# Multimodal Beliefs About Binary Outcomes Aren't Bayesian

Benjamin Bushong[*]
Michigan State

February 19, 2026

## Abstract

Multimodal beliefs about a Bernoulli mean should not persist after sufficient data is seen. I show that when Bayesian agents observe binary outcomes, posterior beliefs become essentially unimodal after modest sample sizes, regardless of how irregular their priors are. For example, with 100 observations, sustaining a secondary mode requires priors that favor that region by factors exceeding $10^8$.

[*] E-mail: bbushong@msu.edu. I thank Harvard University and Harvard Business School for hosting me while this work was produced.

# 1 Introduction

When people learn from binary evidence—successes and failures, wins and losses, defaults and repayments—the data arrive in a simple form. A long sequence of coin flips, medical trial outcomes, or repayment records contains no hidden structure: only the frequency of successes matters. Yet in practice, elicited beliefs about such frequencies may look complicated. From a Bayesian perspective, this complexity should be short-lived. However irregular a prior may be, once evidence accumulates the binomial likelihood punishes beliefs that stray from the empirical frequency. The punishment is sharp: deviations from the data incur exponential likelihood penalties that scale with $n$, while even astronomical prior advantages enter only as log terms. The result is that multimodal priors collapse quickly to a single peaked posterior. In this paper, I make this intuition precise and characterize exactly how many observations are needed before the posterior becomes essentially unimodal, regardless of how pathological the prior might be.

The simple, non-asymptotic result depends on just two interpretable quantities: how far the secondary modes are from the empirical frequency (determining the likelihood penalty), and how strongly the prior favors those regions (the only source of resistance). The punchline is stark: sustaining "meaningful multimodality" after even modest sample sizes requires astronomically tilted priors. For example, with one hundred observations, keeping a secondary peak at half the height of the primary mode at distance $0.2$ from the empirical frequency would require a prior that favors that region by a factor exceeding $10^3$. With two hundred observations, this factor explodes to $10^7$. For a larger separation of $0.3$, these numbers become $10^9$ and $10^{18}$ respectively. This means that if we do observe persistent multimodality in beliefs, it cannot be explained by standard Bayesian updating absent implausibly extreme priors.

These bounds link posterior *shapes* to the classic literature on disagreement. Earlier results establish that posterior means converge quickly under shared data. I show that even the possibility of multimodal posteriors disappears unless priors are tilted in implausibly extreme ways. Morris (1995) argues that persistent disagreement reveals agents cannot share common priors; the bounds described herein quantify just how extreme those prior differences would need to be under standard Bayesian updating. Likewise when Acemoglu et al. (2016) show that slight heterogeneity in signal interpretation can sustain disagreement indefinitely, our results highlight why such heterogeneity is necessary: homogeneous interpretation would force rapid convergence (see also the common-learning results of Cripps et al., 2008). The bounds also speak to the broader polarization literature: if groups observing the same binary outcomes maintain conflicting beliefs beyond our thresholds, then non-Bayesian mechanisms must be at work. A few candidate mechanisms include ambiguity-driven polarization (Baliga et al., 2013), confirmatory bias (Rabin and Schrag, 1999), base-rate neglect (Benjamin et al., 2023), echo-chamber dynamics (Levy and Razin, 2019), and motivated belief exchange in laboratory and field settings (Oprea and Yuksel, 2021; Hart et al., 2009).

My results are a direct extension of the pathwise concentration literature. Diaconis and Freedman (1990) show uniform exponential concentration to the empirical distribution under a full-support condition. Fudenberg et al. (2023) generalize this to priors without full support and provide pathwise exponential bounds (and rates for Berk, 1966). By focusing on Bernoulli outcomes, I turn these concentration ideas into explicit height (unimodality) and mass bounds with closed-form constants and derive mode-clustering diagnostics across agents. That is, prior work shows all the mass in posterior beliefs eventually piles near the true parameter. I complement this to show that no secondary bumps can survive, because even their heights are crushed unless the prior gives them an absurd advantage.

