Barcelona School of Economics

WORKING PAPERS

An Axiomatic Approach to the Law of Small Numbers

BSE Working Paper 1364| September 2022

Jawwad Noor, Fernando Payró

bse.eu/research

An Axiomatic Approach to the Law of Small Numbers*

Jawwad Noor and Fernando Payró

September 30, 2022

Abstract

With beliefs over the outcomes of coin-tosses as our primitive, we formalize the Law of Small Numbers (Tversky and Kahneman (1974)) by an axiom that expresses a belief that the sample mean of any sequence will tend towards the coin's perceived bias along the entire path. The agent is represented by a belief that the bias of the coin is path-dependent and self-correcting. The model is consistent with the evidence used to support the Law of Small Numbers, such as the Gambler's Fallacy. In the setting of Bayesian inference, we show how learning is affected by the interplay between two potentially opposing forces: a belief in the absence of streaks and a belief that the sample mean will tend to the true bias. We show that, unlike other learning results in the literature (Rabin (2002), Epstein, Noor and Sandroni (2010)), the latter force ensures that the agent at least admits the true parameter as possible in the limit, if not learn with certainty that it is true. In an evolutionary setting, we show that agents who believe in the Law of Small Numbers are never pushed out of the evolutionary race by "standard" agents who correctly understand randomness.

JEL Classification: D01, D9.

Keywords: Law of Small Numbers, Belief Biases, Heuristics, Gambler's Fallacy, Learning, Misspecified beliefs, Evolution.

1 Introduction

Economic environments are complex, and cognitive limitations make it hard to navigate them on the basis of conscious deliberation and analysis. As a result economic agents often rely on intuitive judgements to help them arrive at decisions (Noor (2022)). Intuitive judgements, while often correct, also embody systematic biases (Tversky and Kahneman (1971)). We study one class of biases in the context of judgement under uncertainty.

In their classic research on the nature of intuitive judgements under uncertainty, Tversky and Kahneman (1971) hypothesize that "people view a sample randomly drawn from a population as....similar to the population in all essential characteristics". This leads them to coin the term "the Law of Small Numbers" (henceforth LSN) to describe beliefs that the sample means in any finite sample should be concentrated around the population mean, as they would be in a large sample

*Noor is at the Department of Economics, Boston University, 270 Bay State Road, Boston MA 02215. Email: jnoor@bu.edu. Payró is at the Department of Economics and Economic History, Universitat Autonoma de Barcelona and Barcelona School of Economics, Campus de la UAB, Edifici B, Bellaterra, 08193. Email: fernando.payro@uab.cat. The authors would like to thank seminar audiences at Boston University, UPenn, U of Edinburgh, Manchester and Durham, for helpful feedback. Payró gratefully acknowledges financial support from the Ministerio de Economía y Competitividad and Feder (PGC2018-094348-B-I00) and financial support from the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S).The usual disclaimer applies.

according to the Law of Large Numbers. For instance, although any outcome of two flips of a fair coin should be equally likely, believers in LSN would judge

$$P(T,H) = P(H,T) > P(H,H) = P(T,T),$$

believing that it is more likely that the sample mean generated by two flips will be closer to 0.5 than not. LSN does not refer to a specific experimental finding. Rather, it is a hypothesis that unifies several findings, prominent of which are:

1. *Gambler's Fallacy*: Subjects believe that the probability of tails is higher following a streak of heads. That is, they expect that the sample mean in any sequence of coin tosses will tend toward the bias of the coin. While there is a large experimental literature establishing this (see Benjamin (2019) for a review), there is also substantial evidence from the field (Terrell (1994), Suetens et al. (2016), Chen et al. (2016)). For instance, Mueller et al. (2021) find the gambler's fallacy at play in job search.

2. *Local Representativeness*: Tversky and Kahneman (1974) cite evidence where subjects view HTHTTH as more likely than HHHTTT. This is different from Gambler's Fallacy in that it is not about the likelihood of a tails after a stream of heads, but rather it expresses a belief in too many alternations and too few streaks on the path of any sequence. Indeed, evidence shows that subjects tend to believe in a switch rate of approximately 60% (Rapoport and Budescu (1997), Bar Hillel and Wagenaar (1991)). This can be understood as LSN, in that it is rooted in the expectation that essential characteristics of population will be represented locally in every segment of the sequence.

3. *Sample Size Neglect*: 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 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. This reflects LSN in that any sample regardless of size is thought resemble the population in its characteristics. See Benjamin, Moore and Rabin (2018) for an incentivized experiment supporting Sample Size Neglect.

While the Law of Small Numbers is often articulated informally, we seek a formal articulation as an axiom. We understand the Law of Small Numbers as a belief specifically about the evolution of the sample mean generated by a sequence of outcomes. Specifically, in a canonical coin-tossing setting, we understand the Law of Small Numbers as a belief that the sample mean in any random draw will tend towards the perceived bias of the coin *along* the sequence. We refer to this as our LSN axiom. Our main representation result for LSN visualizes the agent as believing in a *self-correcting bias*. The LSN axiom straightforwardly gives rise to the Gambler's Fallacy and to Local Representativeness, and can exhibit Sample Size Neglect to the extent that the sampling distribution may not vary with sample size as much as implied by statistical fact.

We explore the model through an application to Bayesian inference. We establish that in the limit the agent always puts strictly positive probability on the true parameter θ^*. Intuitively, this is because by LSN the agent believes that the sample mean tends to the (unknown) true parameter, and by the Law of Large Numbers the sample mean tends to the (actual) true parameter. This contrasts with Rabin (2002), where the agent may become fully confident in the wrong parameter in the limit. Nevertheless, also unlike Rabin (2002), our agent may not be certain of the true parameter in the limit. The reason is that Local Representativeness leads to misinferences from data (since the data almost surely contains streaks), which may lead her to admit some false parameters as possible.

We consider an evolutionary setting, where a population containing agents who believe in the Law of Small Numbers and agents who understand i.i.d. randomness must decide whether to hunt in a "safe small stakes" hunting ground or a "risky high stakes" hunting ground. Their decision is based on a public signal about the risky hunting ground, which they use to determine their beliefs about a parameter that captures the desirability of the risky ground. We show that almost surely, the LSN agents will eventually be more confident about the true parameter than the IID agent, and consequently, relative to the IID population, a larger proportion of the LSN population will hunt in the "correct" hunting ground. This will guarantee that the LSN population will survive asymptotically. We also establish a lower bound on the probability that the IID population does not survive asymptotically. This lower bound can be arbitrarily close to 1, depending on the parameter.

The paper is organized as follows. We close the introduction with a literature review. Sections 2 and 3 present our model and connect it to the experimental evidence. Section 4 contains an application to Bayesian inference, and contains further connections with the evidence. Section 5 contains an application to evolution. All proofs are relegated to appendices.

Related Literature. The evidence on beliefs can be divided in terms of properties of intuitive beliefs and properties of motivated beliefs. Motivated beliefs are beliefs that respond to some emotional objective. For instance, confirmation bias (Rabin and Schrag (1999)) arises from a desire to confirm past beliefs. Intuitive beliefs are not motivated in such a manner, and instead represent beliefs arising from intuitive information processing (as opposed to conscious deliberation). The Law of Small Numbers, and Kahneman and Tversky's research on beliefs more generally, concerns the properties of intuitive beliefs.

While there exists a large experimental literature on intuitive beliefs particularly in psychology (see Tversky and Kahneman (1974) and Benjamin (2019)), there has been little theoretical work in economics since the seminal work of Rabin (2002). [1]Rabin (2002) models the agent beliefs about sequences generated from a coin of perceived bias $\theta^* \in [0,1]$ as if they are determined by an urn $U^N_{\theta^* N}$ containing N balls with an integer $\theta^* N$ number of balls labelled "heads", from which draws are made *without* replacement. Moreover, this urn is renewed after every 2 periods. 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 two heads is $P(HH) = \theta^* \frac{\theta^* N-1}{N-1}$. Similarly, the belief of getting head first and tails second is $P(HT) = \theta^* \frac{(1-\theta^*)N}{N-1}$. Due to the assumption that the urn is renewed after every 2 flips, the belief in a sequence $(x_1 x_2 y_3 y_4, ...)$ of outcomes of flips is a product:

$$P(x_1 x_2 y_3 y_4 ...) = P(x_1 x_2) \times P(y_1 y_2) \times .....$$

The standard agent (defined by θ^*-i.i.d. beliefs) is an asymptotic special case $N \to \infty$. Studying Bayesian inference by such agents, and exploiting the tractability afforded by the "i.i.d. for pairs" simplification, Rabin (2002) shows that such an agent may learn the wrong parameter in the limit, that they may exhibit the hot hand effect, and that they adjust their posteriors too much in response to observations relative to a Bayesian who has the correct prior. As we will see in section 2.2, Our model is different since Rabin (2002) may violate our LSN axiom due to the "i.i.d. for pairs" property. As noted above, the predictions of the two models differ in a learning environment.

In order to study further how inference with LSN can lead to a hot hand effect, Rabin and Vayanos (2010) study an alternative model where the agent receives a sequence of noisy signals $s_n = \theta + \epsilon_n$ about a state θ but mistakenly believes (as per the gambler's fallacy) that the errors ϵ_n are not i.i.d. and exhibit reversals. The self-correcting bias property of our model is similar in spirit, but the models are formally different: when placed in a learning environment, in our model a signal is 0 or 1 (corresponding to a head or tail) whereas in Rabin and Vayanos (2010) a signal is a real number in $\mathbb{R}$. Although excluded by their model, if we imagine their signals to lie in $[0,1]$ and interpret s_n as the sample mean upto toss n, then models become more similar in spirit. In fact Rabin (2002) can also be viewed as a model of self-correcting bias within each pair of tosses, with a reset after each pair.

Benjamin, Rabin and Raymond (2016) hypothesize that people's beliefs may not respect the Law of Large Numbers (as is implied by Sample Size Neglect) and in particular may believe in a sampling distribution for large samples that has more spread than it should. They write a model where the

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

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. Our model is one of a self-correcting (hence non-i.i.d.) stochastic bias.

In psychology, the evidence is also interpreted as subjects believing in a *switching rate* that is higher than 50% (Rapoport and Budescu (1997), Bar Hillel and Wagenaar (1991)). Rapoport and Budescu (1997) provide a model the essence of which is captured by the following, which we will refer to as the *switching rate model:* 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 (no switch) on the i^{th} toss is $\theta > \frac{1}{2}$ (resp. $1 - \theta < \frac{1}{2}$).[2] The i^{th} outcome is "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. As we will see in section 2.2, this model violates our LSN axiom and therefore is different from our model. The switching rate model is also not directly interpretable as a model with self-correcting bias.

