Disclosure Limits of Separated Demographic Elections: A Fréchet-Bound Characterization

Abstract

We study a voting architecture that measures how demographic groups voted by running one independent, sealed election per demographic dimension and issuing each voter cryptographically unlinkable credentials across dimensions. The design never stores a joint demographic record. This raises a question it does not by itself answer: given the per-dimension marginals the architecture publishes, what can an adversary reconstruct about an individual? We model the released information as the axial marginals of a high-dimensional contingency tensor, show that the joint table is generically non-identifiable from those marginals (the consistent populations form a high-dimensional fiber), and characterize individual disclosure exactly through Fréchet bounds on each cell. The level of disclosure then depends on where the published marginals fall rather than on the architecture alone. We give the collapse conditions precisely, formulate release-maximization under a disclosure target as the design problem, prove a separation-equivalence result relating the sealed elections to a single coordinated authority, and place a differential-privacy layer on the marginals as the residual cap. No step in the analysis relies on the dangerous joint table "never being created." The result is a reporting rule with a computed privacy cost. The binding outcome is published exactly and so is trustworthy without added noise; demographic breakdowns of how groups voted remain publishable while each voter stays unlinkable across dimensions; and differential-privacy noise is spent per cell only where a cell needs it, and not at all for a large, regular electorate, so privacy becomes a per-cell quantity read off the rolls rather than a flat tax on the whole release. This paper concerns what an election may *report*; its companion, *Bounded-Payoff Remote Elections: Continuity, Attestation, and Dual-Root Cast-as-Intended Verification*, concerns how ballots are *cast and checked* on untrusted client devices.

1 Model

Population and attributes. Let there be N voters. Each voter carries d demographic attributes; attribute $\ell \in \{1, \dots, d\}$ takes one of k_ℓ categories. Each voter also casts a ballot; for clarity we take the ballot binary, $\{A, \neg A\}$, and track A-votes (the multi-candidate case replicates the analysis per candidate).

The joint tensor. The complete demographic-by-vote record is the order-d tensor

$$T \in \mathbb{Z}_{\geq 0}^{k_1 \times \cdots \times k_d}, \qquad T[i_1, \dots, i_d] \;=\; \#\{A\text{-voters with attribute profile } (i_1, \dots, i_d)\}.$$

T is the object whose existence the architecture refuses: a single cell of T is exactly an intersectional record (e.g. "Muslim $\wedge$ aged 30–39 $\wedge$ Northeast, voted A"). Write $n = \prod_\ell k_\ell$ for the number of cells.

The marginal map. For axis ℓ and category $c \in \{1, \dots, k_\ell\}$, the *axial marginal* is the slice total

$$m_\ell(c) \;=\; \sum_{\substack{i_1,\dots,i_d \\ i_\ell = c}} T[i_1, \dots, i_d],$$

i.e. the number of A-voters in category c of dimension ℓ, summed over every other attribute. Election ℓ publishes the vector $m_\ell = (m_\ell(1), \dots, m_\ell(k_\ell))$ and nothing else. Let

$$\mu\colon \mathbb{Z}_{\ge 0}^{k_1 \times \cdots \times k_d} \to \mathbb{Z}_{\ge 0}^{k_1 + \cdots + k_d}, \qquad \mu(T) = (m_1, \dots, m_d)$$

be the full axial-marginal map. The *released information* is $\mu(T)$, or a chosen subset of its coordinates.

Remark 1 (Separation is a property of μ, not of T). The architecture's sealing guarantees only that no agent observes T; every agent that publishes observes one block m_ℓ of $\mu(T)$. All disclosure analysis below is therefore a statement about $\mu(T)$. This is the formal reason "the joint is never stored" does not, by itself, bound what an adversary learns: the adversary's input is $\mu(T)$, and many cells of T are functions of $\mu(T)$ to within a computable interval.

2 Non-identifiability of the joint

Definition 1 (Fiber)**.** For an observed marginal vector $\hat\mu$, the *fiber* is the set of populations consistent with it,

$$\mathcal{F}(\hat\mu) \;=\; \mu^{-1}(\hat\mu) \cap \mathbb{Z}_{\ge 0}^n \;=\; \{\, T \ge 0 : \mu(T) = \hat\mu \,\}.$$

$\mathcal{F}(\hat\mu)$ is the set of integer points of a transportation polytope.

Proposition 1 (Generic high-dimensional fiber)**.** *The affine hull of $\mu^{-1}(\hat\mu)$ has dimension*

$$n \;-\; \operatorname{rank}(\mu) \;=\; \prod_\ell k_\ell \;-\; \Big(1 + \sum_\ell (k_\ell - 1)\Big),$$

since the axial marginals contribute $\sum_\ell (k_\ell - 1)$ independent constraints beyond the single shared grand total. For $d \ge 2$ and moderate k_ℓ this dimension is exponentially large in d while the constraint count is only linear in d.

Proof. The marginal map $\mu : \mathbb{R}^n \to \bigoplus_\ell \mathbb{R}^{k_\ell}$, sending a table T to its d axial marginal vectors, is linear: each output coordinate $m_\ell(a) = \sum_{\text{axis } \ell = a} T$ is a sum of entries of T. Hence $\mu^{-1}(\hat\mu)$ is a coset of $\ker\mu$, and its affine hull has dimension $\dim\ker\mu = n - \operatorname{rank}(\mu)$ by rank–nullity. It remains to compute $\operatorname{rank}(\mu) = \dim \operatorname{im}(\mu)$.

Every image tuple $(m_1, \dots, m_d)$ satisfies the $d-1$ independent *equal-total* constraints $\sum_a m_\ell(a) = \sum_b m_{\ell'}(b)$ for all ℓ, ℓ' (each side equals the grand total $\sum T$), so $\operatorname{im}(\mu)$ is contained in the affine-linear subspace $\mathcal{C}$ of tuples with a common total, of dimension $\sum_\ell k_\ell - (d-1)$. Conversely μ is onto $\mathcal{C}$: given any $(m_\ell) \in \mathcal{C}$ with common total $M > 0$, the product (independence) table $T^\times[i_1, \dots, i_d] = M^{-(d-1)} \prod_\ell m_\ell(i_\ell)$ has, for each axis ℓ and category a,

$$\sum_{i_{\ell'} :\, \ell' \ne \ell} T^\times = M^{-(d-1)}\, m_\ell(a) \prod_{\ell' \ne \ell} \Big(\sum_{i_{\ell'}} m_{\ell'}(i_{\ell'})\Big) = M^{-(d-1)}\, m_\ell(a)\, M^{\,d-1} = m_\ell(a),$$

so $\mu(T^\times) = (m_\ell)$ (the case $M = 0$ is the zero tuple, also attained). Therefore $\operatorname{im}(\mu) = \mathcal{C}$ and $\operatorname{rank}(\mu) = \sum_\ell k_\ell - (d-1) = 1 + \sum_\ell (k_\ell - 1)$. Substituting into $n - \operatorname{rank}(\mu)$ with $n = \prod_\ell k_\ell$ gives the stated dimension. Finally $\prod_\ell k_\ell$ grows at least as 2^d while $1 + \sum_\ell (k_\ell - 1)$ is linear in d, so the fiber dimension is exponential in d against a linear constraint count. $\square$

Proposition 1 is the structural payoff of separation: axial marginals leave astronomically many distinct populations producing identical published results, so T is generically *not* identifiable from $\mu(T)$. This is the regime the architecture buys, and it is genuinely different from the small-cell, shared-key regime in which large overlapping marginal releases have been shown to reconstruct microdata. It is also *not* the privacy claim. A fiber of size 10^{20} can still force an individual cell to a single value if every population in the fiber agrees on that cell. Whether a given cell is safe is a per-cell question, answered next.

