CCA: Consistent Targets, Uniform-Compatible Quotas, and Theorem 2 Rounding

Computational replication for Section 6.4

10 October 2026

1 Results at a glance

The supplied CCA data contain 825 respondents and imply a selected-group size of 75. We constructed the two requested consistent instances and drew 10,000 groups independently from the sampling procedure underlying Theorem 2 for each instance. Every saved group contains exactly 75 distinct respondents. The primary outcome is the average, over all 66 listed traits, of the number of seats outside a trait's permitted interval. The maximum quota miss is also recorded separately for every group.

Case 1 keeps the original quotas and chooses profile-dependent probabilities by least-squares calibration to the uniform vector. **Case 2** uses probability 1/11 for every respondent and replaces quotas by the floor and ceiling of each trait's uniform expected representation.

Statistic	Case 1	Case 2
Draws	10,000	10,000
Mean quota violation (66 traits)	0.430420	0.097873
Approx. 95% Monte Carlo interval	[0.429139, 0.431700]	[0.097201, 0.098545]
Mean quota violation (61 nonempty traits)	0.465700	0.105895
Mean of group maximum miss	1.7965	1.2889
Largest observed miss	3	3
Theoretical bound for every group	3	3
Mean number of violated traits	27.0759	6.1332
Mean total shortfall	13.9436	1.2145
Mean total excess	14.4641	5.2451

All quota-miss quantities are measured in seats. The confidence intervals concern Monte Carlo estimation of the mean, not uncertainty about the supplied population.

Group maximum miss	Case 1		Case 2	
	Groups	Share	Groups	Share
0	0	0.00%	4	0.04%
1	2,305	23.05%	7,164	71.64%
2	7,425	74.25%	2,771	27.71%
3	270	2.70%	61	0.61%

The observed maximum is a property of these samples. Theorem 2's bound of three applies to every outcome of the exact-arithmetic construction. The supplied finite samples are not exhaustive lottery supports and are not claimed to reproduce all target probabilities exactly as empirical frequencies.

2 Input data and the two constructions

This experiment follows Section 6.4 and the proof of Theorem 2 in the supplied manuscript, *Meritocratic Lotteries under Intersectional Diversity Quotas*, by Yuval Heller, Ron Peretz, and Amnon Schreiber (the supplied PDF is dated 10 October 2026).

2.1 Input validation

The input files are the supplied `respondents.csv` and `categories.csv`. The original bytes are preserved in `input/`; their SHA-256 hashes are in `run_manifest.json`. Respondents have unique `nationbuilder_id` values and exactly one trait in each of the four columns a, b, c, d. No response is missing and every recorded trait matches a category row. There are 554 distinct complete profiles (a, b, c, d).

The category file lists 10, 3, 5, and 48 traits in the respective dimensions, hence 66 rows in total. Five traits have no respondents: $a1, a2, c5, d11, d43$. They have lower quota zero and never contribute a quota violation. Thus there are 61 nonempty traits. The original quotas in each of a, b, c are exact and sum to 75, fixing the group size $k = 75$. Every respondent affects four trait constraints, so the maximum degree is $D = 4$ and the Theorem 2 bound is $D - 1 = 3$.

2.2 Case 1: profile-dependent target probabilities

Section 6.4 specifies consistency and dependence on the trait profile but does not uniquely specify the target vector. We make that choice explicit. Let T be the set of complete profiles and m_t their multiplicities. The common probability p_t for members of profile t is the unique solution of

$$\begin{aligned} \min_{(p_t)_{t\in T}} \quad & \frac{1}{2}\sum_{t\in T} m_t \left(p_t - \frac{1}{11}\right)^2, \\ \text{subject to} \quad & 0 \le p_t \le 1, \qquad \sum_{t\in T} m_t p_t = 75, \\ & l_e \le \sum_{t:\, t \text{ contains } e} m_t p_t \le u_e \quad \text{for every listed trait } e. \end{aligned}$$

Positive profile multiplicities make the objective strictly convex, so the feasible solution is unique. Weighting by m_t minimizes squared distance over respondents, rather than giving equal weight to differently sized profiles.

The original d bounds sum to 73 at their lower endpoints and 77 at their upper endpoints. The empty traits $d11, d43$ must have expectation zero. Consequently, consistency and total size 75 force $d7, d24$ to have expectation four. All other nonempty trait expectations are already fixed by exact quotas. The implementation derives these tightened expectations by iterating partition-sum bounds.