Noor (2022) models the formation of intuitive beliefs by representing the agent's beliefs as a neural network of associations that is trained by her "experience". He shows that if the agent's experience is defined by the sampling distribution generated by the environment, then the resulting beliefs will exhibit LLN in that properties of large-sample sampling 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. Like our model, Noor (2022)'s model also produces the Hot Hand Effect. While our model explains it as a result of Bayesian learning about the bias of a coin (as in the literature, such as Rabin (2000) and Rabin and Vayanos (2010)), Noor (2022) explains it by means of beliefs trained by a mixture of sampling distributions produced by different coins.

Our learning results contribute to the literature on Bayesian learning with misspecified beliefs (spawned by Berk (1966)), as opposed to learning under non-Bayesian updating (as in Epstein, Noor and Sandroni (2010)). Specifically, our agent is misspecified in that she does not contain the correct (that is, i.i.d) data generating process in her prior. She nevertheless uses Bayes' rule to compute posteriors.

2 LSN Axioms

2.1 Primitives

Consider a canonical coin-tossing environment: the possible realizations of a coin toss in any period i are $\Omega_i = \Omega = \{0, 1\}$, and the space of all realizations of any $n \leq \infty$ tosses is $\Omega^n = \prod_{i=1}^{n} \Omega_i$. Throughout, we 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. The concatenation of two sequences $x^n \in \Omega^n$ and $y^m \in \Omega^m$ is denoted $x^n y^m \in \Omega^{n+m}$. Our results do not hinge on the binariness of Ω, which we maintain for simplicity of exposition, and can be readily extended at least to any finite set Ω.

Our primitive consists of a family of beliefs,

$$\{P^n\}_{n=1}^{\infty},$$

where each P^n is a probability measure (henceforth, *belief*) on the measurable space (Ω^n, Σ^n) defined by the sample space Ω^n and the space $\Sigma^n = 2^{\Omega^n}$ of all subsets $A^n \subset \Omega^n$. As our LSN axioms are ordinal in nature, they can be derived from the betting behavior of Subjective Expected Utility agents. Since the translation to betting behavior is obvious, we take beliefs $\{P^n\}$ directly as our

[2]Similar to Rabin (2002), Rapoport and Budescu (1997)'s model switching probability may 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. What we are calling the switch rate model corresponds to their model when $m = 1$.

primitive and interpret them as behavioral objects. Observe that since our primitive consists of the agent's ex ante beliefs prior to any flip of the coin, our model is one of *prior beliefs.* In particular, it is not a model of updating, and as such it lies outside the literature on non-Bayesian updating (see for instance Epstein (2006), Epstein, Noor and Sandroni (2008, 2010)).

Our focus will be on the *sample mean* number of heads at any point n of a sequence $x = (x_1, x_2, ...)$, denoted

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

Throughout, let $\theta^* \in [0,1]$ denote the probability assigned to heads when there is only 1 flip of the coin, that is,

$$\theta^* := P^1(1).$$

We presume throughout that the coin tosses are objectively i.i.d. The agent correctly understands the bias to be θ^* but we model her incorrect understanding of what an i.i.d. data-generating process looks like.

In Economics it is typical to posit the existence of a belief P^∞ over the infinite horizon sample space $(\Omega^\infty, \Sigma^\infty)$, and to consider its marginals.[3] If we begin with a family of beliefs $\{P^n\}_{n=1}^\infty$, Kolmogorov's extension theorem tells us that a necessary and sufficient condition on $\{P^n\}_{n=1}^\infty$ for the existence of P^∞ is

Axiom 1 *(Marginal Consistency) For any n and any event $A^n \subset \Omega^n$,*

$$P^n(A^n) = P^{n+1}(A^n).$$

Marginal Consistency embodies horizon-independence of beliefs, in the sense that the agent does not think differently about a given event A^n if the horizon is extended by a period. Such horizon-independence is satisfied by standard models. An interesting question is whether the Law of Small Numbers is at odds with Marginal Consistency, and we will see that it is not.

2.2 Axiom: LSN

The aim is to construct a model that accommodates the evidence on the Law of Small Numbers noted in the Introduction. Our model is built on the following interpretation of the informal theory laid out in Kahneman and Tversky (1971, 1974).

- According to the Gambler's Fallacy, when faced with a fair coin, the agent believes that $HHHT$ is more likely than $HHHH$, that is, a tails is more likely than heads following a streak of heads. Kahneman and Tversky (1971) state that the agent "feels that the fairness of the coin entitles him to expect that any deviation in one direction will soon be cancelled out by a corresponding deviation in the other". Our interpretation of this statement is that the agent believes that the *sample mean* is likely to tend towards the bias of the coin, that is, it is more likely that the distance between the sample mean of x^n and the bias θ^*,

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

will be small. This will be a distinctive feature of our model.

[3]As is standard, we identify any n-event $A^n \in \Sigma^n$ with 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^∞ known as the *n-cylinder.* Let $\Sigma^\infty = \sigma(\cup_{n=1}^\infty \Sigma^n)$ denote the σ-algebra generated by all the n-cylinders, $n = 1, 2, ...$ Then P^∞ is a probability measure of a well-defined space $(\Omega^\infty, \Sigma^\infty)$. The marginal belief on (Ω^n, Σ^n) is defined by $P^n(A^n) = P^\infty(A^n\Omega^\infty)$ for each n-event $A^n \in \Sigma^n$.

- According to the Local Representativeness, people regard the sequence HTHTTH as more likely than HHHTTT. Tversky and Kahneman (1974) relate such beliefs with LSN by noting that "people expect 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". As before we interpret the "essential characteristics" of the process to refer to the fact that the process requires the sample mean to eventually concentrate around the population mean θ^*. However, instead of being a statement about the sample mean at the *end* of the sequence, we observe that $HTHTTH$ *dominates* $HHHTTT$ *on path* in the sense that the sample mean is always closer to $\frac{1}{2}$ in the former than the latter *along* the entire sequence (Table 1).

	$HTHTTH$	$HHHTTT$
$n = 1$	1	1
$n = 2$	$\frac{1}{2}$	1
$n = 3$	$\frac{2}{3}$	1
$n = 4$	$\frac{1}{2}$	$\frac{3}{4}$
$n = 5$	$\frac{2}{5}$	$\frac{3}{5}$
$n = 6$	$\frac{1}{2}$	$\frac{1}{2}$

Table 1. The entries are $\overline{x}^n = \dfrac{\sum_{i=1}^n x_i}{n}$ for $x = HTHTTH, HHHTTT$ and $n = 1, ..., 6$.

The following axiom captures the idea that beliefs are driven by a consideration of the distance between sample mean and true bias, and that dominance on path implies higher beliefs.

Axiom 2 *(LSN) For any N and $x, y \in \Omega^\infty$ s.t. $d(x^{N-1}) = d(y^{N-1})$,*

$$d(x^n) \leq d(y^n) \text{ for all } n \leq N \implies P^N(x^N) \geq P^N(y^N).$$

Moreover if $d(x^n) < d(y^n)$ for some $n \leq N$ then $P^N(x^N) > P^N(y^N)$.

The axiom states that if the sample means of the sequences x and y are equally distant from θ^* at throw $N - 1$, and if x dominates y on path for all N throws, then x^N is deemed more likely than y^N. This is satisfied, for instance, in the Gambler's Fallacy with a fair coin where $HHHT$ is deemed more likely than $HHHH$. The axiom also embodies Local Representativeness: for a fair coin, $HTHTH$ will be deemed more likely than $HHTTT$ according to the axiom, since both sequences have the same number of heads in the first 4 throws (and thus the means at the 4^{th} throw are equally distant from $\theta^* = \frac{1}{2}$), but the first strictly dominates the second on path. Similarly, $HTHTTH$ will be deemed more likely than $HHHTTT$ as in the evidence noted above, since the sample means of both sequences at the 5^{th} throw are equally distant from $\frac{1}{2}$. LSN has the flavor of "mean reversion", except that mean reversion is expected at each flip. It captures Kahneman and Tversky (1974)'s notion of a "belief that a random process is self-correcting".[4]

At this point we can already note that our model will differ from those in the literature. Consider the example of HHH vs HHT. The LSN axiom straightforwardly implies that

$$P(HHT) > P(HHH).$$

Intuitively, both the Gambler's Fallacy and Local Representativeness favor HHT over HHH. Due to the "i.i.d. pairs" simplification, Rabin (2002) is not consistent with this: when beliefs have full support and $\theta^* = \frac{1}{2}$, the Rabin (2002) model predicts $P(HHT) = P(HHH)$ since the model requires

[4]One may hypothesize that the evidence pertains to a small number of flips, and due to cognitive constraints, agents may behave differently when faced with much longer sequences. In particular, bounded memory or limited attention may cause them to focus on parts of the sequence or to form a coarser perception of the mean number of heads. We abstract from cognitive constraints in this paper since, as a feature of intuitive processing, LSN is distinct from bounded memory or attention, and thus is worth studying in isolation.

that $\frac{P(HHT)}{P(HHH)} = \frac{P(HH)P(T)}{P(HH)P(H)} = \frac{P(T)}{P(H)} = 1$. The switching rate model also violates the LSN axiom. Our axiom predicts that:

$$P(HHHHHHTH) < P(HHHHHHTT).$$

Intuitively, while Local Representativeness requires the former sequence to be more likely (since there are fewer streaks) and the Gambler's Fallacy requires the latter to be more likely (since the sample mean is closer to $\theta^* = \frac{1}{2}$), our LSN axiom imposes that the latter dominates in this example. The switching rate model requires the reverse ranking because the former sequence contains more switches than the latter.

We close by writing a weaker and a stronger condition that will appear in our applications. The Weak LSN axiom weakens LSN by restricting attention to x, y that have the same sample mean in throw $N-1$ and where x^N that dominates y^N on path:

Axiom 3 *(Weak LSN) For any N and $x, y \in \Omega^\infty$ s.t. $\overline{x}^{N-1} = \overline{y}^{N-1}$,*

$$d(x^n) \leq d(y^n) \text{ for all } n \leq N \implies P^N(x^N) \geq P^N(y^N).$$

Moreover if $d(x^n) < d(y^n)$ for some $n \leq N$ then $P^N(x^N) > P^N(y^N)$.

In contrast, the Strong LSN axiom requires that dominance on path to be respected for all pairs of sequences.

Axiom 4 *(Strong LSN) For any N and $x, y \in \Omega^\infty$,*

$$d(x^n) \leq d(y^n) \text{ for all } n \leq N \implies P^N(x^N) \geq P^N(y^N).$$

Moreover if $d(x^n) < d(y^n)$ for some $n \leq N$ then $P^N(x^N) > P^N(y^N)$.

2.3 Axiom: Weak Independence

The LSN axioms only place structure on beliefs that can be ranked by dominance on path. Such beliefs admit a very general representation (see Section 6.2), but additional structure is desirable for applications.

