Surprised by the Gambler's and Hot Hand Fallacies? A Truth in the Law of Small Numbers

Joshua B. Miller and Adam Sanjurjo *†‡

September 24, 2015

Abstract

We find a subtle but substantial bias in a standard measure of the conditional dependence of present outcomes on streaks of past outcomes in sequential data. The mechanism is a form of selection bias, which leads the empirical probability (i.e. relative frequency) to underestimate the true probability of a given outcome, when conditioning on prior outcomes of the same kind. The biased measure has been used prominently in the literature that investigates incorrect beliefs in sequential decision making—most notably the Gambler's Fallacy and the Hot Hand Fallacy. Upon correcting for the bias, the conclusions of some prominent studies in the literature are reversed. The bias also provides a structural explanation of why the belief in the law of small numbers persists, as repeated experience with finite sequences can only reinforce these beliefs, on average.

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

Keywords: Law of Small Numbers; Alternation Bias; Negative Recency Bias; Gambler's Fallacy; Hot Hand Fallacy; Hot Hand Effect; Sequential Decision Making; Sequential Data; Selection Bias; Finite Sample Bias; Small Sample Bias.

*Miller: Department of Decision Sciences and IGIER, Bocconi University, Sanjurjo: Fundamentos del Análisis Económico, Universidad de Alicante. Financial support from the Department of Decision Sciences at Bocconi University, and the Spanish Ministry of Economics and Competitiveness under project ECO2012-34928 is gratefully acknowledged.

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

‡This draft has benefitted from helpful comments and suggestions, for which we would like to thank David Arathorn, Maya Bar-Hillel, Gary Charness, Vincent Crawford, Martin Dufwenberg, Andrew Gelman, Daniel Goldstein, Daniel Houser, Daniel Kahan, Daniel Kahneman, Mark Machina, Justin Rao, Yosef Rinott, Aldo Rustichini and Daniel Stone. We would also like to thank the many seminar participants at Microsoft Research, George Mason University, the Choice Lab at NHH Norway, USC, UCSD's Rady School of Management, Chapman University, as well as conference participants at the 11th World Congress of The Econometric Society, The 30th Annual Congress of the European Economic Association, and the 14th TIBER Symposium on Psychology and Economics. All mistakes and omissions remain our own.

We shall encounter theoretical conclusions which not only are unexpected but actually come as a shock to intuition and common sense. They will reveal that commonly accepted notions concerning chance fluctuations are without foundation and that the implications of the law of large numbers are widely misconstrued. (Feller 1968)

1 Introduction

Jack takes a coin from his pocket and decides that he will flip it 4 times in a row, writing down the outcome of each flip on a scrap of paper. After he is done flipping, he will look at the flips that immediately followed an outcome of heads, and compute the relative frequency of heads on those flips. Because the coin is fair, Jack of course expects this *empirical* probability of heads to be equal to the true probability of flipping a heads: 0.5. Shockingly, Jack is wrong. If he were to sample one million fair coins and flip each coin 4 times, observing the conditional relative frequency for each coin, on average the relative frequency would be approximately 0.4.

We demonstrate that in a finite sequence generated by i.i.d. Bernoulli trials with probability of success p, the relative frequency of success on those trials that immediately follow a streak of one, or more, consecutive successes is expected to be strictly less than p, i.e. the empirical probability of success on such trials is a biased estimator of the true conditional probability of success. While, in general, the bias does decrease as the sequence gets longer, for a range of sequence (and streak) lengths often used in empirical work it remains substantial, and increases in streak length.

This result has considerable implications for the study of decision making in any environment that involves sequential data. For one, it provides a structural explanation for the persistence of one of the most well-documented, and robust, systematic errors in beliefs regarding sequential data—that people have an *alternation bias* (also known as *negative recency bias*; see Bar-Hillel and Wagenaar [1991]; Nickerson [2002]; Oskarsson, Boven, McClelland, and Hastie [2009]) —by which they believe, for example, that when observing multiple flips of a fair coin, an outcome of heads is more likely to be followed by a tails than by another heads, as well as the closely related *gambler's fallacy* (see Bar-Hillel and Wagenaar (1991); Rabin (2002)), in which this alternation bias *increases* with the length of the streak of heads. Accordingly, the result is consistent with the types of subjective inference that have been conjectured in behavioral models of the law of small numbers, as in Rabin (2002); Rabin and Vayanos (2010). Further, the result shows that data in the *hot hand fallacy* literature (see Gilovich, Vallone, and Tversky [1985] and Miller and Sanjurjo [2014, 2015]) has been systematically misinterpreted by researchers; for those trials that immediately follow a streak of successes, observing that the relative frequency of success is equal to the overall base rate of success, is in fact evidence *in favor* of the hot hand, rather than evidence against it. Tying these two implications together, it becomes clear why the inability of the gambler to detect the fallacy of his belief in alternation has an exact parallel with the researcher's inability to detect his mistake when concluding that experts' belief in the hot hand is a fallacy.

In addition, the result may have implications for evaluation and compensation systems. That a coin is expected to exhibit an alternation "bias" in finite sequences implies that the outcome of a flip can be successfully "predicted" in finite sequences at a rate better than that of chance (if one is free to choose when to predict). As a stylized example, suppose that each day a stock index goes either up or down, according to a random walk in which the probability of going up is, say, 0.6. A financial analyst who can predict the next day's performance on the days she chooses to, and whose predictions are evaluated in terms of how her success rate on predictions in a given month compares to that of chance, can expect to outperform this benchmark by using any of a number of different decision rules. For instance, she can simply predict "up" immediately following down days, or increase her expected relative performance even further by predicting "up" only immediately following longer streaks of consecutive down days.[1]

Intuitively, it seems impossible that in a sequence of fair coin flips, we are expected to observe tails more often than heads on those flips that have been *randomly* "assigned" to immediately follow a heads. While this intuition is appealing, it ignores the sequential structure of the data, and thus overlooks that there is a *selection bias* towards observing tails when forming the group of flips that immediately follow a heads.[2] In order to see the intuition behind this selection bias, it is

[1] Similarly, it is easy to construct betting games that act as money pumps. For example, we can offer the following lottery at a \$5 ticket price: a fair coin will be flipped 4 times. if the relative frequency of heads on those flips that immediately follow a heads is greater than 0.5 then the ticket pays \$10; if the relative frequency is less than 0.5 then the ticket pays \$0; if the relative frequency is exactly equal to 0.5, or if no flip is immediately preceded by a heads, then a new sequence of 4 flips is generated. While, intuitively, it seems like the expected payout of this ticket is \$0, it is actually \$-0.71 (see Table 1). Curiously, this betting game would appear relatively more attractive to someone who believes in the independence of coin flips, as compared to someone with Gambler's Fallacy beliefs.

[2] The intuition also ignores that while the variable that assigns a flip to immediately follow a heads is random, its outcome is also included in the calculation of the relative frequency. This implies that the relative frequency is not independent of which flips are assigned to follow a heads, e.g. for 10 flips of a fair coin if we let $G_H = \{2, 4, 5, 8, 10\}$ be the group of flips that immediately follow a heads, then we know that the first 9 flips of the sequence must be

useful to first recognize that in any observed sequence of coin flips, all heads appear in runs, i.e. blocks of successive flips of heads flanked by tails, or the the beginning or end of the sequence. For example, in the sequence HTTHHHTHTTTHHT, there are four runs of H, which is the expected number for a sequence with $n = 14$ flips and $n_H = 7$ heads.[3] Here the 4 runs of H ($r_H = 4$) consist of two of length 1 ($r_{H1} = 2$), one of length 2 ($r_{H2} = 1$), one of length 3 ($r_{H3} = 1$), and zero of length greater than 3 ($r_{Hj} = 0$, for $j > 3$). In selecting the group of flips that immediately follow an H in a given (realized) sequence, every run of length 1 will lead to the selection of one flip, which must necessarily be T; every run of length 2 will lead to the selection of two flips: one H and one T; every run of length 3 will lead to the selection of three flips: two Hs and one T, and so on. In the sequence given above, the relative frequency of heads on those flips that immediately follow a heads, $\hat{p}(H|H)$, can be computed in the following way, using the number of runs of heads of each length: $\hat{p}(H|H) = (r_{H2} + 2r_{H3})/(r_{H1} + 2r_{H2} + 3r_{H3}) = (1 + 2)/(2 + 2 + 3) = 3/7 = 0.42$.[4] That the relative frequency of heads on those flips that immediately follow a heads is less than the base rate of heads, given n_H ($\hat{p}(H|H) < n_H/n$), is typical for a sequence with $n = 14$ flips and $n_H = 7$ heads. This is because the constraint that only 7 Hs are available to fill the expected 4 runs in the sequence means that in most sequences there are more short runs than long runs, leading to an over-selection of flips from short runs—precisely those in which T is relatively over-represented (given n and n_H).[5]

To make this intuition concrete, we return to the opening example of Jack's 4 coin flips, and in Table 1 consider all 16 possible sequences, grouped according to the number of heads in the sequence (n_H). For each sequence, $\hat{p}(H|H)$ is the empirical probability of heads on those flips that immediately follow a heads, and for each n_H, $E[\ \hat{p}(H|H)\ |\ n_H]$ is the expected value of $\hat{p}(H|H)$ across all sequences with n_H heads. Of course, $\hat{p}(H|H)$ is not measurable for the two sequences that contain no flips following a heads, so these sequences must be discarded. The expectation

$HTHHTTHTH$, and thus, that the relative frequency is either 1/5 or 2/5, with equal chances.

[3] More generally, the expected number of runs of heads is $n_H(n - n_H + 1)/n$.

[4] Notice that if the sequence used in the example were to end with an H, rather than a T, this final run of heads could not be followed by a T, which would reduce the number of selected flips by 1 and lead to $\hat{p}(H|H) = 3/6$. While the bias is mitigated for those sequences that end in a run of H (given the same distribution of runs of each length), in general, the bias maintains even among these sequences. This end point issue is taken into account in Section A.

[5] Here, whether a run is "short" or "long," is relative to n and n_H. For a more extreme example, among the sequences with $n = 10$ and $n_H = 9$, approximately two runs of heads are expected (1.8), and a run of length 8 or less that does not appear at the end of the sequence is relatively short. In particular, a run of 8 Hs leads to the selection of 8 flips, of which one is a T, which results in T being over-represented (1 in 8) relative to the available T (1 in 10). Therefore, H is produced at a fraction less than the overall rate, $\hat{p}(H|H) = 7/8 < 9/10 = n_H/n$. In fact, $E[\hat{p}(H|H)|n_H = 9] = 8/9$.

Table 1: *The empirical probability of heads on those flips that immediately follow one or more heads, $\hat{p}(H|H)$, for each possible sequence of four coin flips. $E[\ \hat{p}(H|H) \mid n_H]$ is the arithmetic average for the set of sequences with a fixed n_H, which are equiprobable. In the bottom row the expected value of each quantity is reported under the assumption that the coin is fair.*

# Heads (n_H)	Sequence	$\hat{p}(H\|H)$	$E[\ \hat{p}(H\|H) \mid n_H]$
0	TTTT	-	-
1	TTTH	-	0
	TTHT	0	
	THTT	0	
	HTTT	0	
2	TTHH	1	$\frac{1}{3}$
	THTH	0	
	THHT	$\frac{1}{2}$	
	HTTH	0	
	HTHT	0	
	HHTT	$\frac{1}{2}$	
3	THHH	1	$\frac{2}{3}$
	HTHH	$\frac{1}{2}$	
	HHTH	$\frac{1}{2}$	
	HHHT	$\frac{2}{3}$	
4	HHHH	1	1
Expected Value (fair coin)		.40	.40

of each quantity is reported in the last row of the table; under the assumption that the coin is fair, these are the arithmetic average of $\hat{p}(H|H)$ and the frequency-weighted arithmetic average $E[\ \hat{p}(H|H) \mid n_H]$, respectively. As shown, the empirical probability of heads on those flips that immediately follow a heads is expected to be 0.4, not 0.5. To understand why, for each $0 < n_H < 4$ we examine the set of sequences with n_H heads, which are all equiprobable. Given a fixed n_H, if one selects a sequence at random, the empirical probability of heads on a flip is clearly expected to be $\hat{p}(H) = n_H/n$, as this is precisely the ratio of heads to flips in each sequence. By contrast, the empirical (conditional) probability $\hat{p}(H|H)$ is expected to be strictly less than n_H/n for each value of n_H. To illustrate, when $n_H = 1$, all runs of H are necessarily of length 1, so the flips selected by these runs over-represent T (1 T in 1 flip) relative to the T available (3 Ts in 4 flips), and therefore $\hat{p}(H|H) = 0$ for each sequence. When $n_H = 2$, half of the sequences contain only runs of H of length 1, and the other half contain only runs of H of length 2. Therefore, the flips selected by these runs either over-represent T (1 T in 1 flip), or equally represent T (1 T in 2 flips), relative to the T available (2 Ts in 4 flips), which, on average, leads to a bias towards selecting flips that over-represent T. When $n_H = 3$, runs of H are of length 1,2, or 3, thus the flips taken from any of these runs over-represent T (at least 1 T in 3 flips) relative to the T available (1 T in 4 flips).[6] These facts reflect the constraint that the total number of Hs in runs of H must equal n_H ($\sum_{i=1}^{n_H} i r_{Hi} = n_H$) places on the joint distribution of runs of each length ($(r_{Hi})_{i=1}^{n_H}$), and imply that for each n_H, $E[\hat{p}(H|H)|n_H] = \frac{n_H-1}{n-1}$, which also happens to be the familiar formula for sampling without replacement. When considering the empirical probability of a heads on those flips that immediately follow a streak of k or more heads, $\hat{p}(H|kH)$ (see Section 2), the constraint that $\sum_{i=1}^{n_H} i r_{Hi} = n_H$ places on the joint distribution of run lengths greater than k leads to a downward bias that is even more severe for $E[\hat{p}(H|kH)|n_H]$.[7]

The law of large numbers would seem to imply that as the number of coins that Jack samples increases, the average empirical probability of heads would approach the true probability. The key to why this is not the case, and to why the bias remains, is that it is not the flip that is treated as the unit of analysis, but rather the sequence of flips from each coin. In particular, if Jack were willing to assume that each sequence had been generated by the same coin, and were to compute the empirical probability by instead pooling together all of those flips that immediately follow a heads, regardless of which coin produced them, then the bias would converge to zero as the number of coins approaches infinity. Therefore, in treating the sequence as the unit of analysis, the average empirical probability across coins amounts to an unweighted average that does not account for the number of flips that immediately follow a heads in each sequence, and thus leads the data to appear consistent with the gambler's fallacy.[8] The implications for learning are stark: to the extent

[6] In order to ease exposition, this discussion ignores that sequences can also end with a run of heads. Nevertheless, what happens at the end points of sequences is not the source of the bias. In fact, if one were to calculate $\hat{p}(H|H)$ in a circular way, so that the first position were to follow the fourth, the bias would be even more severe ($E[\hat{p}(H|H)] = 0.38$).

[7] To give an idea of why $E[\hat{p}(H|kH)|n_H] < E[\hat{p}(H|H)|n_H]$ for $k > 1$, first note that each run of length $j < k$ leads to the selection of 0 flips, and each run of length $j \geq k$ leads to the selection of $j - k + 1$ flips, with $j - k$ Hs and one T. Because n_H is fixed, when k increases from 1 there are fewer possible run lengths that are greater than k, and each yields a selection of flips that is balanced relatively more towards T. In the extreme, when $k = n_H$, there are no runs longer than k, and only a single flip is selected in each sequence with n_H heads, which yields $E[\hat{p}(H|kH)|n_H] = 0$.

[8] Given this, the direction of the bias can be understood by examining Table 1; the sequences that have more flips of heads tend to have a higher relative frequency of repetition. Thus, these higher relative frequency sequences are "underweighted" when taking the average of the relative frequencies across all (here, equiprobable) sequences, which results in an average that indicates alternation.

that decision makers update their beliefs regarding sequential dependence with the (unweighted) empirical probabilities that they observe in finite length sequences, they can never unlearn a belief in the gambler's fallacy.[9,10]

That the bias emerges when the empirical conditional probability—the estimator of conditional probability—is computed for finite sequences, suggests a connection to the well-known finite sample bias of the least squares estimator of autocorrelation in time series data (Stambaugh 1986; Yule 1926). Indeed, we find that this connection is more than suggestive: by letting $\boldsymbol{x} \in \{0,1\}^n$ be a sequence of iid Bernoulli(p) trials, and $\mathbb{P}(x_i = 1|x_{i-1} = \cdots = x_{i-k} = 1)$ the probability of a success in trial i, one can see that conditional on a success streak of length k (or more) on the immediately preceding k trials, the least squares estimator for the coefficients in the associated linear probability model, $x_i = \beta_0 + \beta_1 \mathbb{1}_{[x_{i-1}=\cdots=x_{i-k}=1]}(i)$, happens to be the conditional relative frequency $\hat{p}(x_i = 1|x_{i-1} = \cdots = x_{i-k} = 1)$. Therefore, the explicit formula for the bias that we find below can be applied directly to the coefficients of the associated linear probability model (see Appendix D). Furthermore, the bias in the associated statistical test can be corrected with an appropriate form of re-sampling, analogous to what has been done in the study of time series data (Nelson and Kim 1993), but in this case using permutation test procedures (see Miller and Sanjurjo [2014] for details).[11]

In Section 2 (and Appendix B), we derive an explicit formula for the empirical probability of success on trails that immediately follow a streak of success, for any probability of success p, any streak length k, and any sample size n. While this formula does not appear, in general, to admit a closed form representation, for the special case of streaks of length $k = 1$ (as in the example

[9] For a full discussion see Section 3.

