

# How Experience Confirms the Gambler's Fallacy when Sample Size is Neglected \*

(Working Paper)

Joshua B. Miller and Adam Sanjurjo †‡§

March 23, 2019

## Abstract

The Gambler's Fallacy is the mistaken belief that random sequences have a systematic tendency towards reversal, i.e. that streaks of similar outcomes are more likely to end than continue. Despite broad empirical support for gamblers fallacy beliefs, there exists little formal explanation of *why* such beliefs persist. We present a simple model in which an individual formulates his beliefs about the probability of success given recent success via repeated exposure to random sequences. For each sequence he focuses on the proportion of success given recent success and then updates his beliefs, but (partially) neglects sample size. This results in probability beliefs which, in the limit, are smaller than the true (conditional) probability, i.e. gambler's fallacy beliefs. We discuss the model's novel testable predictions.

**JEL Classification Numbers:** C12; C14; C18;C19; C91; D03; G02.

**Keywords:** Law of Small Numbers; Alternation Bias; Negative Recency Bias; Gambler's Fallacy; Sample Size Neglect.

---

\*The model presented here originally appeared in Section 2 of the former working paper entitled “Surprised by the Gambler's and Hot Hand Fallacies? A Truth in the Law of Small Numbers” (Miller and Sanjurjo 2015). The forthcoming version of that paper, which no longer investigates gambler's fallacy beliefs, is entitled “Surprised by the Hot Hand Fallacy? A Truth in the Law of Small Numbers” (Miller and Sanjurjo 2018).

†Fundamentos del Análisis Económico, Universidad de Alicante.

‡Both authors contributed equally, with names listed in alphabetical order.

§This draft has benefitted from helpful comments and suggestions from Maya Bar-Hillel, Vincent Crawford, Daniel Friedman, Itzhak Gilboa, and Muriel Niederle. We would also like to thank seminar participants at Stanford U., UC Santa Cruz, and U. de Alicante.

# 1 Introduction

*Why, if the gambler’s fallacy is truly fallacious, does it persist? Why is it not corrected as a consequence of experience with random events?* (Nickerson 2002)

A classic result on the human perception of randomness in sequential data is that people believe the outcomes of randomly generated sequences alternate more than they actually do. For example, if a (fair) coin flip lands heads, then a tails is thought to be more likely on the next flip (Bar-Hillel and Wagenaar 1991; Nickerson 2002; Oskarsson, Boven, McClelland, and Hastie 2009; Rabin 2002).<sup>1</sup> Further, as a streak of identical outcomes (e.g. heads) increases in length, it is believed that the alternation rate on the outcome that follows becomes even larger, which is known as the *Gambler’s Fallacy* (Bar-Hillel and Wagenaar 1991).<sup>2</sup> Gambler’s fallacy beliefs are widespread among novice gamblers, with adherents that have included at least one historically eminent mathematician (D’Alembert 1761, pp. 13-14).<sup>3</sup> The fallacy has been commonly attributed to a mistaken belief in the “Law of Small Numbers,” by which large sample properties are incorrectly thought to also hold within small samples (Tversky and Kahneman 1971), so if, for example, several heads flips have occurred in a row, then tails is deemed more likely on the next flip to help “balance things out.”

The opening quote by Nickerson (2002) poses an important question: given that the gambler’s fallacy is an error, why does experience fail to correct it? One explanation is that there may be insufficient incentive, or opportunity to learn, given that people are often mere passive observers of random sequential data, or have little at stake.<sup>4</sup> However, this explanation is unsatisfying as it presupposes no advantage to holding correct beliefs per se, and ignores their option value. Therefore a potentially more satisfying explanation for the persistence of the gambler’s fallacy is one that is capable of addressing how it could be robust to experience.

Based (partially) on the theoretical results of Miller and Sanjurjo (2018), in Section 2 we propose

---

<sup>1</sup>This *alternation bias* is also sometimes referred to as *negative recency bias*.

