# Are baby names a pure-drift meme pool? Re-testing random copying, 1880–2025

**Preprint. Not peer reviewed.** This is the author's manuscript, prepared for submission to a peer-reviewed journal.

**Peter Kalafatis**

Independent researcher, New York, NY, USA. ORCID: 0009-0000-1637-7142. Correspondence: x9m8x7@gmail.com

Version 1, November 2026.

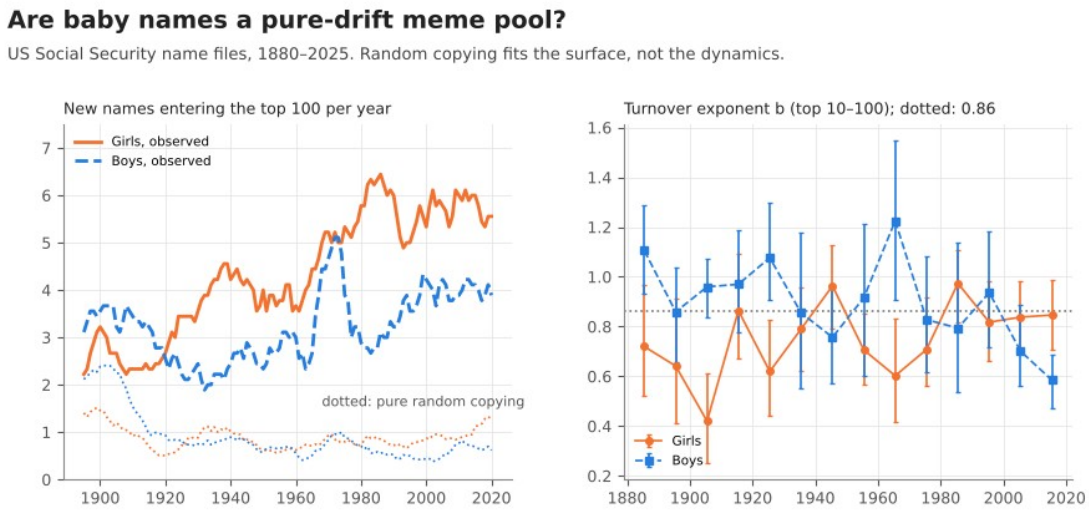
## Abstract

Hahn and Bentley (2003) proposed that American first names behave like neutral variants spread by random copying, and later work used the turnover of top-y popularity lists as a test of drift. I re-test this with US Social Security name files for 1880–2024 (2,149,123 name–sex–year records), replicating turnover on the current release, which adds 2025. Analyses, candidate break years (1946, 1965, 2000) and tests were fixed in a hashed plan before outcomes were computed. No year-by-sex frequency distribution favoured a power law over a lognormal (lognormal favoured in 45 of 290). The top-10-to-100 turnover exponent stayed near the neutral benchmark of 0.86 (95% interval including it in 10 of 14 decades per sex), but this depends on the list sizes used. Census-scale neutral simulations usually matched the exponent but never the concentration of births in the top 100 names. Random copying from each previous year's actual cohort produced only a fifth (girls) to under a third (boys) of observed top-100 turnover; matching it requires an effective copying pool of 1.4–4.8% of births since 1941. Boys' exponent declined

after 2000 (Holm-adjusted p = 0.0075). Drift is a useful null for baby names but not an adequate description.

**Keywords:** cultural evolution; neutral model; random copying; baby names; turnover; memetics; drift

**Social media summary:** US baby names since 1880: random copying fits on the surface, but names churn 3–5 times faster than drift allows.




*Graphical abstract.*

## 1. Introduction

A meme pool in which nothing matters except how often a variant was seen would still change. Popular variants would come and go, rare variants would appear and vanish, and a few would rise to dominance by chance alone. This is the cultural analogue of neutral genetic drift (Kimura & Crow, 1964), and since Neiman's (1995) work on prehistoric ceramic decoration it has served as the default null model for cultural change: before claiming that some variant spread because it was better, more prestigious or more fashionable, show that random copying could not have produced the pattern.