The contribution is threefold. First, I provide the first explicit finite-sample characterization of posterior shape, not just eventual concentration but precise control over multimodality at any given $n$. Second, I establish a diagnostic tool for empirical work: observed belief distributions that showcase meaningful multimodality cannot arise from standard Bayesian updating with common likelihoods, immediately implying either heterogeneous information processing or non-Bayesian information processing. Third, a straightforward extension offers guidance for experimental design: the bounds tell experimenters exactly how much common evidence should suffice to eliminate multimodal disagreement among Bayesian subjects.[1]

The technical approach is deliberately simple. Those who have properly digested Diaconis and Freedman (1990) may find the result straightforward. But I hope to train intuitions outside that literature. By focusing on Bernoulli observations, I obtain sharp, closed-form bounds. The key insight is that posterior ratios decompose into prior ratios plus scaled KL divergences, making the competition between prior and likelihood transparent. While the Bernoulli case is restrictive, it captures the essential tension present in all learning problems: the prior pulls the posterior toward its modes, the likelihood pulls toward the data, and the likelihood always wins once enough data accumulate. The bounds described below quantify how surprisingly little data may be "enough."

[1] In the Appendix, I offer an extension to eliminating mean disagreement which is a simple extension of the main result but requires more stringent assumptions about the

# 2 Formal Argument

I study how quickly a posterior distribution over the Bernoulli mean collapses to a single mode, even when the prior is highly irregular. The key message is simple: the binomial likelihood rapidly overwhelms prior multimodality, and explicit finite-sample bounds make this precise.

## 2.1 Setup

Concretely, I consider an agent with prior density $f$ on the Bernoulli mean $z \in (0, 1)$, where $f(z) > 0$ for all $z \in (0, 1)$. After observing $n$ draws with $S_n$ successes, the posterior density is

$$\pi_n(z) \propto f(z)\, z^{S_n}(1-z)^{n-S_n}.$$

I am interested in the shape of $\pi_n$. In particular: how quickly does the posterior become essentially unimodal? Intuitively, once the sample size is modest, the binomial likelihood overwhelms any prior multimodality. This note makes this intuition precise with explicit, non-asymptotic bounds. The key insight is that the likelihood function becomes increasingly peaked around the empirical frequency $\hat{z}_n = S_n/n$ as $n$ grows, with the strength of this concentration controlled by the Kullback-Leibler (KL) divergence. Even the most pathological prior—with arbitrarily many modes and deep valleys—cannot maintain meaningful multimodality once the likelihood's influence becomes strong enough.

I first define the central goal: a form of "meaningful" unimodality.

**Definition 1.** Let $z_n^*$ denote the global posterior mode. For $r, \eta \in (0, 1)$, the posterior is *multiplicatively* $(r, \eta)$*-unimodal* if

$$\sup_{|z-\hat{z}_n|\geq r} \pi_n(z) \;\leq\; \eta \cdot \pi_n(z_n^*).$$

This condition rules out multimodality up to a scalar factor. That is, this definition limits density outside an $r$-band to at most an $\eta$ fraction of the mode.

## 2.2 Main Result

As I explore this notion of meaningful unimodality, two facts underlie the analysis. The first is that posterior ratios can be written as the sum of a ratio of priors about a given point plus a scaled version of the KL divergence (as alluded to in the introduction of this section). This starting point is a simple-but-useful identity for posterior ratios.

**Observation 1** (Likelihood Ratio Structure)**.** For any $z_1, z_2 \in (0,1)$,

$$\log \frac{\pi_n(z_1)}{\pi_n(z_2)} = \log \frac{f(z_1)}{f(z_2)} + n \cdot [D(\hat{z}_n \,\|\, z_2) - D(\hat{z}_n \,\|\, z_1)],$$

where $D(x\|y)$ is the Kullback–Leibler (KL) divergence between Bernoulli($x$) and Bernoulli($y$):

$$D(x\|y) = x \log \frac{x}{y} + (1-x) \log \frac{1-x}{1-y}.$$

The second key fact is that $D(p\|\cdot)$ is strictly convex and minimized uniquely at $p$. Together these imply that the likelihood imposes an exponential penalty on deviations from $\hat{z}_n$. This exponential penalty forms the key intuition; subsequent sections deal with the relative speed of that penalty when in conflict with an arbitrary prior.