<sup>2</sup>For simplicity, in the following discussion we assume that a decision maker keeps track of the alternation rate of a single outcome (e.g. for heads,  $1 - \hat{p}(H|H)$ ), which seems especially reasonable for applications in which outcomes appear qualitatively different (e.g. rainy/sunny days). On the other hand, in the case of flipping a fair coin there may be no need to discriminate between an alternation that follows heads, or tails, respectively. In this special case, the overall alternation rate, (# alternations for streaks of length 1)/(number of flips – 1), is expected to be 0.5. Nevertheless, it is easy to demonstrate that the overall alternation rate computed for any other streak length ( $k > 1$ ) is expected to be strictly greater than 0.5.

<sup>3</sup>In particular, D’Alembert famously argued in favor of his gambler’s fallacy beliefs. In response to the problem: “When a fair coin is tossed, given that heads have occurred three times in a row, what is the probability that the next toss is a tail?” D’Alembert argued that the probability of a tail is greater than 1/2 because it is unlikely that a probable event will never occur in a finite sequence of trials (D’Alembert 1761, pp. 13-14); see Gorroochurn (2012, p. 124) for a discussion.

<sup>4</sup>In casino games such as roulette, people make active decisions based on events that are sequentially independent. While there is typically no additional cost to placing one’s bets on an event that hasn’t occurred for some time, rather than another event, the fallacy can be costly if it leads one to bet larger amounts (given that expected returns are negative). See Rabin (2002), Ayton and Fischer (2004), Croson and Sundali (2005), and Chen, Moskowitz, and Shue (2014) for further discussion.

a simple model of how a mistaken belief in the gambler’s fallacy can persist. We consider a decision maker (DM) who repeatedly encounters finite length sequences of “successes” and “failures.” DM begins with prior beliefs regarding the conditional probability of “success,” given that a trial immediately follows  $k$  consecutive successes. Naturally, for each encounter with a finite sequence, DM attends to the success rate on the trials that immediately follow  $k$  consecutive successes, and updates accordingly.

Importantly, when updating his prior, we allow for the possibility that DM focuses on the *strength evidence*, i.e. the success rate on the trials that follow a streak of successes, rather than the *weight of evidence*, i.e. the effective sample size used in the calculation of the success rate. This feature of the model is consistent with results on how people weight evidence when updating their beliefs (Griffin and Tversky 1992).<sup>5</sup> We find that if DM does not fully account for the sample size, then DM’s posterior will converge to point mass on a belief that the probability of success given recent success is less than the true probability. The magnitude of DM’s underestimate will depend on the degree to which DM neglects sample size, as well as the length of sequences that DM typically encounters, and streak lengths that DM considers notable.

We are not the first to propose that beliefs may be influenced by the statistical properties of finite samples. For example, in the psychology literature, it has been proposed that associations learned via experience may be influenced by the smallness of samples that people are typically exposed to (Kareev 1995a,b, 2000; Kareev, Lieberman, and Lev 1997).<sup>6</sup> More recently, and closely related, Hahn and Warren (2009) conjecture that the gambler’s fallacy may arise from the small sample properties of the distribution of finite length strings, which relates to the *overlapping words paradox* (Guibas and Odlyzko [1981]; also see Appendix B). In particular, the authors note that in a sequence of length  $n > 4$ , the pattern HHHT is more likely to occur than the pattern HHHH, which may explain why people believe that the probability of tails is greater than 1/2 after three heads in a row. While this conjecture has sparked some debate (Hahn and Warren 2010a,b; Sun, Tweney, and Wang 2010a,b; Sun and Wang 2010), there appears to have been only a single empirical test (Reimers, Donkin, and Le Pelley 2018).<sup>7</sup> In a formal comment based on an earlier version of this

---

<sup>5</sup>In particular, sample size neglect has been documented extensively (Benjamin, Rabin, and Raymond 2014; Kahneman and Tversky 1972), and is sometimes attributed to working memory capacity limitations (Kareev 2000).

<sup>6</sup>In a review article, Kareev (2000) observes that the sampling distribution of the correlation coefficient between any two variables is strongly skewed for small samples, which implies that measures of central tendency in the sampling distribution of the correlation can be substantially different than the true correlation, which can influence belief formation. Interestingly, in earlier work Kareev (1992) observes a finite sample property for the alternation rate in a sequence. In particular, while the expected overall alternation rate for streaks of length  $k = 1$  is equal to 0.5 (when not distinguishing between a preceding heads or tails), people’s experience can be made to be consistent with an alternation rate that is greater than 0.5 if the set of observable sequences that they are exposed to is restricted to those that are subjectively “typical” (e.g. those with an overall success rate close to 0.5). In fact, for streaks of length  $k > 1$ , this restriction is not necessary, as the expected overall alternation rate across all sequences is greater than 0.5.

