Belief in Mean Reversion and the Law of Small Numbers*

Jawwad Noor and Fernando Payró

April 17, 2026

Abstract

Studies find that people systematically underestimate the likelihood of streaks in a random sequence. In a canonical coin-tossing environment, this paper shows that the evidence can be explained by a belief in mean reversion. Such beliefs are represented as if the bias of the coin is history-dependent and "self-correcting". In a Bayesian inference setting, a belief in mean reversion ensures that the agent never rules out the true parameter.

JEL Classification: D01, D9.

Keywords: Law of small numbers, belief biases, gambler's fallacy, learning, misspecified beliefs.

*An earlier version of this paper was circulated under the title "An Axiomatic Approach to the Law of Small Numbers". Noor is at the Department of Economics, Boston University, 270 Bay State Road, Boston MA 02215. Payró is at the Department of Economics and Economic History, Universitat Autònoma de Barcelona and Barcelona School of Economics, Campus de la UAB, Edifici B, Bellaterra, 08193. The authors would like to thank the Editor and three anonymous referees for insightful comments, and seminar audiences at Boston University, UPenn, U of Edinburgh, Manchester, Durham, LSE, UAB, UPF, McGill U and attendees at the SAEe 2022, the East Coast Behavioral and Experimental 2023 workshop, BRIC 2023, CESC 2023, MOVE-ISER 2023 workshop, BSE Summer Forum, RUD 2023 and TUS 2023 for helpful feedback. Payró gratefully acknowledges financial support from the Ministerio de Economía y Competitividad and Feder (PGC2018-094348-B-I00, PID2020-116771GB-I00 and PID2023-147183NB-I00), the Generalitat de Catalunya (2021 SGR 00194) and the Spanish Agencia Estatal de Investigación (AEI), and the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2024-001476-S), funded by MCIN/AEI/10.13039/501100011033. All errors are our own. The usual disclaimer applies.

1 Introduction

A large number of empirical studies show that many people do not seem to understand the nature of randomness, in that their beliefs about the outcomes of an i.i.d. random process tend to be systematically misspecified. Three well-known findings are as follows:

Gambler's Fallacy: Subjects systematically tend to perceive the probability of tails to be higher following a streak of heads: for instance, $P(HHHT) > P(HHHH)$.[1] While there is a large experimental literature establishing the Gambler's Fallacy (reviewed in Benjamin (2019)), there is also substantial evidence from the field. Observing gamblers playing roulette in a casino setting, Croson and Sundali (2005) document that over 65% of subjects bet against the color that appeared 5 times in a row. Studying betting on lottery numbers, Terrell (1994) demonstrates that people are significantly less likely to bet on a lottery number if it was recently a winning number (see Suetens et al. (2016) for a more recent study). Using data from a field experiment, Chen, Moskowitz and Shue (2016) document that loan officers are more likely to reject (resp. accept) a loan application after they have accepted (resp. rejected) the previous application, and estimate that in some treatments up to 9% of decisions are erroneous due to such a sequencing effect. They find a similar bias in decisions by US judges in refugee asylum cases. Jin and Peng (2022) show that various puzzles in finance (such as the disposition effect, where traders hold on to recently losing stock and sell recently winning stock) can be explained by the Gambler's Fallacy.

Belief in Alternation: Subjects exhibit a tendency to expect frequent switching in the evolution of random outcome sequences. For instance, they believe that $P(HHTHTH) > P(HHHHTT)$. Rapoport and Budescu (1997), and Bar-Hillel and Wagenaar (1991) propose that subjects believe in an alternation rate of approximately 60% for a fair coin. In 2014, Spotify changed its perfectly random shuffling algorithm to a non-random algorithm, due to complaints by users that the original shuffle was not random because one artist's songs would sometimes play in succes-

[1] The *Hot Hand Effect* is the opposite finding that subjects sometimes expect a streak to be more likely to continue (Gilovich, Tversky, and Vallone (1985)). The literature reconciles the two findings in the following way. The Gambler's Fallacy appears when the agent is sure about the bias, but the Hot Hand Effect is understood to be the result of updating of uncertainty about the bias: Gilovich, Tversky and Vallone (1985) note that a streak of heads will push the posterior belief towards a higher bias, leading to an expectation that the streak will continue. See Rabin and Vayanos (2010) for a formalization. The Hot Hand Effect has been demonstrated in contexts such as sports (Gilovich, Tversky, and Vallone (1985)), and to our knowledge there is no evidence of a Hot Hand Effect in the context of coin tosses.

sion.[2]

Sample Size Neglect: A striking finding in the literature is that subjects do not recognize that sampling variance decreases with sample size. Kahneman and Tversky (1972) report an experiment where subjects are told that 45 babies are born per day in a large hospital and 15 babies are born per day in a small hospital, and each hospital has recorded the daily gender distribution over a full year. Subjects were asked which hospital had more days with over 60% boy births. Subjects had to respond "larger hospital", "smaller hospital" or "about the same", and the vast majority believed that both hospitals had a similar number of such days, not recognizing that the variance of the sampling distribution should be higher in the small hospital.

The paradigm in psychology explains such beliefs as arising from an incorrect intuition about probabilities, which is driven by the use of particular heuristics. It is hypothesized that people believe that "even small samples are highly representative of the populations from which they are drawn" (Tversky and Kahneman (1974, pg 1125-1126)), a property dubbed *the Law of Small Numbers*.[3] The term "small samples" extends further to segments within any sample: "each segment of the [sequence] is highly representative of the "fairness" of the coin" (Tversky and Kahneman (1971, pg 106)). One feature of a large sample is that it has roughly equal proportions of heads and tails, and this generates a central prediction of the Law of Small Numbers, namely, that subjects will believe that small samples have similar proportions of heads and tails.[4] This generates a belief that streaks are unlikely, since a streak creates a segment that is not highly representative of the fairness of the coin. This in turn explains the Gambler's Fallacy and a Belief in Alternation. Tversky and Kahneman (1974) also hypothesize that, because subjects presume that the properties of a sample represent that of the population, the probability of an event (e.g. 60% boy births) in a sample of any size approximately equals the probability of the event in a large sample, leading to Sample Size Neglect.

Working in a canonical coin-tossing environment, in this paper we show that the above evidence can be tied together through a belief that the outcome of each coin

[2] https://engineering.atspotify.com/2014/02/how-to-shuffle-songs/

[3] The term "Law of Small Numbers" is coined in Tversky and Kahneman (1971) but, in their 1974 paper, the authors appear to prefer to use the term "Local Representativeness" instead. Local Representativeness describes the belief that "the essential characteristics of the process will be represented, not only globally in the entire sequence, but also locally in each of its parts" (Tversky and Kahneman (1974, pg 1125)). In this paper we treat the two terms as interchangeable.

[4] For an example of other predictions, the Law of Small Numbers predicts that subjects will believe that random processes will not generate patterns, such as a recursion of HT (Wagenaar (1970), Kahneman and Tversky (1972)). This is presumably because patterns are considered more representative of non-random processes.

toss will evolve such that the sample mean gravitates towards the bias θ^* of the coin. We formulate a basic notion of *Belief in Mean Reversion* (*Mean Reversion* for short) and show that beliefs satisfying it can be represented as if the bias of the coin is history-dependent and "self-correcting" towards θ^*. We show that Mean Reversion implies the Gambler's Fallacy and Belief in Alternation and that it is consistent with Sample Size Neglect. We also study a simple "local" version of the model where the sample mean is replaced by the "mean in the last *few* tosses", in a manner reminiscent of Tversky and Kahneman (1974) and Rabin (2002).

Having explored the behavioral implications of Mean Reversion, we next study how an agent holding such misspecified beliefs learns from data, focusing on Bayesian inference. We show that a misspecified agent observing infinitely many i.i.d. coin tosses almost surely assigns strictly positive probability to the true bias. Intuitively, this is because the agent believes that the sample mean tends to the (unknown) true parameter, which it almost surely does according to the Law of Large Numbers. This interaction steers the agent towards the true parameter, though her misspecified beliefs may keep her from being sure about its truth. This is in contrast with Rabin (2002), where the agent can become confident about the wrong parameter.

Contributions to the Literature. This paper makes two key contributions.

(i) The Law of Small Numbers (Kahneman and Tversky (1972), Tversky and Kahneman (1974)) is expressly an informal theory about how subjects reason intuitively, and has been criticized as such.[5] Our model of mean reversion (and its local version) can be viewed as a formalization of a key hypothesis in the Law of Small Numbers, namely, that sample proportions will approximately equal population proportions. Indeed, such a belief suggests, and is suggested by, a belief that deviations from population proportions in a sample ought to be corrected in further realizations, that is, a belief in mean reversion. By showing that key evidence for the Law of Small Numbers can be generated by a belief in mean reversion (Propositions 4-7), we establish that *our model captures several core empirical implications of the Law of Small Numbers.* As discussed in Section 4, our work complements but also significantly differs from the seminal work by Rabin (2002).

(ii) We derive novel insights about the Law of Small Numbers. Theorem 1 shows that some forms of belief in mean reversion necessarily violate *Marginal Consistency*, the standard condition that beliefs over sequences of a given length are the marginals of beliefs about sequences of longer length. Intuitively, while expecting mean reversion, an agent may think that large deviations from θ^* in the first 10 tosses are more likely to happen when the coin is known in advance to be tossed 1000 times rather

[5] The lack of formal specification of heuristics has been criticized by Gigerenzer (1996) on the grounds that they offer too much flexibility and risk being unfalsifiable.

than just 10 times and thus, her beliefs over the first n tosses depend on how many times the coin is tossed in general. Theorem 3 on the other hand shows that a belief in local mean reversion precludes a belief in global mean reversion. This reveals that the Law of Small Numbers should not be thought of asserting both a belief in global and local mean reversion, but rather as asserting that beliefs are a function of both the global sample mean and the local sample mean. This makes the theory relatively weak, as it predicts global and local mean reversion only for the small subset of pairs of sequences where there is a dominance in terms of both global and local mean reversion.

The paper is organized as follows. Section 2 states our primitives and main model. Section 3 presents several results, while Section 4 presents a "local" version of our model. Section 5 studies Bayesian inference and Section 6 relates this paper to the literature. All proofs are relegated to appendices.

2 Mean Reversion: Framework

The evidence on beliefs about randomness is both static (ex-ante, is it more likely to get HHHH or HHHT?) and dynamic (given HHH, is it more likely that the next toss is H or T?). The static evidence is, by definition, revealing properties of the agent's ex-ante beliefs about a sequence of coin tosses. In particular the evidence necessitates a model of misspecified beliefs about the data generating process. An important observation is that the dynamic evidence, which reveals properties of posterior beliefs, is entirely consistent with Bayesian updating of an incorrect prior. Parsimony demands that we visualize the evidence in terms of incorrect ex-ante beliefs alone rather than incorrect updating as well. Consequently, the theory we present is static, with the understanding that dynamic applications will assume Bayesian updating.

2.1 Primitives

Consider a canonical coin-tossing environment: the possible realizations of a coin toss in any period i are $\Omega_i = \Omega = \{0, 1\}$. We interpret 1 as heads H, and 0 as tails T, and sometimes use the notation H, T instead of $1, 0$. The space of all realizations of any $1 \leq n \leq \infty$ tosses is $\Omega^n = \prod_{i=1}^n \Omega_i$. Adopt the convention that $\Omega^0 = \emptyset$, and use $x = (x_1, x_2, ...) \in \Omega^\infty$ to denote an infinite sequence and $x^n = (x_1, ..., x_n) \in \Omega^n$ to denote a finite sequence of length n. Our results do not hinge on the binariness of Ω, which we maintain for simplicity of exposition, and can be extended at least to any finite set Ω (see the remarks at the end of Section 2.2).

Our primitive is a family of probability measures $\{P^n\}_{n=1}^{\infty}$, where each P^n is a probability measure on the measurable space (Ω^n, Σ^n) defined by the sample space Ω^n generated by n-tosses and the σ-algebra $\Sigma^n = 2^{\Omega^n}$ of all subsets $A^n \subset \Omega^n$. Throughout, we maintain that the family of beliefs has *full support*: $P^n(A^n) > 0$ for every $A^n \subset \Omega^n$ and all n. We also impose a natural *regularity* assumption: $P^n(x^{n-1}x_n) \leq P^{n-1}(x^{n-1})$ for all x^{n-1}, x_n, that is, the likelihood of a given sequence is always higher than any of its continuations. Define conditional beliefs by:

$$P^n(x_n|x^{n-1}) := \frac{P^n(x^{n-1}x_n)}{P^{n-1}(x^{n-1})}. \tag{1}$$

By the full support and regularity assumptions, conditional beliefs take values in $(0, 1]$.

