

The 3-Majority Dynamics on the Complete Graph

Aakash Kumar, Maria Sofia Bucarelli, Emanuele Natale

October 7, 2026

Abstract

We give a complete, elementary proof that the *3-majority* opinion-dynamics process on n fully-mixing agents, each holding one of two opinions, reaches consensus within $O(\log n)$ rounds with probability $1 - O(1/n)$, provided the initial imbalance between the two opinions exceeds a fixed constant fraction of n (at least 60% vs. 40%). The probability space is finite and uniform throughout; the only analytic ingredient beyond finite sums is a single self-contained exponential-moment (Chernoff-type) concentration lemma for sums of independent $\{0, 1\}$ -valued random variables, proved from the elementary bound $1 + x \leq e^x$ — no measure theory, no PMF/kernels, no appeal to Mathlib’s probability library. This is a genuinely different regime from the companion `rumor_spread` (push protocol) development: the opinion-1 count is *not* monotone in the round index (an unlucky round can shrink it), so a counting argument over “good rounds” does not suffice — honest concentration around the process’s mean trajectory is needed in *both* the growth and saturation phases. This argument has been *fully formalized* (no `sorry`) in Lean 4 + Mathlib; the main theorem is `ThreeMajority.majority3_consensus_whp`.

1 The model

Fix $n \geq 1$ and let $V = \{0, 1, \dots, n-1\}$ be the set of agents, each holding one of two opinions, 1 (“red”) or 0 (“blue”). Write $I \subseteq V$ for the set of agents currently holding opinion 1.

Definition 1 (Round randomness). A *round configuration* is a function $r: V \rightarrow V \times V \times V$ assigning to every agent v an ordered triple of *samples*, each drawn independently and uniformly from V , *with* replacement (an agent may sample itself, or the same agent twice — harmless, since a majority of three Boolean values is always well defined). Let $R_n = V^{3V}$ denote the (finite) set of round configurations, with the uniform probability measure.

Definition 2 (Majority step). For an opinion-1 set $I \subseteq V$ and $r \in R_n$,

$$\text{step}(I, r) = \{v \in V : \text{at least two of } r(v)\text{'s three coordinates lie in } I\}.$$

Each agent independently samples 3 agents and adopts the majority of their current opinions.

Definition 3 (Process). Fix a horizon $T \in \mathbb{N}$ and an initial set $I_0 \subseteq V$. The sample space is $\Omega = R_n^T$ with the uniform (i.i.d.) measure. For $r = (r_0, \dots, r_{T-1}) \in \Omega$ define $I_{t+1} = \text{step}(I_t, r_t)$. We write $X_t = |I_t|$ and $U_t = n - X_t$. In the formalization, a trajectory is a `List` of round configurations and run I l folds `step` over the list l .

Lemma 4 (Monotone coupling). *For a fixed round r , the map $I \mapsto \text{step}(I, r)$ is monotone: $I \subseteq I'$ implies $\text{step}(I, r) \subseteq \text{step}(I', r)$; iterating over a trajectory, run is monotone in the same sense. This is purely combinatorial (no probability involved).*

Remark 5. Unlike the push protocol, where the analogous monotonicity (Lemma 4 there) is load-bearing throughout, here the per-round probabilistic estimates below (Lemmas 25, 28, etc.) are instead stated and proved directly for *any* set I satisfying a hypothesis on $|I|$, using monotonicity of the real polynomial p of Section 4 on $[0, 1]$ rather than a reference-embedding argument via Lemma 4. Lemma 4 is recorded because it is the exact analogue of the push protocol’s structural fact, but it is not otherwise used by the chain of lemmas leading to the main theorem.

2 Main theorem

Throughout, $\log$ denotes the natural logarithm.

Theorem 6 (3-majority reaches consensus in $O(\log n)$ rounds w.h.p.). *Let n be such that $\log n \geq 30$ (in particular $n \geq 2$), let $I_0 \subseteq V$ with $|I_0| \geq \frac{3}{5}n$, and set*

$$T = 10 + (T_{2a}(n) + 2), \quad T_{2a}(n) = \lceil 6 \log n \rceil.$$

Then $\mathbb{P}[I_T = V] \geq 1 - \frac{500}{n}$.