SLSQP identifies the active probability bounds. A rational linear solve then reconstructs the solution and verifies every trait expectation, the total of 75, the probability bounds, and the convex program's first-order optimality conditions exactly. The exact fractions and a checkable multiplier certificate are archived. The sole probability-one coordinate is fixed by its singleton trait constraint before the reduced optimality conditions are checked.

The resulting calibration assigns probability zero to **424 respondents** (264 profiles), and probability one to **respondent 695**, the sole member of trait $d1$, whose quota is one. The smallest positive probability is 0.0000524988; the largest probability below one is 0.9957377835. The sum of squared deviations from 1/11 is 18.8014073707. These are constructed consistent targets. They do not estimate pre-existing merit, response rates, or priorities. The zeros are consequences of this chosen objective; we do not assert that each zero is required by quota consistency alone.

2.3 Case 2: uniform targets and adjusted quotas

For every respondent v, set $p_v = k/n = 75/825 = 1/11$. If trait e contains N_e respondents, its expected count is exactly $N_e/11$. We use

$$l_e^{\text{unif}} = \left\lfloor \frac{N_e}{11} \right\rfloor, \qquad u_e^{\text{unif}} = \left\lceil \frac{N_e}{11} \right\rceil.$$

All floor and ceiling operations are evaluated using rational arithmetic. This changes 60 of the 66 rows. For example, $b1$ has 15 respondents, so its quota changes from $[8, 8]$ to $[1, 2]$; $b2$ changes from $[42, 42]$ to $[22, 23]$; and $b3$, with 561 respondents, changes from $[25, 25]$ to $[51, 51]$. The full audit is in `quota_audit.csv`.

3 Rounding algorithm and experimental protocol

We implement the sampling construction in the proof of Theorem 2. Start each draw from $x = p$. Freeze integral coordinates. For the remaining fractional set F, keep the total-sum equation and every trait-sum equation whose trait contains more than $D = 4$ members of F. Find a nonzero null direction h for these equations, supported on F. If $\alpha, \beta > 0$ are the maximal feasible steps in the positive and negative directions, replace x by

$$x + \alpha h \quad \text{with probability } \frac{\beta}{\alpha + \beta}, \qquad x - \beta h \quad \text{with probability } \frac{\alpha}{\alpha + \beta}.$$

This preserves the current vector in expectation and freezes at least one additional coordinate in exact arithmetic. Iterate until integral. The total-sum equation gives exactly 75 selected respondents. A trait can begin to drift only after it contains at most four fractional coordinates, yielding the strict deviation bound below four from its consistent initial expectation, hence quota miss at most three for integer quotas. The exact-arithmetic induced law has the prescribed individual marginals.

Theorem 2 permits choices of null directions. We fix an implementation for reproducibility: fractional columns are considered in original respondent order; the total row comes first and trait rows retain original category order. Incremental column elimination stops at the first dependency. Each new column uses its largest-magnitude residual entry as pivot, taking the first row in a tie. The direction is normalized to maximum absolute entry one. This rule is not tuned to the observed outcomes.

Each case uses 10,000 fresh paths, with pseudorandom seeds 20261010 and 20261011. The generator is the standard C++ `mt19937_64`; the top 53 bits of each output are converted to a number in $[0, 1)$ by multiplication by 2^{-53}. **No support-compression algorithm is run.** In particular, equal weights on the saved draws form an empirical approximation, not the explicit marginal-preserving lottery of support at most 825 that can be obtained by the theorem's additional compression construction. The reported averages are Monte Carlo estimates for this sampling implementation, not exact expectations over an enumerated support.

Null-direction elimination uses double precision; states, step lengths, and sums use C++ `longdouble`. Coordinates within 10^{-15} of an endpoint are snapped to it. A sole fractional coordinate, which cannot occur under the total equation in exact arithmetic, is corrected only when within 10^{-9} of its uniquely required integer. Every active equation is checked during rounding, and every final integer group and quota miss is checked again independently. The measured maxima are much smaller than the stopping tolerances:

Numerical diagnostic	Case 1	Case 2
Null-direction equality residual	4.22e-15	5.36e-15
Group-size equality drift	1.69e-15	2.67e-14
Active-trait equality drift	1.52e-15	2.01e-14
Final numerical correction	1.69e-15	2.67e-14
Groups requiring final correction	157	9,990

The exact theoretical marginal statement and the floating-point implementation should be distinguished. The report supplies numerical residuals and empirical frequency diagnostics; it does not claim a proof that rounding errors are zero.