First names are an unusually good test case. Each year's choices are recorded nearly completely, the variants are discrete, and most names have no obvious functional advantage. Hahn and Bentley (2003) analysed the SSA's top 1,000 names per decade for 1900–1999 and reported that the frequency distribution of names in each decade was well described by a power law, as a simple random-copying model with innovation predicts, and that the model reproduced the observed distribution closely. Bentley, Hahn and Shennan (2004) extended the argument to prehistoric pottery motifs and patent citations, and Bentley, Lipo, Herzog and Hahn (2007) proposed a second signature: the number of new entries z per period in a top-y list rises roughly in proportion to y, with $z \approx y\sqrt{\mu}$ for innovation fraction μ. Evans and Giometto (2011) then showed that this linear rule is only an approximation; in the Wright–Fisher model their fitted relationship was $z \propto y^{0.86}$ in one parameter region ($N\mu > 0.15y$, smaller N) and $y^{0.88}$ in another (larger N), and they found that Bentley's formula does not hold at all in the Moran model. Acerbi and Bentley (2014) adopted 0.86 as the neutral benchmark, showed in simulation that conformist copying makes the turnover profile convex (exponent above 0.86) and anti-conformist copying makes it concave (below 0.86), and applied this to SSA names: boys' names in 1880–1930 gave an exponent of 1.69, girls' names in the same period 0.81 (indistinguishable from neutral), and both sexes in 1960–2010 gave concave profiles (0.51 and 0.56). Newberry and Plotkin (2022), using a different method on the same kind of data, inferred negative frequency-dependent selection: common names tend to decline and rare names tend to grow, faster than drift would produce.

Baby names have therefore already been used both as an example of drift and as an example of departure from it. Three gaps remain. First, the classic tests used decade lists or two 50-year blocks; whether a drift fit "held" across the whole 145-year record, and whether it broke at identifiable points, has not, as far as I am aware, been tested with pre-specified break years. Second, the benchmark exponent (0.86) was derived for particular simulation settings. Whether it is the right benchmark for populations of one to two million births per year, sampled annually and with the SSA's suppression of rare names, is an empirical

question that can be answered by simulating the neutral model at those settings. Third, the files have been revised and extended since 2014; the current release covers births through 2025.

This paper is a re-test and extension, not a new idea. It belongs to a sub-series on memetics in which earlier papers asked whether memes in human memory copy faithfully enough to count as replicators (Kalafatis, 2026a), whether language models replicate or merely rephrase text passed along a chain (Kalafatis, 2026b), and how attention to internet memes decays (Kalafatis, 2026c). Names are the cleanest large-scale record of copying in the series: if any meme pool should look like pure drift, this one should. The question here is whether it does, for how long, and in what respects.

# 2. Methods

## 2.1 Pre-specified plan

After the data had been downloaded and checked for integrity, but before any power-law fit, turnover statistic, simulation or break test had been computed, I wrote an analysis plan (code/00_analysis_plan.txt) fixing the units, list-size grids, estimators, simulation design, the three candidate break years and their justification, the test, the multiple-comparison correction, and the interpretation rules. Its SHA-256 hash was recorded by the first script of the pipeline (results/plan_hash.txt). Before writing it I had read the published results summarised in the Introduction, including Acerbi and Bentley's (2014) period exponents, so the plan is pre-specified with respect to these data, not with respect to the literature. Two analyses were added after the plan; they are labelled as such wherever they appear (§2.6, §2.7), and all deviations are listed in §2.10.