3 The characterization: Fréchet bounds

3.1 Two marginals

Lemma 1 (Fréchet–Hoeffding cell bound)**.** *Fix a cell determined by category a on axis ℓ and category b on axis ℓ', $\ell \neq \ell'$, and let n_{ab} be the corresponding two-way cell count $\sum_{\text{other axes}} T$. Writing the two published marginals as $m_\ell(a)$ and $m_{\ell'}(b)$ and the grand total as $M = \sum_c m_\ell(c)$, the cell is confined to*

$$\boxed{\max\bigl(0,\ m_\ell(a) + m_{\ell'}(b) - M\bigr) \ \le\ n_{ab}\ \le\ \min\bigl(m_\ell(a),\ m_{\ell'}(b)\bigr).}$$

In normalized form, with $P(\cdot) = (\cdot)/M$,

$$P(a \wedge b) \in \bigl[\ \max(0,\ P(a) + P(b) - 1),\ \min(P(a),\ P(b))\ \bigr].$$

Both endpoints are attained by some population in $\mathcal{F}$; the bound is tight.

Proof. Upper: n_{ab} cannot exceed either of the two slices it lies in, giving $\min(m_\ell(a), m_{\ell'}(b))$. Lower: the complement of category a on axis ℓ holds $M - m_\ell(a)$ of the A-voters; at most that many of the $m_{\ell'}(b)$ voters in category b can avoid category a, so at least $m_{\ell'}(b) - (M - m_\ell(a))$ of them must fall in the cell, and the count is non-negative. Tightness is witnessed by the Fréchet extremal couplings. □

Corollary 1 (Exact reconstruction of an individual cell)**.** *The cell value is determined exactly iff the interval has width zero,*

$$\max\bigl(0,\ m_\ell(a) + m_{\ell'}(b) - M\bigr)\ =\ \min\bigl(m_\ell(a),\ m_{\ell'}(b)\bigr).$$

Proof. By Lemma 1 the cell count n_{ab} ranges over exactly the integer points of $[\max(0, m_\ell(a) + m_{\ell'}(b) - M), \min(m_\ell(a), m_{\ell'}(b))]$, both endpoints attained. The value is determined uniquely iff this interval is a single point, i.e. iff its endpoints coincide, which is the displayed equality; otherwise the interval contains at least two integers and the value is not determined. □

3.2 Many marginals

The architecture publishes more than two marginals, so Lemma 1 applied pairwise is a valid but not generally tight bound. The sharp bound over a released set S of marginals is the value of a linear (integer) program over the fiber.

Lemma 2 (Sharp multi-way bound)**.** *For a target cell e and a released marginal set S, the sharp lower and upper bounds are*

$$\underline{n}_e = \min_{T \in \mathcal{F}_S} T[e], \qquad \overline{n}_e = \max_{T \in \mathcal{F}_S} T[e],$$

the optimal values of linear programs over the marginal polytope $\mathcal{F}_S = \{T \geq 0 : \mu_S(T) = \hat{\mu}_S\}$. When S is decomposable (its dependency graph is chordal), $\underline{n}_e$ and $\overline{n}_e$ admit closed-form generalized Fréchet expressions; otherwise they are computed by the program.

Proof. The objective $T \mapsto T[e]$ is linear. The feasible set $\mathcal{F}_S = \{T \geq 0 : \mu_S(T) = \hat{\mu}_S\}$ is the intersection of an affine subspace with the nonnegative orthant, hence a polyhedron; it is bounded, since every coordinate satisfies $0 \leq T[e'] \leq M$ (each entry is dominated by any axial marginal total containing it), and nonempty, since the observed population lies in it. A linear functional on a nonempty compact polytope attains its minimum and maximum, at vertices, and these extrema are by definition the optimal values $\underline{n}_e, \overline{n}_e$ of the two linear programs. This proves the LP characterization. When the released set S is decomposable, with clique set $\mathcal{K}$ and separator multiset $\mathcal{S}$ of its dependency graph, the extrema have the closed generalized-Fréchet form

$$\overline{n}_e = \min_{C \in \mathcal{K}} m_C[e_C], \qquad \underline{n}_e = \max\Big(0,\ \textstyle\sum_{C \in \mathcal{K}} m_C[e_C] - \sum_{D \in \mathcal{S}} m_D[e_D]\Big),$$

where e_C is the projection of cell e onto the variables of clique C; this is the Dobra–Fienberg bound for decomposable marginal sets (Dobra & Fienberg, PNAS 2000), and Lemma 1 is its two-clique, single-empty-separator instance. For non-decomposable S no such product form exists in general and the program is solved directly. $\square$

Lemma 1 is thus the conceptual key and the exact two-margin case; Lemma 2 is the operational tool the design layer actually optimizes against.

Remark 2 (The design problem is tractable only at low marginal order)*.* For $d = 2$ the sharp width has the closed form of Lemma 3 below. The as-designed release publishes *order-one* (axial) marginals, one attribute per election, the three axis-sum vectors in the $d = 3$ picture, and these are the weak constraints. By the De Loera–Onn universality theorem, a three-way *planar* (two-margin, line-sum) transportation polytope can be made affinely isomorphic to *any* rational polytope, whereas the *axial* (one-margin) polytope is universal only up to a face. Their disclosure-relevant corollary is the sharp one: once a released set includes marginals of order ≥ 2, the range an individual cell entry can attain may contain *arbitrary gaps*, so deciding cell bounds, and hence Problem 1, inherits the full complexity of integer programming, while traversing a fiber requires Markov-basis machinery (Diaconis–Sturmfels). The closed-form safety rule is therefore special to order-one release; raising marginal order forfeits both the formula and the tractability. This is an independent, computational reason, on top of the disclosure reason in Section 10, to keep published marginals low-order.

4 When disclosure happens

Disclosure is the collapse, or near-collapse, of a cell's bound. Three modes arise, in increasing subtlety. Let $w(e) = \overline{n}_e - \underline{n}_e$ be the bound width for cell e.

Theorem 1 (Disclosure taxonomy)**.** *Let $\tau \geq 0$ be a disclosure tolerance (a cell is* disclosed *if $w(e) \leq \tau$). Then:*

(M0) ***Single-marginal leakage.*** *A single published marginal m_ℓ discloses the votes of an externally identifiable group when its slice is small and lopsided: if category c has g members all externally known (e.g. the only members in a locale) and $m_\ell(c) \in \{0, 1, \dots\}$ is within τ of 0 or of g, each member's vote is fixed to within τ. No intersection is required.*

(M1) ***Saturating-margin collapse (data only).*** *By Corollary 1, $w(e) \to 0$ as one released slice approaches the whole electorate, $\max\big(m_\ell(a),\, m_{\ell'}(b)\big) \to M$ (lower bound rises to $\min$) or as $\min(m_\ell(a), m_{\ell'}(b)) \to 0$ (upper bound falls to the lower). This requires no auxiliary information.*

(M2) ***Auxiliary-correlation sharpening.*** *Even where $w(e)$ is large, an adversary holding external correlations $P(a \mid b)$ (e.g. census co-occurrence rates) forms a posterior over $[\underline{n}_e, \overline{n}_e]$ that may concentrate sharply inside the interval. Disclosure here is probabilistic: re-identification with posterior mass exceeding a threshold p, not exact reconstruction.*

Proof. (M0). The slice count $m_\ell(c)$ equals the number of category-c members who voted A, a sum of g binary indicators over the externally identified group. If $m_\ell(c) \le \tau$ then at most τ of the g identified members voted A, so the number of A-votes among them is pinned to $[0, \tau]$, an interval of length τ; symmetrically, if $m_\ell(c) \ge g - \tau$ the count is pinned to $[g - \tau, g]$. Either way each identified member's vote is fixed to within the tolerance, using the single marginal m_ℓ alone and no second axis.