4 Outcome definitions and interpretation

For a realized selected group S and a listed trait e, define

$$c_e(S) = |S \cap e|, \qquad \delta_e(S) = \max\{l_e - c_e(S),\ c_e(S) - u_e,\ 0\}.$$

Thus a count inside its interval has miss zero; missing a quota by two seats has miss two. Shortfalls and excesses are not canceled. The two requested group-level metrics are

$$A(S) = \frac{1}{66}\sum_e \delta_e(S), \qquad M(S) = \max_e \delta_e(S).$$

The first-page average is $R^{-1}\sum_{r=1}^{R} A(S_r)$, where $R = 10{,}000$. The nonempty-trait average instead divides the same total by 61. The approximate 95% Monte Carlo interval is the sample mean plus or minus $1.96 s_A/\sqrt{R}$, using variation *between groups*. Traits within a group are dependent and are not treated as independent observations. The maximum-miss histogram records $M(S_r)$ separately for every draw.

All observed groups satisfied the theoretical quota-miss bound of three. The per-group CSVs retain the exact maximum, average, shortfall, excess, and all 66 counts and misses, allowing the summary to be reconstructed without rerunning the rounding algorithm. The all-trait appendix gives each trait's mean miss and largest observed miss for both cases.

The largest absolute differences between empirical inclusion frequencies and target probabilities were 0.013761 and 0.010509, respectively. These are maxima over 825 respondents, not individual confidence limits. The file `marginal_diagnostics.csv` records each respondent's target, number of selections, frequency error, and binomial standard error $\sqrt{p_v(1-p_v)/R}$. Sampling fluctuation means the archived draw frequencies need not equal the exact-law marginals. There were 10,000 and 10,000 distinct sampled groups; these counts are not support-size estimates.

The lower empirical violations in Case 2 concern a different quota system and a different marginal vector. They do not establish superiority of one algorithm: both cases use the same rounding rule. Choices of consistent targets and of null directions can change average performance. This experiment implements only the two requested CCA constructions and does not reproduce the manuscript's HD results or its comparison with LEXIMIN. The present CCA numbers should be identified as such if added to Section 6.4.

5 Files and replication

- In `case1_original_quotas/`, `respondents.csv` adds the column `selection_probability`. The unchanged quotas are copied alongside it. The exact profile fractions are in `profile_targets.csv`, with their `calibration_certificate.json`. Decimal CSV probabilities use 17-digit representations; the exact fractions are authoritative.
- `case2_uniform/categories.csv`: the requested adjusted quotas. A companion `respondents.csv` gives the common probability 1/11.
- In both case folders: `panel_members.csv` lists each group's 75 original identifiers; `panel_metrics.csv` gives every group's requested outcomes and all trait counts; `trait_results.csv` and `summary.json` give aggregate results.

- `run_experiment.py` constructs the inputs, compiles and calls `theorem2_sampler.cpp`, and writes all results. The independent standard-library program `verify_results.py` checks exact target feasibility and optimality, every group, and all reported integer counts and summary statistics.
- `run_manifest.json` records package/compiler versions, seeds, ordering, and input hashes. `verification.json` records the independent checks. `README.txt` documents every output and all commands.

The archived environment used Python 3.12.14, NumPy 2.3.5, SciPy 1.17.0, and GCC 13.3.0 on x86_64 Linux. To reproduce from the delivered exact targets:

```
python3 -m pip install -r requirements.txt
OPENBLAS_NUM_THREADS=1 OMP_NUM_THREADS=1 \
  python3 run_experiment.py --draws 10000 --seed 20261010
python3 verify_results.py
python3 build_report.py
pdflatex -interaction=nonstopmode -halt-on-error CCA-results-report.tex
pdflatex -interaction=nonstopmode -halt-on-error CCA-results-report.tex
```

Add `--recalibrate` to the experiment command to reconstruct the unique rational calibration from the raw inputs. A C++17 compiler is required for sampling; LaTeX is needed only to rebuild the PDF. The verifier alone requires only Python's standard library. Different floating-point behavior across platforms can alter endpoint choices and later random paths; the delivered panel files provide the exact empirical record underlying this report.

A All-trait quota and violation results

Each interval is lower–upper. “Target 1” is the exact Case 1 expected count; the Case 2 target is always population divided by 11. “Mean” is mean quota miss over the sampled groups; “Max” is the largest observed miss for that trait. Empty traits are included in the primary denominator of 66.