I now present two additional definitions that will help us quantify the amount of separation; this forms a critical step in the main theorem.

**Definition 2.** The *one-sided KL gap* is

$$\underline{\Delta}(r) := \inf_{|z-\hat{z}_n| \geq r} D(\hat{z}_n \| z).$$

This (strictly positive) gap measures the smallest possible likelihood penalty for moving at least $r$ away from the empirical frequency. In other words, no matter how favorable the prior may be, shifting mass outside the $r$-band around $\hat{z}_n$ comes with a built-in cost of at least $\underline{\Delta}(r)$ per observation. The convexity of KL divergence ensures that this cost grows quadratically in $r$ near $\hat{z}_n$ and diverges as $z$ approaches the boundaries 0 or 1. Intuitively, $\underline{\Delta}(r)$ is the "resistance" the likelihood imposes against sustaining secondary peaks: once $n$ is large enough, this resistance dominates any finite prior advantage.

Explicit bounds on the one-sided KL gap include a quadratic bound

$$\underline{\Delta}(r) \geq r^2/[2\hat{z}_n(1-\hat{z}_n)] \;\text{ when }\; r \leq 1/2.$$

This inequality comes from the local curvature of the KL divergence around its minimum at $z = \hat{z}_n$. The denominator $\hat{z}_n(1-\hat{z}_n)$ reflects how "flat" or "sharp" the likelihood is at the empirical frequency: the likelihood is steepest in the middle of the unit interval (when $\hat{z}_n \approx 1/2$) and flattens as $\hat{z}_n$ approaches 0 or 1. Intuitively, this bound says that stepping $r$ away from the empirical frequency always incurs *at least* a quadratic penalty, with the size of the penalty scaling inversely with the local variance of a Bernoulli($\hat{z}_n$).

**Definition 3.** I define the prior contrast of $f$ as

$$R_f(r) := \sup_{|z-\hat{z}_n|\geq r} \frac{f(z)}{f(\hat{z}_n)}.$$

This quantity measures how much more heavily the prior density can weight points outside the $r$-band compared to the empirical frequency. Intuitively, $R_f(r)$ captures the "advantage" the prior is willing to give to a distant region. If $R_f(r)$ is close to 1, the prior is relatively flat and cannot meaningfully sustain far-away bumps. If it is very large, the prior may place huge spikes far from $\hat{z}_n$, temporarily keeping multiple peaks alive. In the main theorem, $R_f(r)$ plays the role of an offsetting factor: the larger the prior contrast, the longer it can delay unimodality—but only logarithmically, since the likelihood's exponential penalty eventually overwhelms even extreme prior tilts.

My main theorem unifies these ideas and shows the finite-sample concentration.

**Theorem 1** (Multiplicative $(r,\eta)$-unimodality)**.** *If*

$$n > \frac{\log R_f(r) + \log(1/\eta)}{\underline{\Delta}(r)},$$

*then for all* $|z-\hat{z}_n| \geq r$, $\pi_n(z) \leq \eta\, \pi_n(z_n^*)$.

*Proof.* See Appendix. ■

This theorem captures the fundamental asymmetry between prior beliefs and accumulating evidence. The threshold has a simple structure: the numerator contains all the prior's ammunition—$\log R_f(r)$ measuring how much the prior favors alternative regions, plus $\log(1/\eta)$ setting our tolerance for residual bumps. The denominator contains the likelihood's weapon: the KL gap $\underline{\Delta}(r)$, measuring the information-theoretic cost of beliefs away from the data. The logarithmic-versus-linear scaling creates an inexorable dynamic. Even if the prior assigns infinitesimal probability—say $10^{-6}$—near the true parameter and probability near 1 elsewhere (a million-fold disadvantage) we need only $n > 14/\underline{\Delta}(r)$ observations to reverse this imbalance. Thus even pathological priors succumb after modest observations.

