WINPEC Working Paper Series No. E2611
September 2026

Local Representativeness and Distorted Bayesian Updating: A Finite-Urn Analysis

Kazumi Shimizu

Waseda INstitute of Political EConomy
Waseda University
Tokyo, Japan

Local Representativeness and Distorted Bayesian Updating: A Finite-Urn Analysis

Kazumi Shimizu*

*Faculty of Political Science and Economics, Waseda University, 1-6-1, Nishiwaseda, Shinjuku-city, Tokyo, 169-8050, Japan. E-mail: skazumi1961@gmail.com.This manuscript is currently under review. This version has not been peer reviewed.

Abstract

This paper develops a theoretical model of belief updating under local representativeness. Building on the finite-urn framework of Rabin (2002), we study a quasi-Bayesian agent who updates by Bayes' rule but mistakenly interprets independent and identically distributed signals as if they were drawn without replacement from a finite-urn. Under this misperception, posterior beliefs are systematically distorted, and these distortions may generate biased economic behavior.

Using a beta prior, we derive closed-form results for the benchmark pure-streak case in which the perceived urn is reset every two draws. This benchmark corresponds to the canonical setting in which gambler's fallacy reasoning is typically elicited, namely prediction after a short run of identical outcomes. After two consecutive successes, the quasi-Bayesian agent becomes more willing to invest than a Bayesian, whereas after two consecutive failures, the same agent becomes more pessimistic and less willing to invest. Notably, these responses can move in the *opposite* direction from what a naive reading of local representativeness alone would suggest, because the effect of the finite-urn misperception on posterior beliefs is shaped by Bayesian updating and normalization.

The paper also provides a behavioral characterization of the finite-urn updating rule under blockwise reset. We show how a qualitative principle of local representativeness, together with a blockwise reset assumption and a within-block predictive structure, yields the generalized finite-urn likelihood kernel used in the analysis. This characterization does not amount to a full axiomatization of dynamic choice, but it clarifies the short-run cognitive logic underlying the model.

Overall, the paper offers a tractable framework for analyzing how local representativeness distorts Bayesian inference, thereby affecting posterior beliefs and economic decision-making.

JEL codes: C11, D03, D91.

Keywords: local representativeness, law of small numbers, Bayesian updating, finite-urn model, belief distortion.

1 Introduction

Individuals routinely rely on probability assessments when making decisions across various domains, including investment strategies, career trajectories, and personal goal-setting. However, human judgment in probability estimation is often systematically flawed. One of the most extensively documented cognitive biases is *local representativeness*, that is, the law of small numbers: the erroneous belief that small samples must accurately represent the general population (Tversky & Kahneman 1971; Kahneman & Tversky 1972).[1] Much of the evidence for local representativeness comes from laboratory experiments and field data (related evidence from both laboratory and field settings is reviewed in DellaVigna (2009)). While Rabin (2002) formalizes local representativeness as "belief in the law of small numbers," Rabin & Vayanos (2010) extend this framework to provide a unified account of both gambler's fallacy–type reversals (when the underlying state is known to be constant) and hot–hand–type persistence (when the state is uncertain and time-varying). Their key insight is that an agent with the gambler's fallacy exaggerates the magnitude of changes in the underlying state but underestimates their duration, so that short streaks generate predicted reversals while long streaks generate predicted continuation.

On the other hand, a growing literature suggests that human judgment can often be usefully described in Bayesian terms. For example, Patel (2011) shows that the mind's ability to make predictions can be grounded in Bayesian theory at a computational level. Griffiths & Tenenbaum (2006) suggest that humans make remarkably accurate predictions about everyday events by employing a form of Bayesian inference. Lin & Garrido (2022) state that Bayes' rule can derive predictions for people's perception, cognition, and learning in complex situations.

Despite a vast body of literature on local representativeness and Bayesian updating, systematic theoretical analysis of how these two phenomena interact within a unified mathematical framework has been limited. This study develops such a framework by formalizing how decision-makers (DMs) who exhibit local representativeness distort their probability assessments through Bayesian updating.

[1] The gambler's fallacy can be regarded as a specific manifestation of the law of small numbers: it arises when the over-application of representativeness leads people to expect short-term reversals. The fallacy has been extensively studied. Clotfelter & Cook (1993) find that lottery participants systematically avoid recently drawn numbers, despite their statistical independence.

1.1 Relation to Prior Work

The finite-urn approach to biased inference under local representativeness originates in Rabin (2002), who showed that the law of small numbers can be modeled by assuming that DMs mistakenly interpret independent signals as if they were drawn without replacement from a finite population. In Rabin's framework, this misperception generates systematic distortions in belief updating and provides a tractable account of the gambler's fallacy.

The influence of this framework is also reflected in standard behavioral economics textbooks. For example, Cartwright (2018) presents the finite-urn logic in Chapter 5 and discusses how the finite-urn model can generate gambler's-fallacy type updating. Similarly, Dhami (2016) discusses Rabin's model in Chapter 1.2.2 and uses it to illustrate how biased inference under uncertainty can produce overconfidence. In this sense, the finite-urn model has become a standard tool in behavioral economics for analyzing belief distortions under local representativeness.

Rabin & Vayanos (2010) subsequently show that the gambler's fallacy—the same psychological bias underlying Rabin's (2002) framework—can, within a richer dynamic model, generate hot-hand-type beliefs as an endogenous implication. In particular, Rabin & Vayanos (2010) demonstrate that the gambler's and hot-hand fallacies need not be viewed as unrelated phenomena. In their framework, the same gambler's fallacy logic can generate different behavioral patterns depending on whether agents are learning about an underlying success rate (or state). When there is no such learning motive, the bias implies mean reversion: after a streak of successes, agents expect failures to become more likely. When agents infer an underlying state from observed signals, however, a streak of successes can lead them to over-infer that the state is favorable (or that the success probability is high), so that the inference effect dominates the depletion effect and generates hot-hand beliefs. More broadly, this line of work shows how a simple misperception of the signal-generating process can produce a rich range of belief distortions.

While Rabin & Vayanos (2010) is closely related, the present paper follows Rabin (2002) more directly for three reasons. First, our focus is on distorted Bayesian updating about a fixed parameter θ, abstracting from latent-state learning; the dynamic state-space structure of Rabin & Vayanos (2010) is therefore not needed for our purposes.

Second, our main analytical contribution—the closed-form finite-urn likelihood kernel and its behavioral characterization under blockwise reset—is naturally developed in a discrete finite-urn environment rather than in the linear-Gaussian state-space framework of Rabin & Vayanos (2010). Third, the blockwise reset structure we introduce is most naturally interpreted within the finite-urn logic of Rabin (2002), where the perceived urn size N and the block length M have direct cognitive interpretations.

Our contribution is threefold.

First, we show that when local representativeness is combined with Bayesian updating, the resulting comparative statics can differ sharply from the naive reversal intuition usually associated with the gambler's fallacy. In the benchmark pure-streak case of two identical outcomes within the same block, the the distorted-updating agent becomes more willing to invest after two consecutive successes and less willing to invest after two consecutive failures. Thus, the finite-urn distortion does not simply translate into a mechanical expectation of short-run reversal; once it interacts with Bayesian updating and normalization, it can generate posterior shifts and economic behavior that move in the opposite direction from that naive intuition.

Second, while Rabin (2002) introduces the finite-urn specification as a tractable modeling device, we provide a behavioral characterization of the corresponding updating rule. More specifically, we show how the generalized finite-urn likelihood kernel can be derived from a qualitative principle of local representativeness, a blockwise reset assumption, and a within-block predictive structure. We emphasize that this is a behavioral characterization, not a complete axiomatization of dynamic choice. Its value lies in clarifying the short-run cognitive logic underlying the model and in making explicit what is assumed and what is derived.

Third, the paper makes the domain of the benchmark analysis more transparent by separating the intensity of the bias from its temporal scope. The perceived urn size N governs how strongly the finite-urn distortion affects updating, while the block length M determines how long that logic is applied before beliefs are reset. This separation clarifies why the benchmark two-signal case is analytically central, while longer within-block histories require more limited treatment and are better interpreted through simulation-based robustness checks rather than as immediate extensions of the closed-form comparative statics.

1.2 Relation to Other Models of Belief Distortion

It is useful to contrast our approach with two influential frameworks that also generate systematic belief distortions: local thinking and diagnostic expectations.

Local thinking (Gennaioli & Shleifer 2010) emphasizes that DMs often rely on a small subset of considerations that are most salient or "come to mind," rather than fully processing all available information. This mechanism captures how attention is selectively allocated across possible states, leading agents to overreact to vivid or representative scenarios. Diagnostic expectations (Bordalo et al. 2018, 2020) extend this idea to forecasting. Here, individuals exaggerate the likelihood of outcomes that appear especially representative or "diagnostic" of the current situation. For example, after observing strong economic growth, people may come to expect that strong growth will continue into the future.

Our model differs from both frameworks at a more basic level. In local thinking and diagnostic expectations, the distortion operates primarily through the *weight placed on states or scenarios*: some states become disproportionately salient or diagnostic relative to others. In our model, by contrast, the distortion operates through the *perceived data-generating process itself*. Agents continue to update using Bayes' rule, but they do so with a misspecified likelihood, because they treat objectively i.i.d. signals as if they were drawn without replacement from a finite urn. Thus, the source of non-Bayesian behavior in our framework is not a direct reweighting of states, but a structural misperception of how signals are generated. This distinction matters because it yields a different class of predictions: rather than implying a general overreaction to salient or diagnostic states, the finite-urn account generates distortions that depend sharply on the recent signal history and on the perceived urn size N. In that sense, our framework contributes to this literature by identifying a distinct mechanism of belief distortion that remains Bayesian in form at the updating stage, but becomes non-Bayesian through a misspecified likelihood.

This difference yields sharper sequence-specific predictions. In our model, distortions arise specifically after short streaks of identical outcomes and vary systematically with the perceived urn size N. For example, after two consecutive successes, the agent overestimates the probability of future success, whereas after two consecutive failures, the agent underestimates it; in both cases, the magnitude of the distortion increases as

N becomes smaller. These comparative statics help distinguish the finite-urn account from attention-based theories, since in our model distortions arise after specific short streaks and become stronger as N becomes smaller.