<sup>7</sup>The focus on fixed length string patterns has a few limitations with regard to testability: (1) some patterns with lower

paper, Sun and Wang (2015) relate the bias found in Miller and Sanjurjo (2018) to this debate, but argue that its implications for human judgement and decision-making are limited. Instead, the authors emphasize the primacy of the waiting time distribution of finite length patterns in infinite sequences, rather than the distribution of sample statistics in finite length sequences.

The explanation provided here can be thought of as complementary to Rabin (2002) and Rabin and Vayanos (2010). In particular, it provides a structural account for why the central behavioral primitive of their model—the belief in the law of small numbers—should persist even when a decision maker has abundant experience. Further, our approach relates to Benjamin et al. (2014) in that it illustrates how a limited sensitivity to sample size can affect inference.

Our model has two main contributions: (i) it provides a plausible account for the persistence of gambler’s fallacy beliefs—even in environments where decision makers have ample opportunity to learn, (ii) it provides novel, and testable, predictions. First, in terms of plausibility, there is ample evidence that people tend to adapt to the natural statistics in their environment (Atick 1992; Simoncelli and Olshausen 2001), with the sample proportion being an example of a statistic that humans find intuitive and tend to assess relatively accurately (Garthwaite, Kadane, and O’Hagan 2005; Jonides and Jones 1992).<sup>8</sup> Second, in terms of testability, our model predicts that the magnitude of bias in peoples’ beliefs should depend on the following measurable and experimentally manipulable factors: (1) the length of sequences ( $\ell$ ), (2) the streak lengths ( $k$ ) that immediately precede the outcomes attended to, and (3) sensitivity to sample size  $w(\cdot)$ . Finally, a practical implication is that because of the limits of unguided real (or simulated) experience, a well-structured education in probability and statistics is needed to correct for this well-known behavioral bias.

## 2 The Model

Suppose DM has beliefs regarding the conditional probability  $\theta = \mathbb{P}(X_i = 1 | \prod_{j=i-k}^{i-1} X_j = 1)$ , with a prior  $\mu(\theta)$  over the support  $[0, 1]$ . When DM encounters a sequence  $\{X_i\}_{i=1}^\ell$ , he attends to those trials that immediately follow  $k$  (or more) successes, defined as  $I' := \{i \in \{k+1, \dots, \ell\} : \prod_{j=i-k}^{i-1} X_j = 1\}$ . Thus, he effectively observes  $\mathbf{Y} := (Y_i)_{i=1}^M = (X_i)_{i \in I'}$ , where  $M := |I'|$ . Whenever a sequence contains trials worthy of attending to (i.e.  $I' \neq \emptyset$ ), DM calculates the proportion of successes  $\hat{p}$  on those trials, weighting it according to his perception of the sample size  $w = w(M)$ .

---

associated proportions e.g. HTHT, have much lower probabilities than patterns with high associated proportions, such as TTHH, (2) for most patterns the difference in the probability is small, even for patterns in which the proportion associated with the pattern varies considerably.

<sup>8</sup>Jonides and Jones (1992, p. 368) write: “Ask about the relative numbers of many kinds of events, and you are likely to get answers that reflect the actual relative frequencies of the events with great fidelity.”

Given  $w$ , DM's posterior distribution for  $\theta$  follows:

$$p(\theta|\mathbf{Y}) = \frac{\theta^{w\hat{p}}(1-\theta)^{w(1-\hat{p})}\mu(\theta)}{\int \theta'^{w\hat{p}}(1-\theta')^{w(1-\hat{p})}\mu(\theta')}$$

Using this simple setup, we now briefly explore under what conditions gambler's fallacy beliefs can persist. Suppose that DM encounters an i.i.d. sequence of Bernoulli random variables  $\{X_i\}_{i=1}^\ell$  in which each trial has probability of success  $p$ . Further, DM is a believer in the law of small numbers, and holds a strong prior towards gambler's fallacy beliefs. In the case that he observes few sequences, experience will have little effect on DM's beliefs, regardless of whether or not he accounts for sample size. In the case that DM observes many sequences, the degree to which his gambler's fallacy beliefs persist will depend on (1) the extent to which he neglects sample size  $w(\cdot)$ , (2) the length of the sequences he is exposed to ( $\ell$ ), and (3) the threshold streak length ( $k$ ) that leads him to attend to outcomes.