The evidence on LSN reveals that people do not have a correct understanding of the properties that i.i.d. tosses of a coin should have. One property of i.i.d. tosses is:

Axiom 5 *(Independence)For any n, any $x^n, y^n \in \Omega^n$ and any $x_{n+1} \in \Omega$ s.t. $P^n(y^n) > 0$ and $P^{n+1}(y^n x_{n+1}) > 0$,*

$$\frac{P^{n+1}(x^n x_{n+1})}{P^{n+1}(y^n x_{n+1})} = \frac{P^n(x^n)}{P^n(y^n)}.$$

We impose that our agent at least understands that i.i.d. tosses imply:

Axiom 6 *(Weak Independence)For any n, any $x^n, y^n \in \Omega^n$ and any $x_{n+1} \in \Omega$ s.t. $P^n(y^n) > 0$ and $P^{n+1}(y^n x_{n+1}) > 0$,*

$$\overline{x}^n = \overline{y}^n \implies \frac{P^{n+1}(x^n x_{n+1})}{P^{n+1}(y^n x_{n+1})} = \frac{P^n(x^n)}{P^n(y^n)}.$$

In providing some structure to our model, Weak Independence places a restriction on the extent of Local Representativeness, which is illustrated as follows. Consider sequences HHHTTT and TTTHHH generated by a fair coin and suppose the agent finds these equally likely, $\frac{P^6(HHHTTT)}{P^6(TTTHHH)} = 1$. It is conceivable that TTTHHHT may be considered more likely than HHHTTTT because there are shorter streaks in the former. However, under Weak Independence the agent must find them equally likely, $\frac{P^7(HHHTTTT)}{P^7(TTTHHHT)} = \frac{P^6(HHHTTT)}{P^6(TTTHHH)} = 1$. Weak Independence requires that the perspective on the n-length streams HHHTTT and TTTHHH with equal sample means should not change in the $n+1$ horizon based on the outcome in flip $n+1$. Repeated application of the axiom implies more generally that the concatenated streams $x^n z^m$ and $y^n z^m$ are ranked by P^{n+m} in accordance with the ranking of x^n and y^n by P^n, whenever $\overline{x}^n = \overline{y}^n$.

3 The Self-Correcting Bias Representation

3.1 Representation Result

For completeness we first present the conditions that characterize an agent who understands that the coin tosses are independent, and also the additional condition that ensure that she understands that the tosses are identical. See Appendix 6.2 for more general versions of the results in this section that drop Marginal Consistency.

Theorem 1 *A family of full support beliefs $\{P^n\}_{n=1}^{\infty}$ satisfies Independence and Marginal Consistency iff it admits a time-varying bias representation: for all n and $x^n \in \Omega^n$,*

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

where $\theta_i \in (0,1)$ for all i and $\theta_1 = \theta^*$. Moreover, $P^n(\Omega^{n-1}1) = \theta^*$ for all $n > 1$ iff $\theta_i = \theta^*$ for all i.

To model an LSN agent, we maintain Marginal Consistency, relax Independence to Weak Independence, and use the LSN structure to discipline the representation. Recall that for any sequence $x \in \Omega^{\infty}$, the sequence truncated at i is denoted $x^i \in \Omega^i$. Adding an outcome of heads (respectively, tails) in the $i+1^{st}$ toss yields $x^i 1 \in \Omega^{i+1}$ (respectively $x^i 0 \in \Omega^{i+1}$) and the sample mean is denoted $\overline{x^i 1}$ (respectively $\overline{x^i 0}$).

Theorem 2 *A family of full support beliefs $\{P^n\}_{n=1}^{\infty}$ satisfies LSN, Weak Independence and Marginal Consistency iff it admits a self-correcting bias representation: for all n and $x^n \in \Omega^n$,*

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

where $\theta_{1,\overline{x}^0} \equiv \theta^*$ and $\theta_{i,\overline{x}^{i-1}} \in (0,1)$, and moreover for all $i > 1$ and $x^{i-1}, y^{i-1} \in \Omega^{i-1}$ s.t. $d(x^{i-1}) = d(y^{i-1})$,

$$d(x^{i-1}1) \leq d(y^{i-1}0) \iff \theta_{i,\overline{x}^{i-1}} \geq 1 - \theta_{i,\overline{y}^{i-1}}$$

$$d(x^{i-1}1) \leq d(y^{i-1}1) \iff \theta_{i,\overline{x}^{i-1}} \geq \theta_{i,\overline{y}^{i-1}}$$

$$d(x^{i-1}0) \leq d(y^{i-1}0) \iff \theta_{i,\overline{x}^{i-1}} \leq \theta_{i,\overline{y}^{i-1}}.$$

These three conditions hold for all $x^{i-1}, y^{i-1} \in \Omega^{i-1}$ (resp. for all $x^{i-1}, y^{i-1} \in \Omega^{i-1}$ s.t. $\overline{x}^{i-1} = \overline{y}^{i-1}$) iff $\{P^n\}$ satisfies Strong LSN (resp. Weak LSN).

The result tells us that LSN and Weak Independence are characterized by a path-dependent bias that is self-correcting in that it varies so as to keep the sample mean near the bias θ^*. Specifically, interpret $\theta_{i,\overline{x}^{i-1}} \in (0,1)$ as the propensity for heads in the i^{th} flip, conditional on a sample mean $\overline{x}^{i-1}$ up to that point. The three noted conditions state that the bias towards heads in the i^{th} is stronger whenever it brings the sample mean at the end of the i^{th} toss closer to the mean. For instance, the first condition, when specialized to $x^{i-1} = y^{i-1}$, tells us that if at the end of a *given* sequence x^{i-1} a heads on the i^{th} toss will bring the sample mean strictly closer to θ^* than a tails, then it must be that the bias towards heads is stronger than that towards tails, $\theta_{i,\overline{x}^{i-1}} > \frac{1}{2}$. The three conditions moreover impose some structure on the intensity of the bias *across* sequences. For instance, the second condition states that, for any x^{i-1}, y^{i-1} that result in the same distance to θ^* at the end of toss $i-1$, if it is the case that a heads after x^{i-1} leads to a mean closer to θ^* than a heads after y^{i-1}, then the bias towards head in the i^{th} toss is stronger after x^{i-1} than it is after y^{i-1}, that is, $\theta_{i,\overline{x}^{i-1}} \geq \theta_{i,\overline{y}^{i-1}}$. The other conditions are interpreted similarly. Observe that the theorem assures us that the Law of Small Numbers is not fundamentally incompatible with Marginal Consistency.

As is evident in the proof of the theorem, the three conditions relating the bias with the distance between the sample mean and true bias can be described in a simpler way, even without requiring Marginal Consistency. We present the most general version of our model next.

Theorem 3 *Beliefs $\{P^n\}$ satisfy LSN and Weak Independence if and only if for each i there exists $g^i : [0,1]^2 \to [0,1]$ that is strictly decreasing in its first argument such that for any n and $x^n \in \Omega^n$,*

$$P^n(x^n) = \theta^{*x_1}(1-\theta^*)^{1-x_1} \prod_{i=2}^{n} g^i(d(x^i), d(x^{i-1})).$$

In the representation, $\{P^n\}$ satisfies Strong LSN (resp. Weak LSN) iff g^i is constant in its second argument (resp. g^i has $\overline{x}^{i-1}$ as its second argument).[5]

The connection with the self-correcting bias representation – after assuming Marginal Consistency – is that

$$\theta_{i,\overline{x}^{i-1}} = g^i(d(x^{i-1}1), d(x^{i-1})) \text{ and } 1-\theta_{i,\overline{x}^{i-1}} = g^i(d(x^{i-1}0), d(x^{i-1})).$$

To say that g^i is constant is to say that $\theta_{i,\overline{x}^{i-1}} = \frac{1}{2}$. To say that g^i is constant in i is to say that the intensity of the bias $\theta_{i,\overline{x}^{i-1}}$ depends only on $d(x^{i-1}1)$ and $d(x^{i-1})$ and not on the toss i per se. An example of a specification of the model is that there exists $\epsilon^* > 0$ and $\alpha > \frac{1}{2}$ such that for all $i > 1$, each g^i is given by

$$g^i_{\epsilon^*}(|\overline{x}^i - \theta^*|) = \begin{cases} \frac{\alpha}{Z^i} & |\overline{x}^i - \theta^*| \leq \epsilon^* \\ \frac{1-\alpha}{Z^i} & \textit{otherwise} \end{cases} \tag{1}$$

where $\{Z^i\}_{i=1,..,n}$ normalize the representation so that P^n is a probability.[6] Thus, when facing sequence x, the agent "rewards" (in the sense of boosting the probability of the sequence) the outcome of flip n by $\alpha > \frac{1}{2}$ if $\overline{x}^n$ is within ϵ^* of θ^*, and otherwise "punishes" it by $1-\alpha$ when determining her belief $P^n(x^n)$.

We turn next to a discussion of how our model connects with the evidence on the Law of Small Numbers.

3.2 Gambler's Fallacy

According to Gambler's Fallacy, following a streak of heads, the agent believes that a tails is more likely to occur. Relatedly, an agent that satisfies Weak LSN exhibits:[7] for any $x^n \in \Omega^n$,

$$\frac{\sum_{i=1}^n x_i}{n} > \theta^* \implies P^{n+1}(0|x^n) \geq P^{n+1}(1|x^n).$$

That, if there are "too many heads" in the first n flips (in that $\frac{\sum_{i=1}^n x_i}{n} > \theta^*$) then she believes that it is more likely for there to be a tails in the $(n+1)^{th}$ flip. This makes a qualified prediction regarding

[5]Furthermore, $\{P^n\}$ satisfies Marginal Consistency iff $g^i(d(x^{i-1}1), d(x^{i-1})) + g^i(d(x^{i-1}0), d(x^{i-1})) = 1$ for all x, i.

[6]Recall that by definition, for $i = 1$ it must be that $g^1_{\epsilon^*}(|1-\theta^*|) = \theta^*$ and $g^1_{\epsilon^*}(|0-\theta^*|) = 1-\theta^*$. Therefore the model is defined by a restriction on g^i only for $i > 1$.

[7]Observe that x^n0 and x^n1 are equivalent on the first n flips. By Weak LSN, the belief over these sequences is completely determined by which outcome for flip $n+1$ takes the sample mean closer to θ^*.

the Gambler's Fallacy: it is not the presence of a streak prior to the last flip that is responsible for exaggerating the belief in a tails, but rather the sample mean of the entire sample. Thus, if the sample mean is too low at the n^{th} flip then there will be no gambler's fallacy despite an immediately preceding streak.