Remark 7. As in the push protocol’s write-up, the threshold $\log n \geq 30$ and the constant 500 are deliberately crude: the proof is organized to keep every inequality an easy direct computation, not to be sharp. Reaching this modest, two-digit threshold on $\log n$ (rather than an astronomically large one) relies on two ingredients in Section 5: a degree-11 instance of the generic exponential-vs-power bound (Lemma 18), and a log-log bound tangent at a reference point near the threshold itself rather than at the fixed point e (Lemma 20). It is known (Doerr–Goldberg–Minder–Sauerwald–Scheideler 2011; Becchetti–Clementi–Natale–Pasquale–Silvestri, SODA 2015) that an initial imbalance of only $\Theta(\sqrt{n \log n})$ already suffices for $O(\log n)$ -round consensus w.h.p.; reaching that sharp threshold requires tracking the process on the scale of its own standard deviation throughout, a substantially more delicate multi-scale argument. Here a constant-fraction initial imbalance keeps every step to a single, reusable Chernoff bound (Section 6) applied at a handful of explicit parameters.

The proof has two phases, as in the push protocol, but for a structurally different reason: **growth** (Section 7), where the opinion-1 fraction self-amplifies from 60% to 75%; and **saturation** (Section 8), where the last dissenting agents are eliminated in three stages. Because the process is not monotone in the round index, *both* phases need the Chernoff bound of Section 6 — unlike the push protocol, where only saturation needed a (much weaker) expectation argument.

3 Finite uniform probability and independence

As in the push protocol, all randomness is uniform over finite types, so instead of measure theory or PMF we work with plain finite sums.

Definition 8 (Average). For a finite type α and $f: \alpha \rightarrow \mathbb{R}$, $\text{avg}(f) := (\sum_a f(a))/|\alpha|$ is the expectation of f under the uniform distribution on α .

Definition 9 (Trajectory expectation). Exactly as in the push protocol: $\text{expList } \alpha \ T \ F$ is the expectation of a functional F of a trajectory of T i.i.d. uniform draws from α , defined by recursion on T (the head draw is averaged out first), so conditioning on the first round is a definitional unfolding.

Lemma 10 (Basic properties). *avg and expList are monotone, additive, homogeneous, and constants pass through unchanged; expList over a concatenated horizon splits as a nested expList (expList_append), used to glue phases in Section 9.*

Lemma 11 (Independence over a product space). *New relative to the push protocol, since that development only ever averaged over a single uniform draw at a time. For finite types β, δ and $g: \beta \rightarrow \mathbb{R}$, $h: \delta \rightarrow \mathbb{R}$: $\text{avg}(p \mapsto g(p_1)h(p_2)) = \text{avg}(g)\text{avg}(h)$ over the product $\beta \times \delta$ (avg_mul_prod), with marginal consequences avgfst_mul/avg_snd_mul. Iterating over a Fin n-indexed product (avg_prod_pi): for $f: \text{Fin } n \rightarrow \gamma \rightarrow \mathbb{R}$,*

$$\text{avg}\left(x \mapsto \prod_i f_i(x_i)\right) = \prod_i \text{avg}(f_i)$$

over $x: \text{Fin } n \rightarrow \gamma$ — the discrete, measure-theory-free form of “coordinate projections on a product space are independent,” with a one-coordinate marginal corollary (avg_eval). This is exactly the independence fact that lets the n agents’ per-round updates be combined into a single concentration bound (Section 6).

4 The majority map

Write $x = X_t/n \in [0, 1]$ for the opinion-1 fraction, and $\text{ind}(I)$ for the real-valued membership indicator of I , so $\text{avg}(\text{ind}(I)) = |I|/n$.