(M1). In the pinched regime $m_\ell(a) + m_{\ell'}(b) \ge M$ the bound of Lemma 1 has width

$$w(e) = \min(m_\ell(a), m_{\ell'}(b)) - \big(m_\ell(a) + m_{\ell'}(b) - M\big) = M - \max(m_\ell(a), m_{\ell'}(b)),$$

which tends to 0 exactly as $\max(m_\ell(a), m_{\ell'}(b)) \to M$, i.e. as one slice fills the electorate; this is the Corollary 1 equality ($\max = M \Rightarrow m_a + m_b - M = \min$). In the slack regime $m_\ell(a) + m_{\ell'}(b) < M$ the lower bound is 0 and $w(e) = \min(m_\ell(a), m_{\ell'}(b))$, which tends to 0 as $\min \to 0$. Both limits are functions of published margins only. (The worked example below realizes $w = M - \max = 600 - 310,\ 320 - 310,\ 310 - 310$ across its three rows.)

(M2). Non-identifiability (Proposition 1) states only that many populations share the published marginals; it places no measure on them. An adversary supplies one: external correlations $P(a \mid b)$ induce a prior over $\mathcal{F}_S$, and the posterior on n_e is that prior restricted to the feasible interval $[\underline{n}_e, \overline{n}_e]$ and renormalized. This posterior can concentrate even when $w(e) > \tau$: Proposition 6 exhibits the extreme case, a release with data-only width $M/2$ in which a single auxiliary interior count moves the posterior to a point. Intermediate auxiliary facts produce intermediate concentration, placing posterior mass $> p$ on a sub-interval narrower than τ. Because this reweights the fiber rather than shrinking it, it is consistent with Proposition 1; disclosure here is probabilistic, measured by posterior mass, and is the only one of the three modes that auxiliary information can drive and that noise (Section 7) is needed to bound. □

(M2) is the formal content of statistical inference of the joint from independent marginals: it does not contradict non-identifiability (Proposition 1); it reweights the fiber.

4.1 Worked example: same architecture, opposite outcomes

Two axes, religion $\in$ {Muslim, Christian} and age $\in$ {young, old}, tracking A-votes. The religion election publishes the row margins, the age election the column margins; the interior is never stored. Apply Lemma 1 to the Muslim $\wedge$ young cell.

Case	m(Muslim)	m(young)	M	Muslim$\wedge$young bound
Safe	310	290	600	$[\max(0,0),\ \min(310,290)] = [0,290]$
Pinched	310	290	320	$[\max(0,280),\ \min(310,290)] = [280,290]$
Collapsed	310	290	310	$[\max(0,290),\ \min(310,290)] = \{290\}$

Nothing changes across the three rows but the grand total M, the same two elections, the same separation, the same published row and column margins. In the top row the cell is unconstrained and the voters are hidden; in the bottom row every Muslim $\wedge$ young A-voter is reconstructed exactly. Disclosure lives in the margin values, not in the architecture.

5 The design problem

The characterization turns the system into an optimization. The architecture chooses *which* marginals to publish; the disclosure target constrains that choice.

Problem 1 (Release maximization)**.** Given the population T, a tolerance τ, and (optionally) an adversary auxiliary model, choose a released marginal set $S \subseteq$ (all axial marginals) that

$$\text{maximizes the published information } |S| \text{ (or a utility weight on } S)$$

$$\text{subject to } w_S(e) > \tau \text{ for every cell } e,$$

where $w_S(e)$ is the sharp width from Lemma 2. The constraint couples the released marginals: whether a marginal is safe to add depends on what is already in S.

Three levers feed Problem 1:

1. **Suppress narrow-interval cells.** The correct primitive is small *width* $w_S(e)$, not small cell value, this captures both the tiny slice (M0) and the saturated large slice (M1) in one condition, where a naive minimum-cell-size rule captures only the former.
2. **Coarsen slices.** Merge categories until no measured group is near-empty or near-saturated (e.g. aggregate the five-member locale into a larger geography), trading resolution for width.
3. **Add calibrated noise.** See Section 7; this is the only lever that bounds (M2).

Remark 3 (Suppression must move to the publication layer)*.* The levers act on the *set* S jointly, but the elections are sealed from one another at the voting layer and so cannot coordinate suppression there. Worse, naive per-election suppression is recomputable: a withheld marginal value can be recovered from its complement plus the grand total. Safe suppression is therefore a property of the full release and must be enforced where the marginals are published, not where they are tallied. This is a genuine architectural tension, the separation that removes the shared key also removes voting-layer coordination, and it should be stated, not hidden.

6 Separation versus a coordinated authority

A natural objection: does running d sealed elections leak more, less, or the same as a single trusted authority that holds T and chooses to release the same marginals S? The answer separates cleanly into a data-only part and an auxiliary part.

Proposition 2 (Data-only equivalence)**.** *Against an adversary with no auxiliary information, the sealed-election architecture and a single coordinated authority that publishes the identical marginal set S induce the* same *disclosure: for every cell e, the sharp width $w_S(e)$ is a function of $\hat{\mu}_S$ alone (Lemma 2) and is independent of which agent published which block of $\hat{\mu}_S$.*

Proof. The feasible region $\mathcal{F}_S$ depends only on the published marginal values and the non-negativity constraints, both identical under the two publishing models. The bounds $\underline{n}_e, \overline{n}_e$ are optima over $\mathcal{F}_S$, hence identical. □

Corollary 2. *On the disclosure axis, separation costs nothing relative to a coordinated authority releasing the same marginals: it adds no reconstructive power to a data-only adversary, while it removes the shared key that drives small-cell microdata reconstruction and removes the single stored copy of T. The price of separation is paid in the* ease of choosing *a safe S (the coordination remark above), not in the disclosure of any fixed S.*

Proof. Immediate from Proposition 2: for a fixed published set S the two models induce the identical feasible region $\mathcal{F}_S$ and hence identical widths $w_S(e)$ for every cell, so neither discloses more to a data-only adversary. The shared decryption key and the single stored copy of T are features of the data-holding model, not arguments of $w_S(e)$; removing them changes no width. The only quantity separation degrades is the difficulty of *selecting* a safe S under Problem 1 without voting-layer coordination, which does not appear in $w_S(e)$ for fixed S. □

Remark 4 (Cross-election inconsistency as a minor privacy term). Because voters may vote differently across the sealed elections, the published margins need not be the marginals of any single coherent T: each m_ℓ is computed over a possibly distinct response pattern. To the extent inconsistency is present at rate δ, the reconstruction target is not a fixed table but a δ-perturbed family, which loosens the bounds slightly. This is a small, *bounded* effect, not a shield: at the empirically expected $\delta \lesssim 2\%$ the table is approximately coherent and the bounds apply approximately. It should be quantified, not leaned on.

7 The residual cap: differential privacy on the marginals

Modes (M1) and especially (M2) show that exact margins can collapse or sharpen cells. The architecture limits *what is collected*; a differential-privacy mechanism on the release bounds *what the publication reveals*, and the two compose.

Definition 2 (Noisy release)**.** Replace each published margin $m_\ell(c)$ by $\tilde{m}_\ell(c) = m_\ell(c) + \mathrm{Lap}(\Delta/\varepsilon_\ell)$, where the L_1 sensitivity of an axis-ℓ marginal vector to one voter is Δ (a single voter moves unit mass within one axis's vector). By basic composition, releasing axes $\ell \in L$ is $\left(\sum_{\ell \in L} \varepsilon_\ell\right)$-DP; advanced composition tightens this.