The *Retrospective Gambler's Fallacy* refers to the belief that outcome of the flip *preceding* a streak of heads is most likely to be a tails (Oppenheimer and Monin (2009)). This is consistent with our model, albeit with some nuance. For instance, observe that when $\theta^* = \frac{1}{2}$, Weak LSN implies

$$P^{n+1}(1_1|1_2 1_3 1_4) < P^{n+1}(0_1|1_2 1_3 1_4),$$

that is, conditional on obtaining a heads on tosses 2-4, our agent will believe it is more likely that there was a tails on the first toss. This follows simply because the stream $(0, 1, 1, 1)$ dominates $(1, 1, 1, 1)$ on path. However, unlike a comparison of sequences of the form $x^n 1$ vs $x^n 0$ used in the Gambler's Fallacy, we do not necessarily have dominance in sequences of the form $(1_1, x_2 ... x_{n+1})$ vs $(0_1, x_2 ... x_{n+1})$. For instance, when $\theta^* = \frac{1}{2}$, the sequence $(1, 0, 1, 1)$ dominates $(0, 0, 1, 1)$ by the end of the the second toss, but is dominated by it at the end of the fourth toss. Consequently, our LSN axioms are silent on the comparison – they permit a Retrospective Gambler's Fallacy but do not necessitate it. [8]

3.3 Local Representativeness

According to the Local Representativeness, people regard the sequence HTHTTH as more likely than HHHTTT. The literature interprets this evidence as revealing that people expect the sample mean to stay close to the true bias on path. Our LSN axioms formalize this using the notion of dominance on path.

An interesting observation is that Local Representativeness and the Gambler's Fallacy, while both serving as evidence for LSN, can in fact contradict each other. Consider the following example discussed in Section 2.2,

$$HHHHHHTH \text{ vs } HHHHHHTT.$$

Local Representativeness presumably requires the former sequence to be more likely, since there are fewer streaks. But the Gambler's Fallacy requires the latter to be more likely, since the sample mean is closer to $\theta^* = \frac{1}{2}$. This observation implies that any model of LSN will have to take a position on which force dominates in any given example. In the above example, our LSN axiom implies that the Gambler's Fallacy will be stronger, since $HHHHHHTT$ dominates $HHHHHHTH$ on path. Rabin (2002) implies that Local Representativeness is stronger, since $\frac{P(HHHHHHTH)}{P(HHHHHHTT)} = \frac{P(TH)}{P(HH)} > 1$. The switching rate model also implies that Local Representativeness dominates here since $HHHHHHTH$ has more switches than $HHHHHHTT$. The models may generate predictions in other directions in other examples.

3.4 Sample Size Neglect and Non-Belief in LLN

As noted in the Introduction, Kahneman and Tversky (1972) report an experiment where subjects beliefs about the sampling distribution appear to be relatively insensitive to sample size., which they term Sample Size Neglect To show that our model can accommodate this, consider the ϵ^*-specification for g in (1) presented in Section 3.1. For any n, all sequences x^n that have a mean $\overline{x}^n$

[8] It is worthwhile to make note also of the finding of *Long-Distance Gambler's Fallacy* in Benjamin, Moore and Rabin (2018). Following a streak of $r = 1, 2, 5$ heads on *consecutive* flips up to the n^{th} flip, their subjects exhibited that a probability of heads on flip $n+1$ was respectively 44%,41% and 39%. But when the streak came from nonconsecutive draws from *random* locations flips, the probability of heads on another randomly chosen flip was 45%,42% and 41% resp. That is, Gambler's Fallacy appeared in randomly chosen subsequences from the original sequence. This is difficult to reconcile in our model, as it is in any of the models in the literature, especially if the number n of flips is large.

within ϵ^* of θ^* will get the same probability $\frac{\alpha}{Z^n}$, whereas those that have a more distant mean will all get probability $\frac{1-\alpha}{Z^n}$. Thus the sampling distribution is a step function for each n, and in particular there is some N beyond which there are always sample means that lie in $[\theta^* - \epsilon^*, \theta^* + \epsilon^*]$. If we take $\alpha \to 1$, then the sampling distribution becomes concentrated on the sample means $[\theta^* - \epsilon^*, \theta^* + \epsilon^*]$ for all $n \geq N$. Therefore it exhibits Sample Size Neglect.

Benjamin, Rabin and Raymond (2016) note that Sample Size Neglect suggests that subjects' believed sampling distributions tend to be too spread out relative to the objective sampling distribution. They term this a *Non-Belief in the Law of Large Numbers*, and evidence for it is cited in Benjamin (2019). The limiting case $\alpha \to 1$ in the above example demonstrates that the agent's sampling distribution does not collapse on θ^*, consistent with Non-Belief in the Law of Large Numbers.

4 Bayesian Inference

In this section we study whether an agent who exhibits LSN can learn the bias after observing an infinite sequence of i.i.d. tosses.

4.1 Model

Let $\Theta = \{\theta_1, ..., \theta_n\}$ 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. Define $d_\theta(x^i) := |\overline{x}^n - \theta|$. Suppose it is given by the following model which satisfies LSN and Weak Independence:

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

where we assume that P^n_θ has full support and therefore that g^i is strictly positive. Moreover, we assume g^i is independent of θ.

Suppose she has a prior $\mu \in \Delta(\Theta)$ over the parameter. 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:

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

Since for each given θ, the family of beliefs$\{P^n_\theta\}$ is permitted to violate Marginal Consistency, the Bayesian posteriors may be computed with respect to ex-ante beliefs $P^n \in \Delta(\Theta \times \Omega^n)$ that are not consistent in the sense that $P^n(x^n)$ need not be equal to $P^{n+1}(x^n\Omega)$. However, if each of these families satisfies Marginal Consistency, our model reduces to a Bayesian Model that is misspecified in that the true data generating process (the i.i.d. model) is not in the support of the prior.

4.2 Results

Suppose the data is generated by p^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 *Assume $\mu \in \Delta(\Theta)$ and each $P^n \in \Delta(\Omega^n)$ have full support. Then,*

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

Without further conditions, it cannot be assured that beliefs will converge. But the theorem establishes that regardless of whether there is convergence, the agent always places a non-vanishing probability on the true parameter. The reason is that the agent believes not just that the sample mean tends to the true parameter at every point of the path, but rather that it must eventually tend to it in the limit by the Law of Large Numbers. The latter is the reason that the evolution of the sample mean keeps the agent from ruling out the true parameter. The former, 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 converges a.s. to a degenerate posterior, which 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.

The following result provides sufficient conditions on $\{P^n\}$ for $P^n(\theta^*|x^n)$ to converge, providing both a case where the agent learns the truth and a case where she fails to rule out some wrong parameters.

Theorem 5 *Suppose $\mu \in \Delta(\Theta)$ and each $P^n \in \Delta(\Omega^n)$ have full support, and that g^i is continuous in its second argument for each i.*

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

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

2. If $g^i \to c$ uniformly faster than $\frac{1}{n^2} \to 0$,[9] where $c > 0$ is a constant function, then $p^{\theta^*}(\lim_{n\to\infty} P^n(\theta^*|x^n) \in (0,1)) = 1$, that is,

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

Claim (i) assumes that $g^i = g$ is strictly increasing in the first argument, which ensures that posteriors are responsive to the sample mean in every period. In this case the agent eventually learns the true parameter θ^*. Claim (ii) assumes that g^i approaches a constant g "fast enough". As noted after Theorem 3, a constant g corresponds to a belief that the bias is constant and equals $\frac{1}{2}$. In the current context, the agent believes that, for every θ, the coin becomes less self-correcting 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 distribution almost surely. In line with Theorem 4, in the limit the agent places strictly positive probability on the true parameter θ^*. Indeed, patterns observed early in the sequence can never lead the agent to discard the truth.

4.3 Hot Hand Effect and Under/Over-inference

We illustrate how our model can accommodate some evidence related to inference. Consider the ϵ^*-specification for g in (1) presented in Section 3.1. Suppose there are two parameters $\Theta = \{\hat{\theta}, 1-\hat{\theta}\}$, where $\hat{\theta} > \frac{1}{2}$. The agent has a uniform prior over parameters.

An intuition familiar from psychology (Gilovich, Vallone and Tversky (1985)) and formalized in economics (Rabin (2002), Rabin and Vayanos (2010)) is that inference with LSN can give rise to a *Hot Hand Effect*, the finding that subjects sometimes expect a streak to be more likely to continue, in contrast to the Gambler's Fallacy. We illustrate this same idea in our model. Intuitively, our

[9] 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.

agent expects the mean of the sequence to remain close to the bias. Therefore, according to the agent, a higher bias will imply longer streaks. Then in an inference setting, observing a longer streak will more strongly suggest that the true bias is high, and the agent will therefore expect the streak to continue, as in the Hot Hand Effect. To illustrate, let 1^n be a sequence of heads of length n, in which case the sample mean is 1. Suppose that this sample mean is within ϵ^* of $\hat{\theta}$ and beyond ϵ^* of $1-\hat{\theta}$, that is, $1-\hat{\theta} < \epsilon^* < 1-(1-\hat{\theta})$. Then given the expressions in Appendix 6.4, we see that

$$\frac{P^n(\hat{\theta}|1^n)}{P^n(1-\hat{\theta}|1^n)} = \frac{\hat{\theta}\alpha^{n-1}}{(1-\hat{\theta})(1-\alpha)^{n-1}}.$$

Since $\alpha > \frac{1}{2}$, the belief in $\hat{\theta}$ grows exponentially with the length of the streak n, which in turn gives rise to a belief that the streak is likely to continue. This illustrates a Hot Hand Effect.

Over- or under-inference refers to how much posteriors move with data relative to those of an IID agent (see Benjamin (2019) for a discussion of the evidence). We show that our model can generate extreme movements or sluggish movement. The g function in the representation determines this. For instance in the above example, suppose that g is close to constant: $\alpha \approx \frac{1}{2}$. Then by the above expression we see that the posteriors after observing a single heads are approximately the same as those after observing a streak of heads:

$$\frac{P^n(\hat{\theta}|1)}{P^n(1-\hat{\theta}|1)} = \frac{\hat{\theta}}{(1-\hat{\theta})} \approx \frac{\hat{\theta}\alpha^{n-1}}{(1-\hat{\theta})(1-\alpha)^{n-1}} = \frac{P^n(\hat{\theta}|1^n)}{P^n(1-\hat{\theta}|1^n)}.$$

Given that the IID agent's posteriors will be different, for $\alpha \approx \frac{1}{2}$ we can have under-inference.

To illustrate over-inference, consider again ϵ^*-specification for g in (1) and suppose there are two parameters $\hat{\theta}, 1-\hat{\theta}$ where $\hat{\theta} > \frac{1}{2}$. Consider beliefs over parameters after observing a heads followed by a tails, $(1,0)$. Given the expressions in Appendix 6.4, the belief in $\hat{\theta}$ satisfies

$$P^{LSN}(\hat{\theta}|1,0) = \frac{\hat{\theta}\alpha}{\hat{\theta}\alpha + (1-\hat{\theta})\alpha} = \hat{\theta} > \frac{1}{2} = P^{IID}(\hat{\theta}|1,0),$$

that is, while the IID agent's posterior equals the prior after observing a heads and a tails, the LSN agent will over-infer that the true parameter is $\hat{\theta}$.