To illustrate the role of sample size sensitivity, let  $w(M) := M^\alpha$  for some  $0 \leq \alpha \leq 1$ . On one extreme, if  $\alpha = 0$  then DM does not discriminate between different sample sizes, weighting all proportions the same. In this case, as the number of sequences increases, DM's beliefs,  $\mu$ , approach point mass on the fully biased (unweighted) expected proportion provided in Miller and Sanjurjo (2018).<sup>9</sup> As in the gambler's fallacy, these beliefs are strictly less than  $p$ , and become more biased as  $k$  increases. On the other extreme, when  $\alpha = 1$  DM fully discriminates between sample sizes, weighting proportions according to their sample sizes. In this case, there is no asymptotic bias in the proportion, so his beliefs will be correct in the limit.<sup>10,11</sup> Perhaps more plausibly, if DM has

<sup>9</sup>To see this, first note that DM will observe a sequence of i.i.d. proportions  $\hat{p}_i$ , with  $E[\hat{p}_i] := \theta^* < p$  (by Theorem 1 of Miller and Sanjurjo (2018)). The strong law of large numbers applies in this case, and  $\bar{p}_n := \sum_{i=1}^n \hat{p}_i/n$  will converge to  $\theta^*$  almost surely (a.s.). After the  $n^{th}$  sequence, DM's posterior odds in favor of  $\theta$  (relative  $\theta^*$ ) become  $\left[ \left( \frac{\theta}{\theta^*} \right)^{\bar{p}_n} \left( \frac{1-\theta}{1-\theta^*} \right)^{1-\bar{p}_n} \right]^n \frac{\mu(\theta)}{\mu(\theta^*)}$ . The posterior probability will converge to point mass on  $\theta^*$  (a.s.) because the posterior odds in favor of  $\theta$  converge to zero (a.s.) for all  $\theta \neq \theta^*$ , which follows because  $\theta \neq \theta^*$  implies  $\left( \frac{\theta}{\theta^*} \right)^{\theta^*} \left( \frac{1-\theta}{1-\theta^*} \right)^{1-\theta^*} < 1$ .

<sup>10</sup>The weighted average satisfies  $\sum_{i=1}^n M_i \hat{p}_i / \sum_{i=1}^n M_i = \sum_{i=1}^n \sum_{j=1}^{M_i} x_{ij} / \sum_{i=1}^n M_i$ , where  $x_{ij}$  is the  $j^{th}$  outcome from the  $i^{th}$  sequence. This weighted average is the maximum likelihood estimator for the transition probability  $p$  from the state “a trial is immediately preceded by  $k$  successes” to itself (with  $\sum_{i=1}^n M_i$  total observations), in the associated irreducible and ergodic  $2^k$ -state Markov chain, and converges to the transition probability  $p$  almost surely (see e.g. Grimmett and Stirzaker (2001, p. 358)). Following the argument in footnote 9, we conclude the DM's step  $n$  posterior odds in favor of  $\theta$  relative to  $p$  converge to 0 (a.s.), which implies the asymptotic posterior probability will have point mass on  $p$  (a.s.).

<sup>11</sup>There are two alternative statistical approaches that do not require an infinite sample of sequences for DM to obtain an unbiased estimate of the conditional probability. To illustrate, if DM were instead to control the number of selected trials by repeating the experiment until he generates exactly  $m$  trials that immediately follow  $k$  consecutive successes, then the proportion would be unbiased. Alternatively, if DM were to eliminate the overlapping nature of the measure, there would be no bias, even though the number of selected trials would still be random. In particular, for a sequence of  $n$  trials, one can take each run of successes, and if it is of even length  $2\ell$ , divide it into blocks of two trials; if it is of odd length  $2\ell - 1$  include the right adjacent tails and divide it into blocks of two trials. In each case, the run of successes contributes  $\ell$  observations.