[10] When learning is deliberate, rather than intuitive, this result also holds if empirical probabilities are instead represented as natural frequencies (e.g. 3 in 4, rather than .75), under the assumption that people neglect sample size (see Kahneman and Tversky (1972), and Benjamin, Rabin, and Raymond (2014)).

[11] In a comment on this paper, Rinott and Bar-Hillel (2015) assert that the work of Bai (1975) (and references therein) demonstrate that the bias in the conditional relative frequency follows directly from known results on the bias of Maximum Likelihood estimators of transition probabilities in Markov chains, as independent Bernoulli trials can be represented by a Markov chain with each state defined by the sequence of outcomes in the previous k trials. While it is true that the MLE of the corresponding transition matrix is biased, the cited theorems do not apply in this case because they require that transition probabilities in different rows of the transition matrix not be functions of each other, and not be equal to zero, a requirement which does not hold in the corresponding transition matrix. Instead, an unbiased estimator of each transition probability will exist, and will be a function of the unconditional relative frequency. Rinott and Bar-Hillel (2015) also provide a novel alternative proof for the case of $k = 1$ that provides reasoning which suggests sampling-without-replacement as the mechanism behind the bias. This reasoning does not extend to the case of $k > 1$, due to the combinatorial nature of the problem. For an intuition why sampling-without-replacement reasoning does not fully explain the bias, see Appendix C.

above) one is provided. For k larger than one, we use analogous reasoning to compute the expected empirical probability for sequences with lengths relevant for empirical work, using results on the sampling distribution developed in Appendix B.

Section 3 begins by detailing how the result provides a structural explanation of the persistence of both alternation bias, and gambler's fallacy beliefs. In addition, results are reported from a simple survey that we conducted in order to gauge whether the degree of alternation bias that has typically been observed from experimental subjects in previous studies is roughly consistent with subjects' experience with finite sequences outside of the laboratory. Then we discuss how the result reveals an incorrect assumption made in the analyses of the most prominent hot hand fallacy studies, and outline how to correct for the bias in statistical testing. Once corrected for, the previous findings are reversed.

2 Result

We find an explicit formula for the expected *conditional* relative frequency of a success in a finite sequence of i.i.d. Bernoulli trials with any probability of success p, any sequence length n, and when conditioning on streaks of successes of any length k. Further, we quantify the bias for a range of n and k relevant for empirical work. In this section, for the simplest case of $k = 1$, we present a closed form representation of this formula. To ease exposition, the statement and proofs of results for k larger than one are treated in Appendix B.

2.1 Expected Bias

Let $X_1, \dots, X_n$ be a sequence of n i.i.d Bernoulli trials with probability of success $p := \mathbb{P}(X_i = 1)$. Let the sequence $\boldsymbol{x} = (x_1, ..., x_n) \in \{0,1\}^n$ be the realization of the n trials, $N_1(\boldsymbol{x}) := \sum_{i=1}^n x_i$ the number of ones, and $N_0(\boldsymbol{x}) := n - N_1(\boldsymbol{x})$ the number of zeros, with their respective realizations n_1 and n_0. We begin with two definitions.

Definition 1 *For $\boldsymbol{x} \in \{0,1\}^n$, $n_1 \geq 0$, and $k = 1, \dots, n_1$, the set of* k/1-streak *successors $I_{1k}(\boldsymbol{x})$ is the subset of trials that immediately succeed a streak of ones of length k or more, i.e.*

$$I_{1k}(\boldsymbol{x}) := \{i \in \{k+1, \dots, n\} : x_{i-1} = \cdots = x_{i-k} = 1\}$$

The *k/1-streak momentum statistic* $\hat{P}_{1k}(\boldsymbol{x})$ is the relative frequency of 1, for the subset of trials that are k/1-streak successors, i.e.

Definition 2 *For $\boldsymbol{x} \in \{0,1\}^n$, if the set of k/1-streak successors satisfies $I_{1k}(\boldsymbol{x}) \neq \emptyset$, then the k/1-streak momentum statistic is defined as:*

$$\hat{P}_{1k}(\boldsymbol{x}) := \frac{\sum_{i \in I_{1k}} x_i}{|I_{1k}(\boldsymbol{x})|}$$

otherwise it is not defined

Thus, the momentum statistic measures the success rate on the subset of trials that immediately succeed a streak of success(es). Now, let P_{1k} be the derived random variable with support $\{p_{1k} \in [0,1] : p_{1k} = \hat{P}_{1k}(\boldsymbol{x}) \text{ for } \boldsymbol{x} \in \{0,1\}^n,\ I_{1k}(\boldsymbol{x}) \neq \emptyset\}$, and distribution determined by the Bernoulli trials. Then the expected value of P_{1k} is equal to the expected value of $\hat{P}_{1k}(\boldsymbol{x})$ across all sequences $\boldsymbol{x}$ for which $\hat{P}_{1k}(\boldsymbol{x})$ is defined, and can be represented as follows:

$$\begin{aligned}
E[P_{1k}] &= E\left[\ \hat{P}_{1k}(\boldsymbol{x})\ \middle|\ I_{1k}(\boldsymbol{x}) \neq \emptyset\ \right] \\
&= C \sum_{n_1=k}^{n} \sum_{\substack{\{\boldsymbol{x}:\ N_1(\boldsymbol{x})=n_1 \\ I_{1k}(\boldsymbol{x})\neq\emptyset\}}} p^{n_1}(1-p)^{n-n_1} \hat{P}_{1k}(\boldsymbol{x}) \\
&= C \sum_{n_1=k}^{n} p^{n_1}(1-p)^{n-n_1} \left[\binom{n}{n_1} - U_{1k}(n,n_1)\right] \cdot E\left[\ \hat{P}_{1k}(\boldsymbol{x})\ \middle|\ I_{1k}(\boldsymbol{x}) \neq \emptyset,\ N_1(\boldsymbol{x}) = n_1\ \right] \\
&= C \sum_{n_1=k}^{n} p^{n_1}(1-p)^{n-n_1} \left[\binom{n}{n_1} - U_{1k}(n,n_1)\right] \cdot E[P_{1k}|N_1 = n_1] \qquad (1)
\end{aligned}$$

where $U_{1k}(n,n_1) := |\{\boldsymbol{x} \in \{0,1\}^n : N_1(\boldsymbol{x}) = n_1 \ \&\ I_{1k}(\boldsymbol{x}) = \emptyset\}|$ is the number of sequences for which $\hat{P}_{1k}(\boldsymbol{x})$ is undefined and C is the constant that normalizes the total probability to 1.[12] The distribution of $P_{1k}|N_1$ derived from $\boldsymbol{x}$ has support $\{p_{1k} \in [0,1] : p_{1k} = \hat{P}_{1k}(\boldsymbol{x}) \text{ for } \boldsymbol{x} \in \{0,1\}^n,\ I_{1k}(\boldsymbol{x}) = \emptyset, \text{ and } N_1(\boldsymbol{x}) = n_1\}$ for all $k \geq 1$ and $n_1 \geq 1$. In Appendix B we determine this distribution for all $k > 1$. Here we consider the case of $k = 1$, and compute the expected value directly. First we establish the main lemma.

[12]More precisely, $C := 1/\left(1 - \sum_{n_1=k}^{n} U_{1k}(n,n_1) p^{n_1}(1-p)^{n-n_1}\right)$. Note that $U_{1k} = \binom{n}{n_1}$ when $n_1 < k$, and $U_{1k} = 0$ when $n_1 > (k-1)(n-n_1)+k$. Also, for $n_1 = k$, $\hat{P}_{1k}(\boldsymbol{x}) = 0$ for all admissible $\boldsymbol{x}$.

Lemma 3 *For $n > 1$ and $n_1 = 1, \ldots, n$*

$$E[P_{11}|N_1 = n_1] = \frac{n_1 - 1}{n - 1} \tag{2}$$

Proof: See Appendix A.

The quantity $E\left[\hat{P}_{11}(\boldsymbol{x}) \,\middle|\, I_{1k}(\boldsymbol{x}) = \emptyset,\ N_1(\boldsymbol{x}) = n_1\right]$ can, in principle, be computed directly by calculating $\hat{P}_{11}(\boldsymbol{x})$ for each sequence of length n_1, and then averaging across these sequences, but the number of sequences is typically too large to perform the complete enumeration needed.[13] The key to the argument is to reduce the dimensionality of the problem by identifying the set of sequences over which $\hat{P}_{11}(\boldsymbol{x})$ is constant (this same argument can be extended to find the conditional expectation of $\hat{P}_{1k}(\boldsymbol{x})$ for $k > 1$ in Appendix B). As discussed in Section 1, we observe that each run of length j that does not appear at the end of the sequence contributes j trial observations to the computation of $\hat{P}_{11}(\boldsymbol{x})$, of which $j - 1$ are 1s, therefore:

$$\begin{aligned}\hat{P}_{11}(\boldsymbol{x}) &= \frac{\sum_{j=2}^{n_1}(j-1)R_{1j}(\boldsymbol{x})}{\sum_{j=1}^{n_1} jR_{1j}(\boldsymbol{x}) - x_n} \\ &= \frac{n_1 - R_1(\boldsymbol{x})}{n_1 - x_n}\end{aligned}$$

where $R_{1j}(\boldsymbol{x})$ is the number of runs of ones of length j, and $R_1(\boldsymbol{x})$ is the total number of runs of ones.[14] For all sequences with $R_1(\boldsymbol{x}) = r_1$, $\hat{P}_{11}(\boldsymbol{x})$ is (i) constant and equal to $(n_1 - r_1)/n_1$ for all those sequences that terminate with a zero, and (ii) constant and equal to $(n_1 - r_1)/(n_1 - 1)$ for all those sequences that terminate with a one. The distribution of r_1 in each of these cases can be found by way of combinatorial argument, and the expectation computed directly, yielding Equation 2.

This result is quantitatively identical to the formula one would get for the probability of success, conditional on n_1 successes and $n_0 = n - n_1$ failures, if one were to first remove one success from an unordered set of n trials with n_1 successes, and then draw one of the remaining $n - 1$ trials at random, without replacement. Sampling-without-replacement reasoning can be used to provide an

[13] For example, with $n = 100$ and $n_1 = 50$, there are $100!/(50!50!) > 10^{29}$ distinguishable sequences, which is greater than the nearly 10^{24} microseconds since the "big bang."

[14] More precisely, $R_{1j}(\boldsymbol{x}) := |\{i \in \{1, \ldots, n\} : \prod_{\ell=i-j+1}^{i} x_\ell = 1$ and, if $i < n$, then $x_{i+1} = 0$ and, if $i > j$, then $x_{i-j} = 0\}|$. $R_1(\boldsymbol{x}) := |\{i \in \{1, \ldots, n\} : x_i = 1$ and, if $i < n$, then $x_{i+1} = 0\}|$.

alternate proof of Lemma 3 (see Appendix A), but this reasoning does not extend to the case of $k > 1$.[15]

Given (2), Equation 1 can now be simplified, with $E[P_{11}]$ expressed in terms of only n and p

Theorem 4 *For* $p > 0$

$$E[P_{11}] = \frac{\left[p - \frac{1-(1-p)^n}{n}\right] \frac{n}{n-1}}{1 - (1-p)^n - p(1-p)^{n-1}} \tag{3}$$

Proof: See Appendix A.

In the following subsection we plot $E[P_{11}]$ as a function of n, for several values of p, and we also plot $E[P_{1k}]$ for $k > 1$.

In Section 3.2 we compare the relative frequency of success for k/1-streak successor trials (repetitions) to the relative frequency of success for k/0-streak successor trials (alternations), which requires explicit consideration of the expected difference between these conditional relative frequencies. We now consider the case in which $k = 1$ (for $k > 1$ see Appendix B), and find the difference to be independent of p. Before stating the theorem, we first define P_{0k} to be the relative frequency of a 0 for k/0-streak successor trials, i.e.

$$\hat{P}_{0k}(\boldsymbol{x}) := \frac{\sum_{i \in I_{0k}} 1 - x_i}{|I_{0k}(\boldsymbol{x})|}$$

where P_{0k} is the derived distribution. We find that for $k = 1$, the expected difference between the relative frequency of success for k/1-streak successor trials, and the relative frequency of success for k/0-streak successor trials, $D_1 := P_{11} - (1 - P_{01})$, depends only on n.[16]

Theorem 5 *Letting* $D_1 := P_{11} - (1 - P_{01})$*, then for any* $0 < p < 1$ *and* $n > 2$*, we have:*

$$E[D_1] = -\frac{1}{n-1}$$

Proof: See Appendix A

[15] For an intuition why the sampling without replacement reasoning that worked when $k = 1$ does not extend to $k > 1$, see Appendix C.

[16] This independence from p does not extend to the case of $k > 1$; see Section 2.2, and Appendix B.2.

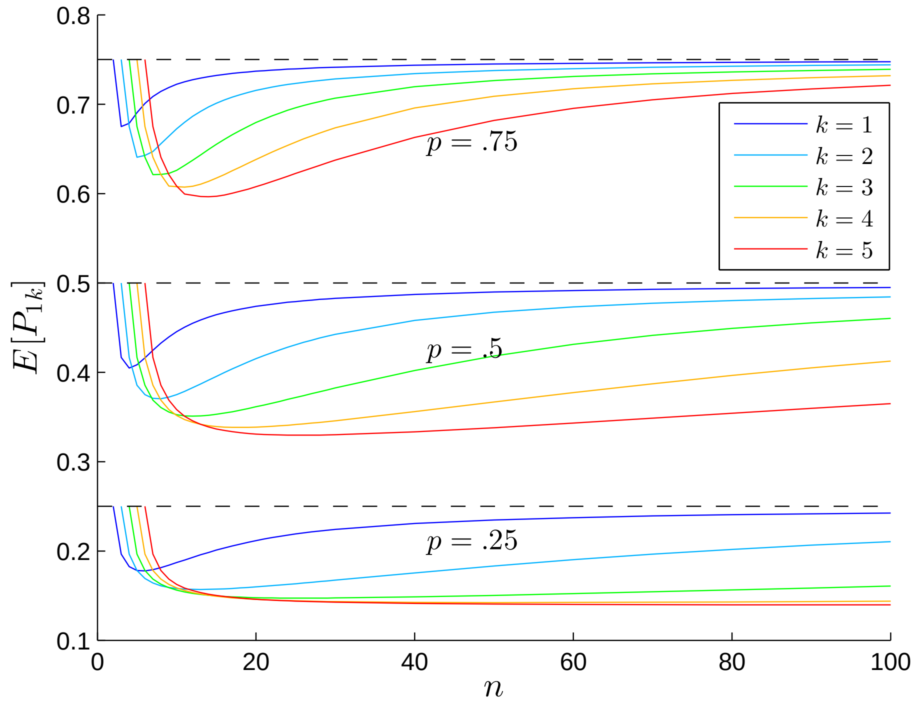


Figure 1: *The expected empirical probability of success on a trial that immediately follows a streak of k or more successes, $E[P_{1k}]$, as a function of the total number of trials, n, for different values of k and p (using the explicit formula provided in Theorem 7, combined with Equation 1).*

2.2 The degree of bias as p, n, and k vary

In Section 1, for sequences consisting of $n = 4$ trials, with the probability of success in each trial $p = .5$, Table 1 reported the empirical probability of a success (heads) on those trials that immediately follow a success. This subsection illustrates how the empirical (conditional) probability depends on p, n and k, more generally.