Lemma 12 (Majority-of-three as a polynomial). *For $a, b, c \in \{0, 1\}$: $\mathbf{1}[a + b + c \geq 2] = ab + bc + ac - 2abc$. Applied pointwise to the three samples of a fixed agent, this expresses the majority indicator as a polynomial in the three membership indicators.*

Definition 13 (Per-agent contribution). $Y_{\text{maj}}(I, v, s) \in \{0, 1\}$ is 1 iff a majority of the sample triple s lies in I (independent of v , kept as a function of v to match the shape the Chernoff bound of Section 6 expects). Then $|\text{step}(I, r)| = \sum_v Y_{\text{maj}}(I, v, r(v))$ (card_step_eq_sum): the per-agent, $\{0, 1\}$ -valued decomposition that Section 6’s tail bounds need.

Lemma 14 (Exact one-round drift). *Conditionally on a fixed I with $|I| = m$, the next count $X_{t+1} = |\text{step}(I, r)|$ satisfies*

$$\mathbb{E}_r[X_{t+1}] = np(x), \quad p(x) := 3x^2 - 2x^3, \quad x = m/n,$$

proved exactly (no approximation) via Lemma 12, Definition 13 and the independence toolkit of Lemma 11 (each pairwise/triple product of membership indicators over distinct coordinates of the product space Tgt3 n factors by avg_mul_prod).

Definition 15 (Dissent contribution). $Y_{\text{dis}} := 1 - Y_{\text{maj}}$, the complementary per-agent dissent contribution, so $n - |\text{step}(I, r)| = \sum_v Y_{\text{dis}}(I, v, r(v))$ with mean $n - np(x)$ (sum_avg_Y_dis); by the algebraic identity $1 - p(x) = (1 - x)^2(1 + 2x)$ this is the quantity controlled in Section 8.

Remark 16. The algebraic identities for p used throughout below — $p(x) - x = x(1 - x)(2x - 1)$, $1 - p(x) = (1 - x)^2(1 + 2x)$, and the growth/saturation regional bounds $p(\frac{1}{2} + \beta) - \frac{1}{2} \geq \frac{5}{4}\beta$ for $0 \leq \beta \leq \frac{1}{4}$ and $u(3 - 2u) \leq \frac{5}{8}$ for $0 \leq u \leq \frac{1}{4}$ — are not isolated as separate Lean lemmas: each use site (growth_round, saturation_round_closed, saturation_mean_quad) re-derives the specific polynomial inequality it needs inline via ring/nlinarith, in the same spirit as the push protocol’s reverse-Markov bound, which is also not isolated as its own named lemma.

5 Elementary inequalities

Lemma 17 (Quadratic-tight logarithm bound). *For $\frac{1}{2} \leq x \leq 1$: $(x-1) - (x-1)^2 \leq \log x$. (Mathlib’s crude linear bound $\log x \geq x-1$ is false for $x < 1$ and, worse, plugging the correct one-sided bound $1 - \frac{1}{x} \leq \log x$ into the Chernoff exponent $k - \mu - k \log(k/\mu)$ gives exactly 0 when $k/\mu \rightarrow 1$ — useless. This quadratic correction, derived from Mathlib’s degree-3 Taylor lower bound for $\exp$ by elementary algebra, is what makes the growth-phase Chernoff bound (Lemma 25) bite.)*

Lemma 18 (Exponential beats any fixed power). *For $x \geq 0$ and any $k \in \mathbb{N}$: $x^k/k! \leq \exp(x)$ — the single degree- k term of the Taylor truncation, extracted by dropping every other (nonnegative, since $x \geq 0$) term of the sum. The cubic case $k = 3$ (`exp_ge_cube`) is the special case used by Lemma 17. Choosing k large enough relative to a given lower bound on $\log n$ lets n be shown to dominate any fixed polynomial in $\log n$ using only a modest such lower bound (Stage 2b, Lemma 35, and the final numeric bound Lemma 40) — a single low-degree truncation, such as the cubic case alone, would instead force $\log n$ to be astronomically large.*

Lemma 19 (Tangent-line log bound). *For $L > 0$ and any reference point $c > 0$: $\log L \leq \log c - 1 + L/c$, from the crude bound $\log t \leq t - 1$ applied to $t = L/c$; tight at $L = c$ (equality there), unlike a bound tangent at a single globally-fixed point, so choosing c near the L actually in play gives a far sharper bound. Also records $\exp(3) \geq 20$ (cubing the standard 10-digit lower bound on e), used to instantiate $c = \exp(3)$ below.*

