"gambler's fallacy" versus" hot hand ". *Psychonomic bulletin & review*, 11:179–184.

Chen, D. L., Moskowitz, T. J., and Shue, K. (2016). Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires. *The Quarterly Journal of Economics*, 131(3):1181–1242.

Chen, H.-L., Jegadeesh, N., and Wermers, R. (2000). The value of active mutual fund management: An examination of the stockholdings and trades of fund managers. *Journal of Financial and quantitative Analysis*, 35(3):343–368.

Clotfelter, C. T. and Cook, P. J. (1993). The "gambler's fallacy" in lottery play. *Management Science*, 39(12):1521–1525.

Criscuolo, P., Dahlander, L., Grohsjean, T., and Salter, A. (2021). The sequence effect in panel decisions: Evidence from the evaluation of research and development projects. *Organization Science*, 32(4):987–1008.

Frydman, C. and Nave, G. (2017). Extrapolative beliefs in perceptual and economic decisions: Evidence of a common mechanism. *Management Science*, 63(7):2340–2352.

Fudenberg, D., Kleinberg, J., Liang, A., and Mullainathan, S. (2022). Measuring the completeness of economic models. *Journal of Political Economy*, 130(4):956–990.

Gilovich, T., Vallone, R., and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive psychology*, 17(3):295–314.

Guryan, J. and Kearney, M. S. (2008). Gambling at lucky stores: Empirical evidence from state lottery sales. *American Economic Review*, 98(1):458–473.

Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental economics*, 14:399–425.

Huber, J., Kirchler, M., and Stöckl, T. (2010). The hot hand belief and the gambler's fallacy in investment decisions under risk. *Theory and decision*, 68:445–462.

Jiao, P. (2017). Belief in mean reversion and the disposition effect: An experimental test. *Journal of Behavioral Finance*, 18(1):29–44.

Lien, J. W. and Yuan, J. (2015). Selling to biased believers: strategies of online lottery ticket vendors. *Economic Inquiry*, 53(3):1506–1521.

Loh, R. K. and Warachka, M. (2012). Streaks in earnings surprises and the cross-section of stock returns. *Management Science*, 58(7):1305–1321.

Misirlisoy, E. and Haggard, P. (2014). Asymmetric predictability and cognitive competition in football penalty shootouts. *Current Biology*, 24(16):1918–1922.

Powdthavee, N. and Riyanto, Y. E. (2015). Would you pay for transparently useless advice? a test of boundaries of beliefs in the folly of predictions. *Review of Economics and Statistics*, 97(2):257–272.

Rabin, M. (2002). Inference by believers in the law of small numbers. *The Quarterly Journal of Economics*, 117(3):775–816.

Rabin, M. and Vayanos, D. (2010). The gambler's and hot-hand fallacies: Theory and applications. *The Review of Economic Studies*, 77(2):730–778.

Rao, K. and Hastie, R. (2023). Predicting outcomes in a sequence of binary events: Belief updating and gambler's fallacy reasoning. *Cognitive Science*, 47(1):e13211.

Roney, C. J. and Trick, L. M. (2009). Sympathetic magic and perceptions of randomness: The hot hand versus the gambler's fallacy. *Thinking & reasoning*, 15(2):197–210.

Sundali, J. and Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgment and Decision making*, 1(1):1–12.

Terrell, D. (1994). A test of the gambler's fallacy: Evidence from pari-mutuel games. *Journal of risk and uncertainty*, 8:309–317.

Terrell, D. (1998). Biases in assessments of probabilities: New evidence from greyhound races. *Journal of Risk and Uncertainty*, 17:151–167.

Tversky, A. and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological bulletin*, 76(2):105.

Weber, M. and Camerer, C. F. (1998). The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior & Organization*, 33(2):167–184.