In this sense, the paper contributes to the literature not only by contrasting with attention-based accounts, but by showing that a likelihood-based distortion generates sharper and more falsifiable sequence-specific predictions.

1.3 Organization of the Paper

The remainder of the paper is organized as follows. Section 2 introduces the finite-urn model and provides its behavioral characterization under blockwise reset. We present Axioms C1 and C2, Structure 1, and Proposition 1, which establish the finite-urn representation result. Section 3 derives the main analytical results, showing how posterior beliefs and investment decisions are systematically distorted after consecutive successes and failures. We provide closed-form expressions for posterior distributions and optimal investment levels, and we conduct comparative statics with respect to the bias parameter N, holding the block length fixed at $M = 2$. Section 4 discusses the implications and limitations of the model, including potential extensions and empirical applications. Technical details and proofs are provided in Appendices A and B.

2 Model

2.1 Finite-Urn Model

In constructing a general model, we usually employ application-neutral language. However, for clarity and motivation, it is useful to begin with a concrete setting inspired by Rabin (2002). Consider an urn containing 10 balls, evenly divided between red and black. The DM knows that the total number of balls is $N = 10$ and that the probability of drawing a red ball is $1/2$. If each ball is replaced after being drawn, then the process is independent and identically distributed (i.i.d.), so after one red draw the probability of another red draw remains $1/2$.

Now suppose instead that the DM mistakenly interprets the same process as if the second draw were made *without replacement* from the same urn. In that case, after one red draw, the perceived probability of drawing red again becomes

$$\frac{(0.5 \times 10) - 1}{10 - 1} = \frac{4}{9},$$

which is smaller than the correct probability 1/2. Equivalently, the DM overestimates the probability of a black draw, assigning it probability 5/9. This captures the basic intuition of *local representativeness*: the mistaken expectation that even a small sample should quickly resemble the population from which it is drawn.

We refer to such an individual as a *quasi-Bayesian*. The key point is that the individual is Bayesian in the sense of updating beliefs using Bayes' rule, but misperceives the signal-generating process. Signals are in fact generated by an i.i.d. process, yet the agent behaves as if they were drawn without replacement from a finite population. By contrast, an agent who correctly interprets the process as i.i.d. and conducts the Bayesian update will be called a *Bayesian*.[2]

To define the setting more formally, let $\theta \in (0, 1)$ denote the true probability of success in an infinite sequence of binary signals S_t, where $s_t = 1$ denotes success and $s_t = 0$ denotes failure. Conditional on θ, the signals are objectively i.i.d. with

$$P(s_t = 1 \mid \theta) = \theta, \qquad P(s_t = 0 \mid \theta) = 1 - \theta.$$

The agent holds a prior density $f(\theta)$ over θ, with $f(\theta) > 0$ on $(0, 1)$. If one interprets success and failure as, for example, the success or failure of an investment project, then the model can be applied directly to economic decision-making.

The distinguishing feature of the quasi-Bayesian is that, although she starts from a prior over θ and updates by Bayes' rule, she evaluates new signals through a finite-urn lens. Specifically, she behaves as if signals were drawn without replacement from an urn of size N, containing θN success signals and $(1 - \theta)N$ failure signals. Importantly, this distortion does not alter the posterior after the first observed signal. After one success, the quasi-Bayesian still updates from the prior in the standard Bayesian way, because the first update uses the likelihood of the observed success relative to the original prior, so the finite-urn correction enters only in the perceived probability of

[2]This terminology is in line with the broader "quasi-Bayesian" perspective often used in behavioral economics; see Camerer & Thaler (2003).

the next signal. Thus, after one success has been observed, she uses

$$\frac{\theta N - 1}{N - 1}, \tag{1}$$

rather than θ, as the perceived conditional probability of another success and hence as the likelihood factor for the next update.

This distortion becomes weaker as N becomes large:

$$\lim_{N \to \infty} \frac{\theta N - 1}{N - 1} = \theta.$$

Hence, larger values of N correspond to milder bias, while smaller values of N correspond to stronger bias. In this sense, N serves as a natural measure of the intensity of the bias. The finite-urn model therefore makes it possible, at least in principle, to study how stronger or weaker local-representativeness distortions affect posterior beliefs and economic behavior.

At the same time, if one were to apply the without-replacement logic indefinitely over a long sequence, the resulting inference would become implausibly strong. For example, after many identical draws the agent would eventually behave as if only the opposite signal remained, thereby generating an unrealistically strong expectation of reversal. To avoid this implication, we assume that the finite-urn logic is applied only over a finite horizon of M signals. Following Rabin (2002), we therefore impose a *blockwise reset* assumption: the agent applies the finite-urn logic only within a finite block of length M, and the perceived urn is reset at the beginning of the next block. The benchmark case in the main analysis sets $M = 2$, because this is the first nontrivial case and already delivers the central comparative statics in closed form, although the same logic can be extended to any fixed finite block length $M \geq 2$. [3]

For the purposes of the main analysis, the essential object is therefore the finite-urn updating rule applied within a block. After one success, the quasi-Bayesian evaluates the next success using (1); after one failure, the analogous correction is

$$\frac{(1 - \theta)N - 1}{N - 1}.$$

[3]This distinction of N and M is useful. The parameter N governs the strength of the finite-urn distortion, while M determines how long the agent applies that logic before resetting the perceived urn.

More generally, within a block, the perceived probability of the next signal depends on how many successes and failures have already been observed in that block.

The finite-urn model above is the workhorse specification used in the analysis of Section 3. Before turning to its implications, however, it is useful to show that the same updating rule can be interpreted as arising from a simple set of behavioral restrictions under blockwise reset.

2.2 Behavioral Characterization of the Finite-Urn Rule

Our aim is deliberately limited. We do not attempt a full preference-based axiomatization of dynamic choice under uncertainty. Rather, we focus directly on belief formation and show how the finite-urn updating rule can be behaviorally characterized by a small set of intuitive within-block restrictions. The resulting formal derivation is reported in Appendix A.

2.2.1 Belief-Updating Environment

Consider a DM who observes a sequence of binary signals

$$\{s_1, s_2, \dots\}, \qquad s_t \in \{0, 1\},$$

where 1 denotes success and 0 denotes failure. Signals are generated by an unknown state $\theta \in \Theta = [0, 1]$, where θ is the true probability of success. Under the objective data-generating process, signals are conditionally i.i.d.:

$$P(s_t = 1 \mid \theta) = \theta.$$

The DM starts from a prior belief μ_0 over Θ. After observing a finite history

$$h_t = (s_1, \dots, s_t),$$

the DM forms a posterior belief $\mu_t(\cdot \mid h_t)$. Let

$$n_1(h_t) = \sum_{\tau=1}^{t} s_\tau,$$
$$n_0(h_t) = \sum_{\tau=1}^{t} (1 - s_\tau),$$

so that $n_1(h_t) + n_0(h_t) = t$.

Under standard Bayesian updating, the posterior is

$$\mu_t^B(\theta \mid h_t) = \frac{L^B(\theta \mid h_t)\mu_0(\theta)}{\int_\Theta L^B(\tilde{\theta} \mid h_t)\mu_0(\tilde{\theta})\, d\tilde{\theta}},$$

where $L^B(\theta \mid h_t)$ denotes the Bayesian likelihood function, that is, the probability of the observed history h_t given θ, viewed as a function of θ. For an ordered history h_t with n_1 successes and n_0 failures,

$$L^B(\theta \mid h_t) = \theta^{n_1}(1 - \theta)^{n_0}.$$

We now introduce a distorted updating rule in which the DM partitions the signal sequence into finite blocks of length M, applies a finite-urn logic within each block, and resets the perceived urn between blocks.

2.2.2 Block Structure

Fix an integer $M \geq 2$, which we interpret as the DM's *processing block length*. The DM partitions the signal sequence into consecutive non-overlapping blocks of length M:

$$B_1 = (s_1, \ldots, s_M), \quad B_2 = (s_{M+1}, \ldots, s_{2M}), \quad \ldots$$

Within each block, the DM behaves as if signals were drawn without replacement from a perceived urn of size N, where $N \geq M$. At the start of each new block, the perceived urn is reset.

Thus, the model has two distinct parameters:

- N: perceived urn size, which governs the intensity of the finite-urn distortion,

- M: block length, which governs how many signals are processed before the perceived urn is reset.

The benchmark case in the main analysis corresponds to $M = 2$, but the same within-block logic extends to any fixed finite $M \geq 2$.

2.2.3 Axiom C1: Local Representativeness Within a Block

The first restriction captures the idea that short sequences are perceived as excessively representative of the underlying population, consistent with the law of small numbers (Tversky & Kahneman 1971).

Axiom 1 (Local Representativeness Within a Block)**.** *Within a block, the DM expects small samples of signals to resemble the underlying population too closely. Consequently, short runs of identical outcomes are perceived as overly imbalanced relative to that population, and the same outcome is therefore viewed as less representative of the remaining signals than under the Bayesian benchmark.*

Axiom 1 is qualitative. It captures the direction of the bias associated with local representativeness, but it does not by itself determine a unique predictive rule. To obtain a tractable specification, we will introduce one concrete within-block structure, Structure 1, that implements this idea in the finite-urn setting.

2.2.4 Axiom C2: Blockwise Reset

The second restriction formalizes the reset assumption.

Axiom 2 (Blockwise Reset)**.** *There exists a fixed finite block length $M \geq 2$ such that the DM processes signals block by block. Within each block, the finite-urn depletion process depends only on the signals observed since the start of that block. At the beginning of each new block, the perceived urn is reset, so that successes and failures observed in previous blocks do not contribute to further depletion in the current block.*

Axiom 2 prevents the DM from carrying depletion indefinitely across a long sequence of signals and thereby preserves the idea that local representativeness is a short-run inference heuristic.