5 Evolutionary Survival of LSN vs IID Agents

We study the survival of LSN agent in a specific evolutionary context. Consider two populations of agents. An "IID agent" has an accurate perception of i.i.d. sequences. An "LSN" agent follows the ϵ^*-specification for g in (1).

Suppose there is a continuum of "safe" hunting grounds where hunting yields a small reward $r = 1$ (a "rabbit"). There is a continuum of "risky" hunting grounds of which a fraction $\theta \in \Theta = \{\hat{\theta}, 1-\hat{\theta}\}$, where $\hat{\theta} > \frac{1}{2}$, contains a large reward, $r = 2$ (a "deer"), and the remaining fraction contain no reward, $r = 0$. The fraction θ is an unknown parameter. In every period, one risky ground is randomly chosen and publicly sampled by both LSN and IID agent types. The agents update their beliefs about θ based on whether a deer is sighted in the sampling hunting ground. Let $x_i = 0$ (resp. $x_i = 1$) denote that the deer was not present (resp. present), in which case we can write the reward in period i as $r = 2x_i$. Each type $A = IID, LSN$ determines the fraction of its population, $k^A_{x^i} \in [0,1]$, that hunts in the risky grounds in period i conditional on having observed x^i signals, which the remainder fraction $1-k^A_{x^i}$ hunting in the safe ground.[10] Letting Λ^A_{i-1} denote the population of type A at the

[10] An interpretation is that each agent has to choose between spending their full day in the safe or the risky hunting ground, and they randomize by flipping a coin with bias $k^A_{x^i}$ and they go to the risky hunting ground if there is a heads. If all agents do this, then fraction $k^A_{x^i}$ goes to the risky hunting ground.

start of period i. The total reward per capita received by type A is

$$c_{x^i}^A := \frac{R_{x^i}^A}{\Lambda_{i-1}^A} = k_{x^i}^A(2\theta) + (1 - k_{x^i}^A).$$

Both types of agents have a common prior over Θ:

$$\mu_{LSN}(\hat{\theta}) = \mu_{IID}(1 - \hat{\theta}) = \frac{1}{2}.$$

Both maximize expected utility using a common strictly increasing strictly concave utility index u to determine the optimal $(k_{x^i}^A, 1 - k_{x^i}^A)$ based on history of deer sightings x^i from the sampled hunting ground. We assume that agents do not observe the outcome of other agents' hunting. Consequently, they cannot deduce θ by observing the fraction of agents that found a deer.

The population of type A agents grows by a factor of $\lambda^{c_{x^i}^A}$ in period i, where $\lambda > 1$. Thus, higher per capita consumption leads to faster growth in the population. Assuming that both populations start with the same size, we are interested in determining which grows faster over time, that is, we are interested in the ratio:

$$\prod_{i=1}^{n} \frac{\lambda^{c_{x^i}^{LSN}}}{\lambda^{c_{x^i}^{IID}}},$$

and in particular how this ratio grows as $n \to \infty$.

We show that:

Proposition 1 *Denote the true parameter as $\theta \in \Theta = \{\hat{\theta}, 1 - \hat{\theta}\}$. Then the following hold for LSN agents with $0 < \epsilon^* < 2\hat{\theta} - 1$.*

(i) The LSN agents are eventually more confident about the true parameter, a.s.:

$$P_{LSN}^n(\theta|x^n) \geq P_{IID}^n(\theta|x^n) \text{ for all sufficiently large } n \in \mathbb{N} \text{ a.s.-}p^\theta$$

(ii) The population of LSN agents never vanishes, a.s.:

$$\liminf_{n\to\infty} \prod_{i=1}^{n} \frac{\lambda^{c_{x^i}^{LSN}}}{\lambda^{c_{x^i}^{IID}}} > 0 \quad \text{a.s-}p^\theta.$$

The first part of the result states that along any realized path x, almost surely, LSN agents will eventually be more confident in the true parameter than the IID agents. The reason is that by LLN the sample mean will eventually be within ϵ^* of the true parameter θ, and the LSN agents will take this as a stronger indication that the true parameter is θ than not, relative to the IID agents. This leads a larger proportion of the LSN population, relative to the IID population, to hunt in the "correct" hunting ground (the correct hunting ground is the risky one iff the true parameter is $\theta = \hat{\theta}$), and therefore to grow faster than the IID population. Accordingly, the second part of the result states that the LSN population is never pushed out of the evolutionary race by the IID population.

Since we have already seen (in Theorem 4) that in general LSN agents may not become sure about the truth in the limit, it is possible to construct settings where LSN agents may not survive relative to IID agents with positive probability. The above result, however, shows us that LSN agents do not possess an inherent evolutionary disadvantage that they carry into all possible settings. As long as they can learn, the fact that their beliefs are misspecified relative to IID agents and cause misinferences does not threaten their survival. In fact it can benefit their survival. It is possible to construct examples in the current setting where in fact LSN agents completely dominate the population with probability strictly greater than $\frac{1}{2}$. For instance:

Proposition 2 *If $\hat{\theta} \geq \frac{3}{4}$ and $\epsilon^* = \hat{\theta} - \frac{1}{2}$, then $P(\liminf_{n\to\infty} \prod_{i=1}^{n} \frac{\lambda^{c_{x^i}^{LSN}}}{\lambda^{c_{x^i}^{IID}}} = \infty) > \frac{1}{2}$.*

We close by noting, however, that in a market setting LSN agents share the same fate when facing IID agents as any agent with misspecified beliefs. Sandroni (2000) shows that agents who eventually make accurate forecasts will push out agents who do not. Even when LSN agents know the true parameter, their one-step ahead belief will be bounded away from the true probability. Such agents will not survive relative to well-specified agents because, intuitively, the misspecified beliefs lead agents to bet (via their demand for assets) on events that have probability 0, to their detriment. However, this result hinges on the market for assets for being complete. We leave it to future research to study the survival of LSN agents in incomplete markets.

6 Appendix

6.1 Appendix: Proof of Theorem 1

Lemma 1 *A family of full support beliefs $\{P^n\}$ satisfies Independence iff*

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

where $\theta_i, \gamma_i \in (0,1)$ and $\theta_1 = 1 - \gamma_1 = \theta^*$. Moreover, Marginal Consistency holds iff $\theta_n + \gamma_n = 1$ for all n. Finally, $P^n(\Omega^{n-1}1) = \theta^*$ for all n iff $\theta_n = 1 - \gamma_n = \theta^*$ for all n.

Proof. By definition, $P^1(x^1) = (\theta^*)^{x_1}(1-\theta^*)^{1-x_1}$. By Independence, $\frac{P^2(1,1)}{P^2(0,1)} = \frac{P^1(1)}{P^1(0)} = \frac{\theta^*}{1-\theta^*}$ and $\frac{P^2(1,0)}{P^2(0,0)} = \frac{P^1(1)}{P^1(0)} = \frac{\theta^*}{1-\theta^*}$ Define $\theta_2 := \frac{P^2(1,1)}{\theta^*} = \frac{P^2(0,1)}{1-\theta^*}$ and $\gamma_2 := \frac{P^2(1,0)}{\theta^*} = \frac{P^2(0,0)}{1-\theta^*}$. Then we have

$$P^2(x_1 x_2) = (\theta^*)^{x_1}(1-\theta^*)^{1-x_1} \times (\theta_2)^{x_2}(\gamma_2)^{1-x_2}.$$

Moreover by Marginal Consistency, $P^2(1,1) + P^2(1,0) = P^1(1)$ and so $\theta^*\theta_2 + \theta^*\gamma_2 = \theta^*$ and in particular $\theta_2 + \gamma_2 = 1$ given that $\theta^* > 0$ by the full support assumption. Proceed inductively. Assume that the representation holds for n. Invoking Independence as above, $\theta_{n+1} := \frac{P^{n+1}(x^n,1)}{P^n(x^n)}$ and $\gamma_{n+1} := \frac{P^{n+1}(x^n,0)}{P^n(x^n)}$ are independent of x^n, and the same argument establishes that P^{n+1} has the desired representation. Moreover, $\theta_{n+1} + \gamma_{n+1} = 1$ holds iff Marginal Consistency holds.

Suppose $\{P^n\}$ satisfy Independence and the property that $P^n(\Omega^{n-1}1) = \theta^*$ for all n. Then using the representation we have that $\theta_n = P^n(\Omega^{n-1}1) = \theta^*$ and $\gamma_n = P^n(\Omega^{n-1}0) = 1 - P^n(\Omega^{n-1}1) = 1 - \theta^*$. ■

6.2 LSN

To simplify exposition, for any $x \in \Omega^\infty$ and n write

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

6.2.1 Representing LSN and Weak Independence

Lemma 2 *P^n satisfies Weak LSN iff for each $r \in [0,1]$ there exists a strictly decreasing function $f^n(\cdot|r)$ on $[0,1]^n$ such that for any $x^n \in \Omega^n$,*

$$P^n(x^n) = f^n(d(x^1), ..., d(x^n)|\overline{x}^{n-1}).$$

P^n satisfies LSN if the dependence of f^n on $\overline{x}^{n-1}$ is replaced with dependence on $d(x^{n-1})$. P^n satisfies Strong LSN iff for each f^n is constant in its last argument.

Proof. For any $x, y \in \Omega^n$ s.t. $d(x^{n-1}) = d(y^{n-1})$ and $d(x^i) \leq d(y^i)$ for all $i \leq n$, LSN implies $P^n(x^n) \geq P^n(x^n)$, with $d(x^i) < d(y^i)$ for some $i \leq n$ implying $P^n(x^n) \geq P^n(x^n)$. Therefore there exists a function $f^n : [0,1]^{n+1} \to [0,1]$ s.t.

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

and f^n is strictly decreasing in all arguments $d(x^i)$ for $i \neq n-1$. Wlog $f^n(\cdot|d(x^{n-1}))$ can be presumed strictly decreasing in all arguments. Conversely, if P^n admits such a representation, then LSN is implied. A similar argument establishes the desired characterization of Weak LSN and Strong LSN. ■

Lemma 3 *A family of full support beliefs $\{P^n\}_{n=1}^{\infty}$ satisfies LSN and Weak Independence iff for each $n > 1$ there exists a continuous $g^n : [0,1]^2 \to [0,1]$ that is strictly decreasing in its first argument such that for any N and $x^N \in \Omega^N$,*

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

where $g^1(d(x^1), \overline{x}^0) := (\theta^*)^{x_1}(1-\theta^*)^{1-x_1}$. If LSN is replaced with Strong LSN then g^n is constant in its last argument.

Proof. Consider a family of full support beliefs. Begin with sufficiency of Weak LSN and Weak Independence. For each n and $\frac{r}{n}$ fix some $y^{n,r} \in \Omega^n$ with $\sum_{i=1}^{n} y_i^{n,r} = r$. By the full support assumption, $P^n(y^{n,r}) > 0$.