Without more structure, conditional beliefs at a given history do not necessarily sum to 1. This property is obtained in a more standard model of beliefs where $\{P^n\}_{n=1}^{\infty}$ are the marginals of some belief P over Ω^∞, modeled as a probability measure P over the usual infinite horizon sample space $(\Omega^\infty, \Sigma^\infty)$.[6] By Kolmogorov's extension theorem, there exists such P over Ω^∞ if and only if $\{P^n\}_{n=1}^{\infty}$ satisfies *Marginal Consistency*, the property that each P^n is the marginal of P^{n+1}:

$$P^n(A^n) = P^{n+1}(A^n) \text{ for any } n \text{ and } A^n \subset \Omega^n.$$

This property states that the likelihood of an event A^n involving the first n tosses does not depend on the number of tosses $m \geq n$ specified before hand. We do not impose Marginal Consistency at the outset since it will prove not to be compatible with some forms of belief in mean reversion, as we show in Theorem 1.

2.2 Model

In experiments, subjects are told that sequences are objectively generated by θ^*-i.i.d. tosses of a coin. But subjects' beliefs are often not consistent with the objective properties of the coin. This admits a natural interpretation where the subjects have a misspecified understanding of the objective random process.[7] To formalize this, suppose the coin has objective bias $\theta^* \in (0, 1)$ and is tossed independently multiple

[6] For any $A^n \subset \Omega^n$, define an n-cylinder by the event $A^n\Omega^\infty = \{(x^n z) \in \Omega^\infty : x^n \in A^n \text{ and } z \in \Omega^\infty\}$ in the infinite horizon space Ω^∞. Then $\Sigma^\infty = \sigma(\cup_{n=1}^{\infty}\Sigma^n)$ is the σ-algebra generated by all the n-cylinders, $n = 1, 2, ...$

[7] In a companion paper, Noor and Payro [18], we take a purely subjective approach to defining a belief in mean reversion, one that does not invoke the existence of an objective random process.

times to generate sequences of heads and tails. Denote the probability of a θ^*-i.i.d sequence x^n by

$$Q_{\theta^*}(x^n) = \prod_{i=1}^{n} (\theta^*)^{x_i} (1-\theta^*)^{1-x_i}.$$

The *sample mean* for a sequence $x^n = (x_1, ..x_n)$ is the mean number of heads, given by

$$\overline{x}^n := \frac{1}{n} \sum_{i \leq n} x_i.$$

Denote the distance between the sample mean and the bias θ^* by

$$d(x^n) = |\overline{x}^n - \theta^*|.$$

Adopt the convention that $d(x^{i-1}) = \overline{x}^{i-1} = 0$ when $i = 1$.

We model the agent as one whose beliefs are driven by $d(x^n)$.[8] Say that x_n is *mean reverting at* x^{n-1} if, after history x^{n-1}, outcome x_n induces a smaller distance to θ^* than $1 - x_n$,

$$d(x^{n-1}x_n) \leq d(x^{n-1}(1 - x_n)).$$

A very basic notion of belief in mean reversion would require simply that, at every history, the agent assigns a weakly higher conditional probability to the mean reverting outcome:

Axiom 1. *(Mean Reversion) For any $n \geq 1$, sequence $x^{n-1} \in \Omega^{n-1}$ and outcomes $x_n, y_n \in \Omega$,*

$$d(x^{n-1}x_n) \leq d(x^{n-1}y_n) \implies P^n(x_n|x^{n-1}) \geq P^n(y_n|x^{n-1}).$$

This axiom is reminiscent of a "belief that a random process is self-correcting" (Tversky and Kahneman (1974)). This is reflected in the following simple representation result. Recall that our notation for the empty history is $\Omega^0 = \emptyset$.

Proposition 1. *Beliefs $\{P^n\}_{n=1}^{\infty}$ satisfies Mean Reversion if and only if for all n and $x^n \in \Omega^n$,*

$$P^n(x^n) = \prod_{i=1}^{n} g^i(d(x^i), x^{i-1}),$$

[8] The use of absolute differences implies that deviations from θ^* from above or from below are treated symmetrically. This is natural for a coin-tossing setup but may not be in others. For instance, symmetry may be violated if there is some utility attached to each outcome - noise in a forest may be associated with danger and silence with safety, or positive vs negative news about a favored political party may not be treated symmetrically.

where the function $g^i(\cdot, x^{i-1}) : [0,1] \to (0,1]$ *is continuous, weakly decreasing for each* i *and* x^{i-1}.

Thus, the likelihood that the agent assigns to a sequence x^n can be written as the product of the (history-dependent) propensity $g^i(d(x^{i-1}x_i), x^{i-1})$ for each outcome x_i in the sequence. Since g^i is decreasing in its first argument, the propensity for x_i is higher than that for $1 - x_i$ whenever it induces a lower distance to θ^*.[9] The next proposition clarifies the empirical meaning of this propensity and also provides a characterization of Marginal Consistency.

Proposition 2. *In the representation for* $\{P^n\}_{n=1}^{\infty}$ *that satisfies Mean Reversion, for any* n *and* $x^n \in \Omega^n$,

$$g^n(d(x^n), x^{n-1}) = P^n(x_n | x^{n-1}).$$

Moreover, $\{P^n\}_{n=1}^{\infty}$ *satisfies Marginal Consistency if and only if in the representation, for all* x^{i-1} *and all* i,

$$g^i(d(x^{i-1}1), x^{i-1}) + g^i(d(x^{i-1}0), x^{i-1}) = 1.$$

Thus, the propensities g^i are uniquely pinned down by conditional beliefs. The noted characterization of Marginal Consistency is obtained from the observation that Marginal Consistency is equivalent to the requirement that conditional beliefs sum to 1: $P^i(1|x^{i-1}) + P^i(0|x^{i-1}) = 1$ for all x^{i-1}. When Marginal Consistency is satisfied, the representation can be written as

$$P^n(x^n) = \prod_{i=1}^{n} (\theta_{x^{i-1}})^{x_i} (1 - \theta_{x^{i-1}})^{1-x_i},$$

where the bias towards heads is $\theta_{x^{i-1}} := g^i(d(x^{i-1}1), x^{i-1})$ and towards tails is $1 - \theta_{x^{i-1}} := g^i(d(x^{i-1}0), x^{i-1})$. As g^i is weakly decreasing, the model requires that the bias is weakly self-correcting: if $d(x^{i-1}1) \leq d(x^{i-1}0)$ then $\theta_{x^{i-1}} \geq 1 - \theta_{x^{i-1}}$.

An example of our model with $\theta^* = \frac{1}{2}$ is the *Friedman Urn* (Friedman (1949)) studied in the statistics literature. Imagine that an urn initially contains one ball labeled "heads" and one ball labeled "tails". Inductively, at any toss n, if a heads

[9] While the model posits that beliefs depend on the distance $|\frac{\sum x^i}{i} - \theta^*|$, it subsumes the case where beliefs depend on the difference $|\sum x^i - i\theta^*|$ between actual number of heads and the expected number of heads. The reason is that g^i depends on i, and so we can think of models where $g^i(|\frac{\sum x^i}{i} - \theta^*|, x^{i-1}) = f^i(i \times |\frac{\sum x^i}{i} - \theta^*|, x^{i-1})$. Since $i \times |\frac{\sum x^i}{i} - \theta^*| = |\sum x^i - i\theta^*|$ we see that g^i now depends on the deviation in head counts.

(resp. tails) is drawn, then it is placed back in the urn and an additional ball labeled "tails" (resp. "heads") is added to the urn. This generates conditional probabilities for any $n \geq 1$ given by

$$P^n(1|x^{n-1}) = \frac{(n-1)(1-\overline{x}^{n-1})+1}{n+1} \text{ and } P^n(0|x^{n-1}) = \frac{(n-1)\overline{x}^{n-1}+1}{n+1},$$

with the convention that $\overline{x}^0 = 1$, so that the probability of a heads on the first toss is $P^1(1) = P^1(0) = \frac{1}{2}$.[10] While Mean Reversion is silent on how conditional probabilities depend on the history, the Friedman Urn requires dependence on a history x^{n-1} through its sample mean $\overline{x}^{n-1}$. Marginal Consistency is satisfied since $P^n(1|x^{n-1}) + P^n(0|x^{n-1}) = 1$.

We conclude by discussing what an extension of our model could look like when the random process has more than two outcomes. Consider a countable outcome space $\Omega \subset \mathbb{R}$, where the bias for outcome $\omega \in \Omega$ is given by θ_ω. For any sequence, write the distance wrt bias θ^*_ω as $d_\omega(x^n) = |\frac{\sum_{i=1}^n I(x_i=\omega)}{n} - \theta^*_\omega|$. One can consider various ways of aggregating $\{d_\omega(x^n)\}_{\omega\in\Omega}$. A simple example is the sup distance:

$$d_\Omega(x^n) = \sup_{\omega\in\Omega} d_\omega(x^n).$$

Mean Reversion can be defined in terms of d_Ω and analogs of our representation results obtain accordingly. Our definition of Mean Reversion for a coin $\Omega = \{H, T\}$ is a special case of such a definition:

$$d(x^n) = d_H(x^n) = |\overline{x}^i - \theta^*|$$

$$= |(1-\overline{x}^i) - (1-\theta^*)| = d_T(x^n).$$

The distance d_Ω is well-defined even when Ω is uncountable, and we can consider representations where $f^n(x^n) := \prod_{i=1}^n g(d_\Omega(x^i), x^{i-1})$ is the density for belief P^n.

[10] The Friedman Urn is a probability distribution P on Ω^∞ that satisfies $P(x^n) = \prod_{i=1}^n P^n(x_i|x^{i-1})$ where each P^n is as defined above. It is readily determined that

$$P^n(1|x^{n-1}) \geq P^n(0|x^{n-1}) \iff \overline{x}^{n-1} \leq \frac{1}{2}.$$

Moreover, heads (resp. tails) is mean reverting iff $\overline{x}^{n-1} \leq \frac{1}{2}$ (resp. $\overline{x}^{n-1} \geq \frac{1}{2}$). This verifies that Mean Reversion holds.

3 Properties

3.1 d-Driven Mean Reversion

Mean Reversion does not place any restriction across histories. A natural special case of Mean Reversion is where the conditional belief about an outcome x_n at history x^{n-1} depends on the history only through the induced distance $d(x^{n-1}x_n)$:

Axiom 2. *(d-Driven Mean Reversion) For any $n > 0$ and sequences $x^n, y^n \in \Omega^n$,*

$$d(x^n) \leq d(y^n) \implies P^n(x_n|x^{n-1}) \geq P^n(y_n|y^{n-1}).$$

To illustrate, take a fair coin $\theta^* = \frac{1}{2}$. Consider the histories HTH and THH, and suppose the next outcome is T in the first case and H in the second, giving rise to $HTHT$ and $THHH$ respectively . We see that the sample mean moves closer to θ^* in the first continuation and farther away in the second, $d(HTHT) < d(THHH)$. It is natural that a mean-reverting tails is expected with greater intensity conditional on HTH than a non-mean-reverting heads after THH, that is, $P(T|HTH) \geq P(H|THH)$. This is the assertion of the axiom.

With $x^{n-1} = y^{n-1}$ subsumed as a special case, we see that this strengthens Mean Reversion. Indeed, we obtain a special case of the representation in Proposition 1 where g^i depends only on the distance $d(x^i)$.

Proposition 3. *$\{P^n\}_{n=1}^{\infty}$ satisfies d-Driven Mean Reversion iff for all n and $x^n \in \Omega^n$,*

$$P^n(x^n) = \prod_{i=1}^{n} g^i(d(x^i)),$$

where each $g^i : [0,1] \to (0,1]$ is continuous, weakly decreasing. Moreover, Marginal Consistency holds if and only if $g^i(d(x^{i-1}1)) + g^i(d(x^{i-1}0)) = 1$ for all x^{i-1}.

Natural and tractable as this model is, we establish a fundamental tension with Marginal Consistency: if both d-Driven Mean Reversion and Marginal Consistency are satisfied, then the agent must view the coin as fair for all tosses $n \geq 3$.

Theorem 1. *If $\{P^n\}_{n=1}^{\infty}$ satisfies d-Driven Mean Reversion and Marginal Consistency, then $P^n(1|x^{n-1}) = P^n(0|x^{n-1}) = \frac{1}{2}$ for all $n \geq 3$ and all histories $x^{n-1} \in \Omega^{n-1}$.*

A model of mean reversion is interesting only if it is nontrivial, in the sense that at each toss n there is at least one history x^{n-1} after which the agent exhibits a strict belief in a correction, $P^n(1|x^{n-1}) \neq P^n(0|x^{n-1})$. Theorem 1 is an impossibility result establishing that, under d-Driven Mean Reversion, it is impossible to satisfy Marginal Consistency and simultaneously exhibit nontrivial mean reversion. The proof is based on the following observation. Take a fair coin $\theta^* = \frac{1}{2}$ and sequences $HTHT$ and $THHH$. Consider the continuations of these sequences, respectively $HTHTT$ and $THHHH$. We see that the mean gets farther from $\theta^* = \frac{1}{2}$ in the second sequence than in the first, $d(HTHTT) < d(THHHH)$. Therefore, d-Driven Mean Reversion requires $P(T|HTHT) \geq P(H|THHH)$. However, d-Driven Mean Reversion also requires $P(H|HTHT) = P(T|THHH)$ since both lead to the same distance to $\theta^* = \frac{1}{2}$. If there is strict mean reversion, given by $P(T|THHH) > P(H|THHH)$, then it must be that

$$P(T|HTHT) + P(H|HTHT) > P(H|THHH) + P(T|THHH).$$

Since the terms on either side of the inequality cannot both sum to 1, Marginal Consistency is necessarily violated.