Proposition 3 (Noise prevents exact collapse)**.** *Under noisy release, the true margins are known only to within $\pm\tilde{O}(1/\varepsilon)$ with high probability, so the Fréchet endpoints of Lemma 1 inflate to high-probability bounds whose width is bounded below by the noise scale. In particular the exact-collapse condition of Corollary 1 holds only with probability decaying in $1/\varepsilon$, and the posterior sharpening of (M2) is bounded by the total privacy budget $\varepsilon = \sum_\ell \varepsilon_\ell$ over the entire release set.*

Proof. Each released margin is $\tilde{m}_\ell(c) = m_\ell(c) + \eta$ with $\eta \sim \mathrm{Lap}(b)$, $b = \Delta/\varepsilon_\ell$, for which $\Pr[|\eta| > t] = e^{-t/b}$. Over the $\sum_\ell k_\ell$ released coordinates a union bound gives, with probability $\geq 1 - \beta$, every $|\tilde{m} - m| \leq b \ln(\sum_\ell k_\ell/\beta) = \tilde{O}(1/\varepsilon)$. The Fréchet endpoints of Lemma 1 are 1-Lipschitz in the margins, so the interval an adversary computes from $\tilde{m}$ differs from the true interval by at most this amount; no width can be certified below the noise scale except on the low-probability event that the noise is atypically small. For exact collapse specifically, the released margins are absolutely continuous random variables, so the measure-zero equality of Corollary 1 holds with probability

0, and the event that the certified width falls within τ' of zero has probability $O(\tau'/b)$, decaying linearly in ε.

For (M2), the bound is the Bayesian form of differential privacy. Releasing the axes $\ell \in L$ is ε-DP with $\varepsilon = \sum_\ell \varepsilon_\ell$ by basic composition; and ε-DP implies that for any adversary prior, any individual u, and any vote value, the posterior-to-prior odds on u's vote after seeing the release lie in $[e^{-\varepsilon}, e^{\varepsilon}]$ (the standard reformulation, e.g. Dwork & Roth, §2.3). Hence no auxiliary-correlation posterior can sharpen any individual's disclosed vote by more than the multiplicative factor e^{ε}, which is exactly the claim that (M2) sharpening is capped by the total budget. □

The three mechanisms have three distinct jobs, and the paper should never let one stand in for another:

Mechanism	Bounds	Leaves open
Separation (architecture)	shared-key reconstruction; stored joint	marginal-induced collapse
Release maximization + suppression	(M0), (M1) for a target τ	(M2)
Differential privacy on margins	(M2); residual (M1) probabilistically	utility loss $\tilde{O}(1/\varepsilon)$

8 A no-noise privacy quantity: the identifiability interval

The previous sections measured disclosure as the width of a cell-*count* bound. We now isolate the quantity that governs an *individual's* privacy without any added noise, prove when it suffices, and show precisely the two quantities it depends on.

Known-composition model. We make the conservative assumption that the demographic composition is public: the cell-size tensor $N \in \mathbb{Z}_{\geq 0}^{k_1 \times \cdots \times k_d}$, with $N[e]$ the number of *people* (regardless of vote) of profile e, is known to the adversary (it is the registration roll or census). What is secret is the vote tensor $T \leq N$ (cellwise) and its marginals $\mu(T)$ are partially released. This hands the adversary strictly more than the architecture does, so guarantees proved here are conservative.

Definition 3 (Identifiability interval)**.** For a released marginal set S and a cell e with sharp count bounds $[\underline{n}_e, \overline{n}_e]$ (Lemma 2) and known size $N_e = N[e] > 0$, the *identifiability interval* is the normalized range

$$I_S(e) \;=\; \left[\, \tfrac{\underline{n}_e}{N_e},\; \tfrac{\overline{n}_e}{N_e} \,\right] \;\subseteq\; [0,1],$$

with *identifiability width* $\iota_S(e) = (\overline{n}_e - \underline{n}_e)/N_e$.

This is the count window of Lemma 1, divided by the cell size. The division is the whole point: it is what converts a statement about counts into a statement about a person.

Definition 4 (Data-only adversary)**.** A *data-only* adversary holds a prior π over populations supported on the fiber $\mathcal{F}_S$ (it believes only consistent populations possible) and treats the N_e members of any cell e as exchangeable (it has no information distinguishing one member of a cell from another). Exchangeability is the formal content of "no auxiliary information about a specific person."

Theorem 2 (No-noise individual privacy). *Let u be any individual whose full profile places them in cell e. Against* every *data-only adversary (Definition 4), the posterior probability that u voted A lies in the identifiability interval:*

$$\Pr_\pi[\,\mathrm{vote}_u = A \mid \hat{\mu}_S\,] \;\in\; I_S(e).$$

Consequently, no data-only adversary can determine u's vote to within margin δ whenever $\underline{n}_e \geq \delta N_e$ and $\overline{n}_e \leq (1-\delta)N_e$.

Proof. Fix a population T in the support of π. By within-cell exchangeability, each of the N_e members of cell e is equally likely to be among the $n_e(T)$ members who voted A, so $\Pr_\pi[\mathrm{vote}_u = A \mid T] = n_e(T)/N_e$. Taking expectation over π,

$$\Pr_\pi[\mathrm{vote}_u = A \mid \hat{\mu}_S] = \frac{\mathbb{E}_\pi[\,n_e(T)\,]}{N_e}.$$

Because π is supported on $\mathcal{F}_S$, we have $\underline{n}_e \leq n_e(T) \leq \overline{n}_e$ pointwise (Lemma 2), hence $\mathbb{E}_\pi[n_e] \in [\underline{n}_e, \overline{n}_e]$ and the posterior lies in $I_S(e)$. The margin statement is immediate. The bound holds for every π of the stated form, so it is worst case over the entire data-only class. □

Theorem 2 exhibits the two protective quantities explicitly. The *width* $\overline{n}_e - \underline{n}_e$ comes from the Fréchet bound; the *size* N_e is the base-rate dilution. A singleton cell ($N_e = 1$) has no dilution, so individual privacy then requires the count itself to remain ambiguous ($\underline{n}_e = 0,\ \overline{n}_e = 1$); this is the formal "only members in the locale" failure.

8.1 A closed-form width and a designer's safety rule

Lemma 3 (Two-way width in closed form). *For a two-way cell with margins m_a, m_b and grand total M, the Fréchet count width of Lemma 1 is*

$$w_{ab} \;=\; \min(m_a,\ m_b,\ M - m_a,\ M - m_b).$$

Proof. $w_{ab} = \min(m_a, m_b) - \max(0, m_a + m_b - M)$. If $m_a + m_b \leq M$ the second term is 0 and $w_{ab} = \min(m_a, m_b)$; since then $\max(m_a, m_b) \leq M - \min(m_a, m_b)$, we have $\min(m_a, m_b) \leq \min(M - m_a, M - m_b)$, so $w_{ab} = \min(m_a, m_b, M - m_a, M - m_b)$. If $m_a + m_b > M$ then $w_{ab} = \min(m_a, m_b) - (m_a + m_b - M) = M - \max(m_a, m_b) = \min(M - m_a, M - m_b)$, which is then the smallest of the four. In both cases the four-way minimum is attained. □

Corollary 3 (No-noise safety rule). *Every two-way cell has width $w_{ab} \geq \tau$ if and only if every published slice count satisfies $\tau \leq m_\ell(c) \leq M - \tau$. That is: a slice is safe exactly when it is neither too small nor too close to unanimous, with the same slack τ at both ends.*

Proof. By Lemma 3, $w_{ab} \geq \tau$ for all pairs iff $\min(m_a, M - m_a) \geq \tau$ for every margin, i.e. $m_\ell(c) \in [\tau, M - \tau]$. □

Remark 5 (Pairwise is necessary, not sufficient, for multi-way). Corollary 3 controls two-margin reconstruction exactly. Adding more marginals shrinks the fiber, so the sharp multi-way width (Lemma 2) satisfies $w_S(e) \leq \min_{\text{pairs}} w_{ab}$. The balance rule is therefore necessary for multi-way safety but not sufficient; the sufficient version is the LP constraint of Problem 1. We do not claim the closed form survives to the multi-way case.