*some* degree of sensitivity to sample size then the asymptotic beliefs will be biased, and will lie somewhere between the two extremes just given, depending on the sensitivity  $0 < \alpha < 1$ .

## References

- ATICK, J. J. (1992): “Could information theory provide an ecological theory of sensory processing?” *Network: Computation in neural systems*, 3, 213–251.
- AYTON, P. AND I. FISCHER (2004): “The hot hand fallacy and the gamblers fallacy: Two faces of subjective randomness?” *Memory & Cognition*, 21, 1369–1378.
- BAR-HILLEL, M. AND W. A. WAGENAAR (1991): “The perception of randomness,” *Advances in Applied Mathematics*, 12, 428–454.
- BENJAMIN, D. J., M. RABIN, AND C. RAYMOND (2014): “A Model of Non-Belief in the Law of Large Numbers,” Working Paper.
- CHEN, D., T. J. MOSKOWITZ, AND K. SHUE (2014): “Decision-Making under the Gamblers Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires,” Working Paper.
- COWAN, N. (2001): “The magical number 4 in short-term memory: A reconsideration of mental storage capacity,” *Behavioral and Brain Sciences*, 24, 87–114.
- CROSON, R. AND J. SUNDALI (2005): “The Gamblers Fallacy and the Hot Hand: Empirical Data from Casinos,” *Journal of Risk and Uncertainty*, 30, 195–209.
- D’ALEMBERT, J. (1761): *Opuscules mathmatiques*, David, Paris.
- GARTHWAITE, P. H., J. B. KADANE, AND A. O’HAGAN (2005): “Statistical Methods for Eliciting Probability Distributions,” *Journal of the American Statistical Association*, 100, 680–700.
- GORROOCHURN, P. (2012): *Classic problems of probability*, New Jersey: John Wiley & Sons.
- GRIFFIN, D. AND A. TVERSKY (1992): “The weighing of evidence and the determinants of confidence,” *Cognitive Psychology*, 24, 411–435.
- GRIMMETT, G. R. AND D. R. STIRZAKER (2001): *Probability and Random Processes*, Oxford University Press.
- GUIBAS, L. J. AND A. M. ODLYZKO (1981): “String overlaps, pattern matching, and nontransitive games,” *Journal of Combinatorial Theory, Series A*, 30, 183–208.
- HAHN, U. AND P. A. WARREN (2009): “Perceptions of randomness: why three heads are better than four,” *Psychological Review*, 116, 454–461.
- (2010a): “Postscript: All together now: ‘Three heads are better than four.’” *Psychological Review*, 117, 711–711.

- (2010b): “Why three heads are a better bet than four: A reply to Sun, Tweney, and Wang (2010).” *Psychological Review*, 117, 706–711.
- JONIDES, J. AND C. M. JONES (1992): “Direct coding for frequency of occurrence.” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 368–.
- KAHNEMAN, D. AND A. TVERSKY (1972): “Subjective Probability: A Judgement of Representativeness,” *Cognitive Psychology*, 3, 430–454.
- KAREEV, Y. (1992): “Not that bad after all: Generation of random sequences.” *Journal of Experimental Psychology: Human Perception and Performance*, 18, 1189–1194.
- (1995a): “Positive bias in the perception of covariation.” *Psychological Review*, 102, 490–502.
- (1995b): “Through a narrow window: working memory capacity and the detection of covariation,” *Cognition*, 56, 263–269.
- (2000): “Seven (indeed, plus or minus two) and the detection of correlations.” *Psychological Review*, 107, 397–402.
- KAREEV, Y., I. LIEBERMAN, AND M. LEV (1997): “Through a narrow window: Sample size and the perception of correlation.” *Journal of Experimental Psychology: General*, 126, 278–287.
- KONOLD, C. (1995): “Confessions of a coin flipper and would-be instructor,” *The American Statistician*, 49, 203–209.
- MILLER, J. B. AND A. SANJURJO (2015): “Surprised by the Gambler’s and Hot Hand Fallacies? A Truth in the Law of Small Numbers,” Working Paper, September 24.
- (2018): “Surprised by the Hot Hand Fallacy? A Truth in the Law of Small Numbers,” *Econometrica*, 86.
- NICKERSON, R. S. (2002): “The production and perception of randomness,” *Psychological Review*, 109, 350–357.
- (2007): “Penney Ante: Counterintuitive Probabilities in Coin Tossing,” *The UMAP Journal*, 28, 503–532.
- OSKARSSON, A. T., L. V. BOVEN, G. H. McCLELLAND, AND R. HASTIE (2009): “What’s next? Judging sequences of binary events,” *Psychological Bulletin*, 135, 262–385.
- RABIN, M. (2002): “Inference by Believers in the Law of Small Numbers,” *Quarterly Journal of Economics*, 117, 775–816.
- RABIN, M. AND D. VAYANOS (2010): “The Gamblers and Hot-Hand Fallacies: Theory and Applications,” *Review of Economic Studies*, 77, 730–778.
- REIMERS, S., C. DONKIN, AND M. E. LE PELLEY (2018): “Perceptions of randomness in binary sequences: Normative, heuristic, or both?” *Cognition*, 172, 11–25.