Step 1: For any r and $x^n x_{n+1} \in \Omega^{n+1}$ s.t. $\frac{\sum_{i=1}^{n} x_i^n}{n} = \frac{r}{n}$,

$$\frac{P^{n+1}(x^n x_{n+1})}{P^n(x^n)} = \frac{P^{n+1}(y^{n,r} x_{n+1})}{P^n(y^{n,r})}.$$

This just restates the conclusion of Weak Independence.

Step 2: Show that there exists a function g^{n+1} on $[0,1]^2$ that is strictly decreasing in its first argument and for any x^{n+1},

$$g^{n+1}(d(x^{n+1}), d(x^n)) = \frac{P^{n+1}(x^n x_{n+1})}{P^n(x^n)}.$$

By Weak LSN (and lemma 2), there exists f^n s.t.

$$P^n(x^n) = f^n(d(x^1), ..., d(x^n)|\overline{x}^{n-1}).$$

If LSN (resp. Strong LSN) holds, then we can take this function to depend on $d(x^{n-1})$ rather than $\overline{x}^{n-1}$ (resp. independent of the last argument $d(x^{n-1})$), and the proof below does not change.

Take any n and r and any $x_{n+1} \in \Omega$. Take the corresponding $y^{n,r}$ but suppress superscript r in the notation for exposition as needed. Then by the representation,

$$\frac{P^{n+1}(y^{n,r} x_{n+1})}{P^n(y^{n,r})} = \frac{f^{n+1}(d(y^1), ..., d(y^n), d(y^n x_{n+1})|\overline{y}^n)}{f^n(d(y^1), ..., d(y^n)|\overline{y}^{n-1})}.$$

Since n, r and thus $y^{n,r}$ are given, we can write the RHS ratio as a function g^{n+1} so that

$$\frac{P^{n+1}(y^{n,r} x_{n+1})}{P^n(y^{n,r})} = g^{n+1}(d(y^{n,r} x_{n+1}), \overline{y}^{n,r}) > 0.$$

By definition of d and by Weak LSN, for any realizations $x_{n+1}, x'_{n+1} \in \Omega$,

$$d(y^{n,r}x_{n+1}) \geq d(y^{n,r}x'_{n+1}) \implies P^{n+1}(y^{n,r}x_{n+1}) \geq P^{n+1}(y^{n,r}x'_{n+1})$$
$$\implies g^{n+1}(d(y^{n,r}x_{n+1}), \overline{y}^{n,r}) \leq g^{n+1}(d(y^{n,r}x'_{n+1}), \overline{y}^{n,r}),$$

where the conclusion is strict if the hypothesis is strict. Therefore, g^{n+1} is strictly decreasing in its first argument.

To complete the step, take any $x^n x_{n+1} \in \Omega^{n+1}$ with $\sum_{i=1}^n x_i^n = r$ and note that it must be that $\overline{x}^n = \overline{y}^{n,r}$ and $d^{n+1}(x^n x_{n+1}) = d^{n+1}(y^{n,r}x_{n+1})$. Then Step 1 yields

$$g^{n+1}(d(x^n x_{n+1}), \overline{x}^n) = g^{n+1}(d(y^{n,r}x_{n+1}), \overline{y}^{n,r}) = \frac{P^{n+1}(y^{n,r}x_{n+1})}{P^n(y^{n,r})} = \frac{P^{n+1}(x^n x_{n+1})}{P^n(x^n)}.$$

This argument goes through even if we impose LSN since $\overline{x}^n = \overline{y}^{n,r}$ implies $d(x^n) = d(y^{n,r})$.

Step 3. Complete the proof of sufficiency.

Define $g^1(d(x^1)) := P^1(x_1) = (\theta^*)^{x_1}(1-\theta^*)^{1-x_1}$. Apply Step 2 iteratively to obtain that for any N and $x^N \in \Omega^N$,

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

yielding the desired functional form.

Finally, observe that g^n is defined over a finite subset of $[0,1]^2$ for each $n > 1$ but is strictly decreasing in its first argument for each given value of the second argument. Moreover, by the full support assumption, it takes on strictly positive values. For each possible value of the second argument, the function can clearly be extended to a continuous strictly decreasing function in the first argument. Moreover, it can be continuously extended in its second argument by exploiting the fact that the mixture of decreasing functions is decreasing.

Step 4: Necessity of Weak Independence.

Suppose now that such a representation (using Weak LSN) exists. Take any N and $x^N, y^N \in \Omega^N$ s.t. $\overline{x}^N = \overline{y}^N$. Then $d^{n+1}(x^N x_{N+1}) = d^{n+1}(y^N x_{N+1})$ (and observe that if we assume LSN rather than Weak LSN then $\overline{x}^N = \overline{y}^N$ implies $d(x^N) = d(y^N)$). By the representation,

$$P^{N+1}(x^N x_{N+1}) = \left[\prod_{n=1}^{N} g^n(d(x^n), \overline{x}^{n-1})\right] \times g^{N+1}(d(x^N x_{N+1}), \overline{x}^N)$$

and similarly

$$P^{N+1}(y^N x_{N+1}) = \left[\prod_{n=1}^{N} g^n(d(y^n), \overline{y}^{n-1}))\right] \times g^{N+1}(d(y^N x_{N+1}), \overline{x}^n)$$

$$= \left[\prod_{n=1}^{N} g^n(d(y^n), \overline{y}^{n-1})\right] \times g^{N+1}(d(x^N x_{N+1}), \overline{x}^n)$$

where the last equality uses $d^{N+1}(y^N x_{N+1}) = d^{N+1}(x^N x_{N+1})$ and $\overline{x}^n = \overline{y}^n$. Therefore, given $P^n(y^n) > 0$ and $P^{n+1}(y^n x_{n+1}) > 0$ by the full support assumption,

$$\frac{P^{N+1}(x^N x_{N+1})}{P^{N+1}(y^N x_{N+1})} = \frac{\prod_{n=1}^{N} g^n(d(x^n), \overline{x}^{n-1})}{\prod_{n=1}^{N} g^n(d(y^n), \overline{y}^{n-1})} = \frac{P^N(x^N)}{P^N(y^N)}.$$

as desired. ■

6.2.2 Proof of Theorems 2 and 3

Lemma 4 *A family of full support beliefs $\{P^n\}_{n=1}^{\infty}$ satisfies Weak LSN and Weak Independence iff it admits a self-correcting bias representation:*

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

where $\theta_{i,\overline{x}^{i-1}}, \gamma_{i,\overline{x}^{i-1}} \in (0,1)$ are such that for any $x^{i-1}, y^{i-1} \in \Omega^{i-1}$ s.t. $\overline{x}^{i-1} = \overline{y}^{i-1}$,

$$d(x^{i-1}1) \leq d(y^{i-1}0) \iff \theta_{i,\overline{x}^{i-1}} \geq \gamma_{i,\overline{y}^{i-1}}.$$

whereas if $\{P^n\}$ satisfies Strong LSN (resp. LSN) then for any $x^{i-1}, y^{i-1} \in \Omega^{i-1}$ (resp. for any $x^{i-1}, y^{i-1} \in \Omega^{i-1}$ s.t. $d(x^{i-1}) = d(y^{i-1})$),

$$d(x^{i-1}1) \leq d(y^{i-1}0) \iff \theta_{i,\overline{x}^{i-1}} \geq \gamma_{i,\overline{y}^{i-1}}$$

$$d(x^{i-1}1) \leq d(y^{i-1}1) \iff \theta_{i,\overline{x}^{i-1}} \geq \theta_{i,\overline{y}^{i-1}}$$

$$d(x^{i-1}0) \leq d(y^{i-1}0) \iff \gamma_{i,\overline{x}^{i-1}} \geq \gamma_{i,\overline{y}^{i-1}}.$$

Moreover, $\gamma_{i,\overline{x}^{i-1}} = 1 - \theta_{i,\overline{x}^{i-1}}$ iff $\{P^n\}$ satisfy Marginal Consistency.

Proof. Begin with the representation in lemma 3 for Weak LSN and Weak Independence. Denote the bias towards heads on the i^{th} throw given a sample mean $\overline{x}^{i-1}$ by

$$\theta_{i,\overline{x}^{i-1}} := g^i(d(x^{i-1}1), \overline{x}^{i-1}) \in [0,1]$$

and similarly for the bias towards tails:

$$\gamma_{i,\overline{x}^{i-1}} := g^i(d(x^{i-1}0), \overline{x}^{i-1}) \in [0,1].$$

The representation can then be written

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

Since g^i is strictly decreasing in its first argument, the functions θ, γ must have the desired monotonicity property that for any $x^{i-1}, y^{i-1} \in \Omega^{i-1}$ s.t. $\overline{x}^{i-1} = \overline{y}^{i-1}$,

$$d(x^{i-1}1) \leq d(y^{i-1}0) \iff \theta_{i,\overline{x}^{i-1}} \geq \gamma_{i,\overline{y}^{i-1}}.$$

Conversely, if we have such a representation, then for any sequences $x^{i-1}1, y^{i-1}1 \in \Omega^{i-1}$ ending in a heads we have

$$\overline{x}^{i-1} = \overline{y}^{i-1} \text{ and } d(x^{i-1}1) \leq d(y^{i-1}1) \implies \theta_{i,\overline{x}^{i-1}} \geq \theta_{i,\overline{y}^{i-1}},$$

and for any sequences $x^{i-1}0, y^{i-1}0 \in \Omega^{i-1}$ ending in tails we have

$$\overline{x}^{i-1} = \overline{y}^{i-1} \text{ and } d(x^{i-1}0) \leq d(y^{i-1}0) \implies \gamma_{i,\overline{x}^{i-1}} \geq \gamma_{i,\overline{y}^{i-1}},$$

with strict inequality on the left side implying strict inequality on the right. So we can define functions $f^i(d(x^{i-1}1), \overline{x}^{i-1}) := \theta_{i,\overline{x}^{i-1}}$ and $h^i(d(y^{i-1}0), \overline{y}^{i-1}) := \gamma_{i,\overline{y}^{i-1}}$ that are both strictly decreasing in their first argument. These functions are connected by the condition that

$$\overline{x}^{i-1} = \overline{y}^{i-1} \text{ and } d(x^{i-1}1) = d(y^{i-1}0) \implies \theta_{i,\overline{x}^{i-1}} = \gamma_{i,\overline{y}^{i-1}},$$

in which case $f^i(d(y^{i-1}0), \overline{y}^{i-1}) = f^i(d(x^{i-1}1), \overline{x}^{i-1}) = \theta_{i,\overline{x}^{i-1}} = \gamma_{i,\overline{y}^{i-1}} = h^i(d(y^{i-1}0), \overline{y}^{i-1})$. Therefore $f(\cdot, \overline{x}^{i-1})$ and $h(\cdot, \overline{x}^{i-1})$ coincide on the intersection of their domains. Consequently, together the functions define a strictly decreasing function $g(\cdot, \overline{x}^{i-1})$ on the union of the domains, and we can write

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

By lemma 3, beliefs satisfy Weak LSN and Weak Independence.