9 Placement: noiseless privacy, k-anonymity, and differential privacy

The identifiability interval is a no-noise, syntactic guarantee. We locate it precisely against differential privacy and against the earlier syntactic notions, because the location is the contribution's precise boundary.

Theorem 3 (Exact release is not differentially private)**.** *Deterministic release of any non-constant marginal (in particular any axial marginal of T) satisfies ε-differential privacy for no finite ε.*

Proof. Let $M(\cdot)$ be a deterministic mechanism and D, D' neighboring populations differing in one voter. For deterministic M, $\Pr[M(D) \in \{o\}] \in \{0, 1\}$. The pure-DP inequality at output $o = M(D)$ reads $1 = \Pr[M(D) = o] \leq e^{\varepsilon} \Pr[M(D') = o]$, which for finite ε forces $\Pr[M(D') = o] > 0$, i.e. $M(D') = M(D)$. Holding for all neighbors, M is constant on each connected component of the neighbor graph, hence constant. Adding or removing one voter changes the count of that voter's slice by one and so changes the axial marginal; the marginal map is therefore non-constant, and no finite ε can hold. □

Theorem 3 settles the direction the framework cannot go: there is no deterministic route to differential privacy. Whatever the identifiability interval provides, it is not DP. It is a member of the *noiseless / Bayesian privacy* family, guarantees that hold under a stated restriction on the adversary's prior, whose lineage runs through k-anonymity and whose modern formalization is noiseless database privacy. The conversions below make the relationship one-directional and conditional, which is exactly the signature of that family.

The placement can be made exact. Theorem 2 is, verbatim, a t-closeness guarantee.

Proposition 4 (The interval is a t-closeness guarantee)**.** *Let $p_0 = M/N$ be the global A-rate. Against every data-only adversary, the posterior A-probability of each individual in cell e lies within*

$$t_e \;=\; \max_{q \in I_S(e)} |q - p_0| \;=\; \max\Big(\Big|\tfrac{\overline{n}_e}{N_e} - p_0\Big|,\; \Big|\tfrac{\underline{n}_e}{N_e} - p_0\Big| \Big)$$

of the base rate. The noiseless release is therefore t-close in the sense of Li et al., with a per-cell, release-computable t_e and global parameter $t = \max_e t_e$.

Proof. Immediate from Theorem 2: the posterior lies in $I_S(e)$, whose points are at distance at most t_e from p_0. □

Remark 6 (The classical notions are conditions on $(N_e, I_S(e))$). The syntactic-privacy lineage is recovered as a sequence of conditions on the cell size and the identifiability interval, of strictly increasing strength:

- **k-anonymity** (Sweeney): $N_e \geq k$, a condition on *size* alone, silent on the interval. It fails the homogeneous cell, where N_e is large but $I_S(e) = \{1\}$.
- **ℓ-diversity** (Machanavajjhala et al.): $I_S(e)$ bounded away from $\{0, 1\}$ (in the binary case, the cell is not near-unanimous), the patch that rules out the homogeneous cell.
- **t-closeness** (Li et al.): $I_S(e)$ within t of p_0 (Proposition 4), the form the noiseless release takes here, now exact and computable from the published marginals.

- **Differential privacy** (Dwork et al.): bounded posterior-to-prior odds for *every* prior, including the non-exchangeable ones that Definition 4 excludes, the only condition that survives the data-only assumption being dropped.

The identifiability interval is thus not a new privacy definition; it is the t-closeness point of a known spectrum, made exact and release-computable for separated marginal elections. Theorem 3 marks where the spectrum stops being reachable without noise.

Proposition 5 (DP implies a width floor; the easy direction)**.** *Release each margin with independent Laplace noise of scale $b = \Delta/\varepsilon$. Then for every cell e, the adversary's posterior on n_e has standard deviation $\Omega(b) = \Omega(\Delta/\varepsilon)$; in particular no cell collapses to a point with positive probability, and the effective identifiability width is bounded below by $\Omega(b)/N_e$ with high probability.*

Proof. The adversary observes $\tilde{m} = m + \eta$ with independent $\eta \sim \mathrm{Lap}(b)$ on each margin, so under any prior its posterior on the true margin vector has each coordinate spread on the scale b (the likelihood is a product of Laplace densities of scale b, with variance $2b^2$ per coordinate). Fix a target cell e whose binding Fréchet endpoint in the pinched regime is the affine map $L = m_\ell(a)+m_{\ell'}(b)-M$ (Lemma 1), with unit coefficients on two distinct, independently noised margins. On the event that this constraint is the active one, which has probability bounded away from 0 throughout the pinched regime, the posterior on L is the convolution of two independent scale-b laws, so $\mathrm{Var}(n_e \mid \tilde{m}) \geq \mathrm{Var}(L \mid \tilde{m}) = 4b^2$ and the posterior standard deviation on the cell count is $\Omega(b) = \Omega(\Delta/\varepsilon)$. The posterior on n_e is therefore absolutely continuous and carries no atom, so exact collapse to a point occurs with probability 0. Dividing the count spread by the cell size N_e converts it to the identifiability width of Section 8, giving a high-probability floor $\Omega(b)/N_e$. ☐

Proposition 6 (Width does not imply DP; auxiliary information is decisive)**.** *There is a family of releases, each satisfying the safety rule of Corollary 3 with arbitrarily large τ, together with a single auxiliary fact of bounded informativeness, under which a target cell collapses to a point.*

Proof. Take a 2×2 table of A-votes with margins $m_M = m_Y = M/2$. By Lemma 3 every cell has width $M/2 = \tau$, the maximum possible, so the release is maximally safe in the data-only sense. The 2×2 table has one degree of freedom: fixing $x = n_{MY}$ determines $n_{MO} = m_M - x$, $n_{CY} = m_Y - x$, $n_{CO} = M - m_M - m_Y + x$. Suppose the adversary learns one auxiliary value, say n_{CO}, from an external source. Then $x = n_{CO} - (M - m_M - m_Y)$ is determined exactly: width 0, the target n_{MY} reconstructed, despite the data-only width being $M/2$. A single finer marginal or one known interior count suffices. ☐

Proposition 6 is the formal statement of why this family needs a prior bound: with unrestricted auxiliary information, no width guarantee survives, which is the composition-and-auxiliary phenomenon that motivated differential privacy in the first place. The positive content of the identifiability interval is therefore exactly conditional: *against data-only adversaries (Definition 4), Theorem 2 gives noise-free privacy; against adversaries with side information, Proposition 6 shows only a noised release (Proposition 5) can bound the leakage.* The two-way relationship is then precise: DP $\Rightarrow$ width unconditionally; width $\Rightarrow$ privacy only under a prior bound, and never $\Rightarrow$ DP.

Remark 7 (Two distinct ways the guarantee fails, only one in the model). Theorem 2 can be defeated in two different ways, and the distinction is the whole boundary of the result. *(i) The window narrows.* Publishing finer or higher-order marginals, or an adversary learning an interior count (Proposition 6), shrinks $I_S(e)$. The theorem still holds, the posterior still lies in the interval — but the interval is now tight. This is reconstruction, it remains a *data-only* phenomenon, and it is bounded by the publication budget of Section 10. *(ii) Exchangeability breaks.* An adversary with information distinguishing individuals *within* a cell, having seen u at a rally, escapes $I_S(e)$ entirely, because the hypothesis of Definition 4 fails. No aggregate mechanism, noised or not, restores privacy of a fact the adversary already independently holds; what differential privacy bounds is the *incremental* leakage the release adds, which is the only thing any release-side mechanism can promise. Mode (i) is inside the model and is the subject of the budget below; mode (ii) is outside it and is the reason "privacy" must always be stated as incremental.

10 The reconstruction threshold

The database reconstruction theorem is the right frame for "how much can we publish before raw release stops being safe," and it cuts in the architecture's favor here, for two reasons that we state precisely and one that we leave open.