I have one minor technical annoyance to discuss. As highlighted above, I write $\hat{z}_n = S_n/n$ for the empirical frequency. However, this can lie outside $(0,1)$. I thus define $z_0 := \mathrm{proj}_{(0,1)}(\hat{z}_n)$ by $z_0 = \hat{z}_n$ if $\hat{z}_n \in (0,1)$, $z_0 = \epsilon$ if $\hat{z}_n = 0$, and $z_0 = 1-\epsilon$ if $\hat{z}_n = 1$. I can then merely replace $\hat{z}_n$ by

$z_0$ in the relevant definitions:

$$\underline{\Delta}(r) \;:=\; \inf_{|z-z_0|\geq r} D(z_0\|z), \qquad R_f(r) \;:=\; \sup_{|z-z_0|\geq r} \frac{f(z)}{f(z_0)}.$$

For ease of exposition I suppress this projection convention; replacing it where applicable is technically required to avoid edge cases.

## 2.3 Extension to Multiple Agents

I now ask what the single-agent shape results imply for a population of Bayesians who all observe the same Bernoulli data. The step from unimodality (an intra-agent property) to disagreement (an inter-agent object) requires a mild uniformity condition on priors.

**Assumption 1** (Common data and bounded prior contrast)**.** All agents $i = 1, 2, \ldots,$ observe the same sample $(S_n, n)$ and thus share the same Bernoulli likelihood. For a given radius $r > 0$, the class of priors under consideration satisfies a bounded contrast

$$R_{\max}(r) \;:=\; \sup_i R_{f_i}(r) \;< \infty, \qquad R_{f_i}(r) \;=\; \sup_{|z-\hat{z}_n|\geq r} \frac{f_i(z)}{f_i(\hat{z}_n)}.$$

I adopt the $z_0$ projection convention from Section 2 when $\hat{z}_n \in \{0, 1\}$.

This assumption is quite permissive. It allows agents to have wildly different priors as long as no one has a prior that places infinitely more weight on distant regions relative to the empirical frequency. Importantly, Assumption 1 places no restrictions on how priors behave near $\hat{z}_n$: agents can disagree sharply about local features while still satisfying the global bound.

The above (relatively minimal) structure supplies a striking result:

**Proposition 1** (Mode Clustering)**.** *Fix $r > 0$. If*

$$n \;>\; \frac{\log R_{\max}(r)}{\underline{\Delta}(r)},$$

*then every agent's posterior mode lies in the $r$-band around the empirical frequency: $|z^*_{n,i} - \hat{z}_n| < r$ for all $i$. Consequently, the maximum pairwise separation of posterior modes is bounded by $2r$: $\max_{i,j} |z^*_{n,i} - z^*_{n,j}| \;<\; 2r$.*

The logic here mirrors our single-agent analysis applied across the population. Once the sample size clears a threshold that depends on the worst-case prior contrast, the data overwhelm every agent's prior simultaneously, forcing all posterior modes into a tight cluster. This doesn't mean agents agree perfectly—their posteriors might still differ in spread, skewness, or other features—but their modal beliefs, and in this sense their "best guesses," must be close together.

This mode-clustering phenomenon provides a useful diagnostic tool. If we observe persistent disagreement with well-separated camps, we can work backward to infer what must be true about the underlying priors or information structure:

**Corollary 1** (Calibrating two-camp splits)**.** *Fix $r > 0$. Suppose two agents observing the same $(S_n, n)$ have posterior modes separated by at least $2r$: $|z^*_{n,i} - z^*_{n,j}| \geq 2r$. Then at least one agent must violate the bounded-contrast benchmark:*

$$\max\{R_{f_i}(r),\, R_{f_j}(r)\} \;\geq\; \exp\{n\,\underline{\Delta}(r)\}.$$

*Equivalently, if $R_{\max}(r) < \exp\{n\,\underline{\Delta}(r)\}$, such a two-camp split is impossible under common data and likelihood.*

This corollary quantifies just how extreme prior disagreement needs to be to sustain polarized beliefs when agents view common data. With a hundred observations and modes separated by just 0.2, one camp would need a prior that favors their preferred region by a factor in the billions. Such astronomical prior tilts strain credulity in most economic applications, suggesting that observed polarization often reflects something beyond standard Bayesian disagreement—perhaps heterogeneous interpretation of what constitutes ”success,” selective attention to different aspects of the data, or non-Bayesian updating mechanisms.