## 2.2 Data and data route

The SSA publishes national counts of first names on Social Security card applications by year of birth and sex as a single archive (names.zip) of yearly files (name, sex, count), restricted to names given to at least five babies of one sex in a year (SSA, n.d.-a). The computing environment used for this study could not

reach ssa.gov. The primary data are therefore the 145 yearly files yob1880.txt–yob2024.txt from a public GitHub mirror (repository aramshiva/names, path db/names/, commit ecc921d17155412d4ed593d5d2b1c4290184c821; SHA-256 hashes of all files in results/raw_file_sha256.json). The mirror is an earlier SSA tabulation than the current one: its bundled readme is dated to SSA records as of 3 March 2024, although it also contains a 2024 file.

To check the mirror and to extend the analysis to 2025, I ran the official file through a short JavaScript pipeline inside the author's Chrome web browser on the SSA site (code/07_browser_replication.js). The browser was driven by the AI assistant (Claude, through its Chrome extension; §2.11), and its outputs, capped at about 1,000 characters per call, were transcribed by hand from the tool output into data/derived/browser_official_release_outputs.json; code/08_official_release.py checks the transcription for internal consistency. The browser downloaded names.zip (7,860,026 bytes; Last-Modified 8 May 2026; SHA-256 cd78e975…3724), unzipped it, hashed each file, compared per-year row counts and birth totals with the mirror, recomputed every primary-grid turnover value (§2.4) for 1881–2025, and returned only these aggregates. Of the 145 files for 1880–2024, 17 were byte-identical to the mirror and 128 differed. The differences are SSA's routine re-tabulation: summed over all years, the absolute differences were 732 rows, 40,286 female and 48,347 male births, and the largest difference in any year's total births was 0.54% (2023). Of the 2,880 primary-grid turnover values for 1881–2024, 28 differed between the two releases, each by one name. These 28 values and the 2025 values were returned from the browser and used to rebuild the primary turnover table for the current release (code/08_official_release.py), which reproduces the browser-computed decade exponents to three decimals. Everything except the turnover analysis of 2025 is computed on the mirror.

Unit of analysis: a name string within one sex in one year. Spellings were not merged (Aiden and Aidan are different names), as in the earlier studies. Within a year and sex, names were ranked by count, with ties broken alphabetically, which is the SSA's own file order.

Three properties of the data matter for interpretation (SSA, n.d.-b). First, the floor of five removes the rarest names; in 2024, 13.7% of listed girls' names and 13.6% of boys' names sit exactly at the floor, and an unknown number of names below it are missing. This truncates the lower tail of every frequency distribution, so distributional fits start at $x_{min} \geq 5$. Second, many people born before 1937 never applied for a card, so early cohorts are survivors who applied later; the early-year counts (90,994 girls and 110,490 boys in 1880) are far below actual births, and boys outnumber girls in the 1880s files. Third, the files are unedited: sex coding can be wrong, and hyphens and spaces are removed.

### 2.3 Signature 1: the frequency distribution

For each year and sex I fitted a discrete power law to name counts by maximum likelihood following Clauset, Shalizi and Newman (2009), using the powerlaw package (Alstott, Bullmore & Plenz, 2014) with $x_{min}$ chosen by minimising the Kolmogorov–Smirnov distance over values $\geq 5$, and compared it with a lognormal by the normalised log-likelihood ratio R. The pre-specified rule was: power law favoured if $R > 0$ and $p < 0.05$; lognormal favoured if $R < 0$ and $p < 0.05$; otherwise not distinguishable. As a sensitivity check the fit was repeated with $x_{min}$ fixed at 5.