The take-away is that, if an agent satisfies Marginal Consistency and exhibits nontrivial mean reversion, then belief in mean reversion must depend on the history in a way that is not entirely captured by the distance to θ^*. For instance, the implication of d-Driven Mean Reversion noted above, $P(H|HTHT) = P(T|THHH)$, will fail if an agent cares about the sequencing of the history. Indeed, $P(H|HTHT) < P(T|THHH)$ would hold if the agent believes that T is more "urgent" after the history $THHH$ that contains a streak of heads.

An alternative reading of the result is that violations of Marginal Consistency may not be unnatural in the context of belief in mean reversion. For instance, one may think it is very likely to get HTH in the first 3 tosses when there are only $n = 3$ tosses. But if the agent is told that the coin will be tossed $n = 100$ times, then they may think that larger deviations from θ^* are more likely, thereby exhibiting $P^3(HTH) > P^{100}(HTH)$ and violating Marginal Consistency. This yields an insight into the Law of Small Numbers: its prediction of a belief in mean reversion can possibly involve a violation of Marginal Consistency.

It is an empirical matter whether d-Driven Mean Reversion or Marginal Consistency are satisfied by agents who exhibit a belief in mean reversion. From a modeling perspective we note that d-Driven Mean Reversion yields a highly tractable model that can facilitate applications (for instance see Theorem 5 below), and that for some applications this may therefore be a reason to drop Marginal Consistency. For instance, a one-parameter specification that could be useful for empirical work is where $P^n(x^n) = \prod_{i=1}^{n} g^i(d(x^i))$ with

$$g^i(d) = \frac{Z_{i-1}}{Z_i} \left(\frac{1}{1+d} \right)^{\lambda_i},$$

where $Z_0 := 1$ and $Z_n = \sum_{x^n \in \Omega^n} \prod_{i=1}^n \left(\frac{1}{1+d(x^i)}\right)^{\lambda_i}$ for each n.[11] Then, according to the model, $\frac{P^i(x^{i-1}1)}{P^i(x^{i-1}0)} = \left(\frac{1+d(x^{i-1}0)}{1+d(x^{i-1}1)}\right)^{\lambda_i}$. Therefore the parameter λ_i is uniquely identified by data on beliefs and parametrizes the intensity of the belief in mean reversion as a function i, enabling an analysis of whether belief in mean reversion gets stronger further down the sequence.

3.2 Gambler's Fallacy

We define Gambler's Fallacy as a belief that sufficiently long streaks are expected to break. Denote the concatenation of two sequences $x^n \in \Omega^n$ and $y^m \in \Omega^m$ by $x^n y^m \in \Omega^{n+m}$.

Definition 1. *$\{P^n\}_{n=1}^\infty$ exhibits the Gambler's Fallacy if for each outcome $\omega \in \Omega$ and history x^n there exists M such that for all $m \geq M$,*

$$P^n(1-\omega|x^n\omega^m) > P^n(\omega|x^n\omega^m).$$

In words, given any initial segment x^n, there is a long enough streak ω^m of outcome $\omega \in \{0,1\}$ after which the agent expects outcome $1-\omega$ to occur with greater probability. It is readily determined that:

Proposition 4. *If $\{P^n\}_{n=1}^\infty$ admits a Mean Reversion representation with g^n strictly decreasing in its first argument for all n, then $\{P^n\}_{n=1}^\infty$ exhibits the Gambler's Fallacy.*

Mean Reversion takes a stand on a foundational question: "what determines a streak?". While a "streak of heads" is commonly defined by the *contiguity* of heads in a sequence, Mean Reversion defines a streak as a sequence of outcomes that leads the sample mean to overshoot θ^*. If contiguity alone is the driver of the Gambler's Fallacy then a subject who is told that 25 tosses have already been performed with 5 heads appearing in the last 5 tosses might believe that a tails is more likely on the 26th toss, *regardless* of the outcomes of the first 20 tosses. Indeed in an experimental design where the first 20 outcomes $x^{20} = (x_1, .., x_{20})$ are concealed, the subject may not pay to see x^{20}. In contrast, the mean reversion perspective says that a streak of heads has occurred only if the sample mean (sufficiently) exceeds $\frac{1}{2}$. This implies that the subject will pay for the information, since the proportion of heads and tails

[11] This equality comes from $\frac{1}{Z_n}\sum_{x^n \in \Omega^n} \prod_{i=1}^n \left(\frac{1}{1+d(x^i)}\right)^{\lambda_i} = \sum_{x^n \in \Omega^n} \prod_{i=1}^n \frac{Z_{i-1}}{Z^i} \left(\frac{1}{1+d(x^i)}\right)^{\lambda_i} = \sum_{x^n \in \Omega^n} P^n(x^n) = 1$.

in x^{20} is not irrelevant. The question of whether streaks are defined by contiguity alone or also by mean reversion is an interesting question for future experimental research.

3.3 Belief in Alternation

While all sequences of a given length n generated by a fair coin are objectively equally likely, subjects tend to believe that some sequences are more likely to occur than others. For instance, subjects exhibit a so-called *misperception of randomness*, believing that it is likely that outcomes will alternate considerably, giving rise to beliefs such as $P^6(HHTHTH) > P^6(HHHHTT)$ (Tversky and Kahneman (1974)). In fact, Rapoport and Budescu (1997) and Bar-Hillel and Wagenaar (1991) find that subjects believe in an alternation rate of approximately 60% for a fair coin. At the same time, perfectly alternating sequences, such as $HTHTHT$ are viewed as less likely. This is typically explained by a *disbelief in patterns*, where subjects regard it as unlikely that a random sequence will possess a discernible pattern, believing that recursions like $HTHTHT$ are less likely than a less patterned sequence such as $HTTHTH$ (Wagenaar (1970), Kahneman and Tversky (1972)).

We investigate whether our model can capture a belief in "intermediate" levels of alternation. For $\theta^* = \frac{1}{2}$, say that $\{P^n\}_{n=1}^{\infty}$ exhibits a *belief in alternation* if for any $n \geq 2$ the constant sequences and the perfectly alternating sequences are not in $\arg\max_{x^n \in \Omega^n} P^n(x^n)$. That is, the most likely sequences exclude the sequences that have no alternation and also the sequences that have maximal alternation, implying that the most likely sequences have an intermediate degree of alternations. This captures misperception of randomness and a limited form of disbelief in patterns (that is, a disbelief only in perfectly alternating pattern).

The next proposition confirms the obvious intuition that a strict belief in mean reversion implies that constant sequences are never among the most likely sequences, establishing the first requirement for a belief in alternation. The proposition also shows that the second requirement for a belief in alternation is consistent with our model.

Proposition 5. *Suppose that $\theta^* = \frac{1}{2}$ and that $\{P^n\}_{n=1}^{\infty}$ admits a Mean Reversion representation with g^i strictly decreasing in its first argument for all i. Then the following statements hold:*

(a) For any $n \geq 2$, the constant sequences are not in $\arg\max_{x^n \in \Omega^n} P^n(x^n)$.

(b) Perfectly alternating sequences need not belong to $\arg\max_{x^n \in \Omega^n} P^n(x^n)$ for all n.

The proof for the first assertion is simply that for any constant sequence, say TTT, Mean Reversion with strict sensitivity to distance implies that TTH will be strictly more likely. The second assertion is proved by constructing an example where

$$P^3(HTH) < P^3(HHT),$$

so that the perfectly alternating sequence HTH is excluded from $\arg\max_{x^3 \in \Omega^3} P^3(x^3)$. In the example, corrections are rewarded more strongly at histories like HH where there is a larger deviation from θ^* compared to histories like HT, and the idea is that if this reward is strong enough then some delayed corrections may be assigned a higher probability than immediate corrections. The fact that this is possible in our model separates it from classic formal models in the literature such as Rapoport and Budescu (1997), Bar-Hillel and Wagenaar (1991) and Rabin (2002), all of which require that perfectly alternating sequences are among the most likely sequences.

We close by showing, for completeness, that there do exist special cases of the model where perfect alternation is always considered most likely, the d-Driven Mean Reversion model being one special case.

Proposition 6. *Suppose $\theta^* = \frac{1}{2}$ and assume $\{P^n\}_{n=1}^{\infty}$ admits a Mean Reversion representation where, for all i, $g^i(d(x^i)|x^{i-1}) = f^i(d(x^1), \ldots, d(x^i))$ for some f^i that is strictly decreasing in each of its arguments. Then the perfectly alternating sequences belong to* $\arg\max_{x^n \in \Omega^n} P^n(x^n)$ *for all n.*

3.4 Non-Belief in the Law of Large Numbers

Sample Size Neglect (Kahneman and Tversky (1972)) suggests a more general phenomenon where, in contrast to the Law of Large Numbers, people tend to believe that the sampling distribution does not concentrate around the population parameter as the sample size grows – this is referred to as *Non-Belief in the Law of Large Numbers* (Benjamin, Rabin and Raymond (2016) and Benjamin, Moore and Rabin (2018)). The literature has different explanations and models for the Gambler's Fallacy and Non-Belief in the Law of Large Numbers (Tversky and Kahneman (1974), Rabin (2002), Benjamin, Rabin and Raymond (2016)). In this section we explore whether these findings can be generated alongside a belief in mean reversion.

One might suspect not. If, as per the Law of Large Numbers, the objective sampling distribution generated by a fair coin collapses on θ^* as $n \to \infty$, then should the agent's subjective sampling distribution not also collapse on θ^* if the coin is believed to be continually self-correcting towards θ^*? Our next result shows that the answer is "not necessarily".

Proposition 7. *If* $\{P^n\}_{n=1}^{\infty}$ *satisfies Marginal Consistency and Mean Reversion with respect to some* θ^**, then the Law of Large Numbers is not implied: that is, it is possible that*

$$\lim_{n\to\infty} P^n(\overline{x}^n = \theta^*) \neq 1.$$

The proof constructs an example of an agent with a belief P over Ω^∞ whose induced marginals $\{P^n\}_{n=1}^{\infty}$ satisfy Mean Reversion in a very specific way: she expects strict mean reversion along some sequences and no mean reversion along others. This gives rise to a limiting distribution for the sample mean that assigns positive probability to both θ^* and $\frac{1}{2}$, thereby violating the Law of Large Numbers.

The Friedman Urn is an example of a model that exhibits strict mean reversion at every history and gives rise to a Law of Large Numbers.[12] Future research might explore strengthenings of Mean Reversion that guarantee a Law of Large Numbers.

4 Locally Self-Correcting Bias

By asserting that the sample mean tends toward θ^* in any finite sample, the Law of Small Numbers implies a belief in mean reversion (Tversky and Kahneman (1971)). However, it goes further in requiring that the sample mean also tend toward θ^* within any "local" segment of a sequence. For instance, while the sequences $HHHTT$ and $HTHTT$ generated by a fair coin both achieve the same sample mean, the Law of Small Number requires that the latter sequence will be considered more likely because the mean in each contiguous pair of tosses (first and second, second and third, etc) is closer to $\frac{1}{2}$ than that in the first sequence. This is an expression of "local representativeness", that is, the sample mean is believed to be close to θ^* not only in the full sample but also in local segments (Tversky and Kahneman (1974), Rabin (2002)). In this section we ask: is it possible for a model to exhibit a belief that *both* the global and local means tend to θ^*?

We capture local representativeness through a "local" version of our model in which, instead of evaluating the global sample mean at toss i computed from the sequence $x_1, \ldots, x_i$, the agent evaluates a local sample mean based only on the "last few" tosses. Our main finding is that it is not possible to satisfy mean reversion in both the global and the local sense simultaneously.

[12] Another (extreme, but simple) example is where $\theta^* = \frac{1}{2}$ and the agent places probability $\frac{1}{2}$ on heads in the first toss, but then deterministically believes in a perfectly alternating continuation, so that $P(HTHT....) = P(THTH....) = \frac{1}{2}$. This agent has an extreme belief in mean reversion, and the sample mean equals $\frac{1}{2}$ after every two tosses, and approximates $\frac{1}{2}$ arbitrarily well on odd tosses, giving rise to a degenerate distribution on $\frac{1}{2}$ as the sample size goes to infinity.