For completeness, we also provide a uniform bound on the entire posterior shape, not just the modes:

**Corollary 2** (Uniform off-band suppression)**.** *Under Assumption 1, for any $r > 0$ and any agent $i$,*

$$\sup_{|z-\hat{z}_n|\geq r} \frac{\pi_{n,i}(z)}{\pi_{n,i}(z^*_{n,i})} \;\leq\; R_{\max}(r)\,\exp\big(-n\,\underline{\Delta}(r)\big).$$

This tells us that not only do modes cluster, but the entire "off-band" region of every agent's posterior gets exponentially suppressed. This is a natural extension of the results in Diaconis and Freedman (1990). The bound is sharp in that it equals the prior contrast times the likelihood penalty, with no additional agent-specific factors. As $n$ grows, the right-hand side vanishes exponentially fast for every agent simultaneously.[2]

[2] Proposition 1 and Corollary 1 speak only to *mode locations*. Stronger distributional closeness (e.g., in total

# 3 Discussion

A first remark concerns the focus on uni- and multi-modality. The main text works with a pointwise condition: no secondary peak can rise above a fraction of the main mode. In the Appendix, I provide an alternative mass-based condition. This instead requires that only a vanishing share of posterior weight can sit far from the dominant peak. Both perspectives are useful. The pointwise bounds highlight how quickly secondary spikes are ruled out; the mass bounds show that even diffuse regions of probability weight shrink exponentially fast once enough data arrive. The conclusions are qualitatively similar and I focus on the pointwise bounds because they eliminate tedious mathematics with similar machinery underlying the conclusions. When paired with the Together, the point and mass bounds reinforce a central message: whether one looks at the height of peaks or the probability mass they carry, multimodality in Bayesian posteriors over binary outcomes is a fragile, short-lived phenomenon.

There is (at least) a trinity of interpretations of the central result of this paper, and I now invite the reader to consider which resonates with them.

First, it could be that the multimodal beliefs I describe are largely a theoretical curiosity—a phenomenon that, while mathematically possible, rarely manifests in actual economic data. The reader might argue that elicited beliefs from surveys, experiments, or market data typically show unimodal distributions, perhaps with some skewness or heavy tails, but not the distinct multiple peaks that would violate our bounds. If this is your view, then the title's claim comes as neither surprise nor particular interest: multimodal beliefs barely exist in practice.

Yet even under this interpretation, I believe that the exercise remains valuable. Reframing results from Diaconis and Freedman (1990) in terms of $n$ provides concrete benchmarks for how quickly Bayesian learners should converge, which has immediate applications. Experimental economists can use these thresholds to design studies: if you want Bayesian subjects to reach 90% posterior concentration around the true parameter, the formulas tell you exactly how many observations to provide. In this sense, my results may serve as a useful "disciplining device". To this end, I provide Table 1 to aid in converting vague intuitions about "sufficient data" into precise quantitative statements.

Second, the reader might believe that multimodal beliefs are indeed observed, but their genesis remains opaque. Consider a financial analyst observing credit markets where distinct clusters of investors maintain conflicting views about default probabilities, or a health economist studying

variation or Wasserstein) would require additional *local* regularity on priors inside the $r$-band (e.g., bounded variation or a local contrast bound on $|z - \hat{z}_n| < r$).

patient beliefs where some remain convinced a treatment works while others are equally certain it doesn't. The analyst observes these belief distributions but not their complete history. How much evidence might underlie them? Could any plausible dataset generate these convictions?

Under this interpretation, the results can become a forensic tool. If we observe multimodal beliefs and maintain the assumption of Bayesian updating, we can work backward to infer constraints on the information structure. Suppose survey respondents show bimodal beliefs about a success rate, with peaks at 0.3 and 0.7. The theorem implies that if these beliefs arose from observing common Bernoulli trials, either: (a) the sample size must be tiny; or (b) the respondents began with astronomically different priors.[3]

We can thus bound the maximum number of common observations that could underlie the observed disagreement. This backward induction is particularly useful in information economics: when we model agents as having observed private signals from a common source, persistent multimodality places sharp upper bounds on how informative those signals could have been.