Lemma 20 (Log-log bound). *For $\log x > 0$: $\log \log x \leq 2 + (\log x)/20$, from Lemma 19 tangent at the reference point $c = \exp(3) \approx 20.09$ (so $\log c = 3$ exactly), combined with $\exp(3) \geq 20$ to replace $1/\exp(3)$ by the cruder but rational $1/20$. Far tighter, once $\log x$ is more than a few units above e , than a bound tangent at the fixed point $c = e$ (the crude bound $\log t \leq t - 1$ applied directly at $t = (\log x)/e$) — exactly the regime Lemma 35 needs. Used only inside Lemma 35’s numeric estimate.*

6 An elementary Chernoff bound

The only concentration tool used in the whole development, and the only place beyond elementary combinatorics where real exponentials appear.

Lemma 21 (MGF bound). *Let $Y_1, \dots, Y_n: \gamma \rightarrow \{0, 1\}$ be the per-agent contributions on the product space $\text{Fin } n \rightarrow \gamma$ (one independent draw of γ per agent), with $\mu = \sum_i \text{avg}(Y_i)$ and $X(x) := \sum_i Y_i(x_i)$. Then for every real t (no sign restriction):*

$$\mathbb{E}[\exp(tX)] \leq \exp(\mu(e^t - 1)).$$

Proved via the independence toolkit (Lemma 11) applied to $x \mapsto \prod_i \exp(tY_i(x_i))$, and the elementary bound $1 + y \leq e^y$ (Mathlib’s `Real.add_one_le_exp`) at each factor.

Lemma 22 (Chernoff tail bounds). *In the setting of Lemma 21: for $t \geq 0$ and any k , $\mathbb{P}[X \geq k] \leq \exp(\mu(e^t - 1) - tk)$ (`avg_tail_ge`, Markov applied to $\exp(tX)$); symmetrically for $t \leq 0$, $\mathbb{P}[X \leq k] \leq \exp(\mu(e^t - 1) - tk)$ (`avg_tail_le`).*

Lemma 23 (Closed-form tail bounds at the optimal t). *Plugging $t = \log(k/\mu)$ into Lemma 22: for $k \geq \mu > 0$,*

$$\mathbb{P}[X \geq k] \leq \exp(k - \mu - k \log(k/\mu)) \quad (\text{avg_tail_ge_log}),$$

and symmetrically for $0 < k \leq \mu$, $\mathbb{P}[X \leq k] \leq \exp(k - \mu - k \log(k/\mu))$ (*avg_tail_le_log*). Both closed forms are mean-scaled: the exponent depends on the true μ directly, not on a crude worst-case range/variance proxy, which matters late in the saturation phase once few at-risk agents remain (Lemma 35). Two generalized variants take only an upper bound $\mu_{\text{ub}} \geq \mu$ (*avg_tail_ge_log_le*, used by the saturation phase) or a lower bound $\mu_{\text{lb}} \leq \mu$ (*avg_tail_le_log_ge*, used by the growth phase) on the true mean — monotonicity of the exponent in μ at the optimal t makes both directions sound, and this is the form each phase actually needs, since the exact mean depends on $x = |I|/n$ while the phases only assume an inequality on $|I|$.

7 Growth phase

Set $\beta_j = \frac{1}{10}(1.1)^j$ for $j = 0, \dots, 10$; $\beta_0 = 0.1$ (the hypothesis $X_0 \geq 0.6n$) and $\beta_{10} > 0.25$.

Definition 24 (Bias schedule). $\text{growthBias}(j) := \frac{1}{10}(\frac{11}{10})^j$, the deterministic bias target after j growth rounds; monotone, $\leq \frac{1}{4}$ for $j \leq 9$, and $> \frac{1}{4}$ at $j = 10$.