4.1 Locally Self-Correcting Bias Model

Define the "segment at n" by:

Definition 2. *(Segments) For any n, a segment at n is the smallest set of contiguous indices $W_n = \{k_n, .., n\} \subseteq \{1, .., n\}$ containing n and satisfying*

$$P^n(x_n|x_1...x_{k_n-1}x_{k_n}...x_{n-1}) = P^n(x_n|y_1...y_{k_n-1}x_{k_n}...x_{n-1}) \text{ for all } x, y \in \Omega^\infty.$$

The definition requires that outcomes $x_1...x_{k_n-1}$, which are *outside* of the segment $\{k_n, .., n\}$, never impact the conditional probability of x_n. Segments are assumed to consist of contiguous tosses and depend only on n (rather than the outcome x_n). This is to keep notation clean, and can be easily relaxed. An empirical evaluation of the nature of segments is an interesting direction for future research. A priori it is not obvious that segments should be expected to exclude any toss. Attentional constraints may lead subjects to focus on the last few tosses, but excessively long streaks earlier in the sequence may conceivably attract attention as well.

Segments are uniquely defined by beliefs. We require conditional beliefs to exhibit a local version of d-Driven Mean Reversion in its strict form. Define the *segment mean* of x^n at n by $\overline{x}^n(W_n) = \frac{\sum_{i \in W_n} x_i}{|W_n|}$ and corresponding *local distance* by

$$d_{W_n}(x^n) = |\overline{x}^n(W_n) - \theta^*|.$$

Axiom 3. *(Local Mean Reversion) For any n and $x^n, y^n \in \Omega^n$ s.t.*

$$d_{W_n}(x^n) \leq d_{W_n}(y^n) \iff P^n(x_n|x^{n-1}) \geq P^n(y_n|y^{n-1}).$$

In contrast to our previous axioms, the current axiom is an if-and-only-if condition. This is needed so that $\{W_n\}$ are the only possible family of segments that are consistent with the following representation:

Theorem 2. *$\{P^n\}_{n=1}^\infty$ satisfies Local Mean Reversion iff for any n and $x^n \in \Omega^n$,*

$$P^n(x^n) = \prod_{i=1}^{n} g^i(d_{W_i}(x^i)),$$

where for each i there is a set $W_i = \{i - r_i, .., i\}$ for some $r_i \geq 0$, and a continuous strictly decreasing function $g^i : [0, 1] \to (0, 1]$. In this representation, $\{W_i\}_{i=1}^\infty$ are the segments generated by $\{P^n\}_{n=1}^\infty$.

Moreover, the family $\{P^n\}_{n=1}^\infty$ also satisfies Marginal Consistency if and only if $g^i(d_{W_i}(x^{i-1}1)) + g^i(d_{W_i}(x^{i-1}0)) = 1$ for all i and the segment lengths satisfy $r_i \leq 2$.

Local Mean Reversion gives rise to a "local" analog of the d-Driven Mean Reversion representation (Proposition 3). While we showed in Theorem 1 that d-Driven Mean Reversion precludes nontrivial mean reversion when Marginal Consistency is imposed, there is no such tension under Local Mean Reversion. However, Marginal Consistency still imposes a sharp restriction on the representation: *segments must at most have a length of 2* (an implication of Theorem 1). The low cardinality of segments can be viewed positively as a confirmation of a potential intuition that attention is limited to a few tosses. It can alternatively be viewed negatively as too stringent to hold empirically.

Our next result makes another substantive claim about the local model. For any family $\{P^i\}_{i=1}^{\infty}$, assume that the segments $\{W_n\}_{n=1}^{\infty}$ are *non-trivial and uniformly bounded* in that (i) $|W_n| > 1$ for infinitely many n, and (ii) there exists k s.t. $|W_n| \leq k$ for all n.[13] We establish the impossibility of satisfying Mean Reversion and Local Mean reversion simultaneously in this setting.

Theorem 3. *Suppose that $\{P^n\}_{n=1}^{\infty}$ has a non-trivial and uniformly bounded set of segments $\{W_n\}_{n=1}^{\infty}$ and that it satisfies Local Mean Reversion. Then Mean Reversion must be violated.*

The intuition is that in long sequences the local sample mean is necessarily divorced from the global sample mean. The proof constructs a history x^{n-1} in which early outcomes contain a surplus of heads while later outcomes contain a surplus of tails. As a result, the global sample mean lies above θ^* and the local sample mean below θ^*. For such a history, Mean Reversion and Local Mean Reversion make opposite predictions about whether $x^{n-1}1$ or $x^{n-1}0$ is more likely, establishing the incompatibility between the two notions.

For a model-free illustration of the incompatibility between the two notions, consider the sequences $HHTHTH$ and $HHHTTT$ generated by a fair coin, $\theta^* = \frac{1}{2}$. The former sequence should be considered more likely when considering local sample means, since the sequence alternates frequently between H and T. But the latter sequence should be deemed more likely when considering the sample mean since it achieves exactly $\theta^* = \frac{1}{2}$. A take-away is that, as a theory that posits the relevance of both the sample mean and the local sample means in determining belief, the Law of Small Numbers has sharp predictions only for the small subset of pairs of sequences where there is a dominance in terms of both global and local mean reversion, such as $HHTHTH$ and $HHHHHT$. Such dominance becomes more rare with longer sequences.

[13] By Theorem 2, condition (ii) is automatically satisfied under Marginal Consistency, which requires $k = 3$.

We conclude by noting that an alternative means of capturing local mean reversion is with a model where , instead of tracking the segment mean, the agent tracks a weighted sample mean such as

$$\frac{\sum_{i=1}^{n} \gamma^{n-i} x_i}{\sum_{i=1}^{n} \gamma^{n-i}}.$$

Mean Reversion obtains as the special case where $\gamma = 1$. Such a model is different in spirit than Tversky and Kahneman (1974)'s Local Representativeness and it excludes Rabin (2002), but it is reminiscent of the history-dependent signal structure in the learning model of Rabin and Vayanos (2010). We suspect that γ is not identified in this model. We leave an analysis of such "weighted mean reversion" to future research.

4.2 Comparison with Rabin (2002)

Rabin (2002) considers an environment where binary sequences are generated from i.i.d. tosses of a coin with some bias $\theta^* \in [0, 1]$. The agent is modelled as having a misspecified belief about the data generating process. Specifically, the belief assigned to a sequence $x^n = (x_1, ..., x_n)$ of outcomes of n tosses (for odd n) is:

$$P(x^n) = P(x_1x_2) \times P(x_3x_4) \times .... \times P(x_{n-1}x_n),$$

where each $P(x_{i-1}x_i)$ is the distribution generated by an urn containing N balls with an integer $\theta^* N$ number of balls labeled "heads", from which draws are made *without* replacement.[14] Sampling without replacement from urns is the modeling tool used to generate the Gambler's Fallacy. The "i.i.d. by pairs" feature of the model makes it tractable for analysis. Due to the local exchangeablility[15] property of sampling without replacement, the model does not adequately capture a belief in alternation – for instance, for $\theta^* = \frac{1}{2}$ it implies that perfect alternation is among the most likely sequences. Moreover, for any θ^*, pairs of outcomes can be switched to generate streaks without changing the probability, such as $P(THTTHT) = P(HTTTTH)$. The model also generates a Law of Large Numbers due to the "i.i.d. by pairs" feature, thereby precluding a Non-Belief in the Law of Large Numbers.[16]

[14] Therefore the belief that the outcome of a single flip of the coin is heads is $P(H) = \theta^*$ while the belief that two flips generate a heads followed by a tails $P(HT) = \theta^* \frac{(1-\theta^*)N}{N-1}$. Since the urn is "renewed" after every two periods, the belief in HTH is $P(HTH) = P(HT)P(H) = \theta^* \frac{(1-\theta^*)N}{N-1} \times \theta^*$.

[15] Exchangeability states that for any n and $k \leq n$, all sequences of length n with k heads are deemed equally likely.

[16] The proof is as follows. For any even $n + 1$ consider the segments $\{i, i + 1\}$ for $i = 1, 3, 5, ..., n$. There are a total of $\frac{n+1}{2}$ segments and each segment generates a segment mean of $\frac{x_i + x_{i+1}}{2}$ that can

Rabin (2002)'s model is a special case of the local model in Theorem 2 when $\theta^* = \frac{1}{2}$, as it corresponds to the special case where (i) for even n the segment is binary $W^n = \{n-1, n\}$ and for odd n it is a singleton $W^n = \{n\}$, (ii) $g^i(\frac{1}{2}) = \frac{1}{2}$ for all odd i and (iii) $g^i(0) = \frac{\frac{1}{2}N}{N-1}$ and $g^i(\frac{1}{2}) = \frac{\frac{1}{2}N-1}{N-1}$ for all even i.[17] Given that this model violates Mean Reversion, it can be interpreted as modeling a disbelief in local streaks that overshadows a belief that sample proportions will match the population. In contrast, a model satisfying Mean Reversion subjugates a disbelief in local streaks to a consideration of sample proportions.

5 Bayesian Inference

In this section we study whether an agent with misspecified beliefs about randomness can learn the bias of a coin after observing an infinite sequence of i.i.d. outcomes.

5.1 Model

Let $\Theta = \{\theta_1, ..., \theta_I\}$ and let $P^n_\theta(x^n)$ denote the ex-ante probability she assigns to a sequence x^n conditional on the true bias being θ. For generality, we allow the family

only take values $\lambda = \frac{1}{2}, 1, 0$. Let $I(\frac{x_i + x_{i+1}}{2} = \lambda)$ denote the indicator function for whether segment $\{i, i+1\}$ generates a mean λ. By the Strong Law of Large Numbers applied to the i.i.d. segment mean realizations, the limit of the mean of any sequence $x \in \Omega^\infty$ is then

$$\lim_{n\to\infty} \overline{x}^{n+1} = \lim_{n\to\infty} \sum_{i=1,3,..,n} \frac{2}{n+1} \frac{x_i + x_{i+1}}{2}$$

$$= \lim_{n\to\infty} \frac{\sum_{i=1,3,..,n} I(\frac{x_i+x_{i+1}}{2} = \frac{1}{2})}{\frac{n+1}{2}} \frac{1}{2} + \frac{I(\frac{x_i+x_{i+1}}{2} = 1)}{\frac{n+1}{2}} 1 + \frac{I(\frac{x_i+x_{i+1}}{2} = 0)}{\frac{n+1}{2}} 0$$

$$= U^N_{\theta^* N}(\{HT, TH\})\frac{1}{2} + U^N_{\theta^* N}(\{HH\})1 + U^N_{\theta^* N}(\{TT\})0.$$

where $U^N_{\theta^* N}$ is the hypergeometric distribution generated by sampling without replacement from an urn with N balls of which $\theta^* N$ are labelled "heads". This limit equals $\frac{1}{2}$ for $\theta^* = \frac{1}{2}$ but in general deviates from θ^*.

[17] More generally, Rabin (2002)'s model is a special case of a slightly more general class of representations, one where $g^i(d_{W_i}(x_{k_i}, ..., x_i), x_{i-1})$ depends not just on the local mean induced by $x_{k_i}, ..., x_i$ (where k_i is defined by the segment $W^i = \{k_i, ..., i\}$) but also on the outcome of the last toss x_{i-1}. Specifically, for any $\theta^* \in (0, 1)$, it corresponds to a model where (i) for even n the relevant segment is binary $W^n = \{n-1, n\}$ and for odd n it is a singleton $W^n = \{n\}$, (ii) $g^i(d_W(H), \cdot) = \theta^*$ for all odd i and (iii) $g^i(d_{W_i}(HT), 1) = \frac{(1-\theta^*)N}{N-1}$, $g^i(d_{W_i}(TH), 0) = \frac{\theta^* N}{N-1}$, $g^i(d_{W_i}(HH), 1) = \frac{\theta^* N-1}{N-1}$ and $g^i(d_{W_i}(TT), 0) = \frac{(1-\theta^*)N-1}{N-1}$ for all even i.

$\{P^n_\theta\}_{n=1}^\infty$ to possibly fail Marginal Consistency (Section 2.1). Define $d_\theta(x^n) := |\overline{x}^n - \theta|$. Suppose that the agent's beliefs satisfy Mean Reversion (Proposition 1), and therefore admit the representation:

$$P^n_\theta(x) = \prod_{i=1}^n g^i(d_\theta(x^i), x^{i-1}).$$

We assume that g^i is continuous, strictly positive (so that P^n_θ has full support) and strictly decreasing in its first argument. Moreover, we assume g^i is independent of θ.