Third—and this is the interpretation I find most compelling—the reader might believe that fully Bayesian updating is itself a fantasy, while multimodal beliefs are very real phenomena waiting to be explained. Perhaps survey evidence routinely reveals polarized beliefs about probabilities. Certainly, experimental subjects maintain conflicting interpretations of identical evidence, and prediction markets sometimes show persistent price dispersion that suggests fundamental disagreement about probabilities. Anecdotally, these patterns seem to appear even in simple, controlled settings where the Bernoulli structure of my framework applies directly.

If multimodal beliefs exist but cannot plausibly be Bayesian, then we need alternative explanations. The present analysis helps discriminate among psychological and behavioral theories. Base-rate neglect, as in Benjamin et al. (2023), offers one path: if agents underweight prior information, they may cluster too tightly around different interpretations of limited data. Confirmation bias (e.g. Rabin and Schrag, 1999), provides another: agents might selectively attend to evidence confirming their priors, effectively observing different subsamples despite exposure to the same data stream. Categorical thinking—where agents round probabilities to focal points like 0, 0.5, or 1—naturally generates multimodality even from smooth Bayesian posteriors. Multiple models or ambiguity aversion could sustain multiple peaks if agents entertain different generative models for the same observations.

[3] It is, in fact, this latter possibility that led to my conviction about such beliefs being non-Bayesian since even astronomically different priors had to come from *somewhere*.

**Table 1:** Sample Size $n$ Required to Eliminate $\eta$-Multimodality with Fixed Prior Contrast $R_f = 2$

| Distance $r$ | Sample Size Required | | | | | |
|---|---|---|---|---|---|---|
| | $\eta = 0.1$ | | $\eta = 0.25$ | | $\eta = 0.5$ | |
| | $\hat{z}_n = 0.1$ | $\hat{z}_n = 0.5$ | $\hat{z}_n = 0.1$ | $\hat{z}_n = 0.5$ | $\hat{z}_n = 0.1$ | $\hat{z}_n = 0.5$ |
| 0.09 | 100 | 187 | 67 | 125 | 45 | 84 |
| 0.18 | 31 | 44 | 21 | 30 | 15 | 20 |
| 0.27 | 16 | 18 | 11 | 12 | 8 | 9 |
| 0.36 | 10 | 9 | 7 | 6 | 5 | 4 |
| 0.45 | 7 | 4 | 5 | 3 | 4 | 2 |

*Notes:* This table shows the minimum sample size $n$ needed to ensure that no secondary posterior mode can exceed height $\eta$ relative to the primary mode, given a prior contrast of $R_f = 2$ (prior favors secondary region by factor of 2). Values computed using $n > [\log(R_f) + \log(1/\eta)]/\underline{\Delta}(r)$. All values are rounded up to the nearest integer.

# References

ACEMOGLU, D., V. CHERNOZHUKOV, AND M. YILDIZ (2016): “Fragility of Asymptotic Agreement under Bayesian Learning,” *Theoretical Economics*, 11, 187–225.

BALIGA, S., E. HANANY, AND P. KLIBANOFF (2013): “Polarization and Ambiguity,” *American Economic Review*, 103, 3071–3083.

BENJAMIN, D., A. BODOH-CREED, AND M. RABIN (2023): “Base-Rate Neglect: Foundations and Implications,” Working paper.

BERK, R. H. (1966): “Limiting Behavior of Posterior Distributions When the Model is Incorrect,” *Annals of Mathematical Statistics*, 37, 51–58.

CRIPPS, M., J. ELY, G. MAILATH, AND L. SAMUELSON (2008): “Common Learning,” *Econometrica*, 76, 909–933.

DIACONIS, P. AND D. FREEDMAN (1990): “On the Uniform Consistency of Bayes Estimates for Multinomial Probabilities,” *The Annals of Statistics*, 18, 1317–1327.

FUDENBERG, D., G. LANZANI, AND P. STRACK (2023): “Pathwise Concentration Bounds for Bayesian Beliefs,” *Theoretical Economics*, 18, 1585–1622.

HART, W., D. ALBARRACÍN, A. H. EAGLY, I. BRECHAN, M. J. LINDBERG, AND L. MERRILL (2009): “Feeling Validated Versus Being Correct: A Meta-Analysis of Selective Exposure to Information,” *Psychological Bulletin*, 135, 555–588.