Together, Axioms 1 and 2 provide a natural behavioral description of the bias and of its short-run scope. To obtain a sharp formal characterization, we now introduce a specific within-block predictive structure that gives the local-representativeness idea a precise formal expression.

2.2.5 Structure 1: Within-Block Predictive Rule under Local Representativeness

Structure 1 (Within-Block Predictive Rule under Local Representativeness)**.** *Fix a block of length M. Suppose that, within the current block, the DM has already observed n_1 successes and n_0 failures, where $n_1 + n_0 < M$. Then the DM forms a belief about the* next *signal as follows:*

$$q_1(\theta \mid n_1, n_0) = \frac{\theta N - n_1}{N - n_1 - n_0},$$
$$q_0(\theta \mid n_1, n_0) = \frac{(1-\theta)N - n_0}{N - n_1 - n_0}.$$

Here, $q_1(\theta \mid n_1, n_0)$ and $q_0(\theta \mid n_1, n_0)$ are the likelihood factors used to update the posterior density of θ after observing, respectively, a success or a failure at the next observation within the block.

In addition, updating is sequentially coherent within each block: the likelihood assigned to any within-block history is equal to the product of these one-step-ahead perceived probabilities along that history.

These expressions mean that, within a block, the DM behaves as if signals were drawn without replacement from an urn of size N. After each observed success, the perceived number of remaining success signals falls by one; after each observed failure, the perceived number of remaining failure signals falls by one.

For example, if the within-block history is $h = (S, S, F)$, then

$$L(\theta \mid S, S, F) = q_1(\theta \mid 0, 0) \cdot q_1(\theta \mid 1, 0) \cdot q_0(\theta \mid 2, 0)$$
$$= \frac{\theta N}{N} \cdot \frac{\theta N - 1}{N - 1} \cdot \frac{(1-\theta)N}{N - 2}.$$

Structure 1 formalizes the idea that, within a block, the DM behaves as if signals were drawn without replacement from a finite population. This finite-urn interpretation directly captures the local-representativeness intuition that short samples should resemble the population, leading to perceived depletion of the observed outcome type.

This specification is also symmetric between success and failure. The same predictive logic applies if we relabel success as failure, replace θ by $1-\theta$, and interchange the success and failure counts. Thus, the model does not impose any intrinsic optimism or pessimism that would overweight successes or failures.

2.2.6 Behavioral Characterization Result

We now state the main result of this subsection.

Proposition 1 (Finite-Urn Representation under Structure 1)**.** *Suppose the DM's updating rule satisfies Axiom 2 and Structure 1 above. Then, within any block of length M, the DM's perceived likelihood kernel for a within-block history with counts (n_1, n_0) is given by the generalized finite-urn (FU) form*

$$L^{\mathrm{FU}}(\theta \mid n_1, n_0) \propto \frac{\binom{\theta N}{n_1}\binom{(1-\theta)N}{n_0}}{\binom{N}{n_1+n_0}}, \tag{2}$$

for all (n_1, n_0) with $n_1 + n_0 \le M$, where generalized binomial coefficients are defined by

$$\binom{x}{k} = \frac{\Gamma(x+1)}{\Gamma(k+1)\Gamma(x-k+1)}.$$

Consequently, posterior beliefs at any point within the block take the form

$$\mu(\theta \mid n_1, n_0) = \frac{L^{\mathrm{FU}}(\theta \mid n_1, n_0)\mu_{\mathrm{in}}(\theta)}{\int_\Theta L^{\mathrm{FU}}(\tilde{\theta} \mid n_1, n_0)\mu_{\mathrm{in}}(\tilde{\theta})\, d\tilde{\theta}},$$

where μ_{in} denotes the belief distribution that the DM holds at the beginning of the current block. For the first block, $\mu_{\mathrm{in}} = \mu_0$; for later blocks, it is the posterior inherited from the end of the previous block.

When θN is an integer, the kernel in (2) coincides with the standard hypergeometric likelihood from sampling without replacement from an urn of size N.

Idea of the derivation. The formal derivation is given in Appendix A. The key step is Structure 1, which pins down the one-step-ahead predictive rule within a block. Sequential coherence then implies that the likelihood of an ordered within-block history is the product of those predictive terms. Collecting the terms associated with all success realizations and all failure realizations yields the generalized finite-urn kernel in (2). Axiom 2 then determines how this within-block logic is applied over time, while Appendix A also shows that the resulting kernel is consistent with the qualitative direction of bias in Axiom 1 and preserves symmetry between success and failure. □

Proposition 1 should be interpreted as a *behavioral characterization* of the within-block updating rule, not as a complete axiomatization of dynamic choice. It shows that once one combines blockwise reset with a specific within-block implementation of local representativeness, the finite-urn kernel emerges naturally.

2.2.7 Interpretation

This characterization has several implications.

First, it clarifies that the finite-urn rule used in Section 3 is not merely an arbitrary algebraic perturbation of Bayes' rule. Rather, it can be viewed as the reduced-form consequence of a specific short-run sampling perception: within each block, the DM behaves as if signals were drawn without replacement from a perceived finite population. In particular, when $M = 2$, the characterization reproduces the same distorted updating components that appear in (5) and (9) of Section 3, namely the correction factor

$$\frac{\theta N - 1}{N - 1}$$

after two consecutive successes, and the analogous factor

$$\frac{(1-\theta)N - 1}{N - 1}$$

after two consecutive failures (see Appendix A). Thus, it is directly consistent with the specification introduced in Section 2.1.

Second, the characterization sharply separates two conceptually distinct objects. The parameter N governs the intensity of the finite-urn distortion by determining how much an observed success or failure changes the perceived probability of the next signal of the same type. By contrast, the parameter M determines how long the DM applies this logic before resetting the perceived urn.

Third, Proposition 1 is stated for any fixed finite block length $M \geq 2$, so the behavioral characterization itself is general. At the same time, the paper's main analytical results focus on the simplest nontrivial benchmark, namely the case $M = 2$. This benchmark already allows the main comparative statics to be derived in closed form, while remaining directly nested in the more general characterization above.

We therefore now turn to the benchmark case $M = 2$ and derive its implications for posterior beliefs and economic behavior.

3 Analysis and Results

The benchmark case studied below is the first nontrivial pure-streak setting, namely two identical outcomes within the same block.[4] This is the canonical environment in which gambler's-fallacy reasoning is typically elicited, since that bias concerns predictions following short runs of identical outcomes. In this sense, the analysis begins with the most basic setting in which the law of small numbers gives rise to reversal-type beliefs. For longer within-block histories, posterior implications depend on the full configuration of successes and failures, not on streak length alone; for that reason, the paper treats those cases only through simulation-based robustness checks.[5]

We now analyze the benchmark case in which the finite block length is $M = 2$. In this setting, the model's implications can be derived in closed form. We assume that the prior probability density function (hereafter, *prior PDF*) perceived by the DM follows a beta distribution. The beta distribution is particularly suitable for this analysis, as varying the parameters a and b allows for the representation of diverse prior belief structures. The distribution of θ is formally defined as follows, where $B(a, b)$ denotes the beta function:

$$\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}, \quad a > 0,\ b > 0. \tag{3}$$

The DM engages in a binary outcome task, such as an investment decision or a job application, where the result is either success or failure. Initially, the DM forms a

[4]When $M = 1$, no finite-urn correction arises, because the first update always coincides with standard Bayesian updating. Thus, $M = 2$ is the first nontrivial case.

[5]See Appendix A for a detailed discussion of why M > 2 leads to a case-by-case analysis that quickly becomes cumbersome.

prior PDF over the success probability θ. Upon observing the outcome, they update the prior PDF using standard Bayesian updating, leading to a posterior cumulative distribution function (hereafter, *posterior CDF*). To examine this process more precisely, we compare the belief-updating mechanisms of a Bayesian and a quasi-Bayesian.

Bayesian Updating after a Success

When the DM observes a successful outcome, they update their probability assessment according to Bayes' theorem:

$$P(\theta|\text{success}) = \frac{P(\theta \wedge \text{success})}{P(\text{success})}.$$

Since $P(\theta \wedge \text{success}) = P(\theta)\,P(\text{success} \mid \theta) = (3) \cdot \theta$ and the probability of success, $P(\text{success})$, is given by:

$$\begin{aligned}
\int_0^1 f(\theta)\theta \, d\theta &= \frac{1}{B(a,b)} \int_0^1 \theta^{(a-1)+1}(1-\theta)^{b-1} d\theta \\
&= \frac{1}{B(a,b)} B(a+1,b) \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \times \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \\
&= \frac{a}{a+b},
\end{aligned}$$

the *posterior PDF* after a single success is:

$$f(\theta|\text{success}) = \frac{a+b}{a}\left[\frac{1}{B(a,b)}\theta^{a}(1-\theta)^{b-1}\right]. \tag{4}$$

At this stage, the posterior belief remains identical for both the Bayesian and the quasi-Bayesian. This is because the first update uses the likelihood of the observed success relative to the original prior, so the finite-urn correction has not yet entered. More generally, in every period the DM updates the posterior density of θ using the newly observed signal as data; what differs under the finite-urn model is not the object of inference, but the likelihood factor used in that update. The distortion appears only in the likelihood factor used for the next update, that is, when the agent evaluates the probability of a second success within the same block.