Proposition 7 (The release sits below the reconstruction regime)**.** *Axial-marginal release publishes $\sum_\ell (k_\ell - 1)$ independent linear functionals of T. Database reconstruction in the sense of Dinur–Nissim requires a number of sufficiently accurate queries growing with the population N. Two facts therefore hold simultaneously:*

(i) ***Query count.*** *For fixed dimension structure, $\sum_\ell (k_\ell - 1)$ is independent of N and is $\ll N$ for any realistic electorate, placing the release below the query-count threshold for blatant reconstruction.*

(ii) ***Query structure.*** *Reconstruction lower bounds are proved for* arbitrary *(e.g. random) linear queries; axial marginals are a fixed, highly redundant, structured family, for which arbitrary-query bounds are not directly applicable and over-estimate the adversary's power.*

Proof. (i) is immediate from the definition of μ and the form of the Dinur–Nissim threshold. (ii) follows because the reconstruction lower bounds quantify over query sets with spreading or incoherence properties that a fixed axial-marginal family does not satisfy; the bounds give existence of *some* hard query set, not hardness of *this* one. □

Proposition 8 (No-noise publication budget)**.** *Release Q exact, linearly independent marginal queries over N voters.*

(a) ***Safe side.*** *If the release keeps every cell at identifiability width $\iota_S(e) \geq \rho$ (Cor. 3 in the two-way case; the LP of Problem 1 in general), then by Theorem 2 every individual's posterior A-probability remains uncertain over a range $\geq \rho$ against all data-only adversaries, with no noise added.*

(b) ***Unsafe side.*** *If $Q = \Omega(N)$, then by the database reconstruction theorem (Dinur–Nissim; textbook form in Dwork–Roth) an adversary holding the exact answers reconstructs a $1 - o(1)$ fraction of individual votes. No noiseless release of $\Omega(N)$ independent exact marginals is private.*

The as-designed system publishes $Q = \sum_\ell (k_\ell - 1)$ per cycle, which is $o(N)$ for any realistic electorate, placing a single cycle in regime (a) whenever its windows are interior. Over a stable electorate, r repeated cycles accumulate rQ queries and approach the budget at $r = \Theta(N/Q)$, the point at which longitudinal release alone, with no change in method, crosses from (a) into (b).

Proof. (a) is Theorem 2 applied cellwise. (b) is the reconstruction theorem: exact answers are the zero-noise case, for which $\Omega(N)$ independent queries suffice to recover almost all records; the marginal queries are linear functionals of the per-voter vote vector and independence is the stated hypothesis. □

Proposition 8 brackets the question from both sides; what it leaves open is the sharp location of the crossing for the *structured* axial family, as opposed to the worst-case independent queries of part (b). For the as-designed release the structure settles it.

Theorem 4 (Order-one release has no accumulation route)**.** *Restrict releases to order-one (axial) marginals of a d-way ($d \geq 2$) tensor with cardinalities $k_1, \ldots, k_d$ fixed independently of the electorate size N. Let $k_{(1)}, k_{(2)}$ be the two largest cardinalities. Then:*

(a) ***No point-identification, ever.*** *The fiber has affine dimension $\prod_\ell k_\ell - \sum_\ell (k_\ell - 1) - 1 \geq (k_{(1)} - 1)(k_{(2)} - 1) > 0$, so the joint is never determined, at any N.*

(b) ***Bounded, N-independent query count.*** *The release fixes exactly $Q = \sum_\ell (k_\ell - 1) + 1$ independent linear functionals of T, a constant in N. Hence $Q/N \to 0$ as the electorate grows, and the $\Omega(N)$-query regime of Proposition 8(b) is* never reached by order-one release*, however many cells the tensor has.*

(c) ***Saturation is the only route.*** *Every cell obeys $n_e \leq \min_\ell m_\ell(i_\ell)$; an individual in cell e is pinned only if the cell's bounds collapse, which by Lemma 2 (Lemma 3 for $d = 2$) requires a margin to saturate. Order-one disclosure therefore occurs only through modes M0/M1, small or near-unanimous slices, and never through the accumulation of many weak queries.*

Proof. (a) $\operatorname{rank} \mu = \sum_\ell (k_\ell - 1) + 1$, since each axis contributes k_ℓ category sums that are dependent only through the shared grand total; the fiber dimension is $\prod_\ell k_\ell - \operatorname{rank} \mu$, and discarding all but the two largest axes lower-bounds it by $(k_{(1)} - 1)(k_{(2)} - 1) \geq 1$. (b) is the same rank count, constant in N by hypothesis, so it cannot grow like N. (c) the upper bound is membership in the smallest containing slice; the lower bound rises to meet it only as some margin approaches saturation, exactly per Lemma 3 in two dimensions and the LP of Lemma 2 in general. □

Theorem 4 resolves the threshold question for the system as designed: there is *no* reconstruction phase transition in N. A single order-one election is safe at every electorate size provided its windows are interior, and the only way to manufacture danger is to shrink the windows (saturation, handled by the safety rule) or to raise the marginal order or cardinality until cells approach singletons. The genuinely open part is therefore narrower than the original conjecture, and concerns only the higher-order regime.

Conjecture 1 (Higher-order reconstruction threshold; open). Theorem 4 closes the order-one case; the open question is the order-≥ 2 regime. Conjecturally there is a function $Q^*(d, k, N, \mathrm{ord})$ such that a release including marginals up to order ord admits no data-only reconstruction of any cell below tolerance τ while the released structure stays below Q^*, and admits reconstruction of some cell above it. The conjectured shape: the transition is driven jointly by the marginal order (order ≥ 2 enters the De Loera–Onn universal regime of Remark 2, where entry ranges acquire arbitrary gaps) and by the product k^d approaching N (cells becoming singletons, $N_e \to 1$, removing the base-rate dilution of Theorem 2). Locating Q^* for the structured higher-order family, as opposed to the arbitrary-query Dinur–Nissim bound, is open and is, in our view, the essential theorem for a full treatment of higher-order releases.

We state Conjecture 1 as a conjecture deliberately. The sufficient safety direction (Corollary 3, Theorem 2) is proved; the sharp transition is not, and we do not claim it.

11 Deployment regimes and noiseless privacy in large electorates

Theorem 2 certifies privacy without noise against data-only adversaries. Whether that is a strong claim or a weak one turns entirely on whether the data-only model is the realistic one for a given election. It is exactly the realistic model for a large, dense electorate, and naming the conditions converts the no-noise guarantee from a curiosity into the system's main practical claim.

Definition 5 (((β, κ)-regular release)**.** A release is (β, κ)-*regular* if every published cell class e has size $N_e \geq \kappa$ and data-only window $I_S(e) \subseteq [\beta, 1-\beta]$ for some $\beta \in (0, \frac{1}{2}]$.

Theorem 5 (Noiseless privacy in the regular regime)**.** *Under a (β, κ)-regular release, against every data-only adversary (Definition 4):*

(a) *each individual's posterior A-probability lies in $[\beta, 1-\beta]$, so no vote is determined beyond the slack β from certainty;*

(b) *any two members of a class are posterior-exchangeable, so each individual hides in an anonymity set of size $\geq \kappa$;*

(c) *the release satisfies noiseless privacy in the sense of Bhaskar et al.: the protection is supplied entirely by the adversary's within-class uncertainty, with no perturbation of the published tallies.*

Proof. (a) is Theorem 2 with the window contained in $[\beta, 1-\beta]$. (b) within a class the marginals are invariant under permuting members, so the posterior is exchangeable over the $N_e \geq \kappa$ members. (c) (a) and (b) bound the adversary's posterior on any individual's secret away from certainty using only the exact released data, which is the noiseless-privacy condition. □

Remark 8 (Why large urban elections are the favorable regime)*.* Two facts about a large, dense electorate make the data-only model accurate rather than optimistic. First, demographic classes are large, κ in the thousands or more, so base-rate dilution is strong and the small-slice failure (M0) does not arise. Second, distinguishing individuals *within* a class, the exchangeability-breaking knowledge of Remark 7(ii), is precisely what anonymity in a large city denies an adversary. The identical release in a village of five fails on both counts. The guarantee is thus not unconditional; it is a statement about a regime, and the regime is the common one for the elections this system targets. Against the realistic large-electorate adversary we claim provable noiseless privacy; against the worst-case adversary we claim only the t-closeness window (Prop. 4) and defer to the differential-privacy layer.