LEVY, G. AND R. RAZIN (2019): “Echo Chambers and Their Effects on Economic and Political Outcomes,” *Annual Review of Economics*, 11, 303–328.

MORRIS, S. (1995): “The Common Prior Assumption in Economic Theory,” *Econometrica*, 63, 803–829.

OPREA, R. AND S. A. YUKSEL (2021): “Social Exchange of Motivated Beliefs,” *Management Science*, 67, 3040–3055.

RABIN, M. AND J. L. SCHRAG (1999): “First Impressions Matter: A Model of Confirmatory Bias,” *Quarterly Journal of Economics*, 114, 37–82.

# A Proofs

**Proof of Observation 1 (Likelihood Ratio Structure)**. This is a well-known decomposition. For completeness, a proof is given below.

*Proof.* The posterior log-ratio is

$$\begin{aligned} \log \frac{\pi_n(z_1)}{\pi_n(z_2)} &= \log \frac{f(z_1)}{f(z_2)} + S_n \log \frac{z_1}{z_2} + (n - S_n) \log \frac{1 - z_1}{1 - z_2} \qquad (1) \\ &= \log \frac{f(z_1)}{f(z_2)} + n \left[ \hat{z}_n \log \frac{z_1}{z_2} + (1 - \hat{z}_n) \log \frac{1 - z_1}{1 - z_2} \right]. \qquad (2) \end{aligned}$$

The likelihood term can be rewritten as

$$n \left[ \hat{z}_n \log \frac{z_1}{z_2} + (1 - \hat{z}_n) \log \frac{1 - z_1}{1 - z_2} \right] = n[D(\hat{z}_n \| z_2) - D(\hat{z}_n \| z_1)], \tag{3}$$

completing the decomposition. ■

**Proof of Theorem 1.** Three brief corollaries to the Likelihood Ratio Structure will aid in the proof of the main theorem.

**Corollary 3** (Concentration relative to $\hat{z}_n$)**.** *For any $r > 0$ and all $z$ with $|z - \hat{z}_n| \geq r$,*

$$\frac{\pi_n(z)}{\pi_n(\hat{z}_n)} \;\leq\; R_f(r)\, \exp\big( - n\, \underline{\Delta}(r) \big).$$

*Proof.* From the Likelihood Ratio Structure Lemma,

$$\log \frac{\pi_n(z)}{\pi_n(\hat{z}_n)} = \log \frac{f(z)}{f(\hat{z}_n)} - n\, D(\hat{z}_n \| z) \;\leq\; \log R_f(r) - n\, \underline{\Delta}(r),$$

whenever $|z - \hat{z}_n| \geq r$. Exponentiate. ■

**Corollary 4** (Mode localization)**.** *If $n > \log R_f(r) / \underline{\Delta}(r)$, then $|z_n^* - \hat{z}_n| < r$.*

*Proof.* If $|z_n^* - \hat{z}_n| \geq r$, Corollary 3 with $z = z_n^*$ gives

$$\frac{\pi_n(z_n^*)}{\pi_n(\hat{z}_n)} \leq R_f(r)\, \exp\{-n\, \underline{\Delta}(r)\} < 1,$$

contradicting maximality of $z_n^*$. ■

**Corollary 5** (Concentration relative to the mode)**.** *For any $r > 0$ and all $z$ with $|z - \hat{z}_n| \geq r$,*

$$\frac{\pi_n(z)}{\pi_n(z_n^*)} \;\leq\; R_f(r)\,\exp\bigl(-n\,\underline{\Delta}(r)\bigr).$$

*Proof.* $\pi_n(z_n^*) \geq \pi_n(\hat{z}_n)$, so $\pi_n(z)/\pi_n(z_n^*) \leq \pi_n(z)/\pi_n(\hat{z}_n)$, and apply Corollary 3. ■

*Completed Proof of Theorem 1.* The proof is now immediate by Corollary 5 and the assumed inequality on $n$. ■

**Proofs from Section 2.3**.

*Proof of Proposition 1.* Apply Corollary 4 to each agent $i$ with $R_{f_i}(r) \leq R_{\max}(r)$. The $2r$ bound follows by the triangle inequality. ■