Second Update: The quasi-Bayesian's Perspective

Suppose the quasi-Bayesian experiences another success and updates their PDF as follows:

$$f(\theta|\text{success}, \text{success}) = (4) \times \frac{\theta N - 1}{N - 1} \tag{5}$$

$$= \frac{a+b}{a} \times \frac{1}{N-1} \times \frac{1}{B(a,b)} \times \left[N\theta^{a+1}(1-\theta)^{b-1} - \theta^{a}(1-\theta)^{b-1}\right], \quad \theta N > 1.$$

Integrating (5) over $\theta \in [0,1]$ for normalization yields:

$$\frac{a+b}{a} \times \frac{1}{N-1} \times \frac{N(a+1)a - (a+b+1)a}{(a+b)(a+b+1)}. \tag{6}$$

Dividing (5) by (6) gives the *posterior PDF* after two consecutive successes:

$$f(\theta|\text{success}, \text{success})$$
$$= \frac{(a+b)(a+b+1)}{N(a+1)a - (a+b+1)a} \times \frac{1}{B(a,b)} \times \left[N\theta^{a+1}(1-\theta)^{b-1} - \theta^{a}(1-\theta)^{b-1}\right].$$

By integrating this expression, we obtain the *posterior CDF* after two consecutive successes, hereafter $F(x)$.

$$F(x) =$$
$$\frac{(a+b)(a+b+1)}{N(a+1)a - (a+b+1)a} \times \frac{1}{B(a,b)} \times \left[NB_e(x;a+2,b) - B_e(x;a+1,b)\right]^{6}.$$

Since N measures the extent of the bias (larger N corresponding to less bias), taking the partial derivative of *posterior CDF* with respect to N reveals how the bias influences the cumulative density of θ:

$$\frac{\partial F(x)}{\partial N} = \frac{1}{a^2\left(N(a+1) - (a+b+1)\right)^2} \times \frac{(a+b)(a+b+1)}{B(a,b)} \times a\theta^{a+1}(1-\theta)^{b} > 0. \tag{7}$$

Let us denote the *posterior CDF* of quasi-Bayesian as $F_{\text{Q\_Bayesian}}(x \leq \theta)$ and that of

[6] $B_e(x;a,b)$ denotes the incomplete beta function: $\int_0^x t^{a-1}(1-t)^{b-1}\,dt$.

Bayesian as $F_{\text{Bayesian}}(x \leq \theta)$. Since expression (7) is positive for $\theta \in (0,1)$ (except in cases where the denominator is zero), it follows that a lower N (i.e., stronger bias) leads to a lower cumulative density of θ. Considering $F_{\text{Q\_Bayesian}}(0) = F_{\text{Bayesian}}(0) = 0$, it follows that, for each θ, $F_{\text{Q\_Bayesian}}(x \leq \theta) < F_{\text{Bayesian}}(x \leq \theta) \leftrightarrow 1 - F_{\text{Q\_Bayesian}}(x \leq \theta) > 1 - F_{\text{Bayesian}}(x \leq \theta) \leftrightarrow F_{\text{Q\_Bayesian}}(x > \theta) > F_{\text{Bayesian}}(x > \theta)$. See Figure 1 for a visual illustration.

Intuitively, when the quasi-Bayesian's posterior CDF lies below the Bayesian benchmark, she regards high-θ states as relatively more likely. Because investment becomes more attractive when θ is higher, this posterior shift tends to make the quasi-Bayesian invest more. To clarify why this CDF comparison determines investment behavior, consider a stylized investment problem where a DM chooses $I \in [0, \bar{I}]$ to maximize expected utility $U(I,\theta) = \theta R(I) - C(I)$, with $R(I)$ the gross return (increasing and concave) and $C(I)$ the cost (increasing and convex). The following result formalizes the connection between posterior beliefs and investment:

Proposition 2 (Investment Monotonicity)**.** *Suppose that $U(I,\theta)$ is concave in I and satisfies*

$$\frac{\partial^2 U}{\partial I\, \partial \theta}(I,\theta) > 0$$

for all (I,θ). Let F_1 and F_2 be two posterior CDFs over θ, and suppose that

$$F_1(\theta) \leq F_2(\theta) \qquad \textit{for all } \theta \in [0,1].$$

Then the optimal investment under F_1 weakly exceeds that under F_2.

This proposition shows that, under standard concavity and monotone-comparative-statics conditions, investment behavior depends on how much probability mass posterior beliefs place on higher values of θ. The proof, based on first-order stochastic dominance, is given in Appendix B.

Applied to our setting, Proposition **??** implies that the quasi-Bayesian invests more than the rational Bayesian after consecutive successes (when the quasi-Bayesian's CDF lies below), and less after consecutive failures (when it lies above). If we regard the rational Bayesian's behavior as optimal, then the quasi-Bayesian overinvests after successes and underinvests after failures.

These different probability judgments and investment decisions arise because the quasi-Bayesian reweights the prior by the *incorrect* likelihood factor

$$\frac{\theta N - 1}{N - 1} < \theta,$$

rather than by the correct Bayesian likelihood θ. Although $\frac{\theta N-1}{N-1} < \theta$ holds pointwise, the distortion is not a uniform shrinking of the prior density. For small values of θ, the factor $\frac{\theta N-1}{N-1}$ can become very small, so the corresponding region of the density is heavily down-weighted. For large values of θ, by contrast, the factor is closer to 1, so the reduction is relatively mild. After normalization, the resulting posterior is therefore shifted toward higher values of θ relative to the Bayesian posterior. This normalization effect explains why the tail probability $1 - F(x \leq \theta)$ can be larger for quasi-Bayesians.

Importantly, the resulting investment response need not coincide with the naive reversal intuition commonly associated with the gambler's fallacy: "success is followed by failure to balance the outcomes in the short run." If DMs *does not* perform a Bayesian update, but only misinterprets θ as $\frac{\theta N-1}{N-1}$ after observing a single success, they are less likely to invest by underestimating the probability of future success. Hence, the investment decisions of quasi-Bayesian may move in the *opposite* direction from what the local representativeness would predict.

Bayesian Updating after a Failure and the Second Update

Essentially the same argument can be applied when the decision maker experiences two consecutive failures. Using Bayes' theorem:

$$P(\theta|\text{failure}) = \frac{P(\theta \wedge \text{failure})}{P(\text{failure})}. \tag{8}$$

By

$$P(\theta \wedge \text{failure}) = P(\theta)P(\text{failure}|\theta) = (3) \cdot (1 - \theta)$$

and

$$P(\text{failure}) = \frac{b}{a+b},$$

the *posterior PDF* after a single failure is:

$$f(\theta|\text{failure}) = \frac{a+b}{b}\left[\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b}\right].$$

On observing the second consecutive failure, the quasi-Bayesian update is as follows:

$$(8) \times \frac{(1-\theta)N-1}{N-1} = \frac{a+b}{b} \times \frac{1}{N-1} \times \frac{1}{B(a,b)} \times \left[N\theta^{a-1}(1-\theta)^{b+1} - \theta^{a-1}(1-\theta)^{b}\right]. \quad (9)$$

After normalization, we obtain the following:

$$f(\theta|\text{failure}, \text{failure})$$

$$= \frac{(a+b)(a+b+1)}{N(b+1)b-(a+b+1)b} \times \frac{1}{B(a,b)} \times \left[N\theta^{a-1}(1-\theta)^{b+1} - \theta^{a-1}(1-\theta)^{b}\right].$$

Integrating this expression yields the *posterior CDF*,

$$F(x) =$$

$$\frac{(a+b)(a+b+1)}{N(b+1)b-(a+b+1)b} \times \frac{1}{B(a,b)} \times \left[NB_e(x;a,b+2) - B_e(x;a,b+1)\right].$$

Taking the partial derivative of the *posterior CDF* with respect to N reveals how bias influences the *posterior CDF*:

$$\frac{\partial F(x)}{\partial N} = \frac{1}{b^2\left(N(b+1)-(a+b+1)\right)^2} \times \frac{(a+b)(a+b+1)}{B(a,b)} \times \left(-b\theta^{a}(1-\theta)^{b+1}\right) < 0.$$

This expression is negative for $\theta \in (0,1)$, implying that as the bias increases (smaller N), the cumulative density of θ increases. This leads to the opposite result of the consecutive success case: the quasi-Bayesian will be less likely to invest than the Bayesian. The quasi-Bayesian therefore overestimates the likelihood of investment failure. See Figure 2 for a visual illustration (for additional robustness checks under alternative priors $(a,b) = (2,5)$ and $(5,2)$, see Appendix C).

Bayesian updating after alternating outcomes

In cases where success follows failure or failure follows success, their effects on the prior PDF cancel out, leading to an identical *posterior CDF* that is independent of N:

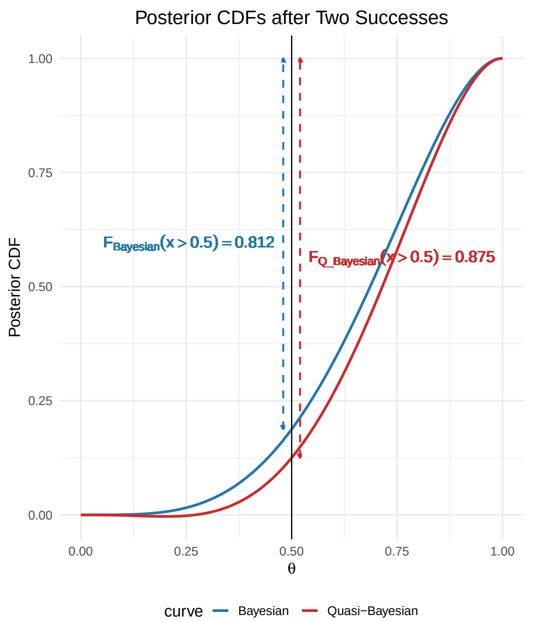


Figure 1: After two consecutive successes

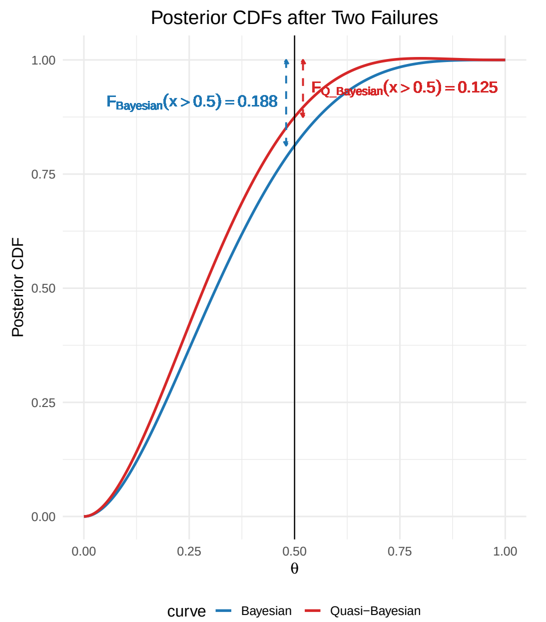


Figure 2: After two consecutive failures

Note: In these figures, the quasi-Bayesian is illustrated using $N = 5$, while the Bayesian benchmark is approximated numerically by taking $N = 10000$. As a benchmark, the threshold value is set at $\theta = 0.5$. For the simulation, we assume that the prior probability density function follows a unimodal beta distribution with parameters $a = 2$ and $b = 2$ in Equation (3).

$$Posterior\ CDF = \frac{(a+b)(a+b+1)}{abB(a,b)} \times B_e(x; a+1, b+1)$$

Since the partial derivative of this *posterior CDF* with respect to N equals zero, we conclude that, as a result of performing the Bayesian update, quasi-Bayesians and Bayesians do not differ in their probability judgments or investment decisions.