12 A tiered construction: exact outcome, protected overlays

A standing objection to noised election reporting is that voters and courts will not accept a perturbed *outcome*: an election whose winner is computed from noisy totals invites suspicion. The objection is real and it is also avoidable, because the binding outcome does not need noise. It is the most aggregated statistic in the system, and aggregation is what makes it safe.

Definition 6 (Tiered release)**. Tier 0 (binding).** The primary election publishes the exact electorate-wide tally, the order-zero marginal, equivalently the winner. **Tier 1 (overlay).** Each demographic dimension publishes its axial marginal: exact where the cell is (β, κ)-regular (Definition 5), and via a calibrated differential-privacy mechanism, or suppressed, where it is not.

Proposition 9 (Soundness of the tiered release). (a) *Tier 0 reveals only the base rate M/N; against a data-only adversary it moves every individual's posterior to the common value M/N and no further, so it carries no targeted disclosure and may be published exactly.*

(b) *Tier 0 adds no constraint beyond Tier 1, since the electorate-wide total is the sum of any single dimension's marginal; publishing it exactly is therefore free given the overlays.*

(c) *On Tier 1, regular cells are noiseless-private (Thm. 5) and irregular cells are bounded by the DP mechanism (Prop. 5); the binding outcome is never noised.*

Proof. (a) by exchangeability over the whole electorate, the posterior on any individual is M/N given only the total. (b) $M = \sum_c m_\ell(c)$ for every axis ℓ, so the total is implied by any published axial marginal and contributes no new linear constraint to the fiber. (c) is Theorems 2–5 and Proposition 5 applied cellwise. □

Remark 9 (Two problems dissolved together). The tiered split answers two objections at once. The noisy-winner objection is met because Tier 0 is exact and noise lives only on non-binding statistical overlays, exactly where approximation is already the norm, exit polls are estimates and no one mistakes them for the count. The *which-election-is-binding* objection — with separate per-dimension elections, which tally is "the result"?, is met because Tier 0 is the binding election and the Tier 1 overlays are measurements; the cross-election inconsistency of the separated design (the δ term of Remark on inconsistency) then perturbs only the measurements, never the outcome. This is the construction with the verifiability of an exact count and the demographic transparency of protected overlays, and it is the form in which we would deploy the system.

12.1 The exact per-cell budget

The tiered split is a policy until it is made arithmetic. It can be: given the published cells, their group sizes, and their rates, exactly which cells may be published raw, which must be perturbed, and by how much, is a finite computation with a definite answer. The result below makes the schedule exact and shows that, in the regime the system targets, almost all of it, and the entire binding outcome, is spent at zero.

Setup. The release is order-one. A published *cell* is a pair (ℓ, c): category c of dimension ℓ. It has public group size $N_{\ell,c}$ (eligible voters in that category, from the roll) and exact A-rate $p_{\ell,c} = m_\ell(c)/N_{\ell,c}$. Each voter v lies in exactly one category per dimension, occupying the d cells $\{(\ell, c_\ell(v))\}_{\ell=1}^{d}$. Fix a target $(\beta, \kappa, \varepsilon_{\max})$: posterior confined to $[\beta, 1-\beta]$, anonymity set $\geq \kappa$, and tolerated worst-case loss $\varepsilon_{\max}$ per voter. A per-cell budget $\varepsilon_{\ell,c} \geq 0$ is assigned, $\varepsilon_{\ell,c} = 0$ meaning the cell is published exactly.

Theorem 6 (Exact tiered disclosure schedule). (a) ***Decomposition.*** *Against a data-only adversary, v's disclosure is a function of the d rates $\{p_{\ell,c_\ell(v)}\}$ of the cells v occupies, and of nothing else. (Theorem 4 removes any joint-accumulation channel; Theorem 2 makes each within-category posterior equal to that category's rate.)*

(b) ***Schedule.*** *Publish Tier 0 and every (β,κ)-regular cell exactly; coarsen every cell with $N_{\ell,c} < \kappa$ into the smallest containing super-category of size $\geq \kappa$; add $\mathrm{Lap}(1/\varepsilon_{\ell,c})$ noise (sensitivity one) to each remaining cell for which a worst-case guarantee is demanded. This schedule gives every* *voter (i) the noiseless (β, κ) guarantee against the data-only adversary, and (ii) a worst-case $B(v)$-differential-privacy guarantee with*

$$B(v) = \sum_{\ell : (\ell, c_\ell(v)) \text{ noised}} \varepsilon_{\ell, c_\ell(v)},$$

the sum over the $\leq d$ noised cells v occupies.

(c) ***Exactness and minimality.*** *Classifying each cell is $O(1)$ and the whole schedule is computed in time linear in the number of cells. Among schedules meeting the data-only target, publishing every regular cell exactly is cost-minimal, since a regular cell needs no action and any added noise is pure utility loss; the only unavoidable costs fall on sub-κ cells (coarsening) and on cells where a worst-case guarantee is separately required (noise).*

(d) ***Corollary (regular electorate).*** *If every published cell is (β, κ)-regular, the common case for a large, dense electorate (Remark 8), the schedule is "publish everything exactly": total noise zero, every demographic rate exact, and the binding outcome trustworthy without qualification. Noise is then spent only on the few sub-κ or separately-flagged cells: the budget is consumed only by the voters who need it, and is exactly zero for everyone else.*

Proof. (a) By Theorem 4 the order-one release neither point-identifies the joint nor admits an accumulation channel, so the only information bearing on v is the published category rates; by Theorem 2 within-category exchangeability makes v's posterior in dimension ℓ equal to $p_{\ell, c_\ell(v)}$. (b)(i) a regular cell has $p_{\ell,c} \in [\beta, 1-\beta]$ and $N_{\ell,c} \geq \kappa$, so Theorem 5 applies cellwise, and coarsening restores $N \geq \kappa$ where it failed; (ii) $\mathrm{Lap}(1/\varepsilon_{\ell,c})$ added to a count of sensitivity one is $\varepsilon_{\ell,c}$-DP, and basic composition over the $\leq d$ noised cells v occupies gives $B(v)$. (c) the predicate "$N_{\ell,c} \geq \kappa$ and $p_{\ell,c} \in [\beta, 1-\beta]$" is checked in constant time per cell; the minimal super-category merge is a single pass; a regular cell admits the zero-cost action, and no schedule meeting the data-only target costs less than zero on it. (d) immediate from (b) and (c). □

Remark 10 (Why the per-voter budget composes over d cells, not the whole release). The release may hold thousands of cells, but a voter occupies one per dimension, so their loss composes over $\leq d$ terms rather than over the whole board. This is what makes the worst-case allocation small rather than global: when a voter's other dimensions are regular and exact, the single sensitive dimension may carry the entire per-voter budget $\varepsilon_{\max}$, so the noise on a flagged cell is set by its worst-occupancy voter, not by the size of the release. In the regular electorate the constraint is slack everywhere and the allocation is the zero schedule of Theorem 6(d).