*Proof of Corollary 1.* If both modes were within $r$ of $\hat{z}_n$, their separation could not exceed $2r$. Thus at least one mode satisfies $|z_{n,k}^* - \hat{z}_n| \geq r$, which by Corollary 4 implies $R_{f_k}(r) \geq \exp\{n\,\underline{\Delta}(r)\}$. ■

*Proof of Corollary 2.* Apply Corollary 5 to agent $i$ and use $R_{f_i}(r) \leq R_{\max}(r)$. ■

# B Relation to Diaconis–Freedman (1990)

There is a sense in which the present exercise is merely an extension of Diaconis and Freedman (1987; 1990). In addition to providing general convergence results (1987), in their 1990 paper, Diaconis and Freedman establish uniform exponential concentration of Bayesian posteriors for multinomial models, including the Bernoulli (coin-tossing) case. Their Theorem 1 shows that if the prior has full support, then along every sample path the posterior assigns exponentially vanishing probability to any fixed closed set $F$ that is bounded away (in Kullback–Leibler divergence) from the empirical distribution. Below I reproduce their theorem in my notation.

**Theorem 2** (Diaconis–Freedman)**.** *Let $Z_t \sim$ Bernoulli$(p)$ i.i.d. and let $S_n = \sum_{t=1}^n Z_t$, $\hat{z}_n = S_n/n$. If the prior density $f$ is strictly positive on $(0,1)$, then for any $r > 0$ there exist constants $C_r, \alpha_r > 0$ such that along almost every sample path*

$$\int_{|z-\hat{z}_n|\geq r} \pi_n(z)\,dz \;\leq\; C_r e^{-\alpha_r n} \quad \textit{for all large } n.$$

Thus DF imply an *integrated mass bound*: posterior probability of being outside any fixed $r$-band decays exponentially fast. This corresponds to the notion of "mass unimodality" referenced in the body.

**Why this does not imply multiplicative unimodality.** My main theorem, Theorem 1, gives a stronger *pointwise height bound*:

$$\sup_{|z-\hat{z}_n|\geq r} \frac{\pi_n(z)}{\pi_n(z_n^*)} \;\leq\; R_f(r)\, e^{-n\underline{\Delta}(r)}. \tag{$\star$}$$

Here $\underline{\Delta}(r)$ is the one-sided KL gap and $R_f(r)$ is the prior contrast. Bound ($\star$) rules out not only *mass* far from $\hat{z}_n$ but also *tall, narrow* secondary peaks.

To see why DF's mass result cannot imply ($\star$), consider the following construction. Fix $z_1$ at distance at least $r$ from $1/2$, and let the prior density be

$$f(z) \;=\; c_0 + \sum_{k=1}^{\infty} \frac{c_k}{\delta_k}\,\mathbf{1}\{|z-z_1| \leq \tfrac{\delta_k}{2}\}, \qquad c_k = e^{-2k\Delta},\;\; \delta_k = e^{-3k\Delta},$$

where $\Delta = D(1/2\|z_1) > 0$ and $c_0 > 0$. This prior has full support, so DF applies. Each "spike" contributes prior mass $c_k$ but density height $c_k/\delta_k = e^{k\Delta}$ at its center.

At sample size $n = k$, the posterior density ratio satisfies

$$\frac{\pi_k(z_1)}{\pi_k(z_k^*)} \;\gtrsim\; \frac{e^{k\Delta}}{c_0} \cdot e^{-\,k(\Delta/2)} \;=\; \tfrac{1}{c_0} e^{k\Delta/2} \;\to \infty.$$

Hence for infinitely many $n$, the posterior at $z_1$ is comparable to or larger than at the global mode. Multiplicative unimodality ($\star$) fails.

At the same time, the posterior *mass* in each spike is $c_k e^{-k\Delta} = e^{-3k\Delta}$, which vanishes exponentially. Thus DF's mass bound holds while our unimodality bound fails.

Absent prior regularity conditions such as bounded $R_f(r)$, DF does not rule out tall, needle-thin peaks that carry negligible mass but break unimodality. Our results therefore complement the DF/FLS tradition by quantifying the *shape* of posteriors, not just their integrated mass.