- SIMONCELLI, E. P. AND B. A. OLSHAUSEN (2001): “Natural Image Statistics And Neural Representation,” *Annual Review of Neuroscience*, 24, 1193–1216.
- SUN, Y., R. D. TWENEY, AND H. WANG (2010a): “Occurrence and nonoccurrence of random sequences: Comment on Hahn and Warren (2009).” *Psychological Review*, 117, 697–703.
- (2010b): “Postscript: Untangling the gambler’s fallacy.” *Psychological Review*, 117, 704–705.
- SUN, Y. AND H. WANG (2010): “Gamblers fallacy, hot hand belief, and the time of patterns,” *Judgement and Decision Making*, 5, 124–132.
- (2015): “Alternation Bias as a Consequence of Pattern Overlap: Comments on Miller and Sanjurjo (2015),” Working Paper, November 8.
- TVERSKY, A. AND D. KAHNEMAN (1971): “Belief in the Law of Small Numbers,” *Psychological Bulletin*, 2, 105–110.

## A Appendix: Evidence that the random sequences that people experience are short

A stylized fact about experimental subjects' perceptions of sequential dependence is that, on average, they believe that random processes, such as a fair coin, alternate at a rate of roughly 0.6, rather than 0.5 (Bar-Hillel and Wagenaar 1991; Nickerson 2002; Oskarsson et al. 2009; Rabin 2002). It is natural to presume that peoples' experience with finite sequences, either by observation or by generation, tends to involve sequences that are sufficiently short for the bias of Miller and Sanjurjo (2018) to be substantial. If so, then this could provide a structural explanation for why the alternation bias and gambler's fallacy have persisted at the observed magnitudes.

In order to get a more precise sense of peoples' experience with sequences of random binary outcomes, we conduct a simple survey. The survey is designed to elicit the typical number of sequential outcomes that people observe when repeatedly flipping a coin, as well as their perceptions of expected conditional probabilities, based on recent outcomes. We recruited 649 subjects to participate in a survey in which they could be paid up to 25 Euros to answer the following questions as best as possible<sup>12</sup>: (1) what is the largest number of successive coin flips that they have observed in one sitting, (2) what is the average number of sequential coin flips that they have observed, and (3) given that they observe a fair coin land heads one (two; three) consecutive times, what do they feel the chances are that the next flip will be heads (tails).<sup>13,14</sup>

---

<sup>12</sup>The exact email invitation to the survey was as follows. "This is a special message regarding an online survey that pays up to 25 Euros for 3 minutes of your time. If you complete this survey ([link](#)) by 02:00 on Friday 05-June Milan time and the Thursday June 4th evening drawing of the California Lottery Pick 3 matches the final 3 digits of your student ID, you will be paid 25 dollars. For details on the Pick 3, see: <http://www.calottery.com/play/draw-games/daily-3/winning-numbers> for detail."