Suppose the parameter space Θ is finite and that the agent has a full-support prior $\mu \in \Delta(\Theta)$ over Θ. Then, her ex-ante beliefs over sequences of length n is given by

$$P^n(x^n) = \sum_\theta P^n_\theta(x^n)\mu(\theta).$$

Let $P^n(\theta|x^n)$ denote her Bayesian posterior after observing x^n:

$$P^n(\theta|x^n) = \frac{P^n_\theta(x^n)\mu(\theta)}{\sum_{\theta' \in \Theta} P^n_{\theta'}(x^n)\mu(\theta')}.$$

5.2 Results

Suppose the data is generated by Q_{θ^*} on $(\Omega^\infty, \Sigma^\infty)$ that is i.i.d. with bias θ^*. We first establish a general property of the model: the agent always places a non-vanishing probability on the true parameter.

Theorem 4. *For any learning model satisfying the assumptions in Section 5.1,*

$$\liminf_n P^n(\theta^*|x^n) > 0, \;\; Q_{\theta^*}\text{-a.s.}.$$

There is no guarantee that posteriors converge along each sequence x. Nevertheless, the theorem establishes that across all streams outside a set of measure 0, the posterior always places a non-vanishing probability on the true parameter. The reason is that the agent believes the sample mean tends to the true parameter at every point of the path, and the Law of Large Numbers ensures that the agent does not rule out the true parameter. Her misspecified belief, however, may keep her from ruling out other parameters when she sees unexpected patterns along the path.

The result stands in contrast with Rabin (2002), where the agent's beliefs always converge a.s. to a degenerate posterior, but may well be degenerate on the wrong parameter. The reason is that in Rabin's model, for any given bias different from

$\frac{1}{2}$, the agent's beliefs predict a different proportion of heads that the one implied by LLN. Therefore, in a learning context, Rabin's agent places probability zero on the true proportion of heads and her beliefs concentrate on the least implausible bias. As Rabin (2002) (with $\theta^* = \frac{1}{2}$) is a special case of Local Mean Reversion, we conclude that such mislearning is possible under Local Mean Reversion while Theorem 4 tells us it is impossible under Mean Reversion.

We provide an illustration where $P^n(\theta^*|x^n)$ converges, providing both a case where the agent learns the truth and a case where she fails to rule out some wrong parameters. We take the tractable d-Driven Mean Reversion model (Proposition 3). The requirement that g^i is strictly decreasing implies that Marginal Consistency is violated.

Theorem 5. *Consider a full support prior $\mu \in \Delta(\Theta)$ and, for each θ, a family $\{P^n_\theta\}_{n=1}^\infty$ where each P^n_θ is represented by*

$$P^n_\theta(x) = \prod_{i=1}^{n} g^i(d_\theta(x^i)),$$

and each g^i is strictly positive, strictly decreasing and continuous. Then the following statements hold.

1. If $g^i \to c$ uniformly faster than $\frac{1}{n^2} \to 0$,[18] where $c > 0$ is a constant function, then

$$0 < \lim_{n\to\infty} P^n(\theta^*|x^n) \neq 1, \; Q_{\theta^*}\text{-a.s.}$$

2. If $g^i = g$ for all $i > 1$, then

$$\lim_{n\to\infty} P^n(\theta^*|x^n) \to 1, \; Q_{\theta^*}\text{-a.s.}$$

Claim (i) assumes that g^i approaches a constant g "fast enough". A constant g corresponds to a belief that the bias towards heads equals the bias towards tails after any history: $g^i(d(x^{i-1}1)) = g^i(d(x^{i-1}0))$. Thus, as g^i approaches a constant function, the agent believes that, for every θ, the coin becomes less self-correcting with n, that is, the belief in Mean Reversion weakens with n. As a result, the progression of the sample mean is viewed as less informative about the true parameter, and the posteriors correspondingly become less responsive to the sample mean as n grows. Indeed, posteriors may be critically shaped by what she sees *early* in any sequence x^n. The result states that the agent's posterior beliefs will not converge to a degenerate

[18]That is, there exists N such that for all $n > N$, $|g^n(a,\theta) - c| < \frac{1}{n^2}$ for all a, θ in the support of g^n.

distribution almost surely. In line with Theorem 4, in the limit the agent places strictly positive probability on the true parameter θ^*.

Claim (ii) assumes that $g^i = g$ does not change with i and so the agent maintains a consistent degree of belief in Mean Reversion. In this case the agent eventually learns the true parameter θ^*.

6 Related Literature

The literature on intuitive likelihood judgments investigates both static beliefs (properties of priors) and dynamic beliefs (properties of updating). While a misspecified prior is necessary to explain the static evidence, it is also sufficient to explain the dynamic evidence using Bayesian updating: if an agent believes ex-ante that HHT is more likely than HHH, then Bayesian updating leads them to believe that T is more likely than H conditional on HH. Consequently, this paper lies outside the literature on non-Bayesian updating (see for instance Epstein, Noor and Sandroni (2008, 2010)). It intersects with the literature on learning with misspecified beliefs (spawned by Berk (1966)) in that (i) the agent has a wrong model of the world and (ii) we have an application to learning (Section 5).

As we have noted already, the seminal paper on the Law of Small Numbers in economics is due to Rabin (2002). We showed that the model intersects with our local model, but is otherwise distinct from our main model. A summary of differences in predictions are as follows: Rabin (2002) implies the Gambler's Fallacy and an extreme version of Belief in Alternation (it permits perfectly alternating sequences to be the most likely sequences), but does not produce Non-Belief in the Law of Large Numbers. Our model produces the Gambler's Fallacy and a Belief in Alternation and is consistent with a Non-Belief in the Law of Large Numbers. In a learning setting, Rabin's model implies the possibility of learning the wrong parameter, whereas our agent never rules out the true parameter.

There has been little theoretical work of the Law of Small Numbers in economics since the seminal work of Rabin (2002).[19] In order to further study how inference with Gambler's Fallacy can lead to the Hot Hand Fallacy, Rabin and Vayanos (2010) study an alternative model where the agent receives a sequence of noisy real-valued signals $s_n = \theta + \epsilon_n$ about a state θ but mistakenly believes that the errors ϵ_n are not i.i.d. and instead exhibit reversals as per the Gambler's Fallacy. Formally, the

[19]See He (2022) for a relatively recent application of the Gambler's Fallacy.

process ϵ_n is modeled as

$$\epsilon_n = \omega_n - \alpha \sum_{i=1}^{n} \gamma^i \epsilon_{n-1-i}$$

where ω_n is i.i.d. normal. Intuitively, a greater number of recent positive realizations of the error make it more likely that the next realization will be negative. This has some flavor of Local Mean Reversion because the highest weights are on recent outcomes. In our model, beliefs are a deterministic function of past outcomes, whereas in Rabin and Vayanos (2010) the stochasticity of ω_n introduces an addition layer of randomness.

Motivated by Sample Size Neglect, Benjamin, Rabin and Raymond (2016) hypothesize that people's beliefs may not respect the Law of Large Numbers and in particular may believe in a sampling distribution for large samples that has more spread than it should. They write an exchangeable model where the agent believes that the outcome of a coin toss is generated by an i.i.d. stochastic bias θ that has the true bias θ^* as its mean.[20] Their model produces Sample Size Neglect for large samples, which corresponds to what they refer to as a Non-Belief in the Law of Large Numbers.

Benjamin, Rabin and Raymond (2016) speak to Sample Size Neglect, while Rabin (2002) and Rabin and Vayanos (2010) speak to the Gambler's Fallacy and Excessive Alternation. A paper that connects this evidence is Noor (2022), which models the formation of intuitive beliefs by representing the agent's beliefs as a neural network of associations that is trained by her "experience". Noor (2022) shows that if the agent's experience is defined by the distribution of sample means generated by the environment, then properties of large-sample distributions are reflected in the agent's beliefs regarding small samples. As a result the model exhibits the Gambler's Fallacy and Sample Size Neglect.

Part of the psychology literature interprets the Gambler's Fallacy in terms of a belief in a *switching rate* that is higher than 50% (Rapoport and Budescu (1997), Bar-Hillel and Wagenaar (1991)). Rapoport and Budescu (1997) informally describe a model, the essence of which is arguably captured in our following formalization of it: presuming that the bias of the coin is perceived to be $\theta^* = \frac{1}{2}$,

$$P^n(x) = \frac{1}{2} \times \prod_{i=1}^{n} \theta^{|x_i - x_{i-1}|}(1-\theta)^{1-|x_i - x_{i-1}|},$$

where the probability of a switch on the i^{th} toss is $\theta > \frac{1}{2}$.[21] The i^{th} outcome is

[20] The counterexample constructed in our Theorem 7 also takes this form.

[21] Similar to Rabin (2002), Rapoport and Budescu (1997)'s model switching probability may

"rewarded" (in the sense of being attributed a higher belief) if $|x_i - x_{i-1}| > 0$, that is, if it differs from the outcome in the previous toss. Mean Reversion predicts that

$$P^8(HHHHHHTH) < P^8(HHHHHHTT),$$

while the model requires the reverse ranking because the former sequence contains more switches than the latter. Given $\theta^* = \frac{1}{2}$, the model satisfies Relative Local Mean Reversion where the relevant segment at each $n \geq 2$ has a fixed length of 2.

A Proofs for Propositions 1, 2 and 3

Propositions 1 and 2 are corollaries of the following two lemmas.

Lemma 1. *Any family of full support beliefs $\{P^n\}_{n=1}^{\infty}$ can be written as*

$$P^n(x^n) = \prod_{i=1}^{n} (\theta_{x^{i-1}})^{x_i} (\gamma_{x^{i-1}})^{1-x_i}.$$

where $\theta_{x^{i-1}} = P^i(1|x^{i-1})$ and $\gamma_{x^{i-1}} = P^i(0|x^{i-1})$. Moreover, $\{P^n\}_{n=1}^{\infty}$ satisfies Marginal Consistency iff $\gamma_{x^{i-1}} = 1 - \theta_{x^{i-1}}$.

Proof. Since beliefs have full support we can always write $P^n(x^n) = P^1(x^1) \prod_{i=2}^{n} \frac{P^i(x^i)}{P^{i-1}(x^{i-1})}$. The first assertion in the lemma follows. The second assertion is established by noting that Marginal Consistency is equivalent to the requirement that the conditional probabilities of outcomes sum to 1. This follows by observing that, for any x^n,

$$P^{n+1}(1|x^n) + P^{n+1}(0|x^n) = \frac{P^{n+1}(x^n 1) + P^{n+1}(x^n 0)}{P^n(x^n)} = \frac{P^{n+1}(x^n)}{P^n(x^n)},$$

and therefore $\frac{P^{n+1}(x^n)}{P^n(x^n)} = 1$ iff $P^{n+1}(1|x^n) + P^{n+1}(0|x^n) = 1$. □

Lemma 2. *A family of full support beliefs $\{P^n\}_{n=1}^{\infty}$ satisfies Mean Reversion iff there exists a family of continuous functions $\{g^i : [0,1] \times \Omega^{i-1} \to (0,1]\}_{i=1}^{\infty}$ such that g^i is continuous and weakly decreasing in its first argument, and for all n and $x^n \in \Omega^n$,*

$$P^n(x^n) = \prod_{i=1}^{n} (\theta_{x^{i-1}})^{x_i} (\gamma_{x^{i-1}})^{1-x_i},$$

depend on whether n is even or odd and is characterized by a parameter m that governs the length of throws in which the agent behaves in the standard way. The model we are describing corresponds to their model when $m = 1$.

where $\theta_{x^{i-1}} := g^i(d(x^{i-1}1), x^{i-1})$ and $\gamma_{x^{i-1}} := g^i(d(x^{i-1}0), x^{i-1})$. Moreover,

(i) $g^n(d(x^n), x^{n-1}) = P^n(x_n|x^{n-1})$ for each x^n, and

(ii) $\{P^n\}_{n=1}^{\infty}$ satisfies Marginal Consistency iff $\gamma_{x^{i-1}} = 1 - \theta_{x^{i-1}}$.

Proof. Mean Reversion implies that for any x^{n-1}, conditional beliefs $P^n(x_n|x^{n-1})$ depend on x_n only through $d(x^n)$. Therefore we can write $g^n(d(x^n), x^{n-1}) = P^n(x_n|x^{n-1})$ and insert this in the representation in Lemma 1 to obtain a representation where $g^n(\cdot, x^{n-1})$ is defined only on a binary domain $\{d(x^{n-1}1), d(x^{n-1}0)\}$. Since $d(x^{n-1}x_n) \leq d(x^{n-1}y_n) \implies g^n(d(x^{n-1}x_n), x^{n-1}) \geq g^n(d(x^{n-1}y_n), x^{n-1})$, it must be that $g^n(\cdot, x^{n-1})$ is weakly decreasing. By the full support assumption and minimal consistency requirement on $\{P^n\}_{n=1}^{\infty}$ (that is, $P^n(x_n|x^{n-1}) \leq 1$ for all x^n) it must be that g^n takes values in $(0, 1]$. The function can be extended to $[0, 1]$ continuously so that it is weakly decreasing.