Figure 1, which was produced by combining Equation 1 with the explicit formula provided in Theorem 7 (Appendix B), shows how the empirical probability of success, on those trials that immediately follow k (or more) successes, changes as n, p, and k vary. Each dotted line represents the true probability of success for $p = 0.25, 0.50$, and 0.75, respectively, with the five solid lines immediately below each dotted line representing the respective expected conditional relative frequencies. Whereas intuition suggests that these expected empirical probabilities should be equal to p, one can see that they are strictly less than the true conditional probability p in all cases. It

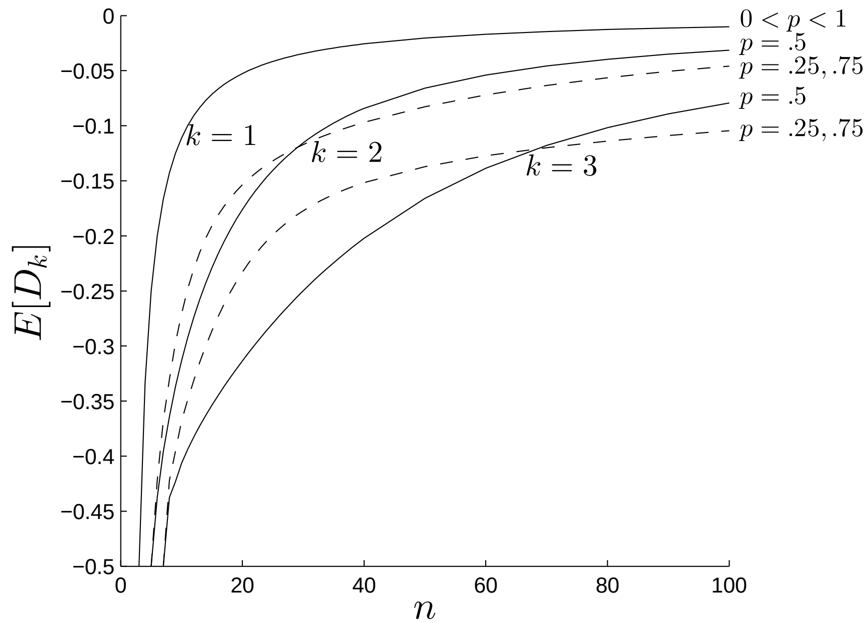


Figure 2: *The expected difference in the empirical probability of a success $E[D_k] := E[P_{1k} - (1 - P_{0k})]$, where P_{1k} is computed for trials that immediately follow a streak of k or more successes, and $(1 - P_{0k})$ is computed for trials that immediately follow a streak of k or more failures, as a function of n, and for three values of k, and various p (using the explicit formula provided in Theorem 8, combined with Equation 1).*[17]

is also shown that as n gets larger, the difference between expected conditional relative frequencies and respective probabilities of success generally decrease. Nevertheless, these differences can be substantial, even for long sequences, as can be seen, for example, in the case of $n = 100$, $p = 0.5$, and $k = 5$, in which the absolute difference is $0.50 - 0.35 = 0.15$, or in the case of $n = 100$, $p = 0.25$, and $k = 3$, in which the absolute difference is $0.25 - 0.16 = 0.09$.

Figure 2 illustrates the expected difference between the empirical probability of a success on those trials that immediately follow a streak of k or more successes (k/1-streak successor trials), and the empirical probability of a success on those trials that immediately follow a streak of k or more failures (k/0-streak successor trails), as a function of the number of trials, for $k = 1, 2, 3$ and various values of p. Whereas intuition suggests that these differences should be zero, it is shown that they are strictly negative in all cases, and can be substantial even for long sequences. Further, the bias revealed by comparing the expected difference to zero is more than twice the bias revealed by comparing the empirical probability to the true probability, as illustrated in Figure 1.

[17] In applying Equation 1, we substitute $U_k(n, n_1)$ for $U_{1k}(n, n_1)$ and $E[D_k|N_1 = n_1]$ for $E[P_{1k}|N_1 = n_1]$.

3 Applications to the Gambler's and Hot Hand Fallacies

Inferring serial dependence of any order from sequential data is an important feature of decision making in a variety of important economic domains, and has been studied in financial markets,[18] sports wagering,[19] casino gambling,[20] and lotteries.[21] The most controlled studies of decision making based on sequential data have occurred in a large body of laboratory experiments (for surveys, see Bar-Hillel and Wagenaar [1991], Nickerson [2002], Rabin [2002], and Oskarsson et al. [2009]; for a discussion of work in the economics literature, see Miller and Sanjurjo [2014]).

First, we explain how the result from Section 2 provides a structural explanation for the persistence of alternation bias and gambler's fallacy beliefs. Then we conduct a simple survey to check whether peoples' experience outside of the laboratory is consistent with the degrees of alternation bias and gambler's fallacy that they exhibit within the laboratory. Lastly, we explain how the result reveals a common error in the statistical analyses of the most prominent hot hand fallacy studies (including the original), which when corrected for, reverses what is arguably the strongest evidence that belief in the hot hand is a "cognitive illusion." In particular, what had previously been considered nearly conclusive evidence of a *lack of* hot hand performance, was instead strong evidence of hot hand performance all along.

3.1 Alternation Bias and Gambler's Fallacy

> *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 in the literature on the human perception of randomly generated sequential data is that people believe outcomes alternate more than they actually do, e.g. for a fair coin, after observing a flip of a tails, people believe that the next flip is more likely to produce a heads than a tails (Bar-Hillel and Wagenaar 1991; Nickerson 2002; Oskarsson et al. 2009; Rabin 2002).[22]

[18] Barberis and Thaler (2003); De Bondt (1993); De Long, Shleifer, Summers, and Waldmann (1991); Kahneman and Riepe (1998); Loh and Warachka (2012); Malkiel (2011); Rabin and Vayanos (2010)

[19] Arkes (2011); Avery and Chevalier (1999); Brown and Sauer (1993); Camerer (1989); Durham, Hertzel, and Martin (2005); Lee and Smith (2002); Paul and Weinbach (2005); Sinkey and Logan (2013)

[20] Croson and Sundali (2005); Narayanan and Manchanda (2012); Smith, Levere, and Kurtzman (2009); Sundali and Croson (2006); Xu and Harvey (2014)

[21] Galbo-Jørgensen, Suetens, and Tyran (2015); Guryan and Kearney (2008); Yuan, Sun, and Siu (2014)

[22] This *alternation bias* is also sometimes referred to as *negative recency bias*.

Further, as a streak of identical outcomes increases in length, people also tend to think 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).[23]

The result presented in Section 2 provides a structural explanation for the persistence of both of these systematic errors in beliefs. Independent of how these beliefs arise, to the extent that decision makers update their beliefs regarding sequential dependence with the relative frequencies that they observe in finite length sequences, no amount of exposure to these sequences can make a belief in the gambler's fallacy go away. The reason why is that for any sequence length n, even as the number of sequences observed goes to infinity, the expected rate of alternation of a given outcome is strictly larger than the (un)conditional probability of the outcome, and this expected rate of alternation generally grows larger when conditioning on streaks of preceding outcomes of increasing length (see Figure 1, Section 2). Thus, experience can, in a sense, train people to have gambler's fallacy beliefs.[24]

A possible solution to the problem is that rather than observing more sequences of size n, one could instead observe *longer* sequences; as n goes to infinity the difference between the conditional relative frequency of, say, a success, and the underlying conditional probability of a success, disappear. Nevertheless, this possibility may be unlikely to fix the problem, for the following reasons: (1) these differences only go away when n is extremely large relative to the lengths of sequences that people are likely to typically observe (as can be seen in Figure 1 and the survey results below), (2) even if one were to observe sufficiently long sequences of outcomes, memory and attention limitations may not allow them to consider more than relatively short subsequences (Bar-Hillel and Wagenaar 1991; Cowan 2001; Miller 1956; Nickerson 2002), thus effectively converting the long sequence into many shorter sub-sequences in which expected differences will again be relatively

[23] The following discussion presumes that a decision maker keeps track of the alternation rate of a particular outcome (e.g for heads, $1-\hat{p}(H|H)$), which is reasonable for many applications. For flips of a fair coin there may be no need to discriminate between an alternation after a flip of heads and an alternation after a flip of tails. In this case, the overall alternation rate, (# alternations for streaks of length 1)/(number of flips -1), is expected to be .5. It is easy to demonstrate that the alternation rate computed for any streak of length $k>1$ is expected to be greater than .5 (the explicit formula can be derived using an argument identical to that used in Theorem 8).

[24] One can imagine a decision maker who believes that $\theta = \mathbb{P}(X_i = 1|X_{i-1} = 1)$ is fixed and has a prior $\mu(\theta)$ with support $[0,1]$. The decision maker attends to trials that immediately follow one (or more) success $I' := \{i \in \{2,\dots,n\} : X_{i-1} = 1\}$, i.e. the decision maker observes the sequence $\mathbf{Y}_{|I'|} = (Y_i)_{i=1}^{|I'|} = (X_{\iota^{-1}(i)})_{i=1}^{|I'|}$, where $\iota : I' \to \{1,\dots,|I'|\}$, with $\iota(i) = |\{j \in I' : j \leq i\}|$. Whenever $|I'| > 0$, the posterior distribution becomes $p(\theta|\mathbf{Y}_{|I'|}) = \theta^{\sum_i Y_i}(1-\theta)^{|I'|-\sum_i Y_i}\mu(\theta)/\int_{\theta'}\theta'^{\sum_i Y_i}(1-\theta')^{|I'|-\sum_i Y_i}\mu(\theta')$. If $X_1,\dots,X_n$ are iid Bernoulli(p), then with repeated exposure to sequences of length n, $\mu(\theta)$ will approach point mass on the value in Equation 3, which is less than p.

large.[25]

Another possibility is that people become able to interpret observed relative frequencies in such a way as to make the difference between the underlying probability and expected conditional relative frequencies small. The way of doing this, as explained briefly in Section 1, is to count the number of observations that were used in computing the conditional relative frequency for each sequence, and then use these as weights in a weighted average of the conditional relative frequencies across sequences. Doing so will make the difference minimal. Nevertheless, this requires relatively more effort, and memory capacity, and at the same time does not intuitively seem to yield any benefit relative to the simpler and more natural approach of taking the standard (unweighted) average relative frequency across sequences. A simpler, and equivalent, correction, is that a person could instead pool observations from all sequences and compute the conditional relative frequency in the new "composite" sample. While simple in theory, this seems unlikely to occur in practice due to similar arguments, such as the immense demands on memory and attention that it would require, combined with the apparent suitability of the more natural, and simpler, alternative approach. Thus, one might conclude that because people are effectively only exposed to finite sequences of outcomes, the natural learning environment is "wicked," in the sense that it does not allow people to calibrate to the true conditional probabilities with experience alone (Hogarth 2010).

Another example of how experience may not be helpful in ridding of alternation bias and gambler's fallacy beliefs is that gambling, in games such as roulette, places no pressure on these beliefs to go away. In particular, while people can learn via reinforcement that gambling is not profitable, they cannot learn via reinforcement that it is disadvantageous to believe in excessive alternation, or in streak-effects, as the expected return is the same for all choices (Croson and Sundali 2005; Rabin 2002).

Thus, while it seems unlikely that experience alone will make alternation bias and gambler's fallacy beliefs disappear, studies have shown that people can learn to perceive randomness correctly

[25] The idea that limitations in memory capacity may lead people to rely on finite samples has been investigated in Kareev (2000). Interestingly, in earlier work Kareev (1992) notes that 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 (the explicit formula that demonstrates this can be derived using an argument identical to that used in Theorem 8).

in experimental environments, when given proper feedback and incentives (Budescu and Rapoport 1994; Lopes and Oden 1987; Neuringer 1986; Rapoport and Budescu 1992). However, to the extent that such conditions are not satisfied in real world settings, and people adapt to the natural statistics in their environment (Atick 1992; Simoncelli and Olshausen 2001), we suspect that the structural limitation to learning, which arises from the fact that sequences are of finite length, may ensure that some degree of alternation bias and gambler's fallacy persist, particularly among amateurs with little incentive to eradicate such beliefs.

It is worth noting that in light of the result presented in Section 2, behavioral models of the belief in the law of small numbers (e.g. Rabin [2002]; Rabin and Vayanos [2010]), in which the subjective probability of alternation exceeds the true probability, and grows as streak lengths increase, not only qualitatively describe behavioral patterns, but also happen to be consistent with the properties of a statistic that is natural to use in environments with sequential data—the relative frequency of success on those outcomes that immediately follow a salient streak of successes.

Survey

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). This expected alternation rate, of course, corresponds precisely with the conditional alternation rate reported for coin flip sequences of length four in Table 1. If it were the case that peoples' experience with finite sequences, either by observation or by generation, tended to involve sequences this short, then this could provide a structural explanation of why the alternation bias and gambler's fallacy have persisted at the observed magnitudes.

In order to get a sense of what people might expect alternation rates to be, based on their experience with binary outcomes outside of the laboratory, we conduct a simple survey, which 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[26]: (1) what is the largest number

[26] 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

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 (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).[27],[28]

The subjects were recruited from Bocconi University in Milan. 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. As mentioned previously, given the result presented in Section 2, for sequences of 4 flips of a fair coin, the true expected conditional alternation rate is 0.6 (as illustrated in Table 1 of Section 1), precisely in line with the average magnitude of alternation bias observed in laboratory experiments.[29] This result suggests that experience outside of the laboratory may have a meaningful effect on the behavior observed inside the laboratory.

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 2 shows subjects' responses for streaks of one, two, and three heads. One

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."

[27]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?

[28]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. 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)?

[29]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.

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]. Thus, subjects' responses are in general directionally consistent with true sample conditional alternation and repetition rates, given that in Section 2 it is demonstrated that the conditional probability of heads, when conditioning on streaks of heads, is consistently below 0.5, and is consistently above or equal to 0.5 when the outcome in question is instead a tails. 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 Figure 2 in Section 2.2), 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.[30,31]

Table 2: *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

[30] 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.

[31] 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" (Nickerson 2002). We observe subjects' responses to lie between their beliefs based on experience, and their "book knowledge."

3.2 The Hot Hand Fallacy

> *This account explains both the formation and maintenance of the erroneous belief in the hot hand: if random sequences are perceived as streak shooting, then no amount of exposure to such sequences will convince the player, the coach, or the fan that the sequences are in fact random."* (Gilovich et al. 1985)

The hot hand fallacy typically refers to the mistaken belief that success tends to follow success (hot hand), when in fact observed successes are consistent with the typical fluctuations of a chance process. The original evidence of the fallacy was provided by the seminal paper of Gilovich et al. (1985), in which the authors found that basketball players' beliefs that a player has "a better chance of making a shot after having just made his last two or three shots than he does after having just missed his last two or three shots" were not supported by the analysis of shooting data. Because these players were experts who, despite the evidence, continued to make high-stakes decisions based on their mistaken beliefs, the hot hand fallacy came to be characterized as a "massive and widespread cognitive illusion" (Kahneman 2011).[32] The strength of the original results has had a large influence on empirical and theoretical work in areas related to decision making with sequential data, both in economics and psychology.

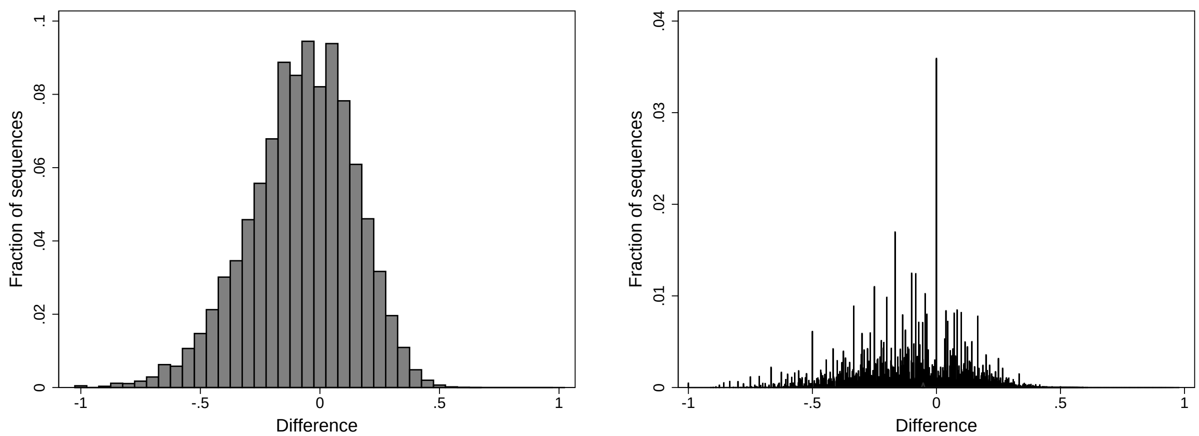
In the Gilovich et al. (1985) study, and the other studies in basketball that follow, player performance records – patterns in hits (successes) and misses (failures) – are checked to see if they "differ from sequences of heads and tails produced by [weighted] coin tosses"(Gilovich et al. 1985). The standard measure of hot hand effect size in these studies is to compare the empirical probability of a hit on those shots that immediately follow a streak of hits to the empirical probability of a hit on those shots that immediately follow a streak of misses, where a streak is typically defined as a shot taken after $3, 4, 5, \ldots$ hits in a row (Avugos, Bar-Eli, Ritov, and Sher 2013; Gilovich et al. 1985; Koehler and Conley 2003).[33] This comparison appears to be a sound one: if a player

[32] The fallacy view has been the consensus in the economics literature on sequential decision making, and the existence of the fallacy itself has been highlighted in the general discourse as a salient example of how statistical analysis can reveal the flaws of expert intuition (e.g. see (Davidson 2013)).

[33] While defining the cutoff for a streak in this way agrees with the "rule of three" for human perception of streaks (Carlson and Shu 2007), when working with data, the choice of the cutoff involves a trade-off. The larger the cutoff k, the higher the probability that the player is actually hot on those shots that immediately follow these streaks, which reduces measurement error. On the other hand, as k gets larger, fewer shots are available, which leads to a smaller sample size and reduced statistical power (and a higher bias in the empirical probability). See Miller and Sanjurjo (2014) for a more thorough discussion, which explicitly considers a number of plausible hot hand models.