Summary of Results

The partial derivative of each *posterior CDF* with respect to N reveals the following key findings:

- After consecutive successes, a smaller perceived urn size N leads to stronger bias and enhances overinvestment.
- After consecutive failures, a smaller perceived urn size N leads to stronger bias and discourages investment.

The main analytical results focus on the first nontrivial benchmark case, namely two identical outcomes within the same block. This case already delivers the core qualitative implications in closed form: overinvestment after two consecutive successes and underinvestment after two consecutive failures. Restricting attention to $M = 2$ is also conceptually useful. Once within-block histories become longer, posterior beliefs under the finite-urn logic depend on the entire local configuration of signals, rather than on streak length alone.

This complication is not merely technical. A large literature in psychology shows that perceptions of randomness are shaped by multiple sequence features, including both run length and alternation rate, and that the effect of one depends on the other (Tversky & Kahneman 1971; Bar-Hillel & Wagenaar 1991; Falk & Konold 1997; Scholl & Greifeneder 2011). In particular, people often perceive short samples as overly representative of the underlying process, tend to regard long runs as insufficiently random, and frequently associate randomness with excessive alternation. As a result, when longer within-block histories combine identical streaks with alternating segments, it becomes less clear which aspect of the history the decision maker treats as diagnostically relevant for inference. This is a behavioral interpretation issue rather than a property of the kernel itself: under Structure 1, the kernel depends only on within-block counts, but for longer histories the economic meaning of a given count configuration becomes less transparent. In such environments, additional belief distortions may arise that are not cleanly attributable to the finite-urn force isolated here. This concern is reinforced by broader evidence on judgments of binary sequences, which suggests that longer event histories can also invite other inferential modes, including hot-hand-type reasoning (Oskarsson et al. 2009; Rabin & Vayanos 2010). For these reasons, the main text takes the two-signal case as its analytic benchmark.

To verify that this benchmark is not misleading, Appendix D reports a robustness check for four consecutive successes and four consecutive failures. The results suggest that the benchmark qualitative implications are preserved and quantitatively amplified in this particular pure-streak setting.[7]

[7] For the special case in which the entire block consists of a pure streak of length k, that is, $(n_1, n_0) = (k, 0)$ or $(0, k)$, the qualitative direction of the benchmark result is preserved: the quasi-Bayesian remains more likely to invest after consecutive successes and less likely to invest after consecutive failures. More generally, however, longer within-block histories depend on the full configuration of successes and failures, not on streak length alone.

4 Conclusion and Discussion

This paper developed a formal model of belief updating under local representativeness. We studied a quasi-Bayesian agent who updates by Bayes' rule but mistakenly interprets i.i.d. signals as if they were drawn without replacement from a finite-urn. Under this misperception, posterior beliefs are systematically distorted, and these distortions generate biased economic behavior. In particular, the model predicts more aggressive investment after consecutive successes and more pessimistic behavior after consecutive failures. Notably, these investment responses can move in the *opposite* direction from what a naive reading of gambler's fallacy intuition alone would suggest. Once the finite-urn misperception is filtered through Bayesian updating, the resulting posterior shift need not coincide with the intuitive expectation of simple reversal. More broadly, the analysis shows that probability biases need not merely shift beliefs in a mechanical direction; once they interact with Bayesian updating and normalization, they can generate economically meaningful and sometimes counterintuitive comparative statics.

A central contribution of the paper is to make this mechanism analytically transparent. The finite-urn specification provides a tractable way to study how local representativeness enters belief updating, while the bias parameter N offers a natural measure of the strength of the distortion. Smaller values of N correspond to stronger finite-urn thinking and larger deviations from the Bayesian benchmark; as N grows large, the quasi-Bayesian approaches the Bayesian. In this sense, the model provides a simple framework for comparing different degrees of bias and tracing their implications for beliefs and behavior.

The paper also clarifies the behavioral characterization of the updating rule. Rather than treating the finite-urn correction as an ad hoc perturbation of Bayes' rule, we show that it can be derived from local representativeness together with a within-block predictive structure under blockwise reset. Although this is not a full axiomatization of dynamic choice, it makes the short-scope logic of the model more transparent.

These results contribute to the theory of decision-making by linking two strands of research that are often studied separately: Bayesian updating and probability distortions rooted in heuristic reasoning. The model highlights that biased behavior can arise even when agents update in a formally Bayesian manner, provided that their mental representation of the signal-generating process is misspecified. In this respect, the source of the distortion lies not only in computation, but also in how the environment itself is cognitively encoded.

4.1 Limitations and Future Research

A first limitation is that the present study remains primarily theoretical. The core analytical results arise from the quasi-Bayesian's use of the finite-urn correction—that is, from multiplying by

$$\frac{\theta N - 1}{N - 1}$$

rather than by θ after a success, and analogously on the failure side. Whether actual DMs apply such a correction, and under what conditions, remains an empirical question. The behavioral characterization developed here clarifies the interpretation of the rule, but it does not by itself establish that this is the unique or psychologically exact representation of local representativeness.

A related limitation concerns the calibration of the urn-size parameter N, which governs the quantitative magnitude of the distortion. The main qualitative implications of the model are robust—smaller N strengthens the bias, whereas large N makes the quasi-Bayesian approach the Bayesian benchmark. The size of the resulting belief distortion and the associated change in investment behavior depend directly on the value of N. In the present paper, N is treated as a reduced-form parameter rather than being estimated or disciplined by outside evidence.

A natural direction for future research is therefore empirical identification. Laboratory belief-updating tasks could manipulate the perceived urn size N and examine how posterior judgments respond to short streaks such as two consecutive successes or two consecutive failures. By eliciting posterior beliefs after such histories, one could test directly whether smaller values of N amplify distortions, while large values of N approximate Bayesian updating.

A second direction is to endogenize N. In the present paper, N is treated as exogenous, but it may vary with cognitive ability, task complexity, experience, attention, or cognitive load. Developing a model in which N is linked explicitly to such observables would deepen the framework and help connect it to broader work on bounded rationality and heuristic reasoning.

A third direction is to extend the model to richer environments. The present analysis focuses on binary signals and investment-like decisions because that setting allows the mechanism to be seen most clearly. Extending the framework to multi-valued signals, non-stationary environments, or strategic interaction would improve its generality and clarify the scope of finite-urn reasoning in more realistic settings.

Overall, the paper offers a positive account of how real DMs depart from Bayesian updating when they interpret short samples through the lens of local representativeness. The finite-urn model, together with its behavioral characterization under blockwise reset, provides a tractable framework for analyzing this mechanism.

Acknowledgments

I benefited from many enjoyable discussions with Prof. Yoshio Kamijo and Prof. Daijiro Kawanaka. I am also grateful to the participants at the Behavioral and Experimental Economics Seminar at Waseda University for their helpful comments on a presentation of this research.

Statements and Declarations

Competing Interests

The author declares no competing interests.

References

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics*, *12*(4), 428–454.

Bordalo, P., Gennaioli, N., Ma, Y., & Shleifer, A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, *110*(9), 2748–2782. doi: 10.1257/aer.20181219

Bordalo, P., Gennaioli, N., & Shleifer, A. (2018). Diagnostic expectations and credit cycles. *Journal of Finance*, *73*(1), 199–227. doi: 10.1111/jofi.12541

Camerer, C., & Thaler, R. H. (2003). In honor of Matthew Rabin: winner of the John Bates Clark medal. *Journal of economic perspectives*, *17*(3), 159–176.

Cartwright, E. (2018). *Behavioral economics* (3rd ed.). Routledge.

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

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic literature*, *47*(2), 315–372.

Dhami, S. (2016). *The foundations of behavioral economic analysis*. Oxford University Press.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, *104*(2), 301–318.

Gennaioli, N., & Shleifer, A. (2010). What comes to mind. *Quarterly Journal of Economics*, *125*(4), 1399–1433. doi: 10.1162/qjec.2010.125.4.1399

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological science*, *17*(9), 767–773.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology*, *3*(3), 430–454.

Lin, C.-H., & Garrido, M. (2022). Towards a cross-level understanding of bayesian inference in the brain. *Neuroscience Biobehavioral Reviews*, *137*. doi: 10.1016/j.neubiorev.2022.104649

Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? judging sequences of binary events. *Psychological Bulletin*, *135*(2), 262–285.

Patel, R. (2011). Individual performance in optimal bayesian inference. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*.

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

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

Scholl, S. G., & Greifeneder, R. (2011). Disentangling the effects of alternation rate and maximum run length on judgments of randomness. *Judgment and Decision Making*, *6*(6), 531–543.

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

A Formal Derivation for the Behavioral Characterization of the Finite-Urn Model

This appendix provides the formal derivation underlying Proposition 1. The purpose is not to provide a full axiomatization of dynamic choice, but to show that combining blockwise reset with the explicit within-block predictive rule in Structure 1 yields a distorted likelihood kernel of the generalized finite-urn form inside each block.

Axiom 1 is qualitative and serves to identify the direction of the bias associated with local representativeness. The actual derivation of the likelihood kernel relies on Axiom 2 and Structure 1. After deriving the kernel, we then verify that it is consistent with the qualitative direction of bias described by Axiom 1 and that it preserves symmetry between success and failure.