The necessity of Mean Reversion follows from the fact that $g^n(\cdot, x^{n-1})$ is decreasing and the representation implies

$$
\begin{aligned}
g^n(d(x^n), x^{n-1}) &= \frac{g^n(d(x^{n-1}x_n), x^{n-1}) \prod_{i=1}^{n-1} g^i(d(x^i), x^{i-1})}{\prod_{i=1}^{n-1} g^i(d(x^i), x^{i-1})} \\
&= \frac{P^n(x^{n-1}x_n)}{P^{n-1}(x^{n-1})} = P^n(x_n|x^{n-1}).
\end{aligned}
$$

This also establishes claim (i). Claim (ii) follows from Lemma 1. □

Necessity in Proposition 3 is straightforward. Sufficiency follows from the following lemma.

Lemma 3. *$\{P^n\}_{n=1}^{\infty}$ satisfies d-Driven Mean Reversion iff it admits a Mean Reversion representation (Lemma 2) with a family of continuous functions $\{g^i : [0, 1] \to (0, 1]\}_{i=1}^{\infty}$ such that g^i is weakly decreasing.*

Proof. Define $E^n = \{d(x^n) \in [0, 1] : x^n \in \Omega^n\}$. d-Driven Mean Reversion implies that conditional beliefs satisfy: for all $n > 0$ and any $x^n, y^n \in \Omega^n$,

$$
d(x^n) \geq d(y^n) \implies \frac{P^n(x^n)}{P^{n-1}(x^{n-1})} \leq \frac{P^n(y^n)}{P^{n-1}(y^{n-1})}.
$$

In particular, since $[d(x^n) = d(y^n)$ implies $\frac{P^n(x^n)}{P^{n-1}(x^{n-1})} = \frac{P^n(y^n)}{P^{n-1}(y^{n-1})}]$ and since beliefs have full support and satisfy the minimal consistency requirement, for each $n > 1$ there exists a function $g^n : E^n \to (0, 1]$ s.t. $\frac{P^n(x^n)}{P^{n-1}(x^{n-1})} = g^n(d(x^n))$. Moreover, the axiom imposes that $d(x^n) \geq d(y^n)$ implies $g^n(d(x^n)) \leq g^n(d(y^n))$, establishing that g^n is weakly decreasing on E^n. It is immediate that g^n can be extended to a weakly decreasing continuous function on $[0, 1]$. The remaining construction of the representation is as in the representation for Mean Reversion. □

B Proof of Theorem 1

Proof. Consider the case where $\theta^* \geq \frac{1}{2}$ (if $\theta^* \leq \frac{1}{2}$ then a similar argument holds where heads and tails are switched) and take any $i \geq 3$. Marginal Consistency implies that for any history x^{i-1} with number of heads $k = \sum x^{i-1}$,

$$1 = P^i(0|x^{i-1}) + P^i(1|x^{i-1}) = g^i(|\frac{k}{i} - \theta^*|) + g^i(|\frac{k+1}{i} - \theta^*|).$$

In particular, for any history x^{i-1} and $k = \sum x^{i-1}$,

$$g^i(|\frac{k+1}{i} - \theta^*|) = 1 - g^i(|\frac{k}{i} - \theta^*|).$$

It is then evident that if we establish that $g^i(|\frac{k}{i} - \theta^*|) = \frac{1}{2}$ for $k = 0$, then an inductive argument implies that $g^i(|\frac{k}{i} - \theta^*|) = \frac{1}{2}$ for all $k \leq i-1$.

To prove this, consider streaks of 1's and 0's. Observe that under Marginal Consistency,

$$g^i(\theta^*) + g^i(\theta^* - \frac{1}{i}) = P^i(0|0^{i-1}) + P^i(1|0^{i-1})$$

$$= 1 = P^i(0|1^{i-1}) + P^i(1|1^{i-1}) = g^i(|\frac{i-1}{i} - \theta^*|) + g^i(1 - \theta^*),$$

and in particular

$$g^i(\theta^*) + g^i(\theta^* - \frac{1}{i}) = g^i(|\frac{i-1}{i} - \theta^*|) + g^i(1 - \theta^*)$$

Note that $\theta^* \geq \frac{1}{2}$ and $i \geq 3$ implies $\theta^* \geq 1 - \theta^*$ and $\theta^* - \frac{1}{i} \geq |\frac{i-1}{i} - \theta^*|$. Thus, the fact that g^i is decreasing implies that these equalities can hold only if $g^i(\theta^*) = g^i(\theta^* - \frac{1}{i}) = g^i(|\frac{i-1}{i} - \theta^*|) = g^i(1 - \theta^*) = \frac{1}{2}$. In particular we obtain $g^i(|\frac{0}{i} - \theta^*|) = g^i(\theta^*) = \frac{1}{2}$, as desired. □

C Proof of Proposition 4

Proof. Fix x^n and WLOG let $\omega = H$. Let N be the smallest integer such that $d(x^n \underbrace{H...H}_{N-n} H) > d(x^n \underbrace{H...H}_{N-n} T)$ and let $M = N - n$. Then, for any $m > M$ we have that $d(x^n H^m H) > d(x^n H^m T)$ and since g^n is strictly decreasing, $P^n(x^n H^m H) < P^n(x^n H^m T)$. □

D Proof of Proposition 5

Proof. (a) Fix $n \geq 2$ and consider a constant sequence $H^{n-1}H$. Then $d(H^{n-1}H) > d(H^{n-1}T)$ and since g^n is strictly decreasing we have that $P^n(H^{n-1}H) < P^n(H^{n-1}T)$. The proof for the constant sequence $T^{n-1}T$ is analogous.

(b) Consider a special case of our model where the intensity of belief in mean reversion is stronger at histories for which the sample mean has deviated farther away from θ^*. This can be captured the following "Increasing Mean Reversion" specification of our general model, where the second argument of g^i depends on history x^{i-1} through $d(x^{i-1})$:

$$P^n(x^n) = \prod_{i=1}^{n} g^i(d(x^i), d(x^{i-1})),$$

where $g^i : [0,1] \times [0,1] \to (0,1)$ is continuous, strictly decreasing in the first argument and strictly increasing in the second. Continue to suppose that $\theta^* = \frac{1}{2}$. This model can generate, for instance,

$$P^3(HTH) < P^3(HHT).$$

To see this, choose $g^1 = \frac{1}{2}$ so that $P^1(H) = P^1(T)$. In the second coin toss, since the distance is the same after both histories H and T, the model implies that $P^2(HT) = P^2(TH) > P^2(HH) = P^2(TT)$, regardless of g^2. For the third coin toss, given that $P^3(HTH) = g^1(0.5)g^2(0, 0.5)g^3(\frac{1}{6}, 0)$ and $P^3(HHT) = g^1(0.5)g^2(0.5, 0.5)g^3(\frac{1}{6}, 0.5)$, we have

$$\frac{P^3(HTH)}{P^3(HHT)} = \frac{g^2(0, 0.5)}{g^2(0.5, 0.5)} \frac{g^3(\frac{1}{6}, 0)}{g^3(\frac{1}{6}, 0.5)}.$$

By the monotonicity properties of g^i we have $\frac{g^2(0,0.5)}{g^2(0.5,0.5)} > 1$ and $\frac{g^3(\frac{1}{6},0)}{g^3(\frac{1}{6},0.5)} < 1$, and so with an appropriate choice of g^i we obtain $P^3(HTH) < P^3(HHT)$. □

E Proof of Proposition 6

Proof. Fix $n \geq 2$ and let $x^{*n} \in \Omega^n$ be a perfectly alternating sequence of length n. Consider the following representation of P^n:

$$P^n(x^n) = \prod_{i=1}^{n} f^i(d(x^1), \ldots, d(x^i)),$$

where each f^i is strictly decreasing in each of its arguments. First we note that P^n satisfies a "Mean Reversion on Path" property: for any $x^n, y^n \in \Omega^n$,

$$d(x^i) \leq d(y^i) \text{ for all } i \leq n \text{ , with strict inequality for some } i$$
$$\implies P^n(x^n) > P^n(y^n).$$

That is, if x^n induces a smaller distance to θ^* than y^n at every toss, then x^n is deemed more likely. This follows from the fact that f^i is strictly decreasing in each of its arguments.

Next, observe that when $\theta^* = \frac{1}{2}$, a perfectly alternating sequence x^{*n} has the property that $d(x^{*i}) = \min_{x^i \in \Omega^i} d(x^i)$ for every $i \leq n$. Now take any $y^n \in \Omega^n$, then either there exists $i \leq n$ such that $d(x^{*i}) < d(y^i)$, or $d(x^{*i}) = d(y^i)$ for all $i \leq n$. In the first case, Mean Reversion on Path implies that $P^n(x^{*n}) > P^n(y^n)$. In the second, by assumption, $P^n(x^n)$ depends only on the distance path $(d(x^1), \ldots, d(x^n))$ and thus, $P^n(x^{*n}) = P^n(y^n)$. Combining the two cases, we have $P^n(x^{*n}) \geq P^n(y^n)$ for every y^n which yields the desired result. □

F Proof of Proposition 7

Proof. Let $\theta^* \neq \frac{1}{2}$ and consider a family of probability distributions defined by the following conditionals:

$$R^n(x_n|x^{n-1}) = \begin{cases} 1 & d(x^n) < d(x^{n-1}(1-x_n)) \\ 0 & d(x^n) > d(x^{n-1}(1-x_n)) \\ \frac{1}{2} & d(x^n) = d(x^{n-1}(1-x_n)) \end{cases} .$$

We first verify that the family satisfies Marginal Consistency. For any x^{n-1} consider three cases, and show that in each case the conditionals sum to 1:

(1) $d(x^{n-1}1) < d(x^{n-1}0) \implies R^n(1|x^{n-1}) + R^n(0|x^{n-1}) = 1 + 0$

(2) $d(x^{n-1}1) > d(x^{n-1}0) \implies R^n(1|x^{n-1}) + R^n(0|x^{n-1}) = 0 + 1$

(3) $d(x^{n-1}1) = d(x^{n-1}0) \implies R^n(1|x^{n-1}) + R^n(0|x^{n-1}) = \frac{1}{2} + \frac{1}{2}$.

Hence, Marginal Consistency holds, and the family is the family of marginals generated by some R over Ω^∞. We interpret an agent described by R as displaying a belief in deterministic mean reversion.

Let $Q_{\frac{1}{2}}$ denote the i.i.d. measure with bias $\frac{1}{2}$, and for any $\alpha \in (0,1)$, define the mixture measure P over Ω^∞ by

$$P = \alpha R + (1-\alpha) Q_{\frac{1}{2}}.$$

We need to show that the family of marginals $\{P^n\}_{n=1}^{\infty}$ generated by P (which by construction satisfies Marginal Consistency) satisfies Mean Reversion. But this follows readily from the fact that both R and $Q_{\frac{1}{2}}$ generate marginals that satisfy Mean Reversion (in the case of $Q_{\frac{1}{2}}$, Mean Reversion is satisfied with equality at each history).

To prove the result, recall that P is a mixture model: first a hypothetical coin with bias α is thrown, and if the coin comes up heads, then R is the data generating process, and if the coin comes up tails, then $Q_{\frac{1}{2}}$ is the data generating process. Thus, for any n, the distribution of the sample mean generated by P is the α-mixture of the corresponding distributions generated by R and $Q_{\frac{1}{2}}$. By a standard LLN, the distribution of the sample mean generated by $Q_{\frac{1}{2}}$ converges to the degenerate distribution on $\frac{1}{2}$ as $n \to \infty$. Since $\theta^* \neq \frac{1}{2}$, even without verifying what happens to the corresponding distribution generated by R,[22] this already implies that the corresponding distribution generated by P does not converge to a distribution that puts probability 1 on θ^*, that is, $\lim_{n\to\infty} P^n(\overline{x}^n = \theta^*) \neq 1$, as desired. □

G Proof of Theorem 2

Proof. Necessity is straightforward. To prove sufficiency, we proceed in steps.

Step 1: Show that if a family of full support beliefs $\{P^n\}_{n=1}^{\infty}$ satisfies Local Mean Reversion with segments $\{W_i\}_{i=1}^{\infty}$ then for each $i \geq 1$ there exists $g^i : [0,1] \to (0,1]$ that is continuous and strictly decreasing and for any n and $x^n \in \Omega^n$,

$$P^n(x^n) = \prod_{i=1}^{n} (\theta_{x^{i-1}})^{x_i} (\gamma_{x^{i-1}})^{1-x_i},$$

where $\theta_{x^{i-1}} = g^i(d_{W_i}(x^{i-1}1))$ and $\gamma_{x^{i-1}} = g^i(d_{W_i}(x^{i-1}0))$.