Figure 3: *The graph of the (exact) discrete probability distribution of* $E[P_{1k} - (1 - P_{0k})|N_1 = n_1]$*, the difference between the conditional relative frequency of a success on those trials that immediately follow a streak of* 3 *or more successes (3/1-streak successor trials), and the conditional relative frequency of a success on those trials that immediately follow a streak of* 3 *or more failures (3/0-streak successor trails), with* $n = 100$, $n_1 = 50$ *(using the explicit formula for the distribution provided in the proof of Theorem 8).*[34]

(a) *Bin width = 0.05* **(b)** *Bin width = 0.0001*



is consistent, i.e. always shoots with the same probability of a hit, then the fact that a given shot attempt is taken immediately following a streak of hits, or misses, is determined by chance. Therefore, it appears that these two sets of shot attempts can be regarded as two treatments that are statistically independent. However, what this reasoning ignores is that within any given sequence of finite length, when conditioning on a streak of three (or more) hits, each run of three hits, MHHHM, leads the researcher to select 1 shot for analysis, which is always a miss; each run of four hits, MHHHHM, leads the researcher to select 2 shots for analysis, one hit and one miss, and so on. Because relatively short runs are considerably more common in finite sequences, there is a bias towards selecting shots that are misses. An analogous argument applies to conditioning on streaks of misses, but in this case the selection bias is towards observing shots that are hits.

The seemingly correct, but mistaken, assumption that in a sequence of coin tosses, the empirical probability of heads on those flips that immediately follow a streak of heads is expected to be equal to the empirical probability of heads on those flips that immediately follow a streak of tails, has striking effects on the interpretation of shooting data. As was shown in Figure 2 of Section 2.2,

[34] The distribution is displayed using a reduced 6 decimal digits of precision. For this precision, the more than 10^{29} distinguishable sequences take on 19,048 distinct values when calculating the difference in relative frequencies. In the computation of the expected value in Figures 1 and 2, each difference is represented with the highest floating point precision available.

the bias in this comparison between conditional relative frequencies is more than double that of the comparison of either conditional relative frequency to the true probability (under the Bernoulli assumption). If players were to shoot with a fixed hit rate (the null Bernoulli assumption), then, given the parameters of the original study, one should in fact expect the difference in these relative frequencies to be -0.08, rather than 0. Moreover, the distribution of the differences will have a pronounced negative skew. In Figure 3 we present the *exact* distribution of the difference between these conditional relative frequencies (for two different bin sizes), following Theorem 8 of Appendix B. The distribution is generated using the target parameters of the original study: sequences of length $n = 100$, $n_1 = 50$ hits, and streaks of length $k = 3$ or more. As can be seen, the distribution has a pronounced negative skew, with 63 percent of observations less than 0 (median = -.06).

The effects of this bias can be seen in Table 3, which reproduces data from Table 4 of Gilovich et al. (1985). The table presents shooting performance records for each of the 14 male and 12 female Cornell University basketball players who participated in the controlled shooting experiment of the original hot hand study (Gilovich et al. 1985). One can see the number of shots taken, overall field goal percentage, relative frequency of a hit on those shots that immediately follow a streak of three (or more) hits, $\hat{p}(hit|3\ hits)$, relative frequency of a hit on those shots that immediately follow a streak of three (or more) misses, $\hat{p}(hit|3\ misses)$, and the expected difference between these quantities. Under the incorrect assumption that these conditional relative frequencies are expected to be equal to each player's overall field goal percentage, it indeed appears as if there is little to no evidence of hot hand shooting. Likewise, under the incorrect assumption that the difference between these two conditional relative frequencies is expected to be zero, the difference of $+0.03$ may seem like directional evidence of a slight hot hand, but it is not statistically significant.[35] The last column corrects for the incorrect assumption by subtracting from each player's observed difference the actual expected difference under the null hypothesis that the player shoots with a constant probability of a hit, given his/her overall field goal percentage (see the end of the section for more on the bias correction). Once this bias correction is made, one can see that 19 of the 25 players directionally exhibit hot hand shooting, and that the average difference in shooting

[35] A common criticism of the original study, as well as subsequent studies, is that they are under-powered, thus even substantial differences are not registered as significant (see Miller and Sanjurjo (2014) for a power analysis, and a complete discussion of the previous literature).

Table 3: *Columns 4-5 reproduce columns 2 and 8 of Table 4 from Gilovich et al. (1985) (note: 3 hits/misses includes streaks of 3, 4, 5, etc.). Column 6 reports the difference between the reported relative frequencies, and column 7 adjusts for the bias (mean correction), based on the player's field goal percentage (probability in this case) and number of shots.*

					$\hat{p}(hit\|3\ hits) - \hat{p}(hit\|3\ misses)$	
Player	# shots	fg%	$\hat{p}(hit\|3\ hits)$	$\hat{p}(hit\|3\ misses)$	GVT est.	bias adj.
Males						
1	100	.54	.50	.44	.06	.14
2	100	.35	.00	.43	−.43	−.33
3	100	.60	.60	.67	−.07	.02
4	90	.40	.33	.47	−.13	−.03
5	100	.42	.33	.75	−.42	−.33
6	100	.57	.65	.25	.40	.48
7	75	.56	.65	.29	.36	.47
8	50	.50	.57	.50	.07	.24
9	100	.54	.83	.35	.48	.56
10	100	.60	.57	.57	.00	.09
11	100	.58	.62	.57	.05	.14
12	100	.44	.43	.41	.02	.10
13	100	.61	.50	.40	.10	.19
14	100	.59	.60	.50	.10	.19
Females						
1	100	.48	.33	.67	−.33	−.25
2	100	.34	.40	.43	−.03	.07
3	100	.39	.50	.36	.14	.23
4	100	.32	.33	.27	.07	.17
5	100	.36	.20	.22	−.02	.08
6	100	.46	.29	.55	−.26	−.18
7	100	.41	.62	.32	.30	.39
8	100	.53	.73	.67	.07	.15
9	100	.45	.50	.46	.04	.12
10	100	.46	.71	.32	.40	.48
11	100	.53	.38	.50	−.12	−.04
12	100	.25	.	.32	.	.
Average		.47	.49	.45	.03	.13

percentage when on a streak of hits vs. misses is +13 percentage points.[36,37] This is a substantial effect size, given that all shooters are included in the average, and that in the 2013-2014 NBA season, the difference between the very best three point shooter in the league, and the median shooter, was +10 percentage points.

Miller and Sanjurjo (2014) provide a statistical testing procedure with a permutation-based re-sampling scheme that is invulnerable to this downward bias (explained below), and then analyze data from all extant controlled shooting studies. They find substantial evidence of hot hand shooting in all studies. Miller and Sanjurjo (2015) apply the same corrected statistical approach to 29 years of NBA Three-Point Contest data, and again find substantial evidence of hot hand shooting, in stark contrast with Koehler and Conley (2003), which has been considered a particularly clean replication of the original study, but which happened to use the same incorrect measure of hot hand effect size as Gilovich et al. (1985).[38]

A bias-free procedure for testing hypotheses

The method used to construct the histogram in Figure 3 can also be used to de-bias the hypothesis test conducted in Gilovich et al. (1985). Under the null hypothesis, a player is a consistent shooter, which means that the player has a fixed probability p of hitting a shot throughout his or her sequence of shots. While this p is unobservable to the researcher, upon observing a player shoot n shots with n_1 hits, the null assumption implies that all sequences have the same probability $p^{n_1}(1-p)^{n-n_1}$ of occurring, regardless of the order of shots. For each unique sequence with n_1 hits, the difference in relative frequencies $D_k := \hat{p}(hit|k\ hits) - \hat{p}(hit|k\ misses)$ can be computed, and the exact distribution of the difference can be derived using the argument in Theorem 8. As in Figure 3, the distribution will have a negative skew.[39] Letting $\boldsymbol{x} \in [0,1]^n$ be a sequence of shot outcomes in which $D_k(\boldsymbol{x})$ is defined, the hot hand hypothesis predicts that $D_k(\boldsymbol{x})$ will be significantly greater than what one would expect by chance, and the critical value c_{α,n_1} for the

[36] The effective adjustment after pooling (0.09) is higher than the expected bias for a $p = .5$ shooter due to the different shooting percentages.

[37] The actual bias when a player has the hot hand can be considerably larger than the bias for a player with a fixed probability of a hit, e.g. if the rows of Table 1 were generated by a Markov chain with initial probability $\mathbb{P}(hit) = .5$ and transition probabilities defined by $\mathbb{P}(hit|1\ hit) = \mathbb{P}(miss|1\ miss) = 0.6$, then $E[\hat{p}(hit|1\ hit) - \hat{p}(hit|1\ miss)] = -0.4125 < -0.33 = E[\hat{p}(H|H) - \hat{p}(H|T)]$.

[38] Koehler and Conley (2003) studies 4 years of NBA Three-Point contest data.

[39] Sequences that do not contain at least one trial that immediately follows each streak type must of course be discarded.

associated statistical test is the smallest c such that $\mathbb{P}(D_k(\boldsymbol{x}) \geq c \mid H_0,\ \sum_{i=1}^{n} x_i = n_1) \leq \alpha$.[40] From an ex-ante perspective, a test of the hot hand at the α level of significance consists of a family of such critical values $\{c_{\alpha,n_1}\}$. It is immediate that $\mathbb{P}(\text{reject}|H_0) \leq \alpha$ because $\mathbb{P}(\text{reject}|H_0) = \sum_{n_1=1}^{n} \mathbb{P}(D_k(\boldsymbol{x}) \geq c_{\alpha,n_1}|H_0,\ \sum_{i=1}^{n} x_i = n_1)\mathbb{P}(\sum_{i=1}^{n} x_i = n_1|H_0) \leq \alpha$.[41]

4 Conclusion

We find that in a finite sequence of data that is generated by repeated realizations of a Bernoulli random variable, the expected *empirical* probability of success, on those observations in the sequence that immediately follow a streak of success realizations, is *strictly less than* the true (fixed) probability of success. The mechanism is form of selection bias that arises due to the sequential structure of the (finite) data. One direct implication of this result is a structural explanation for the persistence of alternation bias and gambler's fallacy beliefs. Another is that empirical approaches of the most prominent studies in the hot hand fallacy literature are incorrect. Once corrected for, the data that was previously interpreted as providing substantial evidence that the belief in the hot hand is fallacy, reverses, and becomes substantial evidence that it is not a fallacy to believe in the hot hand. Finally, the respective errors of the gambler, and the hot hand fallacy researcher, are found to be analogous.

References

Arkes, J. (2011): "Do Gamblers Correctly Price Momentum in NBA Betting Markets?" *Journal of Prediction Markets*, 5, 31–50.

Atick, J. J. (1992): "Could information theory provide an ecological theory of sensory processing?" *Network: Computation in neural systems*, 3, 213–251.

[40] For the quantity $\mathbb{P}(D_k(\boldsymbol{x}) \geq c \mid H_0,\ \sum_{i=1}^{n} x_i = n_1)$ it may be the case that for some c^*, it is strictly greater than α for $c \leq c^*$, and equal to 0 for $c > c^*$. In this case, for any sequence with $\sum_{i=1}^{n} x_i = n_1$ one cannot reject H_0 at an α level of significance.

[41] That observed shot outcomes are exchangeable under the null hypothesis means that a hypothesis test can be conducted for any statistic of the data. Miller and Sanjurjo (2014) outline a hypothesis test procedure that uses a re-sampling scheme with Monte-Carlo permutations of the data to generate the null distribution of any test statistic. In addition, they propose three test statistics (and a composite) statistic that are shown to have greater statistical power than previous measures. Finally, they find that these tests detect significant and substantial hot hand effect sizes across all extant controlled shooting data sets.

Avery, C. and J. Chevalier (1999): "Identifying Investor Sentiment from Price Paths: The Case of Football Betting," *Journal of Business*, 72, 493–521.

Avugos, S., M. Bar-Eli, I. Ritov, and E. Sher (2013): "The elusive reality of efficacy performance cycles in basketball shooting: analysis of players performance under invariant conditions," *International Journal of Sport and Exercise Psychology*, 11, 184–202.

Bai, D. S. (1975): "Efficient Estimation of Transition Probabilities in a Markov Chain," *The Annals of Statistics*, 3, 1305–1317.

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

Barberis, N. and R. Thaler (2003): "A survey of behavioral finance," *Handbook of the Economics of Finance*, 1, 1053–1128.

Benjamin, D. J., M. Rabin, and C. Raymond (2014): "A Model of Non-Belief in the Law of Large Numbers," Working Paper.

Brown, W. A. and R. D. Sauer (1993): "Does the Basketball Market Believe in the Hot Hand? Comment," *American Economic Review*, 83, 1377–1386.

Budescu, D. V. and A. Rapoport (1994): "Subjective randomization in one-and two-person games," *Journal of Behavioral Decision Making*, 7, 261–278.

Camerer, C. F. (1989): "Does the Basketball Market Believe in the 'Hot Hand,'?" *American Economic Review*, 79, 1257–1261.

Carlson, K. A. and S. B. Shu (2007): "The rule of three: how the third event signals the emergence of a streak," *Organizational Behavior and Human Decision Processes*, 104, 113–121.

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.

Davidson, A. (2013): "Boom, Bust or What?" New York Times.

De Bondt, W. P. (1993): "Betting on trends: Intuitive forecasts of financial risk and return," *International Journal of Forecasting*, 9, 355–371.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1991): "The Survival of Noise Traders In Financial-markets," *Journal of Business*, 64, 1–19.

Durham, G. R., M. G. Hertzel, and J. S. Martin (2005): "The Market Impact of Trends and Sequences in Performance: New Evidence," *Journal of Finance*, 60, 2551–2569.

Feller, W. (1968): *An Introduction to Probability Theory and Its Applications*, New York: John Wiley & Sons.

Galbo-Jørgensen, C. B., S. Suetens, and J.-R. Tyran (2015): "Predicting Lotto Numbers A natural experiment on the gamblers fallacy and the hot hand fallacy," *Journal of the European Economic Association*, forthcoming, working Paper.

Gilovich, T., R. Vallone, and A. Tversky (1985): "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology*, 17, 295–314.

Guryan, J. and M. S. Kearney (2008): "Gambling at Lucky Stores: Empirical Evidence from State Lottery Sales," *American Economic Review*, 98, 458–473.

Hogarth, R. M. (2010): "Intuition: A Challenge for Psychological Research on Decision Making," *Psychological Inquiry*, 21, 338–353.

Kahneman, D. (2011): *Thinking, Fast and Slow*, Farrar, Straus and Giroux.

Kahneman, D. and M. W. Riepe (1998): "Aspects of Investor Psychology: Beliefs, preferences, and biases investment advisors should know about," *Journal of Portfolio Management*, 24, 1–21.

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.

——— (2000): "Seven (indeed, plus or minus two) and the detection of correlations." *Psychological Review*, 107, 397–402.

Koehler, J. J. and C. A. Conley (2003): "The "hot hand" myth in professional basketball," *Journal of Sport and Exercise Psychology*, 25, 253–259.

Lee, M. and G. Smith (2002): "Regression to the mean and football wagers," *Journal of Behavioral Decision Making*, 15, 329–342.

Loh, R. K. and M. Warachka (2012): "Streaks in Earnings Surprises and the Cross-Section of Stock Returns," *Management Science*, 58, 1305–1321.

Lopes, L. L. and G. C. Oden (1987): "Distinguishing between random and nonrandom events," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 392–400.

Malkiel, B. G. (2011): *A random walk down Wall Street: the time-tested strategy for sucessful investing*, New York: W. W. Norton & Company.

Miller, G. A. (1956): "The magical number seven, plus or minus two: Some limits on our capacity for processing information." *Psychological Review*, 63, 81–97.

Miller, J. B. and A. Sanjurjo (2014): "A Cold Shower for the Hot Hand Fallacy," Working Paper.

——— (2015): "Is the Belief in the Hot Hand a *Fallacy* in the NBA Three Point Shootout?" Working Paper.

Narayanan, S. and P. Manchanda (2012): "An empirical analysis of individual level casino gambling behavior," *Quantitative Marketing and Economics*, 10, 27–62.

Nelson, C. R. and M. J. Kim (1993): "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance*, 48, 641–661.

Neuringer, A. (1986): "Can people behave "randomly?": The role of feedback." *Journal of Experimental Psychology: General*, 115, 62.

Nickerson, R. S. (2002): "The production and perception of randomness," *Psychological Review*, 109, 350–357.

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.

Paul, R. J. and A. P. Weinbach (2005): "Bettor Misperceptions in the NBA: The Overbetting of Large Favorites and the 'Hot Hand'," *Journal of Sports Economics*, 6, 390–400.

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.

Rapoport, A. and D. V. Budescu (1992): "Generation of random series in two-person strictly competitive games." *Journal of Experimental Psychology: General*, 121, 352–.

Rinott, Y. and M. Bar-Hillel (2015): "Comments on a 'Hot Hand' Paper by Miller and Sanjurjo," Federmann Center For The Study Of Rationality, The Hebrew University Of Jerusalem. Discussion Paper # 688 (August 11).

Riordan, J. (1958): *An Introduction to Combinatorial Analysis*, New York: John Wiley & Sons.