A.1 Blockwise Ordered Histories and Predictive Probabilities

Fix a block of length $M \geq 2$. Consider any ordered within-block history

$$h = (s_1, \dots, s_t), \qquad t \leq M,$$

where each $s_m \in \{0, 1\}$, with 1 denoting success and 0 denoting failure.

For each stage $m = 1, \dots, t$, let

$$n_1^{m-1} = \sum_{\tau=1}^{m-1} \mathbf{1}\{s_\tau = 1\},$$
$$n_0^{m-1} = \sum_{\tau=1}^{m-1} \mathbf{1}\{s_\tau = 0\},$$

denote the numbers of successes and failures observed prior to stage m within the current block. By convention, $n_1^0 = n_0^0 = 0$, since no signals have yet been observed at the beginning of the block.

By Structure 1, after a partial within-block history with counts (n_1^{m-1}, n_0^{m-1}), the

DM's perceived probability of the next signal is

$$q_{s_m}(\theta \mid n_1^{m-1}, n_0^{m-1}) = \begin{cases} \dfrac{\theta N - n_1^{m-1}}{N-(m-1)}, & \text{if } s_m = 1, \\ \dfrac{(1-\theta)N - n_0^{m-1}}{N-(m-1)}, & \text{if } s_m = 0. \end{cases}$$

Sequential coherence within the block therefore implies that the perceived likelihood assigned to the full *ordered* within-block history is

$$L^{\text{ord}}(\theta \mid h) = \prod_{m=1}^{t} q_{s_m}(\theta \mid n_1^{m-1}, n_0^{m-1}). \tag{10}$$

In other words, the likelihood of an entire within-block history is obtained by multiplying the DM's perceived probability of each successive signal, conditional on the signals observed so far within that block.

For example, if the within-block history is $h = (S, S, F)$, where S denotes success and F denotes failure, then

$$\begin{aligned} L^{\text{ord}}(\theta \mid S, S, F) &= q_1(\theta \mid 0, 0) \cdot q_1(\theta \mid 1, 0) \cdot q_0(\theta \mid 2, 0) \\ &= \frac{\theta N}{N} \cdot \frac{\theta N - 1}{N-1} \cdot \frac{(1-\theta)N}{N-2}. \end{aligned}$$

Thus, the likelihood of observing success, then success, then failure is the product of the perceived probability of each step in sequence.

A.2 From Ordered Histories to Success and Failure Counts Within a Block

Let the total numbers of successes and failures in h be (n_1, n_0), with $n_1 + n_0 = t \leq M$. We now show that the product in (10) depends only on these counts, not on the order in which the signals occur.

Each time a success appears, the numerator contributes one term of the form

$$\theta N,\ \theta N - 1,\ \theta N - 2,\ \ldots,\ \theta N - (n_1 - 1),$$

exactly once each across the within-block history. Likewise, each time a failure appears, the numerator contributes one term of the form

$$(1-\theta)N,\ (1-\theta)N-1,\ (1-\theta)N-2,\ \ldots,\ (1-\theta)N-(n_0-1),$$

exactly once each. The denominator contributes

$$N,\ N-1,\ N-2,\ \ldots,\ N-(n_1+n_0-1),$$

again exactly once each.

Therefore,

$$L^{\mathrm{ord}}(\theta \mid h) = \frac{\prod_{j=0}^{n_1-1}(\theta N - j)\ \prod_{k=0}^{n_0-1}((1-\theta)N - k)}{\prod_{r=0}^{n_1+n_0-1}(N-r)}.$$

Since the right-hand side depends only on (n_1, n_0), we may write the finite-urn (FU) likelihood as

$$L^{\mathrm{FU}}(\theta \mid n_1, n_0) = \frac{\prod_{j=0}^{n_1-1}(\theta N - j)\ \prod_{k=0}^{n_0-1}((1-\theta)N - k)}{\prod_{r=0}^{n_1+n_0-1}(N-r)}. \tag{11}$$

Thus, within a block, the perceived likelihood depends only on the numbers of successes and failures and not on their order.

A.3 Generalized Finite-Urn Form

Using the generalized binomial coefficient

$$\binom{x}{k} = \frac{\Gamma(x+1)}{\Gamma(k+1)\Gamma(x-k+1)},$$

where x is a positive real number and k is a nonnegative integer, we have the falling-factorial identity

$$\prod_{j=0}^{k-1}(x-j) \equiv k!\binom{x}{k}.$$

Applying this identity to (11) gives

$$L^{\mathrm{FU}}(\theta \mid n_1, n_0) = \frac{n_1!\binom{\theta N}{n_1}\; n_0!\binom{(1-\theta)N}{n_0}}{(n_1+n_0)!\binom{N}{n_1+n_0}}$$
$$= \frac{n_1!n_0!}{(n_1+n_0)!} \cdot \frac{\binom{\theta N}{n_1}\binom{(1-\theta)N}{n_0}}{\binom{N}{n_1+n_0}}.$$

The prefactor

$$\frac{n_1!n_0!}{(n_1+n_0)!}$$

does not depend on θ. Hence it is irrelevant for posterior updating and can be absorbed into the proportionality constant. Therefore, for posterior purposes,

$$L^{\mathrm{FU}}(\theta \mid n_1, n_0) \propto \frac{\binom{\theta N}{n_1}\binom{(1-\theta)N}{n_0}}{\binom{N}{n_1+n_0}}. \tag{12}$$

This is the generalized finite-urn likelihood kernel used in Proposition 1.

Remark. When θN is an integer, (12) coincides with the standard hypergeometric likelihood for drawing $n_1 + n_0$ balls without replacement from an urn with θN success balls and $(1-\theta)N$ failure balls. When θN is not an integer, the same formula should be interpreted as a generalized hypergeometric-form likelihood kernel rather than a literal finite-population probability.

A.4 Posterior Updating Within a Block

Let μ_{in} denote the prior entering the current block. Given that prior, the DM's posterior after observing (n_1, n_0) inside the block is obtained by updating with the perceived kernel:

$$\mu(\theta \mid n_1, n_0) = \frac{L^{\mathrm{FU}}(\theta \mid n_1, n_0)\mu_{\mathrm{in}}(\theta)}{\int_\Theta L^{\mathrm{FU}}(\tilde{\theta} \mid n_1, n_0)\mu_{\mathrm{in}}(\tilde{\theta})\, d\tilde{\theta}}.$$

At the start of the next block, Axiom 2 requires the perceived urn to be reset. Thus, depletion does not carry across block boundaries.

A.5 Qualitative Implications of the Derived Kernel

We next verify that the generalized finite-urn kernel is consistent with the qualitative direction of bias described by Axiom 1 and that it preserves symmetry between success and failure.

(i) Streak Discounting Implied by the Derived Kernel

Consider first a within-block streak of $k \geq 2$ successes. Setting $n_1 = k$ and $n_0 = 0$ in (11), and using the convention that an empty product is equal to 1, yields

$$L^{\mathrm{FU}}(\theta \mid k, 0) = \frac{\prod_{j=0}^{k-1}(\theta N - j)}{\prod_{r=0}^{k-1}(N - r)}.$$

The corresponding Bayesian benchmark is

$$L^{B}(\theta \mid k, 0) = \theta^{k}.$$

It follows that the associated correction factor is

$$\begin{aligned} R(\theta, k, 0; N) &= \frac{L^{\mathrm{FU}}(\theta \mid k, 0)}{\theta^{k}} \\ &= \frac{\prod_{j=0}^{k-1}(\theta N - j)}{\theta^{k} \prod_{r=0}^{k-1}(N - r)} \\ &= \prod_{j=0}^{k-1} \frac{\theta N - j}{\theta(N - j)}. \end{aligned}$$

The first factor (corresponding to $j = 0$) is equal to 1. For every $j \geq 1$,

$$\frac{\theta N - j}{\theta(N - j)} < 1 \qquad \text{for all } \theta \in (0, 1).$$

Therefore,

$$R(\theta, k, 0; N) < 1 \qquad \text{for all } k \geq 2.$$

An analogous argument applies to a within-block streak of $k \geq 2$ failures. In that case, the Bayesian benchmark is $(1-\theta)^k$, so the relevant correction factor is

$$R(\theta, 0, k; N) := \frac{L^{\mathrm{FU}}(\theta \mid 0, k)}{(1-\theta)^k},$$

which likewise satisfies

$$R(\theta, 0, k; N) < 1 \qquad \text{for all } k \geq 2.$$

Thus, relative to the Bayesian benchmark, short runs of identical outcomes are assigned lower likelihood under the FU kernel. This supports the qualitative implication of Axiom 1, namely, that a repeated outcome is viewed as less representative of the remaining signals than under Bayesian updating.

(ii) Symmetry

Swapping θ with $1-\theta$ and interchanging successes and failures yields

$$\begin{aligned} L^{\mathrm{FU}}(1-\theta \mid n_0, n_1) &\propto \frac{\binom{(1-\theta)N}{n_0}\binom{\theta N}{n_1}}{\binom{N}{n_1+n_0}} \\ &= L^{\mathrm{FU}}(\theta \mid n_1, n_0). \end{aligned}$$

Thus the kernel is symmetric between success and failure.