The corresponding arguments for LSN and Strong LSN are analogous.

Finally, compute that for any x^N,

$$\begin{aligned} &\frac{\sum_{x_{N+1}} P^{N+1}(x^N x_{N+1})}{P^N(x^N)} \\ = &\frac{\prod_{n=1}^{N} g^n(d(x^n), d(x^{n-1})) \times \sum_{x_{N+1}} \left[g^{N+1}(d(x^N x_{N+1}), d(x^N))\right]}{\prod_{n=1}^{N} g^n(d(x^n), d(x^{n-1}))} \\ &= \sum_{x_{N+1}} \left[g^{N+1}(d(x^N x_{N+1}), d(x^N))\right] \\ &= \theta_{i,d(x^{i-1}1),d(x^{i-1})} + \gamma_{i,d(x^{i-1}0),d(x^{i-1})}. \end{aligned}$$

It follows that Marginal Consistency holds iff $\frac{\sum_{x_{N+1}} P^{N+1}(x^N x_{N+1})}{P^N(x^N)} = 1$ iff $\gamma_{i,d(x^{i-1}0),d(x^{i-1})} = 1 - \theta_{i,d(x^{i-1}1),d(x^{i-1})}$, as was to be shown. ■

6.3 Proof of Theorems 4 and 5

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

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

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 assumptions, 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^*,

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

By the Law of Large Numbers, $p^{\theta^*}(x^\infty | \lim_{n\to\infty} \overline{x}^n = \theta^*) = 1$. 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,

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

Because g 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

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

for all $i > N$. Hence,

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

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

Lemma 6 *Suppose $\mu \in \Delta(\Theta)$ and each $P^n \in \Delta(\Omega^n)$ have full support. Assume LSN and Weak Independence, and consider a representation where g^i is strictly increasing in its first argument for each i.*

1. If $g^i = g$ for all $i >$ and g is strictly increasing in its first argument and continuous in its second, then $p^{\theta^*}(\lim_{n\to\infty} P^n(\theta^*|x^n) \to 1) = 1$, that is,

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

2. If $g^i \to c$ uniformly faster than $\frac{1}{n^2} \to 0$ for all θ, where g^i is strictly increasing in its first argument and continuous in its second argument for all i and $c > 0$ is a constant function, then

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

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

By an identical argument to the one in Lemma 5, for $\epsilon = \min_{\theta \in \Theta \setminus \{\theta^*\}} |\theta - \theta^*|$, there exists N such that for all $i > N, |\overline{x^i} - \theta^*| < \frac{\epsilon}{4}$, and

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

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

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

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

This implies thatTherefore, to prove the result, it suffices to show that $\prod_{i=N}^n \frac{g(\frac{\epsilon}{2}, \overline{x}^i)}{g(\frac{\epsilon}{4}, \overline{x}^i)} \to 0$.

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

A sufficient condition for the result $\lim_{n\to\infty} P^n(\theta^*|x^n) \to 1$ *a.s.*-p^{θ^*} is that $\prod_{i=N}^n \frac{g^i(\frac{\epsilon}{2}, \overline{x}^i)}{g^i(\frac{\epsilon}{4}, \overline{x}^i)} \to 0$. To see this observe that, given the preceding, $\prod_{i=N}^n \frac{g^i(\frac{\epsilon}{2}, \overline{x}^i)}{g^i(\frac{\epsilon}{4}, \overline{x}^i)} \to 0$ implies

$$\frac{P_\theta^n(x^n)}{P_{\theta^*}^n(x^n)} < \prod_{i=1}^{N-1} \frac{g^i(|\overline{x}^i - \theta|, \overline{x}^i)}{g^i(|\overline{x}^i - \theta^*|, \overline{x}^i)} \times \prod_{i=N}^{n} \frac{g^i(\frac{\epsilon}{2}, \overline{x}^i)}{g^i(\frac{\epsilon}{4}, \overline{x}^i)} \to 0$$

and so $\frac{P_\theta^n(x^n)}{P_{\theta^*}^n(x^n)} \to 0$, and in particular $\lim_{n\to\infty} P^n(\theta^*|x^n) \to 1$ *a.s.*-p^{θ^*}.

So consider the first assumption in the lemma. Since the assumption restricts $g^i = g$ for $i > 1$, we take $N > 1$. We show that $\prod_{i=N}^{n} \frac{g(\frac{\epsilon}{2},\overline{x}^i)}{g(\frac{\epsilon}{4},\overline{x}^i)} \to 0$. Since g is continuous in its second argument and since $\overline{x}^i \to \theta^*$, we have $\frac{g(\frac{\epsilon}{2},\overline{x}^i)}{g(\frac{\epsilon}{4},\overline{x}^i)} \to \frac{g(\frac{\epsilon}{2},\theta^*)}{g(\frac{\epsilon}{4},\theta^*)} < 1$. In particular there exists M and $\overline{\epsilon} > 0$ s.t. $\frac{g(\frac{\epsilon}{2},\overline{x}^i)}{g(\frac{\epsilon}{4},\overline{x}^i)} < 1 - \overline{\epsilon}$ for all $i > M$. But then

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

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

as desired.

Step 2: Establish the result under the second 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 $g \leq 1$. 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

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

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

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

which we shall proof happens a.s-p^{θ^*}. 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 $\frac{g^n(|\overline{x}^n-\theta^*|,\overline{x}^n)}{g^n(|\overline{x}^n-\theta|,\overline{x}^n)} > 1$ (since g^n is strictly increasing 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 $n > N := max\{M, K\}$,[11]

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

[11] 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 $P(\theta^*|x^n) \not\to 1$ a.s.-p^{θ^*}. 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 $\frac{g^n(|\overline{x}^n-\theta^*|,\overline{x}^n))}{g^n(|\overline{x}^n-\theta|,\overline{x}^n)} > 1$ for all $n > N$, it must be that $\lim_{n\to\infty}\frac{P^{\theta^*}(x^n)}{P^{\theta}(x^n)} = \prod_{i=1}^{\infty}\frac{g^i(u(|\sum_{j\le i}\frac{x_i}{i}-\theta^*|))}{g^i(u(|\sum_{j\le i}\frac{x_i}{i}-\theta))} > 0$. Thus, $P(\theta^*|x^n) > 0$ a.s.-p^{θ^*}. ■

6.4 Evolution

6.4.1 Proof of Proposition 1

We start with a convenient observation about the normalizing constants Z^i_θ and $Z^i_{1-\theta}$ in the representation.

Lemma 7 *Let Z^i_θ be the constant associated with the representation 1 for parameter θ. Then, $Z^i_\theta = Z^i_{1-\theta}$ for all $i > 1$ and all $\theta \in [0,1]$.*

Proof. WLOG fix $\theta \ge \frac{1}{2}$. For each sequence $x^n \in \Omega^n$,let $y(x^n)$ be the sequence obtained by replacing the ones in $x^n$with zeros and the zeros with ones. It is easy to see that $|\overline{x}^n - \theta| = |\overline{y(x^n)} - (1-\theta)|$. Hence, $|\overline{x}^n - \theta| \le \epsilon^* \iff |\overline{y(x^n)} - (1-\theta)| \le \epsilon^*$ which implies $P^n_\theta(x^n) = P^n_{1-\theta}(y(x^n))$ by the representation (1). In particular, $P^n_\theta(1^n) = P^n_{1-\theta}(0^n)$. We show that for any $n \ge 2$,

$$\prod_{i=2}^{n}\frac{1}{Z^i_\theta} = \prod_{i=2}^{n}\frac{1}{Z^i_{1-\theta}}.$$

Consider two cases:

(i) $|1-\theta| > \epsilon^*$.

Then $\theta(1-\alpha)^{n-1}\prod_{i=2}^{n}\frac{1}{Z^i_\theta} = P^n_\theta(1^n) = P^n_{1-\theta}(0^n) = (1-(1-\theta))(1-\alpha)^{n-1}\prod_{i=2}^{n}\frac{1}{Z^i_{1-\theta}} \implies \prod_{i=2}^{n}\frac{1}{Z^i_\theta} = \prod_{i=2}^{n}\frac{1}{Z^i_{1-\theta}}$.

(ii) $|1-\theta| \le \epsilon^*$.

Then $\theta\alpha^{n-1}\prod_{i=2}^{n}\frac{1}{Z^i_\theta} = P^n_\theta(1^n) = P^n_{1-\theta}(0^n) = (1-(1-\theta))\alpha^{n-1}\prod_{i=2}^{n}\frac{1}{Z^i_{1-\theta}} \implies \prod_{i=2}^{n}\frac{1}{Z^i_\theta} = \prod_{i=2}^{n}\frac{1}{Z^i_{1-\theta}}$.

Using the equalities $\prod_{i=2}^{n}\frac{1}{Z^i_\theta} = \prod_{i=2}^{n}\frac{1}{Z^i_{1-\theta}}$ for all $n \ge 2$, a proof by induction yields $Z^i_\theta = Z^i_{1-\theta}$ for all $i > 1$. ■

Next, observe that by Bayesian updating, if there is N s.t. $\overline{x}^n > \frac{1}{2}$ and $|\overline{x}^n - \theta| < \epsilon^*$ for all $n > N$, then

$$P^n_{LSN}(\theta|x^n) = \frac{P^n_{LSN}(\theta|x^N)\alpha^{n-N}}{P^n_{LSN}(\theta|x^N)\alpha^{n-N} + P^n_{LSN}(1-\theta|x^N)(1-\alpha)^{n-N}}$$

and as usual,

$$P^n_{IID}(\theta|x^n) = \frac{P^n_{IID}(\theta|x^N)(\theta)^{k_{n-N}}(1-\theta)^{n-N-k_{n-N}}}{P^n_{IID}(\theta|x^N)(\theta)^{k_{n-N}}(1-\theta)^{n-N-k_{n-N}} + P^n_{IID}(1-\theta|x^N)(1-\theta)^{k_{n-N}}(\theta)^{n-N-k_{n-N}}},$$

where k_{n-N} is the number of heads that occur after the N^{th} throw (which satisfies $2k_{n-N} > n-N$ since that $\overline{x}^n > \frac{1}{2}$ for all $n > N$). Now we are ready to prove the proposition.

Proof of (i): We only establish the case in which $\theta = \hat{\theta} > \frac{1}{2}$ since the proof $\theta = 1-\hat{\theta}$ is analogous. Let $P^n_{LSN}(\theta|x^n)$ and $P^n_{IID}(\theta|x^n)$ be the posterior beliefs after observing signals x^n of the LSN and IID agents respectively.