Simoncelli, E. P. and B. A. Olshausen (2001): "Natural Image Statistics And Neural Representation," *Annual Review of Neuroscience*, 24, 1193–1216.

Sinkey, M. and T. Logan (2013): "Does the Hot Hand Drive the Market?" *Eastern Economic Journal*, Advance online publication, doi:10.1057/eej.2013.33.

Smith, G., M. Levere, and R. Kurtzman (2009): "Poker Player Behavior After Big Wins and Big Losses," *Management Science*, 55, 1547–1555.

Stambaugh, R. F. (1986): "Bias in Regressions with lagged Stochastic Regressors," Working Paper No. 156.

——— (1999): "Predictive regressions," *Journal of Financial Economics*, 54, 375–421.

Sundali, J. and R. Croson (2006): "Biases in casino betting: The hot and the gamblers fallacy," *Judgement and Decision Making*, 1, 1–12.

Xu, J. and N. Harvey (2014): "Carry on winning: The gambler's fallacy creates hot hand effects in online gambling," *Cognition*, 131, 173 – 180.

Yuan, J., G.-Z. Sun, and R. Siu (2014): "The Lure of Illusory Luck: How Much Are People Willing to Pay for Random Shocks," *Journal of Economic Behavior & Organization*, forthcoming.

Yule, G. U. (1926): "Why do we Sometimes get Nonsense-Correlations between Time-Series?–A Study in Sampling and the Nature of Time-Series," *Journal of the Royal Statistical Society*, 89, 1–63.

A Appendix: Section 2 Proofs

Proof of Lemma 3

For $n_1 = 1$, clearly $\hat{P}_{11}(\boldsymbol{x}) = 0$ for all $\boldsymbol{x}$, and the identity is satisfied. For $n_1 > 1$ this quantity cannot be computed directly by calculating its value for each sequence because the number of admissible sequences is typically too large.[42] In order to handle the case of $n_1 > 1$, we first define $R_1(\boldsymbol{x})$ as the number of runs of ones, i.e. the number of subsequences of consecutive ones in sequence $\boldsymbol{x}$ that are flanked by zeros or an end point.[43] The key observation is that for all sequences with $R_1(\boldsymbol{x}) = r_1$, $\hat{P}_{11}(\boldsymbol{x})$ is (i) constant and equal to $(n_1 - r_1)/n_1$ across all those sequences that terminate with a zero, and (ii) constant and equal to $(n_1 - r_1)/(n_1 - 1)$ across all those sequences that terminate with a one. The number of sequences in each of these cases can be counted using a combinatorial argument.

Any sequence with r_1 ones can be constructed, first, by building the runs of ones of fixed length with an ordered partition of the n_1 ones into r_1 cells (runs), which can be performed in $\binom{n_1-1}{r_1-1}$ ways by inserting $r_1 - 1$ dividers into the $n_1 - 1$ available positions between ones, and second, by placing the r_1 ones into the available positions to the left or the right of a zero among the n_0 zeros to form the final sequence. For the case in which $x_n = 0$ there are n_0 available positions to place the runs, and therefore $\binom{n_0}{r_1}$ possible placements, while in the case in which $x_n = 1$ (which must end in a run of ones) there are n_0 available positions to place the $r_1 - 1$ remaining runs, and therefore $\binom{n_0}{r_1-1}$ possible placements. Note that for $n_1 > 1$, we have $U_{11}(n, n_1) = 0$ and $r_1 \leq n_1$, therefore

$$E[P_{11}|N_1 = n_1] = \frac{1}{\binom{n}{n_1}} \sum_{x_n \in \{0,1\}} \sum_{r_1=1}^{\min\{n_1, n_0 + x_n\}} \binom{n_1 - 1}{r_1 - 1} \binom{n_0}{r_1 - x_n} \frac{n_1 - r_1}{n_1 - x_n}$$

[42] For example, with $n = 100$, $n_1 = 50$ and $k = 1$ there $100!/(50!50!) > 2^{50} > 10^{15}$ such sequences.

[43] The number of runs of ones can be defined explicitly to be the number of trials in which a one occurs and is immediately followed by a zero on the next trial or has no following trial, i.e. $R_1(\boldsymbol{x}) := |\{i \in \{1, \ldots, n\} : x_i = 1 \text{ and, if } i < n \text{ then } x_{i+1} = 0\}|$

For the case in which $x_n = 0$, the inner sum satisfies:

$$\begin{aligned}\sum_{r_1=1}^{\min\{n_1,n_0\}} \binom{n_1-1}{r_1-1}\binom{n_0}{r_1}\frac{n_1-r_1}{n_1} &= \sum_{r_1=1}^{\min\{n_1,n_0\}} \binom{n_1-1}{r_1-1}\binom{n_0}{r_1}\left(1-\frac{r_1}{n_1}\right)\\ &= \binom{n-1}{n_0-1} - \frac{1}{n_1}\sum_{r_1=1}^{\min\{n_1,n_0\}} \binom{n_1-1}{r_1-1}\binom{n_0}{r_1} r_1\\ &= \binom{n-1}{n_0-1} - \frac{n_0}{n_1}\sum_{r_1=1}^{\min\{n_1,n_0\}} \binom{n_1-1}{r_1-1}\binom{n_0-1}{r_1-1}\\ &= \binom{n-1}{n_0-1} - \frac{n_0}{n_1}\sum_{x=0}^{\min\{n_1-1,n_0-1\}} \binom{n_1-1}{x}\binom{n_0-1}{x}\\ &= \binom{n-1}{n_0-1} - \frac{n_0}{n_1}\binom{n-2}{n_1-1}\end{aligned}$$

The left term of the second line follows because it is the total number of sequences that can be formed in the first $n-1$ positions with n_0-1 zeros and $n_1 = n - n_0$ ones. The final line follows from an application of Vandermonde's convolution.[44]

For the case in which $x_n = 1$, the inner sum can be reduced using similar arguments:

[44] Vandermonde's convolution is given as

$$\sum_{k=\max\{-m,n-s\}}^{\min\{r-m,n\}} \binom{r}{m+k}\binom{s}{n-k} = \binom{r+s}{m+n}$$

from which one can derive the following identity, which we apply

$$\sum_{k=\max\{-m,-n\}}^{\min\{\ell-m,s-n\}} \binom{\ell}{m+k}\binom{s}{n+k} = \sum_{k=\max\{-m,-n\}}^{\min\{\ell-m,s-n\}} \binom{s}{n+k}\binom{\ell}{(\ell-m)-k} = \binom{\ell+s}{\ell-m+n}$$

$$\sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_1-1}{r_1-1}\binom{n_0}{r_1-1}\frac{n_1-r_1}{n_1-1} = \frac{n_1}{n_1-1}\binom{n-1}{n_0} - \frac{1}{n_1-1}\sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_1-1}{r_1-1}\binom{n_0}{r_1-1}r_1$$

$$= \binom{n-1}{n_0} - \frac{1}{n_1-1}\sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_1-1}{r_1-1}\binom{n_0}{r_1-1}(r_1-1)$$

$$= \binom{n-1}{n_0} - \frac{n_0}{n_1-1}\sum_{r_1=2}^{\min\{n_1,n_0+1\}} \binom{n_1-1}{r_1-1}\binom{n_0-1}{r_1-2}$$

$$= \binom{n-1}{n_0} - \frac{n_0}{n_1-1}\sum_{x=0}^{\min\{n_1-2,n_0-1\}} \binom{n_1-1}{x+1}\binom{n_0-1}{x}$$

$$= \binom{n-1}{n_0} - \frac{n_0}{n_1-1}\binom{n-2}{n_1-2}$$

Combining both cases we have:

$$\begin{aligned} E[P_{11}|N_1=n_1] &= \frac{1}{\binom{n}{n_1}}\left[\binom{n-1}{n_0-1} - \frac{n_0}{n_1}\binom{n-2}{n_1-1} + \binom{n-1}{n_0} - \frac{n_0}{n_1-1}\binom{n-2}{n_1-2}\right] \\ &= \frac{1}{\binom{n}{n_1}}\left[\binom{n}{n_0} - \frac{n_0}{n-1}\binom{n}{n_1}\right] \\ &= \frac{n_1-1}{n-1} \end{aligned}$$

■

Alternate proof of Lemma 3:

Recall that for a sequence that ends in a zero P_{11} will be computed based on n_1 observations, and for a sequence that ends in a one P_{11} will be computed based on $n_1 - 1$ observations. Each case can be addressed separately, as before, noting again that $E[P_{11}|N_1 = n_1] = \sum_{x_n\in\{0,1\}} E[P_{11}|N_1 = n_1, X_n = x_n]\mathbb{P}(X_n = x_n|N_1 = n_1)$, where $\mathbb{P}(X_n = 1|N_1 = n_1) = n_1/n$. Below we demonstrate that $E[P_{11}|N_1 = n_1, X_n = x_n] = (n_1 - 1)/(n - 1)$ for $x_n \in \{0, 1\}$, which proves the result.

For the case in which $x_n = 1$, note that $E[P_{11}|N_1 = n_1, X_n = 1]$ is the probability of selecting a trial that is a 1 in the two-step procedure which consists of first selecting one sequence at random, for $N_1 = n_1$ and $X_n = 1$, and then selecting at random one of the $n_1 - 1$ trials that immediately follow a 1, and observing the outcome. Because each sequence with $N_1 = n_1$ and $X_n = 1$ contributes the same number of trial observations $(n_1 - 1)$ to the computation of P_{11}, this procedure is equivalent to selecting one trial at random from the $\binom{n-1}{n_1-1}(n_1 - 1)$ trials that immediately follow a 1 (across all sequences with $N_1 = n_1$ and $X_n = 1$). It is equiprobable that the selected trial be located in any of the positions from 2 to n. If the trial is located in one of the positions from 2 to $n-1$, then there are $n_1 - 2$ 1s available (having removed the 1s from the selected trial and the end point) among the $n-2$ trials available, and therefore the probability that the trial is a 1 is equal to $(n_1 - 2)/(n - 2)$. On the other hand, if the trial is located in position n, then the trial must be a 1. Therefore $E[P_{11}|N_1 = n_1, X_n = 1] = (n - 2)[(n_1 - 2)/(n - 2)][1/(n - 1)] + 1/(n - 1) = (n_1 - 1)/(n - 1)$.

For the case in which $x_n = 0$, analogous to the previous argument, $E[P_{11}|N_1 = n_1, X_n = 0]$ is the probability of selecting a trial that is a 1 in the procedure of selecting one trial at random from the $\binom{n-1}{n_1}n_1$ trials that immediately follow a 1 (across all sequences with $N_1 = n_1$ and $X_n = 0$). As before, it is equiprobable that the selected trial be located in any of the positions from 2 to n. If the trial is located in one of the positions from 2 to $n - 1$, then there are $n_1 - 1$ 1s available (having removed the 1 from the selected trial) among the $n - 2$ trials available, and therefore the probability that the trial is a 1 is equal to $(n_1 - 1)/(n - 2)$. On the other hand, if the trial is located in position n, then the trial must be a 0. Therefore $E[P_{11}|N_1 = n_1, X_n = 0] = (n-2)[(n_1-1)/(n-2)][1/(n-1)] = (n_1-1)/(n-1) = (n_1-1)/(n-1)$.
■

Proof of Lemma 4

The full derivation of Equation 3 is as follows.

$$
\begin{aligned}
E[P_{11}] &= C \sum_{n_1=1}^{n} p^{n_1}(1-p)^{n-n_1} \left[\binom{n}{n_1} - U_{11}(n, n_1) \right] \cdot E[P_{11}|N_1 = n_1] \\
&= \frac{\sum_{n_1=2}^{n} \binom{n}{n_1} p^{n_1}(1-p)^{n-n_1} \frac{n_1-1}{n-1}}{1-(1-p)^n - p(1-p)^{n-1}} \\
&= \frac{\frac{1}{n-1}\left[\left(np - np(1-p)^{n-1}\right) - \left(1-(1-p)^n - np(1-p)^{n-1}\right) \right]}{1-(1-p)^n - p(1-p)^{n-1}} \\
&= \frac{\left[p - \frac{1-(1-p)^n}{n} \right] \frac{n}{n-1}}{1-(1-p)^n - p(1-p)^{n-1}}
\end{aligned}
$$

where the second line follows because $U_{11}(n,n_1)=0$ for $n_1>1$ and $C=1/[1-(1-p)^n-p(1-p)^{n-1}]$. Clearly $E[P_{11}]<p$.

■

Proof of Lemma 5

We show that for $n>2$ and $n_1=1,\dots,n-1$,

$$E[D_1|N_1=n_1]=-\frac{1}{n-1}$$

If $1<n_1<n-1$ then $D_1=P_{11}-(1-P_{01})$ is defined for all sequences. Therefore, by linearity of the expectation, and Lemma 3, we have:

$$\begin{aligned}E[D_1|N_1=n_1]=&E[P_{11}|N_1=n_1]-E[(1-P_{01}|N_1=n_1]\\=&\frac{n_1-1}{n-1}-\left(1-\frac{n_0-1}{n-1}\right)\\=&-\frac{1}{n-1}\end{aligned}$$

If $n_1=1$ then there are $n-1$ possible sequences in which D_1 is defined (i.e. with 1 not in the final position). For the sequence in which 1 is in the first position, $D_1=0$. For the other $n-2$ sequences, $D_1=-1/(n-2)$. Therefore, $E[D_1|N_1=1]=-1/(n-1)$. The argument for $n_1=n-1$ is analogous, with D_1 undefined for the sequence in which there is a 0 in the last position, equal to 0 for the sequence in which there is 0 in the first position, and equal to $-1/(n-2)$ for all other sequences.

That the conditional expectation is independent of N_1 implies that $E[D_1]$ is independent of p, and we have the result.

■

B Appendix: Quantifying the Bias for $k > 1$

In this section, for $k > 1$, we obtain the expected relative frequency of a 1 for k/1-streak successor trials, $E[P_{1k}]$, and the expected difference in the relative frequency of 1, between k/1-streak successor trials and k/0-streak successor trials, $E[P_{1k} - (1 - P_{0k})]$. Similar to what was done in the proof of the $k = 1$ case, representing the relative frequency in terms of runs allows us to reduce the dimensionality of the problem by identifying the set of sequences over which $\hat{P}_{1k}(\boldsymbol{x})$ is constant. We begin with some basic definitions.

Given the sequence $\boldsymbol{x} = (x_1, \ldots, x_n)$, recall that a run of 1s is a subsequence of consecutive 1s in $\boldsymbol{x}$ that is flanked on each side by a 0 or an endpoint.[45] Define runs of 0s analogously to runs of 1s. Let $R_{1j}(\boldsymbol{x}) = r_{1j}$ be the number of runs of 1s of exactly length j for $j = 1, \ldots, n_1$. Let $R_{0j}(\boldsymbol{x}) = r_{0j}$ be defined similarly. Let $S_{1j}(\boldsymbol{x}) = s_{1k}$ be the number of runs of 1s of length j or more, i.e. $S_{1j}(\boldsymbol{x}) := \sum_{i=j}^{n_1} R_{1j}(\boldsymbol{x})$ for $j = 1, \ldots, n_1$, with $S_{0j}(\boldsymbol{x}) = s_{0j}$ defined similarly. Let $R_1(\boldsymbol{x}) = r_1$, be the number of runs of 1s, i.e. $R_1(\boldsymbol{x}) = S_{11}(\boldsymbol{x})$, and $R_0(\boldsymbol{x}) = r_0$ be the number of runs of 0s. Let $R(\boldsymbol{x}) = r$ be the total number of runs, i.e. $R(\boldsymbol{x}) := R_1(\boldsymbol{x}) + R_0(\boldsymbol{x})$. Further, let the k/1-streak *frequency* statistic $F_{1k}(\boldsymbol{x}) = f_{1k}$ be defined as the number of (overlapping) 1-streaks of length k, i.e. $F_{1k}(\boldsymbol{x}) := \sum_{j=k}^{n_1} (j - k + 1) R_{1j}(\boldsymbol{x})$, with $F_{0k}(\boldsymbol{x}) = f_{0k}$ defined analogously. Notice that $f_{1k} = |I_{1k}(\boldsymbol{x})|$ if $\exists i > n - k$ with $x_i = 0$, and $f_{1k} = |I_{1k}| + 1$ otherwise. Also note that $n_1 = f_{11} = \sum_{j=1}^{n_1} j r_{1j}$ and $n_0 = f_{01} = \sum_{j=1}^{n_0} j r_{0j}$.

To illustrate the definitions, consider the sequence of 10 trials 1101100111. The number of 1s is given by $n_1 = 7$. For $j = 1, \ldots, n_1$, the number of runs of 1s of exactly length j is given by $r_{11} = 0$, $r_{12} = 2$, $r_{13} = 1$ and $r_{1j} = 0$ for $j \geq 4$; the number of runs of 1s of length j or more is given by $s_{11} = 3$, $s_{12} = 3$, $s_{13} = 1$ and $s_{1j} = 0$ for $j \geq 4$. The total number of runs is $r = 5$. The k/1-streak *frequency* statistic satisfies $f_{11} = 7$, $f_{12} = 4$, $f_{13} = 1$, and $f_{1j} = 0$ for $j \geq 4$. Finally, the k/1-streak momentum statistic satisfies $p_{11} = 4/6$, $p_{12} = 1/3$, with p_{1j} undefined for $j \geq 3$.