Local Mean Reversion requires that for any x^i and y^i, $d_{W_i}(x^i) = d_{W_i}(y^i)$ if and only if $\frac{P^i(x^{i-1}1)}{P^{i-1}(x^{i-1})} = \frac{P^i(y^{i-1}1)}{P^{i-1}(y^{i-1})}$ and $d_{W_i}(x^i) < d_{W_i}(y^i)$ if and only if $\frac{P^i(x^{i-1}1)}{P^{i-1}(x^{i-1})} < \frac{P^i(y^{i-1}1)}{P^{i-1}(y^{i-1})}$. Therefore there exists a function $g^i(d_{W_i}(x^i)) = \frac{P^i(x^{i-1}1)}{P^{i-1}(x^{i-1})}$ that is strictly decreasing. Use Lemma 1 to obtain the desired representation.

Step 2: Show that if a family of full support beliefs $\{P^n\}_{n=1}^{\infty}$ satisfies Local Mean Reversion with segments $\{W_i\}_{i=1}^{\infty}$ and is represented by $\{g^n : [0,1] \to (0,1]\}_{n=1}^{\infty}$ as in

[22] We expect that the distribution converges to the distribution that is degenerate at θ^*. Note that when θ^* is an irrational number (in which case $d(x^n) = d(x^{n-1}(1-x_n))$ never holds at any x^n), R generates a deterministic sequence that corresponds to the binary expansion of θ^*, and the distribution of the sample mean clearly converges to θ^*.

Step 1 where $\theta_{x^{i-1}} = g^i(d_{V_i}(x^{i-1}1))$ and $\gamma_{x^{i-1}} = g^i(d_{V_i}(x^{i-1}0))$ and $V_i = \{i - k_n, .., i\}$ is a continuous set of indices containing i, then $V_i = W_i$ for each i.

Consider $\{V_i\}_{i=1}^{\infty}$ used in the representation. We need to show that these satisfy the definition of segments. Contiguity is satisfied by definition. Take any n and let $k_n := |V_n|$. Then for all x, y,

$$\frac{P^n(x_1...x_{k_n-1}, x_{k_n}...x_{n-1}x_n)}{P^{n-1}(x_1...x_{k_n-1}, x_{k_n}...x_{n-1})} = g^n(d_{V_n}(x_1...x_{k_n-1}, x_{k_n}...x_{n-1}x_n))$$

$$= g^n(d_{V_n}(y_1...y_{k_n-1}, x_{k_n}...x_{n-1}x_n)) = \frac{P^n(y_1...y_{k_n-1}, x_{k_n}...x_{n-1}x_n)}{P^{n-1}(y_1...y_{k_n-1}, x_{k_n}...x_{n-1})},$$

establishing that outcomes on tosses outside V_n do not matter. It remains to show that there is no $k < k_n$ for which this equality holds for all x, y. Suppose $\theta^* \geq \frac{1}{2}$. Since g^n is strictly decreasing, it must be that a sequence ending in a k_n-segment $T_{n-k_n}, ..., T_n$ generates a strictly larger g^n value than a sequence that has a heads replacing any one of the tails in this segment. But this violates the noted equality. The argument for $\theta^* \leq \frac{1}{2}$ is analogous.

Step 3: If Marginal Consistency and Local Mean Reversion hold, then $|W_i| \leq 2$.

For Marginal Consistency to hold, it has to be the case that $g^n(d_{W_n}(x^{n-1}1)) + g^n(d_{W_n}(x^{n-1}0)) = 1$ for each x^{n-1}. Theorem 1 implies that if $|W_n| \geq 3$, then $g^n = \frac{1}{2}$ regardless of x^{n-1} for all $n \geq 3$. Hence, if there is n s.t. $|W_n| \geq 3$ then g^n cannot be strictly decreasing. This contradicts Local Mean Reversion. □

H Proof of Theorem 3

Proof. Suppose by way of contradiction that Mean Reversion holds. By uniform boundedness, the segment length is bounded by some k. Suppose $\theta^* \geq \frac{1}{2}$ (the case where $\theta^* \leq \frac{1}{2}$ is handled by switching heads and tails in the sequences used in the argument below). Observe that $\frac{n-k}{n} - \theta^* \nearrow 1 - \theta^*$ and so there is some N s.t. $\frac{n-k}{n} - \theta^* > 0$ for all $n \geq N$.

Take any $n \geq N$ such that $|W_n| > 1$ and denote $|W_n| = m \leq k$. Consider $x^n = (H, ..., H, T_{n-m}...T_{n-1}, T_n)$ and $y^n = (H, ..., H, T_{n-m}...T_{n-1}H_n)$. By construction, the local mean at n is closer to θ^* in the latter for any $m \leq k$. To verify this first note due to the common history before toss $n - k$, we have $d_{W_i}(y^i) = d_{W_i}(x^i)$ for all $i < n$. At the final toss, we have $\overline{x}^n(W_n) = 0$ and $\overline{y}^n(W_n) = \frac{1}{m} > 0$ and since $\theta^* \geq \frac{1}{2}$ it follows that $d_{W_n}(y^n) < d_{W_n}(x^n)$. The local model therefore implies that $P^n(x^n) < P^n(y^n)$. However, it must also be that $d(y^n) > d(x^n)$. To see this, first note that since $m \leq k$ we have $\frac{n-m}{n} - \theta^* \geq \frac{n-k}{n} - \theta^* > 0$. Therefore $d(x^n) = \frac{n-m}{n} - \theta^*$

and indeed $d(y^n) = \frac{n-m+1}{n} - \theta^*$ as well. But then $d(x^n) = \frac{n-m}{n} - \theta^* \geq \frac{n-k}{n} - \theta^* > 0$ and $d(y^n) = \frac{n-m+1}{n} - \theta^* = \frac{n-m}{n} - \theta^* + \frac{1}{n} > \frac{n-m}{n} - \theta^* = d(x^n)$, that is, $d(y^n) > d(x^n)$ as desired. Given the common history before toss n, we obtain a contradiction to Mean Reversion. □

I Proof of Theorems 4 and 5

Lemma 4. *Assume $\mu \in \Delta(\Theta)$ and each $P^n \in \Delta(\Omega^n)$ have full support. Then,*

$$\liminf_n P^n(\theta^*|x^n) > 0 \ Q_{\theta^*}\text{-a.s.}$$

Proof. Fix any sequence such that $\lim_n \overline{x}^n = \theta^*$ and let x^{n_k} denote a subsequence that converges to lim inf of $P^n(\theta^*|x^n)$:

$$\lim_{k\to\infty} P^{n_k}_{\theta^*}(x^{n_k}) = \liminf_n P^n(\theta^*|x^n).$$

For any $\theta \neq \theta^*$, this subsequence generates a sequence $\{P^{n_k}_\theta(x^{n_k})\}$ in $[0,1]$, and a further subsequence must lead to convergence of $P^{n_k}_\theta(x^{n_k})$. Since there are finitely many θ, we can wlog suppose that $P^{n_k}_\theta(x^{n_k})$ are convergent for all $\theta \in \Theta$. Due to the full support assumption, it must be that $\sum_{\theta\in\Theta} P^{n_k}_\theta(x^{n_k})\mu(\theta) > 0$ and in particular the posteriors $P^{n_k}(\theta|x^{n_k})$ are well-defined.

Suppose by way of contradiction that $\liminf_n P^n(\theta^*|x^n) = 0$. Thus $P^{n_k}_{\theta^*}(x^{n_k}) \to 0$. It cannot be that $P^{n_k}_\theta(x^{n_k}) \to 0$ for all $\theta \in \Theta$, otherwise we obtain the contradiction that $1 = \sum_{\theta\in\Theta} P^n(\theta|x^n) \to 0$. Let $\theta \in \Theta$ be such that $\lim_{k\to\infty} P^{n_k}_\theta(x^{n_k}) > 0$ and consider the likelihood ratio of θ and θ^*,

$$\frac{P^n_\theta(x^n)}{P^n_{\theta^*}(x^n)} = \frac{\prod_{i=1}^n g^i(|\overline{x}^i - \theta|, x^{i-1})}{\prod_{i=1}^n g^i(|\overline{x}^i - \theta^*|, x^{i-1})} = \prod_{i=1}^n \frac{g^i(|\overline{x}^i - \theta|, x^{i-1})}{g^i(|\overline{x}^i - \theta^*|, x^{i-1})}.$$

By the Law of Large Numbers, $Q_{\theta^*}(x^\infty | \lim_{n\to\infty} \overline{x}^n = \theta^*) = 1$. Hence, it is enough to consider such sequences. Fix $\epsilon = \min_{\theta\neq\theta'} |\theta - \theta'|$ and $x \in \Omega^\infty$ such that $\lim_{n\to\infty} \overline{x}^n = \theta^*$. Let N be such that for all $n > N$, $|\overline{x}^n - \theta^*| < \frac{\epsilon}{4}$. Then, $|\overline{x}^n - \theta| > \frac{\epsilon}{2}$. Further,

$$\prod_{i=1}^n \frac{g^i(|\overline{x}^i - \theta|, x^{i-1})}{g^i(|\overline{x}^i - \theta^*|, x^{i-1})} = \prod_{i=1}^{N-1} \frac{g^i(|\overline{x}^i - \theta|, x^{i-1})}{g^i(|\overline{x}^i - \theta^*|, x^{i-1})} \times \prod_{i=N}^n \frac{g^i(|\overline{x}^i - \theta|, x^{i-1})}{g^i(|\overline{x}^i - \theta^*|, x^{i-1})}.$$

Because g^i is weakly decreasing in its first argument, and since $|\overline{x}^n - \theta^*| < \frac{\epsilon}{4}$ and $|\overline{x}^n - \theta| > \frac{\epsilon}{2}$, then

$$g^i(|\overline{x}^i - \theta^*|, x^{i-1}) \geq g^i(|\overline{x}^i - \theta|, x^{i-1})$$

for all $i > N$. Hence,

$$\lim_{n_k \to \infty} \frac{P_\theta^{n_k}(x^{n_k})}{P_{\theta^*}^{n_k}(x^{n_k})} \leq \prod_{i=1}^{N-1} \frac{g^i(|\overline{x}^i - \theta|, x^{i-1})}{g^i(|\overline{x}^i - \theta^*|, x^{i-1})}$$

which contradicts the hypothesis that $P_{\theta^*}^{n_k}(x^{n_k}) \to 0$ and in particular contradicts $\liminf_n P^n(\theta^*|x^n) = 0$. □

Theorem 5 is a corollary of the following lemma, which uses the model with $g^i(d(x^i), \overline{x}^{i-1})$ rather than $g^i(d(x^i))$.

Lemma 5. *Suppose $\mu \in \Delta(\Theta)$ and each $P^n \in \Delta(\Omega^n)$ have full support. Assume the representation*

$$P_\theta^n(x) = \prod_{i=1}^{n} g^i(d_\theta(x^i), \overline{x}^{i-1})$$

where $g^i(\cdot, \overline{x}^{i-1})$ is continuous and is strictly decreasing in its first argument for each i.

1. If $g^i \to c > 0$ uniformly faster than $\frac{1}{n^2} \to 0$ for all θ, where c is a constant function, then

$$0 < \lim_{n \to \infty} P^n(\theta^*|x^n) \neq 1 \ Q_{\theta^*}\text{-a.s.}$$

2. If $g^i = g$ for all $i > 1$, then $Q_{\theta^*}(\lim_{n\to\infty} P^n(\theta^*|x^n) \to 1) = 1$, that is,

$$\lim_{n \to \infty} P^n(\theta^*|x^n) \to 1 \ Q_{\theta^*}\text{-a.s.}$$

Proof. Because we are only considering finitely many θ's, it is enough to show that $\frac{P_\theta^n(x^n)}{P_{\theta^*}^n(x^n)} \to 0$ Q_{θ^*}-a.s. for all $\theta \in \Theta \backslash \theta^*$.

By an identical argument to the one in Lemma 4, for $\epsilon = \min_{\theta \in \Theta\backslash\{\theta^*\}} |\theta - \theta^*|$, there exists N such that for all $i > N$, $|\overline{x^i} - \theta^*| < \frac{\epsilon}{4}$, and

$$\frac{P_\theta^n(x^n)}{P_{\theta^*}^n(x^n)} = \prod_{i=1}^{N-1} \frac{g^i(|\overline{x}^i - \theta|, \overline{x}^{i-1})}{g^i(|\overline{x}^i - \theta^*|, \overline{x}^{i-1})} \times \prod_{i=N}^{n} \frac{g^i(|\overline{x}^i - \theta|, \overline{x}^{i-1})}{g^i(|\overline{x}^i - \theta^*|, \overline{x}^{i-1})}.$$

Further, because g^i is strictly decreasing in its first argument, and since $|\overline{x}^n - \theta^*| < \frac{\epsilon}{4}$ and $|\overline{x}^n - \theta| > \frac{\epsilon}{2}$,

$$\prod_{i=1}^{N-1} \frac{g^i(|\overline{x}^i - \theta|, \overline{x}^{i-1})}{g^i(|\overline{x}^i - \theta^*|, \overline{x}^{i-1})} \times \prod_{i=N}^{n} \frac{g^i(|\overline{x}^i - \theta|, \overline{x}^{i-1})}{g^i(|\overline{x}^i - \theta^*|, \overline{x}^{i-1})} < \prod_{i=1}^{N-1} \frac{g^i(|\overline{x}^i - \theta|, \overline{x}^{i-1})}{g^i(|\overline{x}^i - \theta^*|, \overline{x}^{i-1})} \times \prod_{i=N}^{n} \frac{g^i(\frac{\epsilon}{2}, \overline{x}^{i-1})}{g^i(\frac{\epsilon}{4}, \overline{x}^{i-1})}.$$

Notice that $\frac{g^i(\frac{\epsilon}{2},\overline{x}^{i-1})}{g^i(\frac{\epsilon}{4},\overline{x}^{i-1})} < 1$ for all i and the term $\prod_{i=1}^{N-1} \frac{g^i(|\overline{x}^i-\theta|,\overline{x}^{i-1})}{g^i(|\overline{x}^i-\theta^*|,\overline{x}^{i-1})}$ does not depend on n.