Remark 11 (What this buys the deployment). Theorem 6 is the arithmetic form of the design intuition that the outcome must be trusted while the overlays must be private. The binding tally is in Tier 0 and is always exact, so no voter is asked to accept a perturbed winner; the overlays are exact wherever the cell is regular, which in a large electorate is nearly everywhere; and a calibrated, *computed* amount of noise is spent only on the genuinely exposed cells, small categories, and sensitive ones where a worst-case guarantee is demanded. The privacy cost of the election is therefore not a fixed global tax but a per-cell quantity read off the roll and the rates, and it can be certified zero for a well-behaved election. This is the exact calculation the tiered construction was promising.

13 What is proved and what is posed

To bound the contribution precisely:

- **Proved.** Non-identifiability of the joint from axial marginals (Prop. 1); the exact two-margin cell characterization (Lemma 1, Cor. 1); the sharp multi-way bound as an LP (Lemma 2); the disclosure taxonomy (Thm. 1); data-only separation equivalence (Prop. 2); the no-noise individual-privacy guarantee against all data-only adversaries (Thm. 2) and its identity with a release-computable t-closeness bound (Prop. 4); the closed-form two-way width and the resulting $[\tau, M - \tau]$ safety rule (Lemma 3, Cor. 3); the impossibility of deterministic DP (Thm. 3); the one-directional conversions (Prop. 5, Prop. 6); the two-sided no-noise publication budget (Prop. 8); the order-one no-accumulation theorem, which resolves the threshold question for the as-designed release (Thm. 4); the noiseless-privacy guarantee in the (β, κ)-regular regime (Thm. 5); the soundness of the tiered exact-outcome / protected-overlay construction (Prop. 9); and the exact per-cell disclosure budget, with its zero-cost schedule for a regular electorate (Thm. 6).

- **Posed / to develop.** The sharp *higher-order* reconstruction threshold $Q^*(d, k, N, \mathrm{ord})$ (Conj. 1; the order-one case is proved); tractable solution of Problem 1 (hardness, approximation, or decomposable special cases); a tight quantification of the inconsistency term δ; and the auxiliary-adversary version of Prop. 2.

- **Explicitly not claimed.** That the architecture makes individual reconstruction impossible (it does not; Prop. 6); that the no-noise guarantee is differential privacy (it is not; Thm. 3); or that the closed-form width survives to the multi-way case (it does not; it is an upper bound there). The identifiability interval certifies privacy without noise only against data-only adversaries, and a noised release is required beyond that class. How ballots are cast, verified, and protected on untrusted clients, and the cryptographic guarantees of the tally, are the subject of the companion paper, *Bounded-Payoff Remote Elections*, which supplies the election this reporting rule consumes.

Relation to prior frameworks. The identifiability interval is a no-noise, syntactic guarantee in the lineage of k-anonymity (Sweeney 2002), formalized for the modern, prior-bounded setting by noiseless privacy (Bhaskar, Bhowmick, Goyal, Laxman & Thakurta, ASIACRYPT 2011) and the Bayesian / distributional privacy line (Pufferfish: Kifer & Machanavajjhala, ACM TODS 2014). Its conditional nature is exactly the composition-and-auxiliary-information phenomenon of Ganta, Kasiviswanathan & Smith (KDD 2008), which is why Prop. 6 holds and why differential privacy (Dwork, McSherry, Nissim & Smith, TCC 2006) remains the only unconditional option once the data-only assumption is dropped.

References

[1] J. Abowd and M. Hawes. Confidentiality protection in the 2020 US Census of Population and Housing. *Annual Review of Statistics and Its Application*, 10:119–144, 2023.

[2] B. Adida. Helios: web-based open-audit voting. In *USENIX Security Symposium*, pages 335–348, 2008.

[3] K. J. Arrow. *Social Choice and Individual Values*. Wiley, New York, 1951.

[4] J. Benaloh. Verifiable secret-ballot elections. PhD thesis / Technical Report YALEU/DCS/TR-561, Yale University, 1987.

[5] D. Bernhard, V. Cortier, D. Galindo, O. Pereira, and B. Warinschi. SoK: a comprehensive analysis of game-based ballot privacy definitions. In *IEEE Symposium on Security and Privacy (S&P)*, pages 499–516, 2015.

[6] R. Bhaskar, A. Bhowmick, V. Goyal, S. Laxman, and A. Thakurta. Noiseless database privacy. In *ASIACRYPT*, pages 215–232, 2011.

[7] D. Black. On the rationale of group decision-making. *Journal of Political Economy*, 56(1):23–34, 1948.

[8] D. Chaum. Blind signatures for untraceable payments. In *CRYPTO*, pages 199–203, 1982.

[9] M. R. Clarkson, S. Chong, and A. C. Myers. Civitas: toward a secure voting system. In *IEEE Symposium on Security and Privacy (S&P)*, pages 354–368, 2008.

[10] R. Cramer, R. Gennaro, and B. Schoenmakers. A secure and optimally efficient multi-authority election scheme. In *EUROCRYPT*, pages 103–118, 1997.

[11] J. A. De Loera and S. Onn. The complexity of three-way statistical tables. *SIAM Journal on Computing*, 33(4):819–836, 2004.

[12] J. A. De Loera and S. Onn. All rational polytopes are transportation polytopes and all polytopal integer sets are contingency tables. In *Integer Programming and Combinatorial Optimization (IPCO)*, LNCS 3064, pages 338–351. Springer, 2004.

[13] P. Diaconis and B. Sturmfels. Algebraic algorithms for sampling from conditional distributions. *Annals of Statistics*, 26(1):363–397, 1998.

[14] I. Dinur and K. Nissim. Revealing information while preserving privacy. In *ACM Symposium on Principles of Database Systems (PODS)*, pages 202–210, 2003.

[15] A. Dobra and S. E. Fienberg. Bounds for cell entries in contingency tables given marginal totals and decomposable graphs. *Proceedings of the National Academy of Sciences*, 97(22):11885–11892, 2000.

[16] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference (TCC)*, pages 265–284, 2006.

[17] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[18] S. E. Fienberg. Fréchet and Bonferroni bounds for multi-way tables of counts with applications to disclosure limitation. In *Statistical Data Protection (SDP'98)*, pages 115–129. Eurostat, 1999.

[19] S. R. Ganta, S. P. Kasiviswanathan, and A. Smith. Composition attacks and auxiliary information in data privacy. In *ACM SIGKDD (KDD)*, pages 265–273, 2008.

[20] A. Gibbard. Manipulation of voting schemes: a general result. *Econometrica*, 41(4):587–601, 1973.

[21] W. Hoeffding. Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5:179–233, 1940.

[22] A. Juels, D. Catalano, and M. Jakobsson. Coercion-resistant electronic elections. In *ACM Workshop on Privacy in the Electronic Society (WPES)*, pages 61–70, 2005.

[23] D. Kifer and A. Machanavajjhala. Pufferfish: a framework for mathematical privacy definitions. *ACM Transactions on Database Systems*, 39(1):3:1–3:36, 2014.

[24] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. ℓ-diversity: privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data*, 1(1):3:1–3:52, 2007.

[25] K. O. May. A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica*, 20(4):680–684, 1952.

[26] R. D. McKelvey. Intransitivities in multidimensional voting models and some implications for agenda control. *Journal of Economic Theory*, 12(3):472–482, 1976.

[27] H. Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980.

[28] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: privacy beyond k-anonymity and ℓ-diversity. In *IEEE International Conference on Data Engineering (ICDE)*, pages 106–115, 2007.

[29] M. A. Satterthwaite. Strategy-proofness and Arrow's conditions. *Journal of Economic Theory*, 10(2):187–217, 1975.

[30] L. Sweeney. k-anonymity: a model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002.

[31] G. Magarshak. Bounded-payoff remote elections: continuity, attestation, and dual-root cast-as-intended verification. Companion manuscript, 2026. The casting-and-integrity half of the system, organized around the aggregation rule's Lipschitz constant; cites the present paper for the disclosure and reporting half, and Intercloud (arXiv:2605.22830) for the consensus layer.