[45] More precisely, it is a subsequence with successive indices $j = i_1 + 1, i_1 + 2, \ldots, i_1 + k$, with $i_1 >= 0$ and $i_1 + k <= n$, in which $x_j = 1$ for all j, and (1) either $i_1 = 0$ or if $i_1 > 0$ then $x_{i_1} = 0$, and (2) either $i_1 + k = n$ or if $i_1 + k < n$ then $i_1 + k + 1 = 0$

B.1 Expected Empirical Probability

In this section we obtain the expected value of the k/1-streak momentum statistic $E[P_{1k}]$ for $k > 1$. As in Section 2, we first compute $E[P_{1k}|N_1 = n_1]$. That $E[P_{1k}|N_1 = n_1]$ was shown to be equal to $(n_1 - 1)/(n - 1)$ for $k = 1$ in Lemma 3 suggests the possibility that a similar formula exists for $k > 1$: $(n_1 - k)/(n - k)$, in the spirit of sampling without replacement. That this formula does not hold in the case of $k > 1$ can easily be confirmed by setting $k = 2$, $n_1 = 4$, and $n = 5$.[46] As in Section 2, it is not possible to determine $\hat{P}_{1k}(\boldsymbol{x})$ directly by computing its value for each sequence, as the number of admissible sequences is typically too large.

We observe that the number of length k/1-streak successors satisfies $|I_{1k}(\boldsymbol{x})| = F_{1k}(x_1, \ldots, x_{n-1})$, i.e. it is equal to the frequency of length k 1-streaks in the sub-sequence that does not include the final term. Further we note that $F_{1k+1}(x_1, \ldots, x_n)$ is the number of length k/1-streak successors that are themselves equal to 1. Therefore the k/1-streak momentum statistic $\hat{P}_{1k}(\boldsymbol{x})$ can be represented as

$$\hat{P}_{1k}(\boldsymbol{x}) = \frac{F_{1k+1}(x_1, \ldots, x_n)}{F_{1k}(x_1, \ldots, x_{n-1})} \quad \text{if } F_{1k}(x_1, \ldots, x_{n-1}) > 0$$

where $\hat{P}_{1k}(\boldsymbol{x})$ is undefined otherwise. Further, because $F_{1k}(x_1, \ldots, x_{n-1}) = F_{1k}(x_1, \ldots, x_n) - \prod_{i=n-k+1}^{n} x_i$, it follows that

$$\hat{P}_{1k}(\boldsymbol{x}) = \frac{F_{1k+1}(\boldsymbol{x})}{F_{1k}(\boldsymbol{x}) - \prod_{i=n-k+1}^{n} x_i} \quad \text{if } F_{1k}(\boldsymbol{x}) > \prod_{i=n-k+1}^{n} x_i$$

A classic reference for non-parametric statistical theory **?** contains a theorem (Theorem 3.3.2, p.87) for the joint distribution $(R_{11}, \ldots, R_{1n_1})$, conditional on N_1 and R_1, from which, in principle, $E[P_{1k}(\boldsymbol{x})|N_1 = n_1]$ can be calculated directly.[47] Unfortunately, the calculation does not appear to be computationally feasible for the sequence lengths of interest here. As a result, we instead follow an approach similar to that in the proof of Lemma 3, making the key observation that for all sequences with $R_{1j}(\boldsymbol{x}) = r_{1j}$ for $j = 1, \ldots, k-1$ and $S_{1k}(\boldsymbol{x}) = s_{1k}$, the k/1-streak momentum

[46] If $k = 2$ then for $n_1 = 4$ and $n = 5$, $E[P_{1k}|N_1 = n_1] = (0/1+1/1+1/2+2/2+2/3)/5 = 19/30 < 2/3 = (n_1-k)/(n-k)$ (see Footnote 15 for intuition).

[47] In fact, the theorem is not quite correct; the distribution presented in the theorem is for $(R_{11}, \ldots, R_{1n_1})$ conditional only on N_1 (unconditional on R_1). For the distribution conditional on R_1 and N_1 it is straightforward to show that

$$\mathbb{P}(R_{11} = r_{11}, \ldots, R_{1n_1} = r_{1n_1}|N_1 = n_1, R_1 = r_1) = \frac{r_1!}{\binom{n_1-1}{r_1-1}\prod_{j=1}^{n_1} r_{1j}!}$$

statistic $\hat{P}_{11}(\boldsymbol{x})$ is (i) constant and equal to $(f_{1k} - s_{1k})/f_{1k}$ for those sequences that have a 0 in one of the final k positions, and (ii) constant and equal to $(f_{1k} - s_{1k})/(f_{1k} - 1)$ for those sequences that have a 1 in each of the final k positions. This is true because $f_{1k+1} = f_{1k} - s_{1k}$, and $f_{1k} = n_1 - \sum_{j=1}^{k-1} jr_{1j} - (k-1)s_{1k}$. Notice that for each case $\hat{P}_{1k}(\boldsymbol{x}) = G(R_{11}(\boldsymbol{x}), \ldots, R_{1k-1}(\boldsymbol{x}), S_{1k}(\boldsymbol{x}))$ for some G, and therefore, by finding the joint distribution of $(R_{11}, \ldots, R_{1k-1}, S_{1k})$, conditional on N_1, it is possible to obtain $E[P_{1k}|N_1 = n_1]$. With $\binom{n}{n_1}$ sequences $\boldsymbol{x} \in \{0,1\}^n$ that satisfy $N_1(\boldsymbol{x}) = n_1$, the joint distribution of $(R_{11}(\boldsymbol{x}), \ldots R_{1k-1}(\boldsymbol{x}), S_{1k}(\boldsymbol{x}))$ is fully characterized by the number of distinguishable sequences $\boldsymbol{x}$ that satisfy $R_{11}(\boldsymbol{x}) = r_{11}, \ldots R_{1k-1}(\boldsymbol{x}) = r_{1k-1}$, and $S_{1k}(\boldsymbol{x}) = s_{1k}$, which we obtain in the following lemma. In the lemma's proof we provide a combinatorial argument that we apply repeatedly in the proof of Theorem 7.

Lemma 6 *The number of distinguishable sequences $\boldsymbol{x} \in \{0,1\}^n$, $n \geq 1$, with $n_1 \leq n$ 1s, $r_{1j} \geq 0$ runs of 1s of exactly length j for $j = 1, \ldots k-1$, and $s_{1k} \geq 0$ runs of 1s of length k or more satisfies:*

$$C_{1k} = \frac{r_1!}{s_{1k}! \prod_{j=1}^{k-1} r_{1j}!} \binom{n_0 + 1}{r_1} \binom{f_{1k} - 1}{s_{1k} - 1}$$

where $r_1 = \sum_{j=1}^{k-1} r_{1j} + s_{1k}$ and $f_{1k} = n_1 - \sum_{j=1}^{k-1} jr_{1j} - (k-1)s_{1k}$. Further, let $\binom{n}{k} = n!/k!(n-k)!$ if $n \geq k \geq 0$ and $\binom{n}{k} = 0$ otherwise, except for the special case $\binom{-1}{-1} = 1$.[48]

Proof:

Any sequence with $r_{11}, \ldots, r_{1k-1}$ runs of 1s of fixed length, and s_{1k} runs of 1s of length k or more can be constructed in three steps by (1) selecting a distinguishable permutation of the $r_1 = \sum_{j=1}^{k-1} r_{1j} + s_{1k}$ cells that correspond to the r_1 runs, which can be done in $r_1!/(s_{1k}! \prod_{j=1}^{k-1} r_{1j})$ unique ways, as for each j, the $r_{1j}!$ permutations of the r_{1j} identical cells across their fixed positions do not generate distinguishable sequences (nor for the s_{1k} identical cells), (2) placing the r_1 1s into the available positions to the left or the right of a 0 among the n_0 0s; with $n_0 + 1$ available positions, there are $\binom{n_0+1}{r_1}$ ways to do this, (3) filling the "empty" run cells, by first filling the r_{1j} run cells of length j with exactly jr_{1j} 1s for $j < k$, and then by filling the s_{1k} indistinguishable (ordered) run cells of length k or more by (a) adding exactly $k-1$ 1s to each cell, (b) with the remaining f_{1k}

[48]Note with this definition of $\binom{n}{k}$, we have $C_{1k} = 0$ if $r_1 > n_0 + 1$, or $\sum_{j=1}^{k-1} jr_{1j} + ks_{1k} > n_1$ (the latter occurs if $s_{1k} > \lfloor \frac{n_1 - \sum_j jr_{1j}}{k} \rfloor$, or $r_{1\ell} > \lfloor \frac{n_1 - \sum_{j \neq \ell} jr_{1j} - ks_{1k}}{\ell} \rfloor$ for some $\ell = 1, \ldots, k-1$, where $\lfloor \cdot \rfloor$ is the floor function). Further, because $r_1 > n_1$ implies that latter condition, it also implies $C_{1k} = 0$.

1s (the number of 1s that succeed some streak of $k-1$ or more 1s), taking an ordered partition of these 1s into a separate set of s_{1k} cells, which can be performed in $\binom{f_{1k}-1}{s_{1k}-1}$ ways by inserting $s_{1k}-1$ dividers into the $f_{1k}-1$ available positions between 1s, and finally (c) adjoining each cell of the separate set of (nonempty and ordered) cells with its corresponding run cell (with exactly $k-1$ 1s), which guarantees that each s_{1k} cell has at least k 1s.

■

Below we state the main theorem, which provides the formula for the expected value of the k/1-streak momentum statistic, conditional on the number of 1s:

Theorem 7 *For n, n_1 and k such that $1 < k \leq n_1 \leq n$*

$$E[P_{1k}|N_1 = n_1] = \frac{1}{\binom{n}{n_1} - U_{1k}} \sum_{\substack{r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j} < n_1 - k \\ s_{1k} \geq 1}} C_{1k} \left[\frac{s_{1k}}{n_0+1} \left(\frac{f_{1k}-s_{1k}}{f_{1k}-1} \right) + \frac{n_0+1-s_{1k}}{n_0+1} \left(\frac{f_{1k}-s_{1k}}{f_{1k}} \right) \right]$$

where f_{1k} and C_{1k} depend on $n_0, n_1, r_{11}, \ldots, r_{1k-1}$, and s_{1k}, and are defined as in Lemma 6.[49] *U_{1k} is defined as the number of sequences in which the k/1-streak momentum statistic is undefined, and satisfies*

$$\begin{aligned} U_{1k} = & \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_0+1}{r_1} \sum_{\ell=0}^{\lfloor \frac{n_1-r_1}{k-1} \rfloor} (-1)^\ell \binom{r_1}{\ell} \binom{n_1-1-\ell(k-1)}{r_1-1} \\ & + \delta_{n_1 k} + \sum_{r_1=2}^{\min\{n_1-k+1,n_0+1\}} \binom{n_0}{r_1-1} \sum_{\ell=0}^{\lfloor \frac{n_1-k-r_1+1}{k-1} \rfloor} (-1)^\ell \binom{r_1-1}{\ell} \binom{n_1-k-1-\ell(k-1)}{r_1-2} \end{aligned}$$

Proof:

For all sequences $\boldsymbol{x} \in \{0,1\}^n$ with n_1 1s, we have three possible cases for how the k/1-streak momentum statistic is determined by r_{1j} $j < k$ and s_{1k}: (1) $\hat{P}_{1k}(\boldsymbol{x})$ is not defined, which arises if (i) $f_{1k} = 0$ or (ii) $f_{1k} = 1$ and $\sum_{i=n-k+1}^{n} x_i = k$, (2) $\hat{P}_{1k}(\boldsymbol{x})$ is equal to $(f_{1k} - s_{1k})/(f_{1k}-1)$, which arises if $f_{1k} \geq 2$ and $\sum_{i=n-k+1}^{n} x_i = k$ or (3) $\hat{P}_{1k}(\boldsymbol{x})$ is equal to $(f_{1k}-s_{1k})/f_{1k}$, which arises if $f_{1k} \geq 1$ and $\sum_{i=n-k+1}^{n} x_i < k$. In case 1i, with $f_{1k} = 0$, the number of terms, which we denote

[49] Note that $\sum_{j=1}^{k-1} jr_{1j} < n_1 - k$ implies $f_{1k} > s_{1k} \geq 1$, which guarantees $f_{1k} \geq 2$.

U^1_{1k}, satisfies:

$$
\begin{aligned}
U^1_{1k} &:= \sum_{\substack{r_{11},\dots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j}=n_1 \\ s_{1k}=0}} C_{1k} \\
&= \sum_{\substack{r_{11},\dots,r_{1k-1} \\ \sum_{j=1}^{k-1} jr_{1j}=n_1}} \frac{r_1!}{\prod_{j=1}^{k-1} r_{1j}!}\binom{n_0+1}{r_1} \\
&= \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_0+1}{r_1} \sum_{\substack{r_{11},\dots,r_{1k-1} \\ \sum_{j=1}^{k-1} jr_{1j}=n_1 \\ \sum_{j=1}^{k-1} r_{1j}=r_1}} \frac{r_1!}{\prod_{j=1}^{k-1} r_{1j}!} \\
&= \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_0+1}{r_1} \sum_{\ell=0}^{\lfloor \frac{n_1-r_1}{k-1} \rfloor} (-1)^\ell \binom{r_1}{\ell}\binom{n_1-1-\ell(k-1)}{r_1-1}
\end{aligned}
$$

where the last line follows by first noting that the inner sum of the third line is the number of compositions (ordered partitions) of n_1-k into r_1-1 parts, which has generating function $(x+x^2+\cdots x^{k-1})^{r_1}$ (Riordan 1958, p. 124). Therefore, the inner sum can be generated as the coefficient on x^{n_1} in the multinomial expansion of $(x+x^2+\cdots x^{k-1})^{r_1}$. The inner sum of binomial coefficients in the fourth line corresponds to the coefficient on x^{n_1} in the binomial expansion of an equivalent representation of the generating function $x^{r_1}(1-x^{k-1})^{r_1}/(1-x)^{r_1} = (x+x^2+\cdots x^{k-1})^{r_1}$. The coefficient in the binomial expansion must agree with the coefficient in the multinomial expansion.[50]

In case 1ii, with $f_{1k}=1$ and $\sum_{i=n-k+1}^{n} x_i = k$, in which case $\hat{P}_{1k}(\boldsymbol{x})$ is also undefined, all sequences that satisfy this criteria can be constructed by first forming a distinguishable permutation of the r_1-1 runs of 1s not including the final run of k 1s, which can be done in $r_1!/(\prod_{j=1}^{k-1} r_{1j}!)$ ways, and second placing these runs to the left or the right of the available n_0 0s, not including the right end point, which can be done in $\binom{n_0}{r_1-1}$ ways with the n_0 positions. Summing over all possible

[50] The binomial expansion is given by:

$$
x^{r_1}\frac{(1-x^{k-1})^{r_1}}{(1-x)^{r_1}} = x^{r_1}\left[\sum_{t_1=0}^{r_1}\binom{r_1}{t_1}(-1)^{t_1}x^{t_1(k-1)}\right]\cdot\left[\sum_{t_2=0}^{+\infty}\binom{r_1-1+t_2}{r_1-1}x^{t_2}\right]
$$

therefore the coefficient on x^{n_1} is $\sum(-1)^{t_1}\binom{r_1}{t_1}\binom{r_1-1+t_2}{r_1-1}$ where the sum is taken over all t_1, t_2 such that $r_1+t_1(k-1)+t_2=n_1$.

runs, the number of terms U_{1k}^2 satisfies:

$$
\begin{aligned}
U_{1k}^2 &:= \sum_{\substack{r_{11},\dots,r_{1k-1},s_{1k}\\ \sum_{j=1}^{k-1} jr_{1j}=n_1-k\\ s_{1k}=1}} \frac{s_{1k}}{n_0+1} C_{1k} \\
&= \sum_{\substack{r_{11},\dots,r_{1k-1},s_{1k}\\ \sum_{j=1}^{k-1} jr_{1j}=n_1-k\\ s_{1k}=1}} \frac{(r_1-1)!}{\prod_{j=1}^{k-1} r_{1j}!} \binom{n_0}{r_1-1} \\
&= \delta_{n_1 k} + \sum_{r_1=2}^{\min\{n_1-k+1,n_0+1\}} \binom{n_0}{r_1-1} \sum_{\substack{r_{11},\dots,r_{1k-1}\\ \sum_{j=1}^{k-1} jr_{1j}=n_1-k\\ \sum_{j=1}^{k-1} r_{1j}=r_1-1}} \frac{(r_1-1)!}{\prod_{j=1}^{k-1} r_{1j}!} \\
&= \delta_{n_1 k} + \sum_{r_1=2}^{\min\{n_1-k+1,n_0+1\}} \binom{n_0}{r_1-1} \sum_{\ell=0}^{\lfloor \frac{n_1-k-r_1+1}{k-1} \rfloor} (-1)^\ell \binom{r_1-1}{\ell} \binom{n_1-k-1-\ell(k-1)}{r_1-2}
\end{aligned}
$$

and we assume that $\sum_{j=m}^{n} a_j = 0$ if $m > n$. The Kronecker delta in the third line appears because when $s_{1k} = 1$ and $\sum_{j=1}^{k-1} jr_{1j} = n_1 - k$, there is only one sequence for which the k/1-streak momentum statistic is undefined. The last line follows because the inner sum of the third line can be generated as the coefficient on x^{n_1-k} in the multinomial expansion of $(x + x^2 + \cdots x^{k-1})^{r_1-1}$, which, as in determining U_{1k}^1, corresponds to the coefficient on the binomial expansion. Taking case 1i and 2ii together, the total number of sequences in which $\hat{P}_{1k}(\boldsymbol{x})$ is undefined is equal to $U_{1k} = U_{1k}^1 + U_{1k}^2$