Lemma 25 (Single growth round). *Starting from opinion-1 fraction $\geq \frac{1}{2} + \beta$ ($\frac{1}{10} \leq \beta \leq \frac{1}{4}$), the probability of failing to reach the next bias target (guaranteed growth factor $\frac{11}{10}$, comfortably below the true drift factor $\frac{5}{4}$ from Lemma 14) is at most $\exp(-n/10^7)$. Proved by computing a lower bound $\mu_{\text{lb}} = n(\frac{1}{2} + \frac{5}{4}\beta)$ on the true mean via monotonicity of p on $[\frac{1}{2} + \beta, 1]$ (an inline *nlinarith* computation, not a separate lemma), then applying the lower-tail bound *avg_tail_le_log_ge* at $k = n(\frac{1}{2} + \frac{11}{10}\beta)$, and bounding the resulting exponent using Lemma 17.*

Lemma 26 (Union bound over growth rounds). *Starting from bias $\text{growthBias}(i)$, after j further rounds (with $i + j \leq 10$) the probability of failing to reach $\text{growthBias}(i + j)$ is at most $j \exp(-n/10^7)$, by induction on j via the *expList-conditioning* idiom (mirroring the push protocol's *Main.lean*).*

Lemma 27 (Phase 1). *Starting from opinion-1 fraction $\geq \frac{3}{5}$, after 10 rounds the opinion-1 fraction exceeds $\frac{3}{4}$, except with probability $\leq 10 \exp(-n/10^7)$.*

8 Saturation phase

Write $U_t = n - X_t$. By Lemma 27, $U_{10} \leq n/4$ except with probability $\leq 10 \exp(-n/10^7)$. Getting U_t all the way to *exactly* 0 needs three stages, since a fixed-ratio Chernoff bound needs the mean itself to be $\Omega(\log n)$ for a polynomially-small per-round failure probability.

Lemma 28 (General per-round upper-tail bound). *Given a reference dissent bound $U_t \leq M \leq n/4$ and any target $k \geq \frac{5}{8}M$ (an upper bound $\mu_{\text{ub}} = \frac{5}{8}M$ on the true one-round mean, via $1 - p(x) = (1 - x)^2(1 + 2x) \leq \frac{5}{8}(1 - x)$ for $x \geq \frac{3}{4}$), the probability the next dissent count exceeds k is at most $\exp(k - \frac{5}{8}M - k \log(k/(\frac{5}{8}M)))$ (*saturation_round_closed*); the same conclusion holds for any directly-supplied upper bound μ_{ub} on the true mean (*saturation_round_generic*, used by Stage 2b where the quadratic estimate below is the tight one).*

Lemma 29 (Quadratic mean bound). *For any dissent bound $U_t \leq M \leq n$, the true one-round mean is at most $3M^2/n$ (the crude bound $1 - p(x) \leq 3(1 - x)^2$, dropping the $\frac{5}{8}$ refinement). Used both by Stage 2b and by Stage 2c (at $M = 10$).*

Lemma 30 (Numeric per-round bound). *Lemma 28's exponent, simplified via Mathlib's quadratic-tight Padé bound for $\log$ (*Real.le_log_one_add_of_nonneg*) to the closed form $-\lambda^2/(k + \frac{5}{8}M)$, $\lambda = k - \frac{5}{8}M$.*

Lemma 31 (Stage 2a per-round contraction). *Guaranteed contraction factor 0.7 (vs. the true $\frac{5}{8}$): failure probability $\leq \exp(-M/250)$.*

Definition 32 (Deterministic dissent target). $\text{satAfter}(n, M, i) := \max(500 \log n, (0.7)^i M)$, the target dissent bound after i Stage 2a rounds, clamped below at the floor $500 \log n$; written in closed form (not recursively) so shifts by one round agree definitionally on the unclamped branch.

Lemma 33 (Stage 2a, union bound). *Starting from a dissent bound M ($500 \log n \leq M \leq n/4$), offset i rounds in, the probability the dissent count exceeds $\text{satAfter}(n, M, i + j)$ after j further rounds is at most $j n^{-2}$, by induction on j via the same `expList`-conditioning idiom as Lemma 26, with the tail direction flipped (dissent shrinking rather than opinion-1 count growing).*

Lemma 34 (Stage 2a, wrapped). *With $T_{2a}(n) := \lceil 6 \log n \rceil$: starting from any dissent bound $\leq n/4$, after $T_{2a}(n)$ rounds the dissent count exceeds the floor $500 \log n$ with probability at most $T_{2a}(n) \cdot n^{-2}$.*