(iii) Large-N Limit

Fix (n_1, n_0) and let $m = n_1 + n_0 \leq M$. In what follows, we consider the limit as $N \to \infty$, holding n_1, n_0, and m fixed. For $\theta \in [\varepsilon, 1-\varepsilon]$ with $\varepsilon > 0$, we use Landau's $O(\cdot)$ notation; in particular, $O(N^{-1})$ denotes a term bounded in absolute value by C/N for some constant C independent of N, uniformly over θ in that interval. Standard falling-factorial expansions then give

$$\begin{aligned}
\binom{\theta N}{n_1} &= \frac{(\theta N)^{n_1}}{n_1!}\left(1+O(N^{-1})\right), \\
\binom{(1-\theta)N}{n_0} &= \frac{((1-\theta)N)^{n_0}}{n_0!}\left(1+O(N^{-1})\right), \\
\binom{N}{m} &= \frac{N^m}{m!}\left(1+O(N^{-1})\right).
\end{aligned}$$

Substituting these expansions into (12) yields

$$\begin{aligned}
L^{\mathrm{FU}}(\theta \mid n_1, n_0) &\propto \frac{\frac{(\theta N)^{n_1}}{n_1!}\frac{((1-\theta)N)^{n_0}}{n_0!}}{\frac{N^m}{m!}}\left(1+O(N^{-1})\right) \\
&= \frac{m!}{n_1!n_0!}\theta^{n_1}(1-\theta)^{n_0}\left(1+O(N^{-1})\right) \\
&\propto \theta^{n_1}(1-\theta)^{n_0}\left(1+O(N^{-1})\right).
\end{aligned}$$

Thus the distortion relative to the Bayesian benchmark vanishes uniformly on compact subsets of $(0,1)$ as $N \to \infty$, up to a θ-independent normalizing constant.

A.6 Benchmark Case $M = 2$

The benchmark case used in the main text corresponds to $M = 2$. We now verify how the generalized finite-urn kernel derived above leads, in this benchmark case, to the specific updating rules used in the main text.

Two Consecutive Successes

For $(n_1, n_0) = (2, 0)$, we have

$$\begin{aligned}
L^{\mathrm{FU}}(\theta \mid 2, 0) &\propto \frac{\binom{\theta N}{2}}{\binom{N}{2}} \\
&= \frac{\theta N(\theta N - 1)}{N(N-1)} \\
&= \theta \cdot \frac{\theta N - 1}{N - 1}.
\end{aligned}$$

Hence, with a Beta(a, b) prior,

$$\begin{aligned} f(\theta \mid SS) &\propto L^{\mathrm{FU}}(\theta \mid 2, 0) \cdot \frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1} \\ &\propto \theta \cdot \frac{\theta N - 1}{N-1} \cdot \frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1}. \end{aligned}$$

Therefore, the blockwise characterization reproduces the same *finite-urn correction factor*

$$\frac{\theta N - 1}{N - 1}$$

that appears in equation (5) of the main text, with the additional factor θ coming from the first success.

Two Consecutive Failures

Similarly, for $(n_1, n_0) = (0, 2)$,

$$\begin{aligned} L^{\mathrm{FU}}(\theta \mid 0, 2) &\propto \frac{\binom{(1-\theta)N}{2}}{\binom{N}{2}} \\ &= (1-\theta) \cdot \frac{(1-\theta)N - 1}{N-1}, \end{aligned}$$

so the blockwise characterization reproduces the same distorted updating component used in the failure-failure analysis: the first failure contributes the usual Bayesian factor $(1-\theta)$, while the second failure contributes the finite-urn correction factor

$$\frac{(1-\theta)N - 1}{N - 1}$$

that appears in equation (9) of the main text.

Alternating Outcomes

For $(n_1, n_0) = (1, 1)$,

$$\begin{aligned} L^{\mathrm{FU}}(\theta \mid 1, 1) &\propto \frac{\binom{\theta N}{1}\binom{(1-\theta)N}{1}}{\binom{N}{2}} \\ &\propto \theta(1-\theta), \end{aligned}$$

where the omitted proportionality factor depends on N but not on θ. Hence, after normalization, the posterior under alternating outcomes coincides with the Bayesian benchmark.

A.7 Larger Block Lengths $M \geq 3$

The same within-block derivation applies verbatim for any fixed finite $M \geq 2$. As shown in Section A.2, the perceived likelihood depends only on the within-block counts (n_1, n_0), not on the order of the signals. Thus, once the within-block predictive rule is imposed, histories with the same numbers of successes and failures generate the same perceived likelihood kernel.

As M increases, however, the analysis becomes less transparent for two related reasons. First, the number of possible count configurations (n_1, n_0) grows with M. For example, when $M = 3$, one must distinguish the cases

$$(n_1, n_0) = (3, 0),\ (2, 1),\ (1, 2),\ (0, 3).$$

For each configuration, one must therefore examine the implied posterior distribution and its comparative statics with respect to N. This case-by-case analysis quickly becomes cumbersome as M grows, even though the underlying within-block logic remains unchanged.

Second, more generally, as M grows, longer histories may combine repeated outcomes and reversals in ways that make the behavioral meaning of a given configuration less transparent, even if the kernel itself depends only on counts.

For these reasons, the main text takes $M = 2$ as the analytic benchmark. Appendix D then reports a robustness check for $M = 4$, restricted to the pure-streak case $k = 4$ (four consecutive successes or four consecutive failures), where the qualitative interpretation remains sharp. These results show that the main qualitative implications are preserved in that particular pure-streak setting.

A.8 Summary

Taken together, the qualitative axiom identifies the direction of the bias, while blockwise reset and the explicit within-block predictive rule pin down the exact distorted likelihood kernel and preserve symmetry between success and failure.

The argument above establishes the following points.

1. Under Structure 1, the DM's perceived likelihood for an ordered within-block history is the product of stagewise predictive probabilities.

2. That product depends only on the total within-block counts (n_1, n_0) and therefore induces a count-based generalized finite-urn kernel.

3. The resulting posterior is obtained by standard likelihood-based updating with that distorted kernel, conditional on the prior entering the block.

4. The kernel implies streak discounting within a block, preserves symmetry between success and failure, and converges to the Bayesian benchmark as N becomes large.

5. For the benchmark case $M = 2$, the characterization reproduces the same correction factors used in the main text, up to normalization.

6. For larger block lengths, the same within-block logic continues to apply, but the number of relevant count configurations (n_1, n_0) grows, making a complete closed-form classification increasingly cumbersome. Moreover, longer within-block histories may combine repeated outcomes and alternations in ways that make it less clear which aspect of the sequence is behaviorally salient.

This completes the formal derivation.

B Formal Analysis of the Investment Decision Problem

This appendix provides the formal details underlying the investment monotonicity result (Proposition 2) stated in Section 3.

B.0.1 Setup

To connect posterior beliefs to investment behavior, suppose that after observing a history of signals, the DM holds a posterior density $f(\theta)$ over $\theta \in [0, 1]$. The DM then chooses an investment level $I \in [0, \bar{I}]$ to maximize expected utility:

$$I^*(f) = \arg\max_{I \in [0,\bar{I}]} \int_0^1 U(I, \theta)\, f(\theta)\, d\theta,$$

where

$$U(I, \theta) = \theta R(I) - C(I), \tag{13}$$

with $R'(I) > 0$, $R''(I) < 0$, $C'(I) > 0$, and $C''(I) > 0$.

Example: Linear-Quadratic Specification. A convenient special case is:

$$U(I, \theta) = \theta \cdot I - \frac{1}{2} I^2,$$

which yields the optimal investment rule $I^*(\theta) = \theta$ when θ is known with certainty.

B.0.2 Optimal Investment under Uncertainty

When θ is unknown, the DM chooses I to maximize expected utility:

$$I^*(f) = \arg\max_{I \in [0,\bar{I}]} \int_0^1 U(I, \theta)\, f(\theta)\, d\theta.$$

If the optimum is interior, and if expected utility is sufficiently smooth that the derivative with respect to I can be taken inside the integral, the first-order condition is

$$\int_0^1 \frac{\partial U}{\partial I}(I^*, \theta)\, f(\theta)\, d\theta = 0.$$

Under (13), this becomes

$$\int_0^1 [\theta R'(I^*) - C'(I^*)]\, f(\theta)\, d\theta = 0.$$

We now state a simple monotonicity result.

B.0.3 Proof of Proposition 2

Proposition 2 states that if $U(I,\theta)$ is concave in I and satisfies

$$\frac{\partial^2 U}{\partial I\,\partial\theta}(I,\theta) > 0 \qquad \text{for all } (I,\theta) \in [0,\bar{I}] \times [0,1], \tag{14}$$

then first-order stochastic dominance of posterior beliefs implies monotone comparative statics in optimal investment. In particular, if f_1 and f_2 are two posterior densities with corresponding CDFs F_1 and F_2, and if

$$F_1(\theta) \le F_2(\theta) \qquad \text{for all } \theta \in [0,1],$$

then

$$I^*(f_1) \ge I^*(f_2).$$

Proof. Condition (14) implies that, for any fixed I, the marginal utility of investment,

$$\frac{\partial U}{\partial I}(I,\theta),$$

is increasing in θ. Intuitively, when θ is higher, an additional unit of investment yields a higher marginal return.

Now define

$$G(I,f) := \int_0^1 \frac{\partial U}{\partial I}(I,\theta)\, f(\theta)\, d\theta,$$

which is the expected marginal utility of investment under posterior density f.

Because $U(I,\theta)$ is concave in I, its marginal utility $\frac{\partial U}{\partial I}(I,\theta)$ is weakly decreasing in I for every θ. Integrating with respect to f preserves this monotonicity, so $G(I,f)$ is weakly decreasing in I.