In case 2, in which $\hat{P}_{1k}(\boldsymbol{x})$ is defined with $\sum_{i=n-k+1}^{n} x_i = k$ and $f_{1k} \geq 2$, it must be the case that $\sum_{j=1}^{k-1} jr_{1j} < n_1 - k$, and all sequences that satisfy this criteria can be constructed in three steps analogous to those used in Lemma 6 by (1) selecting a distinguishable permutation of the $r_1 - 1$ remaining runs, (2) placing the $r_1 - 1$ 1s into the n_0 available positions to the left or the right of a 0, (3) filling the “empty” run cells. For a given $(r_{11}, \dots, r_{1k-1}, s_{1k})$ the total number of sequences satisfying this criteria is:

$$
\frac{(r_1-1)!}{(s_{1k}-1)! \prod_{j=1}^{k-1} r_{1j}!} \binom{n_0}{r_1-1} \binom{f_{1k}-1}{s_{1k}-1} = \frac{s_{1k}}{n_0+1} C_{1k}
$$

In case 3, in which $\hat{P}_{1k}(\boldsymbol{x})$ is defined with $\sum_{i=n-k+1}^{n} x_i < k$ and $f_{1k} \geq 1$, it must be the case that $\sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k$, as before all sequences that satisfy this criteria can be constructed in three steps as before, and we consider two subcases, sequences that terminate in a 1 (i.e. a run of 1s of length less than k) and sequences that terminate in a 0 (i.e. a run of 0s). For those sequence that terminate in a 1, for a given $(r_{11}, \ldots, r_{1k-1}, s_{1k})$ the total number of sequences satisfying this criteria is:

$$\left(\frac{r_1!}{s_{1k}! \prod_{j=1}^{k-1} r_{1j}!} - \frac{(r_1-1)!}{(s_{1k}-1)! \prod_{j=1}^{k-1} r_{1j}!}\right) \binom{n_0}{r_1 - 1} \binom{f_{1k}-1}{s_{1k}-1} = \frac{r_1 - s_{1k}}{n_0+1} C_{1k}$$

with $(r_1-1)!/((s_{1k}-1)! \prod_{j=1}^{k-1} r_{1j}!)$ being the number of sequences that terminate in a run of 1s of length k or more. For those sequences that terminate in a 0, for a given $(r_{11}, \ldots, r_{1k-1}, s_{1k})$ the total number of sequences satisfying this criteria is:

$$\frac{r_1!}{s_{1k}! \prod_{j=1}^{k-1} r_{1j}!} \binom{n_0}{r_1} \binom{f_{1k}-1}{s_{1k}-1} = \frac{n_0 + 1 - r_1}{n_0+1} C_{1k}$$

therefore, the sum of the k/1-streak momentum statistic across all sequences for which it is defined satisfies:

$$\begin{aligned} E[P_{1k}|N_1 = n_1] \left[\binom{n}{n_1} - U_{1k}\right] = & \sum_{\substack{r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j} < n_1 - k \\ s_{1k} \geq 1}} C_{1k} \frac{s_{1k}}{n_0+1} \frac{f_{1k} - s_{1k}}{f_{1k} - 1} \\ & + \sum_{\substack{r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k \\ s_{1k} \geq 1}} C_{1k} \frac{r_1 - s_{1k}}{n_0+1} \frac{f_{1k} - s_{1k}}{f_{1k}} \\ & + \sum_{\substack{r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k \\ s_{1k} \geq 1}} C_{1k} \frac{n_0 + 1 - r_1}{n_0+1} \frac{f_{1k} - s_{1k}}{f_{1k}} \end{aligned}$$

and this reduces to the formula in the theorem because the final two terms can be combined, and then can be summed over only runs that satisfy $\sum_{j=1}^{k-1} jr_{1j} < n_1 - k$, and finally combined with the first term (because $f_{1k} - s_{1k} = 0$ if $\sum_{j=1}^{k-1} jr_{1j} = n_1 - k$).[51]

[51] While the first term has the closed form representation $\sum C_{1k} \frac{s_{1k}}{n_0+1} \frac{f_{1k}-s_{1k}}{f_{1k}-1} = \binom{n_1-1}{k} / \binom{n-1}{k}$, this does not appear to be so for the other terms. $E[P_{1k}|N_1 = n_1]$ cannot have a closed form representation in any case, as the term U_{1k} does not allow one.

■

B.2 Expected Difference in Proportions

The exact formula for the expected difference between the relative frequency of 1 for k/1-streak successor trials and the relative frequency of 1 for k/0-streak successor trials can be obtained with an approach similar to that of the previous section. The difference satisfies $D_k := P_{1k} - (1 - P_{0k})$, and there are three categories of sequences for which D_k is defined: (1) a sequence that ends in a run of 0s of length k or more, with $f_{0k} \geq 2$ and $f_{1k} \geq 1$, and the difference equal to $D_k^1 = (f_{1k} - s_{1k})/f_{1k} - (s_{0k} - 1)/(f_{0k} - 1)$, (2) a sequence that ends in a run of 1s of length k or more, with $f_{0k} \geq 1$ and $f_{1k} \geq 2$, and the difference equal to $D_k^2 := (f_{1k} - s_{1k})/(f_{1k} - 1) - s_{0k}/f_{0k}$, (3) a sequence that ends in a run of 0s of length $k-1$, or less, or a run of 1s of length $k-1$, or less, with $f_{0k} \geq 1$ and $f_{1k} \geq 1$, and the difference equal to $D_k^3 := (f_{1k} - s_{1k})/f_{1k} - s_{0k}/f_{0k}$. For all other sequences the difference is undefined.

Theorem 8 *For n, n_1, n_0 and k such that $n_0 + n_1 = n$, and $1 < k \leq n_0, n_1 \leq n$, the expected difference in the relative frequency of 1 for k/1-streak successors and the relative frequency of 1 for* *k/0-streak successors,* $D_k := P_{1k} - (1 - P_{0k})$*, satisfies*

$$E[D_k \mid N_1 = n_1] = \frac{1}{\binom{n}{n_1} - U_k} \left[\sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} < n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k, s_{1k} \geq 1 \\ r_0 \geq r_1}} C_k \left[\frac{s_{0k}}{r_0} D_k^1 + \frac{r_0 - s_{0k}}{r_0} D_k^3 \right] \right.$$
$$\left. + \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} \leq n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} < n_1 - k, s_{1k} \geq 1 \\ r_1 \geq r_0}} C_k \left[\frac{s_{1k}}{r_1} D_k^2 + \frac{r_1 - s_{1k}}{r_1} D_k^3 \right] \right]$$

where $D_k^1 = (f_{1k} - s_{1k})/f_{1k} - (s_{0k} - 1)/(f_{0k} - 1), D_k^2 := (f_{1k} - s_{1k})/(f_{1k} - 1) - s_{0k}/f_{0k}, D_k^3 := (f_{1k} - s_{1k})/f_{1k} - s_{0k}/f_{0k}$*, and*

$$C_k := \frac{r_0!}{s_{0k}! \prod_{i=1}^{k-1} r_{0i}!} \frac{r_1!}{s_{1k}! \prod_{i=1}^{k-1} r_{1i}!} \binom{f_{0k} - 1}{s_{0k} - 1} \binom{f_{1k} - 1}{s_{1k} - 1}$$

and U_k *(see expression* $*$ *on page 47) is the number of sequences in which there are either no k/1-streak successors, or no k/0-streak successors.*

Proof:

Note that for the case in which $|r_1 - r_0| = 1$, C_k is the number of sequences with $N_1 = n_1$ in which the number of runs of 0s, and runs of 1s satisfy run profile $(r_{01}, \ldots, r_{0k-1}, s_{0k}; r_{11}, \ldots, r_{1k-1}, s_{1k})$; for the cases in which $r_1 = r_0$, C_k is equal to half the number of these sequences (because each sequence can end with a run of 1s, or a run of 0s). The combinatorial proof of this formula, which we omit, is similar to the one used in the proof of Lemma 6.

The sum total of the differences, across all sequences for which the difference is defined and $N_1 = n_1$ is

$$E[D_k \mid N_1 = n_1] \cdot \left[\binom{n}{n_1} - U_k\right]$$
$$= \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} < n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k, s_{1k} \geq 1 \\ r_0 \geq r_1}} \frac{s_{0k}}{r_0} C_k D_k^1 \quad + \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} \leq n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} < n_1 - k, s_{1k} \geq 1 \\ r_1 \geq r_0}} \frac{s_{1k}}{r_1} C_k D_k^2$$
$$+ \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} \leq n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k, s_{1k} \geq 1 \\ r_0 \geq r_1}} \frac{r_0 - s_{0k}}{r_0} C_k D_k^3 \quad + \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} \leq n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k, s_{1k} \geq 1 \\ r_1 \geq r_0}} \frac{r_1 - s_{1k}}{r_1} C_k D_k^3$$

where the first sum relates to those sequences that end in a run of 0s of length k or more (whence $r_0 \geq r_1$, the multiplier s_{0k}/r_0 and $\sum_{j=1}^{k-1} jr_{0j} < n_0 - k$);[52] the second sum relates to those sequences that end in a run of 1s of length k or more (whence $r_1 \geq r_0$, the multiplier s_{1k}/r_1 and $\sum_{j=1}^{k-1} jr_{1j} < n_1 - k$); the third sum relates to those sequences that end on a run of 0s of length k-1 or less (whence $r_0 \geq r_1$, the multiplier $(r_0 - s_{0k})/r_0$) and $\sum_{j=1}^{k-1} jr_{0j} < n_0 - k$);[53] and the fourth sum relates to those sequences that end on a run of 1s of length $k-1$ or less (whence $r_1 \geq r_0$, the multiplier $(r_1 - s_{1k})/r_1$) and $\sum_{j=1}^{k-1} jr_{1j} < n_1 - k$). These four terms can be combined into the following two

[52] Note $\sum_{j=1}^{k-1} jr_{0j} < n_0 - k \iff f_{0k} \geq 2$.

[53] The multiplier $(r_0 - s_{0k})/r_0$ arises because the number of distinguishable permutations of the 0 runs that end with a run of length $k-1$ or less is equal to the total number of distinguishable permutations of the 0 runs minus the number of distinguishable permutations of the 0 runs that end in a run of length k or more, i.e.

$$\frac{r_0!}{s_{0k}! \prod_{i=1}^{k-1} r_{0i}!} - \frac{(r_0 - 1)!}{(s_{0k} - 1)! \prod_{i=1}^{k-1} r_{0i}!} = \frac{r_0 - s_{0k}}{r_0} \frac{r_0!}{s_{0k}! \prod_{i=1}^{k-1} r_{0i}!}$$

terms:

$$E[D_k \mid N_1 = n_1] \cdot \left[\binom{n}{n_1} - U_k \right] = \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} j r_{0j} < n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} j r_{1j} \leq n_1 - k, s_{1k} \geq 1 \\ r_0 \geq r_1}} C_k \left[\frac{s_{0k}}{r_0} D_k^1 + \frac{r_0 - s_{0k}}{r_0} D_k^3 \right] + \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} j r_{0j} \leq n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} j r_{1j} < n_1 - k, s_{1k} \geq 1 \\ r_1 \geq r_0}} C_k \left[\frac{s_{1k}}{r_1} D_k^2 + \frac{r_1 - s_{1k}}{r_1} D_k^3 \right]$$

which can readily be implemented numerically for the finite sequences considered here.[54] The total number of sequences for which the difference is undefined, U_k, can be counted in a way that is analogous to what was done in the proof of Theorem 7, by using an application of the inclusion-exclusion principle:

$$\begin{aligned} U_k := &\sum_{\substack{r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} j r_{1j} = n_1 \\ s_{1k} = 0}} C_{1k} + \sum_{\substack{r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} j r_{1j} = n_1 - k \\ s_{1k} = 1}} \frac{s_{1k}}{n_0 + 1} C_{1k} + \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ \sum_{j=1}^{k-1} j r_{0j} = n_1 \\ s_{0k} = 0}} C_{0k} + \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ \sum_{j=1}^{k-1} j r_{0j} = n_1 - k \\ s_{0k} = 1}} \frac{s_{0k}}{n_1 + 1} C_{0k} \\ &- \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} j r_{0j} = n_0, s_{0k} = 0 \\ \sum_{j=1}^{k-1} j r_{1j} = n_1, s_{1k} = 0 \\ |r_0 - r_1| \leq 1}} \left(2 \cdot \mathbb{1}_{\{r_1 = r_0\}} + \mathbb{1}_{\{|r_1 - r_0| = 1\}} \right) C_k \\ &- \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} j r_{0j} = n_0 - k, s_{0k} = 1 \\ \sum_{j=1}^{k-1} j r_{1j} = n_1, s_{1k} = 0 \\ r_0 \geq r_1}} \frac{s_{0k}}{r_0} C_k - \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} j r_{0j} = n_0, s_{0k} = 0 \\ \sum_{j=1}^{k-1} j r_{1j} = n_1 - k, s_{1k} = 1 \\ r_1 \geq r_0}} \frac{s_{1k}}{r_1} C_k \end{aligned}$$

[54] In the numerical implementation one can consider three sums $r_0 = r_1 + 1$, $r_1 = r_0 + 1$, and for the case of $r_1 = r_0$ the sums can be combined.

where C_{0k} is a function of $r_{01}, \ldots, r_{0k-1}, s_{0k}; n_0, n_1$ and defined analogously to C_{1k}. We can simplify the above expression by first noting that the third term, which corresponds to those sequences in which there are no k/1-streak successors and no k/0-streak successors, can be reduced to a sum of binomial coefficients:

$$\begin{aligned}
&\sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k}\\ r_{11},\ldots,r_{1k-1},s_{1k}\\ \sum_{j=1}^{k-1} jr_{0j}=n_0,s_{0k}=0\\ \sum_{j=1}^{k-1} jr_{1j}=n_1,s_{1k}=0\\ |r_0-r_1|\leq 1}} \left(2\cdot \mathbb{1}_{\{r_1=r_0\}} + \mathbb{1}_{\{|r_1-r_0|=1\}}\right) C_k \\
&= \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k}\\ r_{11},\ldots,r_{1k-1},s_{1k}\\ \sum_{j=1}^{k-1} jr_{0j}=n_0,s_{0k}=0\\ \sum_{j=1}^{k-1} jr_{1j}=n_1,s_{1k}=0\\ |r_0-r_1|\leq 1}} \left(2\cdot \mathbb{1}_{\{r_1=r_0\}} + \mathbb{1}_{\{|r_1-r_0|=1\}}\right) \frac{r_0!}{s_{0k}!\prod_{i=1}^{k-1} r_{0i}!}\frac{r_1!}{s_{1k}!\prod_{i=1}^{k-1} r_{1i}!} \\
&= \sum_{\substack{r_{01},\ldots,r_{0k-1}\\ r_{11},\ldots,r_{1k-1}\\ \sum_{j=1}^{k-1} jr_{0j}=n_0\\ \sum_{j=1}^{k-1} jr_{1j}=n_1\\ |r_0-r_1|\leq 1}} \left(2\cdot \mathbb{1}_{\{r_1=r_0\}} + \mathbb{1}_{\{|r_1-r_0|=1\}}\right) \frac{r_0!}{\prod_{i=1}^{k-1} r_{0i}!}\frac{r_1!}{\prod_{i=1}^{k-1} r_{1i}!} \\
&= \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \sum_{r_0=\max\{r_1-1,1\}}^{\min\{r_1+1,n_0\}} \left(2\cdot \mathbb{1}_{\{r_1=r_0\}} + \mathbb{1}_{\{|r_1-r_0|=1\}}\right) \sum_{\substack{r_{01},\ldots,r_{0k-1}\\ r_{11},\ldots,r_{1k-1}\\ \sum_{j=1}^{k-1} jr_{0j}=n_0\\ \sum_{j=1}^{k-1} jr_{1j}=n_1\\ \sum_{j=1}^{k-1} r_{0j}=r_0\\ \sum_{j=1}^{k-1} r_{1j}=r_1}} \frac{r_0!}{\prod_{i=1}^{k-1} r_{0i}!}\frac{r_1!}{\prod_{i=1}^{k-1} r_{1i}!} \\
&= \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \sum_{r_0=\max\{r_1-1,1\}}^{\min\{r_1+1,n_0\}} \left(2\cdot \mathbb{1}_{\{r_1=r_0\}} + \mathbb{1}_{\{|r_1-r_0|=1\}}\right) \sum_{\substack{r_{01},\ldots,r_{0k-1}\\ \sum_{j=1}^{k-1} jr_{0j}=n_0\\ \sum_{j=1}^{k-1} r_{0j}=r_0}} \frac{r_0!}{\prod_{i=1}^{k-1} r_{0i}!} \sum_{\substack{r_{11},\ldots,r_{1k-1}\\ \sum_{j=1}^{k-1} jr_{1j}=n_1\\ \sum_{j=1}^{k-1} r_{1j}=r_1}} \frac{r_1!}{\prod_{i=1}^{k-1} r_{1i}!} \\
&= \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \sum_{r_0=\max\{r_1-1,1\}}^{\min\{r_1+1,n_0\}} \left(2\cdot \mathbb{1}_{\{r_1=r_0\}} + \mathbb{1}_{\{|r_1-r_0|=1\}}\right) \sum_{\ell_0=0}^{\lfloor \frac{n_0-r_0}{k-1} \rfloor} (-1)^{\ell_0}\binom{r_0}{\ell_0}\binom{n_0-1-\ell_0(k-1)}{r_0-1} \\
&\qquad \times \sum_{\ell_1=0}^{\lfloor \frac{n_1-r_1}{k-1} \rfloor} (-1)^{\ell_1}\binom{r_1}{\ell_1}\binom{n_1-1-\ell_1(k-1)}{r_1-1}
\end{aligned}$$