<sup>13</sup>The subjects were provided a visual introduction to heads and tails for a 1 Euro coin. Then, the two questions pertaining to sequence length were asked in random order: (1) Please think of all the times in which you have observed a coin being flipped, whether it was flipped by you, or by somebody else. To the best that you can recall, what is the maximum number of coin flips that you have ever observed in one sitting?, (2) Please think of all the times in which you have observed a coin being flipped, whether it was flipped by you, or by somebody else. Across all of the times you have observed coin flips, to the best that you can recall, how many times was the coin flipped, on average?

<sup>14</sup>The three questions pertaining to perceived probability were always presented at the end, in order, with each subject assigned either to a treatment in which they were asked for the probability of heads (repetition), or the probability of tails (alternation). The precise language used was: (3) (a) Imagine that you flip a coin you know to be a fair coin, that is, for which heads and tails have an equal chance. First, imagine that you flip the coin one time, and observe a heads. On your second flip, according to your intuition, what do you feel is the chance of flipping a heads (T2: tails) (in percentage terms 0-100)? (b) Second, imagine that you flip the coin two times, and observe two heads in a row. On your third flip, according to your intuition, what do you feel is the chance of flipping a heads (T2: tails) (in percentage terms 0-100)? (c) Third, imagine that you flip the coin three times, and observe three heads in a row.

The subjects were recruited from Bocconi University in Milan, Italy. All subjects responded to each of the three questions. We observe that the median of the maximum number of sequential coin flips that a subject has seen is 6, and the median of the average number of coin flips is 4. For sequences of 4 flips of a fair coin, the true expected conditional alternation rate is 0.6, which is precisely in line with the average magnitude of alternation bias observed in laboratory experiments.<sup>15</sup>

For the third question, which regarded perceptions of sequential dependence, subjects were randomly assigned into one of two treatments. Subjects in the first treatment were asked about the probability of a heads immediately following a streak of heads (repetition), while subjects in the second were asked about the probability of a tails immediately following a streak of heads (alternation). Table 1 shows subjects' responses for streaks of one, two, and three heads. One can see that the perceived probability that a streak of a single head will be followed by a head  $\mathbb{P}(H|\cdot)$  [tails  $\mathbb{P}(T|\cdot)$ ] is 0.49 [0.50], a streak of two heads 0.45 [0.53], and a streak of three heads 0.44 [0.51]. Further, between subjects, average perceptions satisfy

$$\mathbb{P}(T|H) - \mathbb{P}(H|H), \mathbb{P}(T|HH) - \mathbb{P}(H|HH), \mathbb{P}(T|HHH) - \mathbb{P}(H|HHH) > 0$$

and with 649 subjects, all three differences are significant. Thus, subjects' perceptions are consistent with the true positive differences between expected alternation and repetition rates (see (Miller and Sanjurjo 2018)), when these perceptions are based on the finite sequences of outcomes that a subject has observed in the past. Notice also that average perceived expected conditional alternation and repetition rates lie somewhere between those that subjects have observed in the past, and the true (unconditional) probabilities of alternation and repetition (0.5), and more closely resemble the latter than the former.<sup>16,17</sup>

---

On your fourth flip, according to your intuition, what do you feel is the chance of flipping a heads (T2: tails) (in percentage terms 0-100)?

<sup>15</sup>In addition, peoples' maximum working memory capacity has been found to be around four (not seven), in a meta-analysis of the literature in Cowan (2001), which could translate into sequences longer than four essentially being treated as multiple sequences of around four.

<sup>16</sup>Given that subjects were assigned to treatments at random, and that the sample size is sufficiently large for significance tests to be robust to adjustments for multiple comparisons, it is a mystery why, for example, on average  $\mathbb{P}(H|HHH) < \mathbb{P}(H|HH)$  and  $\mathbb{P}(T|HHH) < \mathbb{P}(T|HH)$ . A possible explanation is that the elicitation of the probability of repetition is different than the elicitation of the probability of alternation.