Next, let f_1 and f_2 be two posterior densities with corresponding CDFs F_1 and F_2, and suppose that F_1 first-order stochastically dominates F_2, i.e., $F_1(\theta) \le F_2(\theta)$ for all θ. This means that, relative to f_2, the distribution f_1 places less probability mass on low values of θ and more probability mass on high values. By the standard implication of first-order stochastic dominance, for any weakly increasing function h,

$$\int h(\theta) f_1(\theta)\, d\theta \ge \int h(\theta) f_2(\theta)\, d\theta.$$

That is, a posterior distribution that is shifted toward higher values of θ yields a weakly higher expectation for every increasing function of θ. Since $\frac{\partial U}{\partial I}(I, \theta)$ is increasing in θ, it follows that

$$G(I, f_1) \geq G(I, f_2) \qquad \text{for every } I \in [0, \bar{I}].$$

By (14), the optimal investment level is characterized by the first-order condition

$$G(I, f) = 0.$$

Because $G(I, f)$ is weakly decreasing in I, while $G(I, f_1) \geq G(I, f_2)$ for every I, the value of I at which $G(I, f) = 0$ is weakly larger under f_1 than under f_2. Therefore,

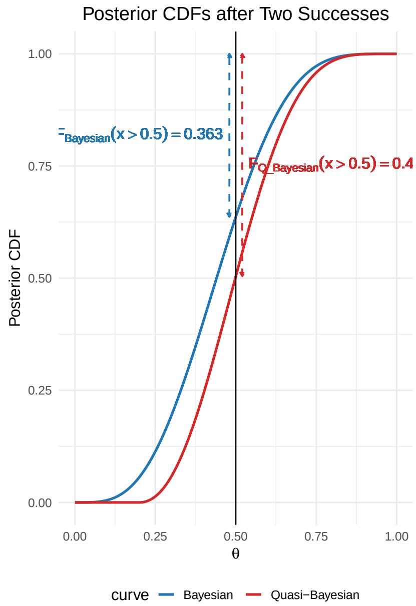
$$I^*(f_1) \geq I^*(f_2).$$

□

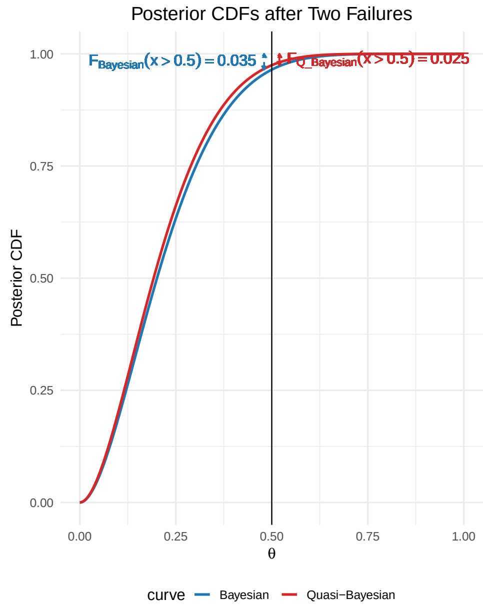
B.0.4 Implications for Distorted Beliefs

This proposition shows that posterior beliefs determine investment behavior through how much probability mass they place on higher values of θ. In particular, if the quasi-Bayesian's posterior CDF lies below the Bayesian's posterior CDF, then the quasi-Bayesian assigns relatively more probability mass to high-θ states and therefore invests more. If instead it lies above, the quasi-Bayesian invests less. Figures 1 and 2 provide a visual illustration of the distorted-belief implications discussed above.

C Robustness Checks in (a, b) = (2, 5) and (5, 2)

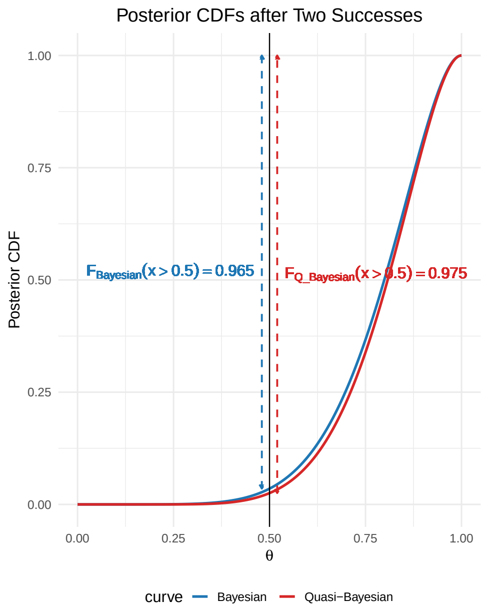


(a) After Two Successes (a=2, b=5)

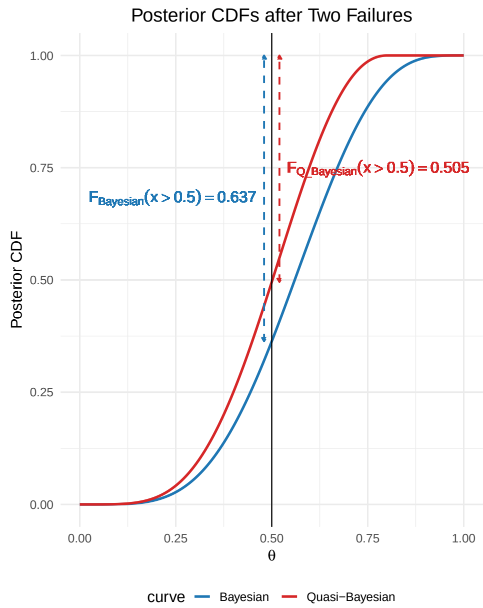


(b) After Two Failures (a=2, b=5)

Note: In these figures, the quasi-Bayesian corresponds to $N = 5$, whereas the Bayesian corresponds to $N = 10000$. As a benchmark, $\theta = 0.5$. The prior follows Beta$(a, b) = (2, 5)$.



(c) After Two Successes (a=5, b=2)



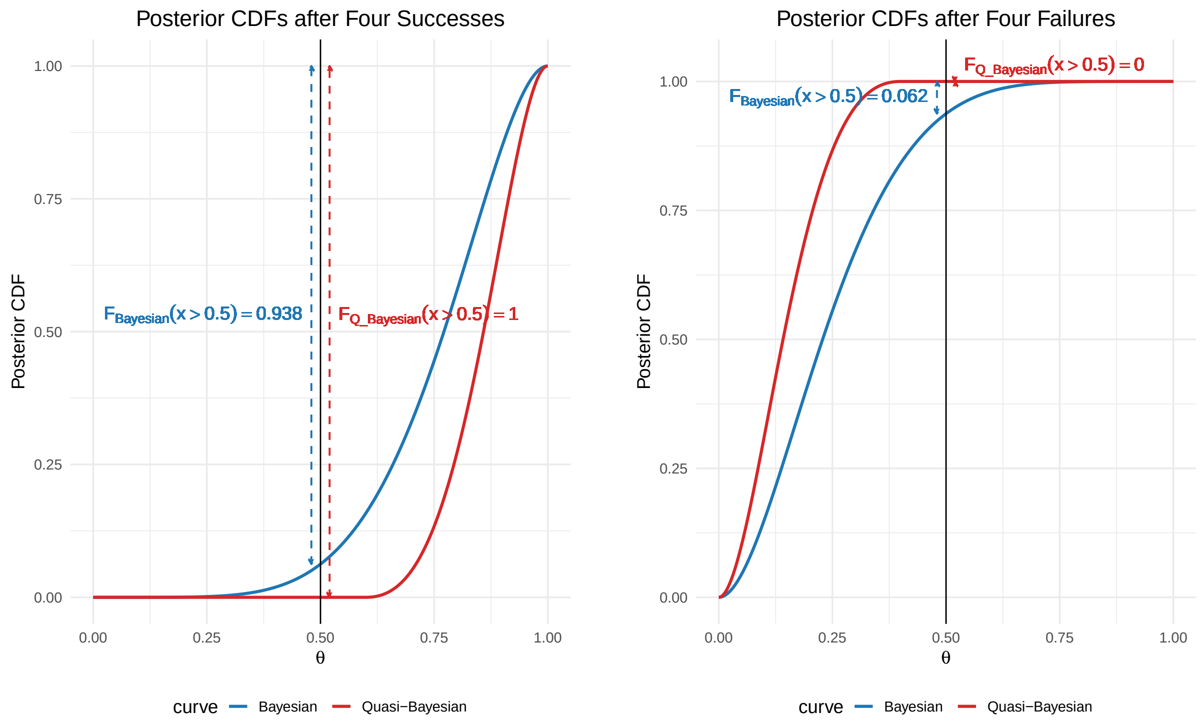
(d) After Two Failures (a=5, b=2)

Note: In these figures, the quasi-Bayesian corresponds to $N = 5$, whereas the Bayesian corresponds to $N = 10000$. As a benchmark, $\theta = 0.5$. The prior follows Beta$(a, b) = (5, 2)$.

Figure 3: Posterior CDFs: Quasi-Bayesian vs. Bayesian under two identical results.

D Robustness Check for the Pure-Streak Case with $M = 4$

This appendix reports a numerical robustness check for a longer block length, focusing on the pure-streak case in which all four signals within the block are identical. Thus, throughout this appendix, we set $M = 4$ and examine only the two configurations $(n_1, n_0) = (4, 0)$ and $(n_1, n_0) = (0, 4)$, corresponding to four consecutive successes and four consecutive failures, respectively. The purpose is not to provide an exhaustive simulation of all within-block configurations when $M = 4$, but only to check whether the benchmark qualitative implications continue to hold in this longer pure-streak environment.



(a) After Four Successes

(b) After Four Failures

Note: In these figures, the quasi-Bayesian corresponds to $N = 5$, whereas the Bayesian corresponds to $N = 10000$. As a benchmark, $\theta = 0.5$. The prior is $\text{Beta}(a, b) = (2, 2)$, and the within-block histories considered here are the two pure-streak cases $(n_1, n_0) = (4, 0)$ and $(0, 4)$.

Figure 4: Posterior CDFs for the pure-streak case with $M = 4$.

For the pure-streak success case $(n_1, n_0) = (4, 0)$, the perceived finite-urn likelihood kernel is

$$L^{\mathrm{FU}}(\theta \mid 4, 0) \propto \frac{\binom{\theta N}{4}}{\binom{N}{4}} = \frac{\theta N(\theta N-1)(\theta N-2)(\theta N-3)}{N(N-1)(N-2)(N-3)}. \tag{15}$$

An analogous expression applies to the pure-streak failure case $(n_1, n_0) = (0, 4)$. Relative to the Bayesian benchmark, these kernels apply successive finite-urn corrections after each identical outcome within the block. This illustrates how the benchmark mechanism extends to a longer block in this particular pure-streak case, although it does not by itself characterize all other $M = 4$ histories.

The figures show that, in these two pure-streak cases, the main qualitative implications remain unchanged. After four consecutive successes, the quasi-Bayesian posterior CDF lies below the Bayesian posterior CDF over the relevant range, implying higher investment by Proposition **??**. After four consecutive failures, the quasi-Bayesian posterior CDF lies above the Bayesian posterior CDF, implying lower investment.