For the final two negative terms in the formula for U_k, we can apply a similar argument in order to represent them as a sum of binomial coefficients. For the first four positive terms we can use the argument provided in Theorem 7 to represent them as sums of binomial coefficients, and therefore U_k reduces to a sum of binomial coefficients:

$$
\begin{aligned}
U_k = & \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_0+1}{r_1} \sum_{\ell=0}^{\lfloor \frac{n_1-r_1}{k-1} \rfloor} (-1)^\ell \binom{r_1}{\ell} \binom{n_1-1-\ell(k-1)}{r_1-1} \qquad (*) \\
& + \delta_{n_1 k} + \sum_{r_1=2}^{\min\{n_1-k+1,n_0+1\}} \binom{n_0}{r_1-1} \sum_{\ell=0}^{\lfloor \frac{n_1-k-r_1+1}{k-1} \rfloor} (-1)^\ell \binom{r_1-1}{\ell} \binom{n_1-k-1-\ell(k-1)}{r_1-2} \\
& + \sum_{r_0=1}^{\min\{n_0,n_1+1\}} \binom{n_1+1}{r_0} \sum_{\ell=0}^{\lfloor \frac{n_0-r_0}{k-1} \rfloor} (-1)^\ell \binom{r_0}{\ell} \binom{n_0-1-\ell(k-1)}{r_0-1} \\
& + \delta_{n_0 k} + \sum_{r_0=2}^{\min\{n_0-k+1,n_1+1\}} \binom{n_1}{r_0-1} \sum_{\ell=0}^{\lfloor \frac{n_0-k-r_0+1}{k-1} \rfloor} (-1)^\ell \binom{r_0-1}{\ell} \binom{n_0-k-1-\ell(k-1)}{r_0-2} \\
& - \Bigg[\sum_{r_1=1}^{\min\{n_1,n_0+1\}} \sum_{r_0=\max\{r_1-1,1\}}^{\min\{r_1+1,n_0\}} \left(2 \cdot \mathbb{1}_{\{r_1=r_0\}} + \mathbb{1}_{\{|r_1-r_0|=1\}}\right) \times \\
& \qquad \times \sum_{\ell_0=0}^{\lfloor \frac{n_0-r_0}{k-1} \rfloor} (-1)^{\ell_0} \binom{r_0}{\ell_0} \binom{n_0-1-\ell_0(k-1)}{r_0-1} \sum_{\ell_1=0}^{\lfloor \frac{n_1-r_1}{k-1} \rfloor} (-1)^{\ell_1} \binom{r_1}{\ell_1} \binom{n_1-1-\ell_1(k-1)}{r_1-1} \Bigg] \\
& - \Bigg[\delta_{n_0 k} + \sum_{r_0=2}^{\min\{n_0-k+1,n_1+1\}} \sum_{r_1=\max\{r_0-1,1\}}^{\min\{r_0,n_1\}} \sum_{\ell_0=0}^{\lfloor \frac{n_0-k-r_0+1}{k-1} \rfloor} (-1)^{\ell_0} \binom{r_0-1}{\ell_0} \binom{n_0-k-1-\ell_0(k-1)}{r_0-2} \times \\
& \qquad \times \sum_{\ell_1=0}^{\lfloor \frac{n_1-r_1}{k-1} \rfloor} (-1)^{\ell_1} \binom{r_1}{\ell_1} \binom{n_1-1-\ell_1(k-1)}{r_1-1} \Bigg] \\
& - \Bigg[\delta_{n_1 k} + \sum_{r_1=2}^{\min\{n_1-k+1,n_0+1\}} \sum_{r_0=\max\{r_1-1,1\}}^{\min\{r_1,n_0\}} \sum_{\ell_1=0}^{\lfloor \frac{n_1-k-r_1+1}{k-1} \rfloor} (-1)^{\ell_1} \binom{r_1-1}{\ell_1} \binom{n_1-k-1-\ell_1(k-1)}{r_1-2} \times \\
& \qquad \times \sum_{\ell_0=0}^{\lfloor \frac{n_0-r_0}{k-1} \rfloor} (-1)^{\ell_0} \binom{r_0}{\ell_0} \binom{n_0-1-\ell_0(k-1)}{r_0-1} \Bigg]
\end{aligned}
$$

■

C Appendix: Why it often is the case that $E[P_{1k}|N_1 = n_1] < \frac{n_1-k}{n-k}$.

For an intuition why the sampling without replacement reasoning that worked when $k = 1$ does not extend to $k > 1$, we first note that $E[P_{1k}|N_1 = n_1]$ is given by:

$$E[\hat{P}_{1k}(\boldsymbol{x})|N_1(\boldsymbol{x}) = n_1] = E\left[\left.\frac{\sum_{j=k}^{n_1}(j-k)R_{1j}(\boldsymbol{x})}{\sum_{j=k}^{n_1}(j-k+1)R_{1j}(\boldsymbol{x}) - \prod_{i=n-k+1}^{n} x_i}\right| N_1(\boldsymbol{x}) = n_1\right] \qquad (\dagger)$$

(as in Appendix B), where, for simplicity, we assume that $n_1 > (k-1)n_0 + k$ (to avoid the possibility of missing observations). This formula can be operationalized as the probability of observing a 1 in the following two-step procedure, which we call "Procedure A": (1) selecting, at random, one of the equiprobable $\binom{n}{n_1}$ sequences, and (2) from that sequence, selecting, at random, one of the trials that immediately follows a streak of k or more 1s (this procedure is analogous to that discussed in the alternate proof of Lemma 3). We now explain why the probability generated by this procedure is typically strictly less than $\frac{n_1-k}{n-k}$. First, imagine (counterfactually) that for each sequence $\boldsymbol{x} = (x_1, \ldots, x_n)$ that satisfies $N_1(\boldsymbol{x}) = n_1$, we have $R_{1i}(\boldsymbol{x}) = E[R_{1i}(\boldsymbol{x})|N_1(\boldsymbol{x}) = n_1]$ for $i = 1, \ldots n_1$, i.e. the joint profile of run lengths is constant across sequences, and the number of runs of each length is equal to its expected value, conditional on n_1 successes (assuming divisibility). Further, assume that $\prod_{i=n-k+1}^{n} x_i = E[\prod_{i=n-k+1}^{n} x_i|N_1(\boldsymbol{x}) = n_1]$, i.e. replace, with its expected value, the variable that equals one when a single trial must be removed from the denominator (because the sequence ends in a run of 1s of length $j \geq k$, leaving the final trial in the run without a successor). In this counterfactual case, we will see below that † can now be operationalized as the probability of observing a 1 in the simpler "Procedure B": selecting a trial at random (uniformly) from the pooled set of trials that (a) appear in one of the $\binom{n}{n_1}$ sequences with n_1 successes and (b) immediately follow k or more 1s. To see this, first note that because each sequence has the same joint profile of run lengths, the right hand side in the equation † becomes the left hand side in the identity ‡ below (i.e. $\hat{P}_{1k}(\boldsymbol{x})$ is constant across sequences).

$$\frac{\sum_{j=k}^{n_1}(j-k)E[R_{1j}(\boldsymbol{x})|N_1(\boldsymbol{x})=n_1]}{\sum_{j=k}^{n_1}(j-k+1)E[R_{1j}(\boldsymbol{x})|N_1(\boldsymbol{x})=n_1]-E[\prod_{i=n-k+1}^{n}x_i|N_1(\boldsymbol{x})=n_1]}=\frac{n_1-k}{n-k} \tag{‡}$$

That ‡ holds can be confirmed by first noting that, on the left hand side, $\binom{n}{n_1}$ is a divisor within each conditional expectation, and therefore can be canceled, which yields a ratio in which the corresponding denominator of the left hand side consists of the number of trials that (a) appear in one of the $\binom{n}{n_1}$ sequences with n_1 successes, and (b) immediately follow k or more 1s, while the corresponding numerator consists of the number of these trials in which a 1 is observed. Therefore, Procedure B is equivalent to selecting a trial at random from one of the trials in the denominator of this ratio, and the probability of observing a 1 is equal to the left hand side of ‡. Now, in Procedure B, the selected trial appears in some sequence $\boldsymbol{x}$, and is equally likely to be in any position $i > k$. It then follows that $\mathbb{P}(x_i = 1) = \frac{n_1-k}{n-k}$, because $x_j = 1$ for the preceding k positions, and the outcomes in the remaining $n-k$ positions are exchangeable. Therefore, if all sequences were to have the same joint profile of run lengths we would expect a bias of $\frac{n_1-k}{n-k} - \frac{n_1}{n} < 0$. In reality, however, the joint profile of run lengths varies across sequences, and, instead, it is Procedure A that operationalizes † as a probability. For an intuition why Procedure A has a downward bias that is (generally) of greater magnitude than that of Procedure B, we first observe that, relative to the more common sequence $\boldsymbol{x}$ with a low empirical probability (e.g. $\hat{P}_{1k}(\boldsymbol{x}) < \frac{n_1}{n}$), the less common sequence $\boldsymbol{x'}$ with a high empirical probability (e.g. $\hat{P}_{1k}(\boldsymbol{x'}) > \frac{n_1}{n}$) has, on average, fewer, and longer runs, which yields it more trials that immediately follow k or more 1s, and more trials in which a 1 is observed. Because Procedure A involves selecting a *sequence* at random, this means that for the trials in which a 1 is observed, it must be less probable for them to be selected (ex-ante) by Procedure A relative to Procedure B because these trials tend to be concentrated in relatively less common sequences, whereas in Procedure B, which sequences these trials appear in is immaterial, as all trials are pooled together across sequences. As a result, each of these trials has a proportional $\frac{n_1-k}{n-k}$ chance of being selected in Procedure B. This implies that the probability of observing a 1 in Procedure B overestimates the probability of observing a 1 in Procedure A.

D The relationship to finite sample bias when estimating autocorrelation in time series data

It is well known that standard estimates of autocorrelation are biased in finite samples (Yule 1926), as well as for regression coefficients in time series data (Stambaugh 1986, 1999). Below we show that the bias in the estimate of the probability of a 1, conditional on a k/1-streak, has a direct relationship with the bias in the least squares estimator for a corresponding linear probability model.

Theorem 9 *Let* $\boldsymbol{x} \in \{0,1\}^n$ *with* $I_{1k}(\boldsymbol{x}) \neq \emptyset$*. If* $\boldsymbol{\beta}_k(\boldsymbol{x}) = (\beta_{0k}(\boldsymbol{x}), \beta_{1k}(\boldsymbol{x}))$ *is defined to be the solution to the least squares problem,* $\boldsymbol{\beta}_k(\boldsymbol{x}) \in \operatorname{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^2} \|\boldsymbol{x} - [\mathbf{1}\ \boldsymbol{d}]^\intercal \boldsymbol{\beta}\|^2$ *where* $\boldsymbol{d} \in \{0,1\}^n$ *is defined so that* $d_i := \mathbb{1}_{I_{1k}(\boldsymbol{x})}(i)$ *for* $i = 1, \ldots, n$*, then*[55]

$$\beta_{0k}(\boldsymbol{x}) + \beta_{1k}(\boldsymbol{x}) = \hat{P}_{1k}(\boldsymbol{x})$$

Proof:

If $\boldsymbol{\beta}_k(\boldsymbol{x})$ minimizes that sum of squares then $\beta_{1k}(\boldsymbol{x}) = \sum_{i=1}^n (x_i - \bar{x})(d_i - \bar{d}) / \sum_{i=1}^n (d_i - \bar{d})^2$. First, working with the numerator, letting $I_{1k} \equiv I_{1k}(\boldsymbol{x})$ we have

$$\begin{aligned}
\sum_{i=1}^n (x_i - \bar{x})(d_i - \bar{d}) &= \sum_{i\in I_{1k}} (x_i - \bar{x})(d_i - \bar{d}) + \sum_{i\in I_{1k}^C} (x_i - \bar{x})(d_i - \bar{d}) \\
&= \left(1 - \frac{|I_{1k}|}{n}\right) \sum_{i\in I_{1k}} (x_i - \bar{x}) - \frac{|I_{1k}|}{n} \sum_{i\in I_{1k}^C} (x_i - \bar{x}) \\
&= \left(1 - \frac{|I_{1k}|}{n}\right) \sum_{i\in I_{1k}} x_i - \frac{|I_{1k}|}{n} \sum_{i\in I_{1k}^C} x_i - \left(1 - \frac{|I_{1k}|}{n}\right) |I_{1k}|\bar{x} + \frac{|I_{1k}|}{n}(n - |I_{1k}|)\bar{x} \\
&= |I_{1k}| \left(1 - \frac{|I_{1k}|}{n}\right) \left(\frac{\sum_{i\in I_{1k}} x_i}{|I_{1k}|} - \frac{\sum_{i\in I_{1k}^C} x_i}{n - |I_{1k}|} \right)
\end{aligned}$$

[55] When $I_{1k}(\boldsymbol{x}) = \emptyset$ the solution set of the least squares problem is infinite, i.e. $\operatorname{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^2} \|\boldsymbol{x} - \beta_0\|^2 = \{(\beta_0, \beta_1) \in \mathbb{R}^2 : \beta_0 = (1/n)\sum_{i=1}^n x_i\}$. If we treat $\boldsymbol{\beta}_k(\boldsymbol{x})$ as undefined in this case, then the bias from using the conditional relative frequency is equal to the finite sample bias in the coefficients of the associated linear probability model. If instead we define $\beta_{1k}(\boldsymbol{x}) = 0$, then the bias in the coefficients of the associated linear probability model will be less than the bias in the conditional relative frequency.

second, with the denominator of $\beta_{1k}(\boldsymbol{x})$ we have

$$\begin{aligned}
\sum_{i=1}^{n}(d_i - \bar{d})^2 &= \sum_{i \in I_{1k}} \left(1 - \frac{|I_{1k}|}{n}\right)^2 + \sum_{i \in I_{1k}^C} \left(\frac{|I_{1k}|}{n}\right)^2 \\
&= |I_{1k}| \left(1 - \frac{|I_{1k}|}{n}\right)^2 + (n - |I_{1k}|) \left(\frac{|I_{1k}|}{n}\right)^2 \\
&= |I_{1k}| \left(1 - \frac{|I_{1k}|}{n}\right)
\end{aligned}$$

therefore we have

$$\beta_{1k}(\boldsymbol{x}) = \frac{\sum_{i \in I_{1k}} x_i}{|I_{1k}|} - \frac{\sum_{i \in I_{1k}^C} x_i}{n - |I_{1k}|}$$

now

$$\begin{aligned}
n\beta_{0k}(\boldsymbol{x}) &= n(\bar{x} - \beta_1(\boldsymbol{x})\bar{d}) \\
&= \sum_{i=1}^{n} x_i - \left(\frac{\sum_{i \in I_{1k}} x_i}{|I_{1k}|} - \frac{\sum_{i \in I_{1k}^C} x_i}{n - |I_{1k}|}\right) |I_{1k}| \\
&= \sum_{i \in I_{1k}^C} x_i + \frac{|I_{1k}| \sum_{i \in I_{1k}^C} x_i}{n - |I_{1k}|} \\
&= \frac{n \sum_{i \in I_{1k}^C} x_i}{n - |I_{1k}|}
\end{aligned}$$

and summing both coefficients, the result follows.

■

Note that the bias in the coefficients follows from the bias in the estimate of the conditional probability, i.e. $E[\hat{P}_{1k}(\boldsymbol{x})] < p$ implies $E[\beta_{1k}(\boldsymbol{x})] < 0$ and $E[\beta_{0k}(\boldsymbol{x})] > p$.[56]

[56]Note that $p = E\left[(1/n)\sum_{i=1}^{n} x_i\right]$, and the sum can be broken up and re-arranged as in the theorem.