Lemma 35 (Stage 2b). *Provided $\log n \geq 30$: whenever $U_t \leq 500 \log n$, $\mathbb{P}[U_{t+1} > 10] \leq \exp(-\log n) = 1/n$. One round, jumping from the $\Theta(\log n)$ floor directly to the fixed constant 10: since the mean is already tiny ($\Theta((\log n)^2/n)$), this single round has an astronomically small failure probability once n is large enough to beat the relevant polynomial in $\log n$, which a degree-11 instance of Lemma 18, combined with the tighter tangent-line bound of Lemma 20, already secures at this modest threshold.*

Lemma 36 (Stage 2c). *Whenever $U_t \leq 10$: $\mathbb{P}[U_{t+1} \geq 1] \leq 300/n$, via ordinary Markov applied to the nonnegative-integer U_{t+1} ($\mathbb{E}[U_{t+1}] \leq 300/n$ by Lemma 29 at $M = 10$) — no concentration needed for the very last step.*

9 Proof of the main theorem

Lemma 37 (Stages 2b+2c combined). *Two rounds, from $U_t \leq 500 \log n$ to exactly $U = 0$, with total failure probability $\leq \exp(-\log n) + 300/n$, by chaining Lemma 35 then Lemma 36 via `expList`-conditioning.*

Lemma 38 (All three saturation stages combined). *$T_{2a}(n) + 2$ rounds, from $U_t \leq n/4$ to exactly $U = 0$, with total failure probability $\leq T_{2a}(n) \cdot n^{-2} + \exp(-\log n) + 300/n$.*

Theorem 39 (Main failure bound, raw form). *For $\log n \geq 30$ and $|I_0| \geq \frac{3}{5}n$: over $10 + (T_{2a}(n) + 2)$ rounds,*

$$\mathbb{P}[I_T \neq V] \leq 10e^{-n/10^7} + (T_{2a}(n) n^{-2} + e^{-\log n} + 300/n),$$

by gluing Lemma 27 (growth) to Lemma 38 (saturation), exactly as the push protocol's `prNotAllInformed_le` glues its own two phases.

Theorem 40 (Clean numeric bound). *Under the same hypotheses, the four error terms of Theorem 39 are each bounded by $O(1/n)$ using Lemma 18 (to show n dominates the relevant polynomials in $\log n$ once $\log n \geq 30$), giving the single clean bound $\mathbb{P}[I_T \neq V] \leq 500/n$.*

With $T = 10 + (T_{2a}(n) + 2)$, Theorem 6 (`majority3_consensus_whp`) follows from Theorem 40 by $\mathbb{P}[I_T = V] = 1 - \mathbb{P}[I_T \neq V] \geq 1 - 500/n$.

10 The Lean formalization

The argument above is fully formalized (no sorry) in the accompanying repository (library `ThreeMajority`, namespace `ThreeMajority`). File structure (dependency order): `Prob.lean` (Section 3: the same finite-uniform `avg/expList` machinery as `rumor_spread/RumorSpread/Prob.lean`, copied since the two Lean projects are independent packages, plus the new independence toolkit `avg_prod_pi` and friends), `Model.lean` (Section 1), `Bounds.lean` (Section 5), `Chernoff.lean` (Section 6), `OneRound.lean` (Section 4), `Growth.lean` (Section 7), `Saturation.lean` (Section 8), `Main.lean` (Section 9 and the main theorem). Design constraint preserved from `rumor_spread`: no measure theory, no PMF/ENNReal, no Mathlib `ProbabilityTheory` kernels — everything is finite sums, plus `Real.exp/Real.log` where the Chernoff bound genuinely needs them (structurally central here, unlike the push protocol, where `Real.exp` appears only in a numeric side lemma).

References

- [1] B. Doerr, L.A. Goldberg, L. Minder, T. Sauerwald, C. Scheideler. *Stabilizing consensus with the power of two choices*. SPAA 2011.
- [2] L. Becchetti, A. Clementi, E. Natale, F. Pasquale, R. Silvestri. *Plurality consensus in the gossip model*. SODA 2015.