Step 1: Establish the result under the first assumption in the lemma.

Assume that $g^n \to c > 0$ uniformly faster than $\frac{1}{n^2} \to 0$. Then there exists K such that $\inf_{n>K} g^n > 0$ for all $n > K$. Moreover, for any $a, b \in [0,1]$ and sample mean θ such that $a > b > 0$, it must be that for all $n > K$, and

$$|\frac{g^n(a,\theta)}{g^n(b,\theta)} - 1| = |\frac{g^n(a,\theta) - g^n(b,\theta)}{g^n(b,\theta)}| < |\frac{g^n(a,\theta) - c}{g^n(b,\theta)}| + |\frac{g^n(b,\theta) - c}{g^n(b,\theta)}|$$

$$< \frac{1}{n^2}\frac{2}{g^n(b,\theta)} < \frac{k}{n^2}$$

for some constant $k = \frac{2}{\inf_{n>K} g^n}$. In fact $k > 1$ since $g^n \leq 1$. Hence, for all $n > K$,

$$\frac{g^n(a,\theta)}{g^n(b,\theta)} < 1 + \frac{k}{n^2}.$$

Fix any $x \in \Omega^\infty$ and consider

$$\lim_{n\to\infty} \frac{P^{\theta^*}(x^n)}{P^{\theta}(x^n)} = \lim_{n\to\infty} \prod_{i=1}^{n} \frac{g^i(|\overline{x}^i - \theta^*|, \overline{x}^{i-1})}{g^i(|\overline{x}^i - \theta|, \overline{x}^{i-1})} = \prod_{i=1}^{\infty} \frac{g^i(|\overline{x}^i - \theta^*|, \overline{x}^{i-1})}{g^i(|\overline{x}^i - \theta|, \overline{x}^{i-1})}.$$

This product exists if and only if there exists N such that for all $m > N$,

$$\sum_{n=m}^{\infty} \ln(\frac{g^n(|\overline{x}^n - \theta^*|, \overline{x}^{n-1})}{g^n(|\overline{x}^n - \theta|, \overline{x}^{n-1})}) < \infty,$$

which we shall prove happens Q_{θ^*}-a.s. Since the law of large numbers implies $\overline{x}^n \to \theta^*$, there is M s.t. $|\overline{x}^n - \theta^*| < |\overline{x}^n - \theta|$ and thus $\frac{g^n(|\overline{x}^n-\theta^*|,\overline{x}^{n-1})}{g^n(|\overline{x}^n-\theta|,\overline{x}^{n-1})} > 1$ (since g^n is strictly decreasing in its first argument) for all $n \geq M$. Also, as we saw earlier, there is K such that $\frac{g^n(a,\theta)}{g^n(b,\theta)} < 1 + \frac{k}{n^2}$ for all $n > K$ and any a, b, θ. It follows that for all $n > N := max\{M, K\}$,[23]

$$\sum_{n=N+1}^{\infty} \ln(\frac{g^n(|\overline{x}^n - \theta^*|, \overline{x}^{n-1})}{g^n(|\overline{x}^n - \theta|, \overline{x}^{n-1})}) < \sum_{n=N+1}^{\infty} \ln(1 + \frac{k}{n^2}) < \infty.$$

[23] To see why the inequality $\sum_{n=1}^{\infty} ln(1+\frac{k}{n^2}) < \infty$ in the expression holds, let $f(x) = ln(1+\frac{k}{x^2})$ and note that it is decreasing on $(0,\infty)$. Then

$$\sum_{n=1}^{\infty} ln(1 + \frac{k}{n^2}) < f(1) + \int_1^{\infty} f(x) = f(1) + (2\sqrt{k})tan^{-1}(\sqrt{k}) - ln(k+1) < \infty.$$

Therefore, we establish $\lim_{n\to\infty} \frac{P^{\theta^*}(x^n)}{P^{\theta}(x^n)} < \infty$, and in particular $P(\theta^*|x^n) \not\to 1$, Q_{θ^*}-a.s. Moreover, since $\frac{P^{\theta^*}(x^n)}{P^{\theta}(x^n)} > 0$ for any n by the full support assumption, and since we have shown that $\frac{g^n(|\overline{x}^n-\theta^*|,\overline{x}^{n-1}))}{g^n(|\overline{x}^n-\theta|,\overline{x}^{n-1})} > 1$ for all $n > N$, it must be that $\lim_{n\to\infty} \frac{P^{\theta^*}(x^n)}{P^{\theta}(x^n)} = \prod_{i=1}^{\infty} \frac{g^i(|\overline{x}^i-\theta^*|,\overline{x}^{i-1})}{g^i(|\overline{x}^i-\theta|,\overline{x}^{i-1})} > 0$. Thus, $P(\theta^*|x^n) > 0$, Q_{θ^*}-a.s.

Step 2: Establish the result under the second assumption in the lemma.

Observe that a sufficient condition for the result $\lim_{n\to\infty} P^n(\theta^*|x^n) \to 1$ Q_{θ^*}-a.s. is that $\prod_{i=N}^{n} \frac{g^i(\frac{\epsilon}{2},\overline{x}^{i-1})}{g^i(\frac{\epsilon}{4},\overline{x}^{i-1})} \to 0$. Since the second assumption restricts $g^i = g$ for $i > 1$, we take $N > 1$. We show that $\prod_{i=N}^{n} \frac{g(\frac{\epsilon}{2},\overline{x}^{i-1})}{g(\frac{\epsilon}{4},\overline{x}^{i-1})} \to 0$. Since g is continuous in its second argument and since $\overline{x}^i \to \theta^*$, we have $\frac{g(\frac{\epsilon}{2},\overline{x}^{i-1})}{g(\frac{\epsilon}{4},\overline{x}^{i-1})} \to \frac{g(\frac{\epsilon}{2},\theta^*)}{g(\frac{\epsilon}{4},\theta^*)} < 1$. In particular there exists M and $\overline{\epsilon} > 0$ s.t. $\frac{g(\frac{\epsilon}{2},\overline{x}^{i-1})}{g(\frac{\epsilon}{4},\overline{x}^{i-1})} < 1 - \overline{\epsilon}$ for all $i > M$. But then

$$\prod_{i=N}^{n} \frac{g(\frac{\epsilon}{2},\overline{x}^{i-1})}{g(\frac{\epsilon}{4},\overline{x}^{i-1})} = \prod_{i=N}^{max\{M,N\}} \frac{g(\frac{\epsilon}{2},\overline{x}^{i-1})}{g(\frac{\epsilon}{4},\overline{x}^{i-1})} \times \prod_{i=max\{M,N\}+1}^{n} \frac{g(\frac{\epsilon}{2},\overline{x}^{i-1})}{g(\frac{\epsilon}{4},\overline{x}^{i-1})}$$

$$< \prod_{i=N}^{max\{M,N\}} \frac{g(\frac{\epsilon}{2},\overline{x}^{i-1})}{g(\frac{\epsilon}{4},\overline{x}^{i-1})} \times \prod_{i=max\{M,N\}+1}^{n} (1-\overline{\epsilon}) \to 0,$$

as desired. □

References

[1] Bar-Hillel, Maya. and Willem A. Wagenaar (1991): "The Perception of Randomness", *Advances in Applied Mathematics* 12(4), pp. 428–454.

[2] Benjamin, Daniel. (2019): "Errors in Probabilistic Reasoning and Judgment Biases." Chapter for the *Handbook of Behavioral Economics* (edited by Doug Bernheim, Stefano DellaVigna, and David Laibson). Elsevier Press.

[3] Benjamin, Daniel, Don A. Moore and Matthew Rabin (2018): "Biased Beliefs About Random Samples: Evidence from Two Integrated Experiments", working paper.

[4] Benjamin, Daniel., Matthew. Rabin and Colin Raymond (2016): "A Model of Non-Belief in the Law of Large Numbers", *Journal of the European Economic Association* 14(2), pp 515-544.

[5] Berk, Richard (1966): “Limiting Behavior of Posterior Distributions when the Model is Incorrect”, *Annals of Mathematical Statistics* 37, pp. 51-58.

[6] Chen, Daniel, Tobias J. Moskowitz and Kelly Shue (2016): “Decision Making under the Gambler’s Fallacy: Evidence from Asylum Judges, Loan Officers and Baseball Umpires”, *Quarterly Journal of Economics* 131, pp. 1181-1242.

[7] Croson, Rachel and James Sundali (2005): "The Gambler’s Fallacy and the Hot Hand: Empirical Data from Casinos”, *Journal of Risk and Uncertainty* 30 (3), pp. 195-209.

[8] Epstein, Larry (2006): “An Axiomatic Model of Non-Bayesian Updating”, *Review of Economic Studies* 73, pp. 413-436.

[9] Epstein, Larry, Jawwad Noor and Alvaro Sandroni (2008): “Non-Bayesian Updating: A Theoretical Framework”, *Theoretical Economics* 3(2), pp. 193-229.

[10] Epstein, Larry, Jawwad Noor and Alvaro Sandroni (2010): “Non-Bayesian Learning”, *B.E. Journal of Theoretical Economics (Advances)* 10(1), article 3.

[11] Friedman, Bernard (1949): “A Simple Urn Model”, *Communications on Pure and Applied Math* 2, pp. 59-70.

[12] Gigerenzer, Gerd (1996): “On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky” Psychological Review 103, pp. 592-596.

[13] Gilovich, Thomas; Tversky, Amos and Vallone, Robert. (1985): "The Hot Hand in Basketball: On the Misperception of Random Sequences", *Cognitive Psychology* 17 (3), pp. 295–314.

[14] He, Kevin (2022): “Mislearning from Censored Data: The Gambler’s Fallacy and Other Correlational Mistakes in Optimal-Stopping Problems”, *Theoretical Economics* 17(3), pp. 1269–1312.

[15] Jin, Lawrence and Peng, Cameron (2022): “Trading under the Law of Small Numbers”, mimeo.

[16] Kahneman, Daniel and Amos Tversky (1972): “Subjective Probability: A Judgement of Representativeness”, *Cognitive Psychology* 3(3), pp.430-454.

[17] Noor, Jawwad (2022): “Intuitive Priors”, mimeo.

[18] Noor, Jawwad and Fernando Payro (2026): "A Subjective Theory of Belief in Mean Reversion", mimeo. 7

[19] Rabin, Matthew (2002), "Inference by Believers in the Law of Small Numbers", *Quarterly Journal of Economics* 117, pp. 775–816.

[20] Rapoport, Amnon and David Budescu (1997): "Randomization in Individual Choice Behavior", *Psychological Review* 104(3), pp. 603–617.

[21] Rabin, Matthew and David Vayanos (2010): "The Gambler's and Hot-Hand Fallacies: Theory and Applications", *Review of Economic Studies* 77, pp. 730–778.

[22] Suetens, Sigrid, Claus Galbo-Jorgensen and Jean-Robert Tyran (2016): "Predicting Lotto Numbers: a Natural Experiment on the Gambler's Fallacy and the Hot-Hand Fallacy", *Journal of the European Economic Association* 14, pp. 584-607.

[23] Terrell, Dek (1994): "A Test of the Gambler's Fallacy: Evidence from Pari-Mutuel Games", *Journal of Risk and Uncertainty* 8, pp. 309-317.

[24] Tversky, Amos and Daniel Kahneman (1971), "Belief in the Law of Small Numbers", *Psychological Bulletin* 76(2), pp. 105–110.

[25] Tversky, Amos and Daniel Kahneman (1974), "Judgement under Uncertainty: Heuristics and biases", *Science* 185, pp. 1124–1131.

[26] Wagenaar, William (1970): "Subjective Randomness and the Capacity to Generate Information" in A. F. Sanders (Ed.) *Attention and performance III, Acta Psychologica* 33, pp. 233-242.