By LLN, there exists N such that $\overline{x}^n > \frac{1}{2}$ and $|\overline{x}^n - \theta| < \epsilon^*$ for all $n > N$. Moreover, by Bayesian updating,

$$\frac{P^n_{LSN}(\theta|x^n)}{P^n_{LSN}(1-\theta|x^n)} = \frac{P^n_{LSN}(\theta|x^N)}{P^n_{LSN}(1-\theta|x^N)}\frac{\alpha^{n-N}}{(1-\alpha)^{n-N}} \text{ and } \frac{P^n_{IID}(\theta|x^n)}{P^n_{IID}(1-\theta|x^n)} = \frac{P^n_{IID}(\theta|x^N)}{P^n_{IID}(1-\theta|x^N)}\frac{(\theta)^{k_{n-N}}(1-\theta)^{n-N-k_{n-N}}}{(1-\theta)^{k_{n-N}}(\theta)^{n-N-k_{n-N}}},$$

where k_{n-N} is the number of heads that occur after the N^{th} throw. Then

$$\frac{P^n_{LSN}(\theta|x^n)}{P^n_{LSN}(1-\theta|x^n)} / \frac{P^n_{IID}(\theta|x^n)}{P^n_{IID}(1-\theta|x^n)} = \frac{P^n_{LSN}(\theta|x^N)}{P^n_{LSN}(1-\theta|x^N)}\frac{\alpha^{n-N}}{(1-\alpha)^{n-N}} / \frac{P^n_{IID}(\theta|x^N)}{P^n_{IID}(1-\theta|x^N)}\frac{(\theta)^{k_{n-N}}(1-\theta)^{n-N-k_{n-N}}}{(1-\theta)^{k_{n-N}}(\theta)^{n-N-k_{n-N}}}$$

$$= \left[\frac{P^n_{LSN}(\theta|x^N)}{P^n_{LSN}(1-\theta|x^N)} / \frac{P^n_{IID}(\theta|x^N)}{P^n_{IID}(1-\theta|x^N)}\right]\left[\frac{\alpha}{1-\alpha}\right]^{n-N}\left[\frac{\theta}{1-\theta}\right]^{n-N-2k_{n-N}} \to \infty,$$

since $\frac{\alpha}{1-\alpha}, \frac{\theta}{1-\theta} > 1$.

Hence, there exists N' such that for all $n > N'$,

$$\frac{P^n_{LSN}(\theta|x^n)}{P^n_{LSN}(1-\theta|x^n)} > \frac{P^n_{IID}(\theta|x^n)}{P^n_{IID}(1-\theta|x^n)},$$

which implies $P^n_{LSN}(\theta|x^n) > P^n_{IID}(\theta|x^n)$ for all $n > N'$, as desired.

Proof of (ii): Let $P^n_{LSN}(\theta|x^n)$ and $P^n_{IID}(\theta|x^n)$ be the posterior beliefs after observing signals x^n of the LSN and IID agents respectively.

Step 1: Show that

$$P^n_{LSN}(\theta|x^n) \geq P^n_{LSN}(\theta|x^n) \iff k^{LSN}_{x^n} \geq k^{IID}_{x^n}.$$

The optimal choice of agent A solves

$$U_{x^i}(k) = [\theta P^n(\theta|x^n) + (1-\theta)P^n(1-\theta|x^n)]u(k2+(1-k)) + [(1-\theta)P^n(\theta|x^n) + \theta P^n(1-\theta|x^n)]u(1-k).$$

The FOC is therefore

$$2[\theta P^n(\theta|x^n) + (1-\theta)P^n(1-\theta|x^n)]u'(k+1) = [(1-\theta)P^n(\theta|x^n) + \theta P^n(1-\theta|x^n)]u'(1-k),$$

which rearranges to

$$2\frac{u'(k+1)}{u'(1-k)} = \frac{1-\theta+\frac{1}{2}P^n(1-\theta|x^n)}{1-\theta+\frac{1}{2}P^n(\theta|x^n)}.$$

Since u is strictly concave, the LHS is strictly decreasing in k. Therefore $P^n_{LSN}(\theta|x^n) \geq P^n_{IID}(\theta|x^n) \iff k^{LSN}_{x^n} \geq k^{IID}_{x^n}$.

Step 2: Prove the result.

By step 1 and part (i) of the proposition, the LSN agent will eventually spend more time on the risky ground than the safe ground a.s. Furthermore, when the true parameter is $\theta = \hat{\theta} > \frac{1}{2}$ (the argument for $\theta = 1-\hat{\theta}$ is analogous), there will eventually be more heads than tails a.s. by LLN. Hence, a.s., the LSN agent will eventually have a higher return than the IID agent.

6.4.2 Proof of Proposition 2

The proof relies on the following statistical fact: in an infinite sequence of coin tosses, the probability of observing more heads than tails at every toss is equal to $2\theta - 1$ whenever $\theta > \frac{1}{2}$.

Lemma 8 *Assume $\theta > \frac{1}{2}$ and let $E = \{x^\infty | \overline{x}^n > \frac{1}{2}$ for all $n\}$. Then $p^\theta(E) = 2\theta - 1$.*

Proof. From the *Gambler's Ruin Problem,* we know that conditional on the first throw being heads, the probability of always having more heads than tails is equal to $1 - \frac{1-\theta}{\theta}$.[12] Hence, the probability of *always* having more heads than tails is equal to $\theta(1 - \frac{1-\theta}{\theta}) = 2\theta - 1$. ■

Suppose that $\theta = \hat{\theta}$ (the argument is analogous for $\theta = 1 - \hat{\theta}$). Consider

$$E = \{x^{\infty} | \overline{x^n} > \frac{1}{2} \text{ for all n} \in \mathbb{N}\},$$

the event where the sample mean exceeds $\frac{1}{2}$ at every i. Under our assumptions on $\hat{\theta}$ and ϵ^*, for any sequence $x \in E$, we will have that $|\overline{x}^n - \theta| < \epsilon^*$ for all n. Hence from the proof of Proposition 1, we see that for each $x \in E$, the LSN population grow at a faster rate than the IID population in every period, that is $\frac{\lambda^{c^{LSN}_{x^i}}}{\lambda^{c^{IID}_{x^i}}} > 1$ for each i. Therefore $\lim_{n\to\infty} \prod_{i=1}^{n} \frac{\lambda^{c^{LSN}_{x^i}}}{\lambda^{c^{IID}_{x^i}}} > 1$.

By lemma 8, $p^{\theta}(E) = 2\theta - 1$. Consider also the event $F = \{1\} \times \{0\} \times E$, where the first toss yields a head, the second a tails and the subsequent stream belongs to E. This event has probability $(1-\theta)\theta p^{\theta}(E) > 0$. Note that after the heads on the first toss, $P^{LSN}(\theta|1) = P^{IID}(\theta|1)$, after the tails on the second toss, $P^{LSN}(\theta|1,0) = \frac{\theta\alpha}{\theta\alpha + \theta(1-\alpha)} = \theta > \frac{1}{2} = P^{IID}(\theta|1,0)$, and then for all subsequent tosses we have $P^{LSN}(\theta|x^i) > P^{IID}(\theta|x^i)$ since $\overline{x}^n > \frac{1}{2}$ for all $n > 2$. That is, the LSN agents are weakly more confident about the true parameter than the IID agents for $i = 1$ and strictly so for all $i > 1$. Consequently, by step 1, $\lim_{n\to\infty} \prod_{i=1}^{n} \frac{\lambda^{c^{LSN}_{x^i}}}{\lambda^{c^{IID}_{x^i}}} > 1$ for each $x \in F$. Conclude that $p^{\theta}(lim_{n\to\infty} \prod_{i=1}^{n} \frac{\lambda^{c^{LSN}_{x^i}}}{\lambda^{c^{IID}_{x^i}}} > 1) \geq p^{\theta}(E \cup F) > \frac{1}{2}$.

References

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

[2] Benjamin, D. (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, D., D. Moore and M. Rabin (2018): "Biased Beliefs About Random Samples: Evidence from Two Integrated Experiments", working paper.

[4] Benjamin, D., M. Rabin and C. 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, R. (1966): "Limiting Behavior of Posterior Distributions when the Model is Incorrect", *Annals of Mathematical Statistics* 37, pp. 51-58.

[6] Chen, D., T. Moskowitz and K. 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] Diaconis, P. (1977): "Finite Forms of de Finetti's Theorem on Exchangeability", *Synthese* 36, pp 271-281.

[8] Diaconis, P. and D. Freedman (1980): "Finite Exchangeable Sequences", *The Annals of Probability* 8(4), pp. 745-764.

[12] In the Gambler's Ruin Problem literature, this probability is referred to as the probability of "never going broke" and $1 - \frac{1-\theta}{\theta}$ is the value for the case in which the gambler starts with one dollar.

[9] Epstein, L. (2006): "An Axiomatic Model of Non-Bayesian Updating", *Review of Economic Studies* 73, pp. 413-436.

[10] Epstein, L., J. Noor and A. Sandroni (2008): "Non-Bayesian Updating: A Theoretical Framework", *Theoretical Economics* 3(2), pp. 193-229.

[11] Epstein, L., J. Noor and A. Sandroni (2010): "Non-Bayesian Learning", *B.E. Journal of Theoretical Economics (Advances)* 10(1), article 3.

[12] He, K. (2022): "Mislearning from Censored Data: The Gambler's Fallacy and Other Correlational Mistakes in Optimal-Stopping Problems", *Theoretical Economics* 17(3), pp. 1269–1312.

[13] Kahneman, D. and A. Tversky (1972): "Subjective Probability: A Judgement of Representativeness", *Cognitive Psychology* 3(3), pp.430-454.

[14] Kerns, G. and G. Szekely (2006): "Definetti's Theorem for Abstract Finite Exchangeable Sequences", *Journal of Theoretical Probability* 19(3), pp. 589-608.

[15] Mueller, A., J. Spinnewijn and G. Topa (2021): "Job Seeker's Perceptions and Employment Prospects: Heterogeneity, Duration Dependence and Bias", *American Economic Review* 111(1), pp. 324-63.

[16] Noor, J. (2022): "Intuitive Priors", mimeo.

[17] Oppenheimer, D. and B. Monin (2009): "The Retrospective Gambler's Fallacy: Unlikely Events, Constructing the Past, and Multiple Universes", *Judgement and Decision-Making* 4(5), pp. 326-334.

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

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

[20] Rabin, M. and J. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias", *Quarterly Journal of Economics* 114(1), pp. 37-82.

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

[22] Sandroni, A. (2000): "Do Markets Favor Agents able to Make More Accurate Predictions?" *Econometrica* 8(6), pp. 1303-1341.

[23] Sedlmeier, P. and G. Gigerenzer (1997): "Intuitions about Sample Size: The Empirical Law of Large Numbers", *Journal of Behavioral Decision Making* 10(1), pp. 33-51.

[24] Suetens, S., C. Galbo-Jorgensen and J. 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.

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

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

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