### 2.4 Signature 2: the turnover profile

For each sex and each consecutive pair of years $(t-1, t)$, $z_y(t)$ is the number of names in year t's top-y list that were not in year $t-1$'s top-y list. The primary grid was $y = 10, 20, \ldots, 100$. Sensitivity grids were $y = 1, \ldots, 30$ (the range Acerbi and Bentley used for SSA names) and $y = 50, 100, 200, \ldots, 1{,}000$ (closer to Bentley et al.'s 2007 lists). The turnover exponent b was estimated from $\log E[z_y] = \log a + b \log y$ by a Poisson generalised linear model with log link, pooling all year-pairs in a window. Windows were the 14 decades 1881–1890 to 2011–2020 plus the short window 2021–2024 (2021–2025 for the current release); b was also estimated for each single year-pair (annual series $b_t$). Because $z_y$ values for nested lists in the same year are not independent, uncertainty for decade exponents comes from a cluster bootstrap that resamples year-pairs within the decade (1,000 resamples, percentile intervals), not from the

model-based standard errors. Benchmarks are b = 0.86 (Evans & Giometto, 2011; Acerbi & Bentley, 2014) and b = 1 (Bentley et al., 2007). The 0.86 value is a Wright–Fisher fit valid where $N\mu > 0.15y$ (always true at the scales here); Evans and Giometto's large-N fit gives 0.88, and in the Moran model neither rule holds, which is one reason the neutral model is also simulated directly (§2.5).

## 2.5 Neutral simulation at the data's own scale (planned)

The neutral model is a Wright–Fisher process with infinite-alleles innovation: each year N newborns of one sex are named; with probability μ a newborn receives a name never used before; otherwise it copies the name of a uniformly random member of the previous year's cohort (Hahn & Bentley, 2003; Acerbi & Bentley, 2014). I simulated it separately for each decade and sex at the data's census scale, N = the decade's mean annual number of births in the files (from 109,594 boys per year in the 1880s to 2,055,807 in the 1950s). Two calibrations of μ were pre-specified. Calibration A sets μ to the decade's mean annual share of births whose name appears in the files for the first time ever within that sex; this is an imperfect proxy because a newly invented name only enters the files once it reaches five babies. The 1880s are excluded from calibration A because almost every name is "new" relative to the start of the record. Calibration B chooses μ by bisection on log μ so that the simulated number of distinct names with at least five bearers matches the decade's observed mean. Each run started from the observed counts of the year before the decade, rescaled to N; ran a burn-in of $4/\mu$ generations (the convention of Acerbi & Bentley, 2014), capped at 20,000; and then recorded 11 consecutive years. The same floor of five and the same tie-breaking rule were applied to simulated counts before computing turnover, b, the share of the top 100 names and the power-law exponent of the end state. There were 10 replicate runs per decade, sex and calibration, with fixed seeds. An observed decade is "consistent with neutral" on b if its b lies inside the 2.5–97.5 percentile range of the replicate b values, and likewise for the top-100 turnover rate $z_{100}/100$.

## 2.6 One-step neutral forecast (added after the plan)

The equilibrium simulation asks what drift does to a population that has had time to settle. A simpler and more direct question is what drift would do to the population that actually existed. For every year-pair from 1891 to 2024 I took the observed year t−1 cohort as the parent pool and drew year t by random copying: a binomial number of innovators at the calibration-A rate, and a multinomial sample of the remaining births from the parent frequencies, with N equal to year t's observed births. I computed the forecast $z_y$ on the primary grid over 200 replicates per year-pair and compared it with the observed $z_y$. This forecast makes no equilibrium assumption and uses no free parameter beyond μ, which has almost no effect on top-100 lists.

## 2.7 Effective copying pool (added after the plan)

If random copying from all births changes the top lists too slowly, random copying from a smaller effective pool will change them faster. For each decade and sex I repeated the one-step forecast with year t frequencies drawn as a multinomial sample of size $N_e$ (innovation at the calibration-A rate) for $N_e$ = 10^3, 10^3.5, …, 10^7 (40 replicates per year-pair), found the $N_e$ at which mean forecast $z_{100}$ equals the observed decade mean by log-linear interpolation, and then reran the forecast at exactly that $N_e$ to obtain its turnover exponent b. This gives the size of copying pool that drift would need to match the level of turnover, and asks whether drift with that pool also matches its shape.