<sup>17</sup>These findings suggest a possible answer to the following puzzle: "Study of the gambler's fallacy is complicated by the fact that people sometimes make predictions that are consistent with the assumption that they believe that independent events are contingent even when they indicate, when asked, that they believe them to be independent"

**Table 1:** Average perceived repetition rate  $P(H|\cdot)$ , and alternation rate  $P(T|\cdot)$ , when conditioning on streak length of Hs, for 649 survey participants

	Streak Length of Hs			
	H	HH	HHH	obs.
$P(H \cdot)$	.49	.45	.44	304
$P(T \cdot)$	.50	.53	.51	345

## B The hot hand selection bias and pattern overlap

Here we discuss the connection between the streak selection bias of Miller and Sanjurjo (2018), which underlies our model, and the finite sample behavior of sub-string patterns, which underly the model of Hahn and Warren (2009). The counter-intuitive behavior of sub-string patterns is often referred to as the *overlapping words paradox*, and is explored at length in Guibas and Odlyzko (1981).

The *overlapping words paradox* states that for a sequence (string) of finite length  $n$ , the probability that a pattern (word) appears, e.g. `HTTHHH`, depends not only on the length of the pattern relative to the length of the sequence, but also on how the pattern *overlaps* with itself (Guibas and Odlyzko 1981).<sup>18</sup> For example, while the expected number of (potentially overlapping) occurrences of a particular three flip pattern—TTT, HTT, THT, etc.—in a sequence of four flips of a fair coin does not depend on the pattern, its probability of occurrence does.<sup>19</sup> The pattern HHH can overlap with itself, so can have up to two occurrences in a single sequence, whereas the pattern HHT cannot overlap with itself, so can have at most one occurrences. Because the expected number of occurrences of each pattern must be equal, this implies that the pattern HHT is distributed across more sequences (4) than the pattern HHH (3), meaning that any given sequence is more likely to contain the pattern HHT than the pattern HHH. It is this latter fact that Hahn and Warren (2009) argue informs people’s beliefs in alternation. In particular, because  $\mathbb{P}(HHH \text{ in 4 flips}) = 3/16 < 4/16 = \mathbb{P}(HHT \text{ in 4 flips})$ , people should believe that

---

(Nickerson 2002). We observe subjects’ responses to lie between their beliefs based on experience, and their “book knowledge.”

<sup>18</sup>For a simpler treatment which studies a manifestation of the paradox in the non-transitive game known as “Penney’s” game, see Konold (1995) and Nickerson (2007).

<sup>19</sup>That all fixed length patterns are equally likely ex-ante is straightforward to demonstrate. For a given pattern of heads and tails of length  $\ell$ ,  $(y_1, \dots, y_\ell)$ , the expected number of occurrences of this pattern satisfies  $E[\sum_{i=\ell}^n 1_{[(X_{i-\ell+1}, \dots, X_i) = (y_1, \dots, y_\ell)]}] = \sum_{i=\ell}^n E[1_{[(X_{i-\ell+1}, \dots, X_i) = (y_1, \dots, y_\ell)]}] = \sum_{i=\ell}^n 1/2^\ell = (n - \ell + 1)/2^\ell$ .

$$\mathbb{P}(H|HH) < 1/2.$$

The insights provided by Guibas and Odlyzko (1981)'s investigation of the overlapping substring patterns and their finite sample properties can also provide a (partial) intuition for why the proportion of heads on flips that immediately follow two heads is expected to be less than 1/2. Let  $x_i \in \{H, T\}^n$ , the proportion be denoted by  $\hat{P}_{H|HH}(\mathbf{x})$ , which satisfies:

$$\hat{P}_{H|HH}(\mathbf{x}) = \frac{\#_{HHH}(\mathbf{x})}{\#_{HHH}(\mathbf{x}) + \#_{HHT}(\mathbf{x})}$$

where  $\#_{HHH}(\mathbf{x})$  and  $\#_{HHT}(\mathbf{x})$  consist of the number of (overlapping) instances of HHH and HHT in the sequence  $\mathbf{x}$ , respectively. Because HHT are distributed across more sequences, there are more often fewer instances of HHH than HHT in a given sequence. Therefore,  $\hat{P}_{H|HH}(\mathbf{x})$  is strictly less than 1/2 more often than it is strictly greater than 1/2. Of course, this intuition is incomplete, as it provides only partial information regarding the sampling distribution of  $\hat{P}_{H|HH}(\mathbf{x})$ . In particular, it does not guarantee that the expected value of  $\hat{P}_{H|HH}(\mathbf{x})$  is less than 1/2.