		Case 1: original quotas				Case 2: uniform		
Trait	Pop.	Quota	Target 1	Mean	Max	Quota	Mean	Max
a1	0	0–0	0	0.0000	0	0–0	0.0000	0
a2	0	0–0	0	0.0000	0	0–0	0.0000	0
a3	31	2–2	2	0.3734	3	2–3	0.0950	2
a4	40	2–2	2	0.3483	2	3–4	0.1014	2
a5	160	10–10	10	0.4834	3	14–15	0.0972	3
a6	155	11–11	11	0.3899	2	14–15	0.0837	1
a7	142	14–14	14	0.3831	2	12–13	0.2011	3
a8	139	16–16	16	0.3890	2	12–13	0.0959	2
a9	79	8–8	8	0.4848	3	7–8	0.1006	2
a10	79	12–12	12	0.5535	3	7–8	0.0909	2
b1	15	8–8	8	0.0067	1	1–2	0.0000	0
b2	249	42–42	42	0.3148	1	22–23	0.0029	1
b3	561	25–25	25	0.3123	1	51–51	0.0070	1
c1	307	31–31	31	0.4140	1	27–28	0.1380	2
c2	394	21–21	21	0.4419	2	35–36	0.0907	2
c3	79	5–5	5	0.3425	2	7–8	0.0916	1
c4	45	18–18	18	0.3708	2	4–5	0.0424	1
c5	0	0–0	0	0.0000	0	0–0	0.0000	0
d1	1	1–1	1	0.0000	0	0–1	0.0000	0
d2	30	2–2	2	0.5796	3	2–3	0.1352	3
d3	25	1–1	1	0.5517	3	2–3	0.0595	2
d4	11	1–1	1	0.3757	2	1–1	0.4353	3
d5	31	2–2	2	0.6036	3	2–3	0.1677	3
d6	25	3–3	3	0.6623	2	2–3	0.0707	2
d7	39	3–4	4	0.2773	2	3–4	0.0959	2
d8	10	2–2	2	0.5827	3	0–1	0.1925	3
d9	18	2–2	2	0.4019	2	1–2	0.1206	3
d10	19	1–1	1	0.2688	2	1–2	0.1245	3
d11	0	0–1	0	0.0000	0	0–0	0.0000	0
d12	49	3–3	3	0.5232	3	4–5	0.0854	2
d13	6	1–1	1	0.5533	2	0–1	0.0718	3
d14	10	1–1	1	0.5057	3	0–1	0.1838	3
d15	5	1–1	1	0.5538	3	0–1	0.0577	3
d16	39	2–2	2	0.3278	2	3–4	0.0912	2
d17	13	1–1	1	0.4732	3	1–2	0.1013	2
d18	4	1–1	1	0.5136	3	0–1	0.0420	2
d19	36	3–3	3	0.4789	3	3–4	0.0678	2
d20	5	1–1	1	0.5939	2	0–1	0.0647	2
d21	19	1–1	1	0.4428	3	1–2	0.1370	3
d22	11	1–1	1	0.5045	3	1–1	0.4158	3
d23	19	1–1	1	0.4915	3	1–2	0.1411	3
d24	17	3–4	4	0.3238	2	1–2	0.0825	3
d25	15	2–2	2	0.5473	3	1–2	0.0402	2
d26	26	3–3	3	0.5547	3	2–3	0.0731	2
d27	14	1–1	1	0.5501	3	1–2	0.0623	2
d28	15	1–1	1	0.5794	3	1–2	0.0850	2
d29	3	1–1	1	0.5517	2	0–1	0.0224	2
d30	4	1–1	1	0.6211	3	0–1	0.0458	2
d31	48	3–3	3	0.5355	3	4–5	0.0913	2
d32	9	1–1	1	0.5919	3	0–1	0.1614	3

Trait	Pop.	Quota 1	Target 1	Mean 1	Max 1	Quota 2	Mean 2	Max 2
d33	9	1–1	1	0.5685	3	0–1	0.1384	3
d34	40	3–3	3	0.5225	3	3–4	0.1223	3
d35	3	1–1	1	0.4685	2	0–1	0.0216	2
d36	43	2–2	2	0.5245	3	3–4	0.1798	2
d37	21	2–2	2	0.4211	2	1–2	0.1735	2
d38	30	3–3	3	0.6208	3	2–3	0.1379	2
d39	8	1–1	1	0.5620	2	0–1	0.1088	2
d40	10	1–1	1	0.5262	3	0–1	0.1815	3
d41	9	1–1	1	0.5308	3	0–1	0.1600	3
d42	41	2–2	2	0.4775	3	3–4	0.1345	3
d43	0	0–1	0	0.0000	0	0–0	0.0000	0
d44	15	1–1	1	0.5584	2	1–2	0.0516	3
d45	3	1–1	1	0.5414	2	0–1	0.0215	2
d46	10	1–1	1	0.3481	3	0–1	0.1877	3
d47	5	1–1	1	0.5757	3	0–1	0.0689	2
d48	2	1–1	1	0.4320	1	0–1	0.0077	1