## 2.8 Break tests (pre-specified)

Candidate break years were fixed in the plan with one reason each: 1946, the start of the post-war baby boom; 1965, the Immigration and Nationality Act, at the start of the late-20th-century rise in name diversity; and 2000, when about half of US households had internet access (50.5% in September 2001; National Telecommunications and Information Administration [NTIA], 2002). For each candidate τ and sex I fitted $v_t = c_0 + c_1(t-\tau) + d_0 D_t + d_1 D_t(t-\tau)$, with $D_t = 1$ for $t \geq \tau$, and tested $d_0 = d_1 =$

0 with a Wald test using Newey–West heteroskedasticity- and autocorrelation-consistent covariance with five lags (Chow, 1960; Newey & West, 1987). The six tests per outcome (three years by two sexes) were Holm-corrected (Holm, 1979). The primary outcome was the annual exponent b_t; secondary outcomes were the annual top-100 turnover rate z_100(t)/100 and the annual power-law exponent α_t. As an exploratory, not inferential, step I also scanned τ from 1895 to 2010 and report the year with the largest Wald statistic (cf. Andrews, 1993); no p-value is attached to it.

### 2.9 Software and reproducibility

All analysis ran in Python 3.13 (numpy 2.5.3, pandas 3.0.5, scipy 1.18.1, statsmodels 0.15.0, powerlaw 1.5, matplotlib 3.11.2). The pipeline is numbered (code/00–12, with 11_tables.py building the tables and 12_patch_docx.py formatting the Word file); every number quoted in this manuscript is written by code/10_numbers.py to results/manuscript_numbers.json. Random seeds are fixed in each script.

### 2.10 Deviations from the analysis plan

(1) Two analyses were added after the plan (§2.6, §2.7). (2) The plan did not specify how uncertainty in decade b would be estimated; I used a year-pair cluster bootstrap because nested lists make Poisson standard errors too small. (3) The planned equilibrium simulation was first run without recording the end-state frequency statistics; it was rerun from scratch with those statistics added, with the same seeds and settings, before its results were examined. (4) Calibration B was tuned with short runs (burn-in capped at 3,000 generations) and then run with the full 4/μ burn-in; the resulting number of names stayed within 9.6% of the target. (5) The plan said calibration B would choose μ “from a grid”; it was implemented as a bisection on log μ (nine steps). (6) The power-law fit with x_min fixed at 5 is a sensitivity check that the plan did not list. (7) The plan set the burn-in at 4/μ generations; the code caps it at 20,000 generations, a cap no run reached (maximum 19,326). (8) The plan’s own verdict rule says the neutral fit “broke” if any candidate

break is significant after Holm correction. By that rule the fit broke for boys' names (2000) and held for girls' names; §4.3 discusses how much that label should carry.

### 2.11 Use of AI

This study was carried out with the assistance of Claude (Anthropic; model Claude Opus 5.5, model id claude-opus-5-5) in November 2026. Under the author's direction, Claude searched for and retrieved the data, wrote and ran the analysis plan, code, simulations and figures, checked citations against Crossref and publishers' pages, and drafted the manuscript. The author reviewed the code, results and text and takes full responsibility for the paper. AI is not an author.

## 3. Results

### 3.1 The record: more names, flatter tops

The files hold 2,149,123 name–sex–year records for 1880–2024, covering 371,964,739 births. The number of distinct names rose from 942 (girls) and 1,058 (boys) in 1880 to peaks of 20,581 girls' names in 2007 and 14,620 boys' names in 2008, then eased to 17,661 and 14,243 in 2024. Concentration fell in step: the 100 most common names accounted for 76.5% of girls and 80.1% of boys in 1880, 67.9% and 77.6% in 1950, and 31.0% and 38.0% in 2024.

### 3.2 Signature 1: heavy-tailed, but not demonstrably a power law

Name-count distributions are heavy-tailed in every year, spanning about four orders of magnitude, and on log–log axes they are close to straight lines over most of their range (Figure 1). Fitted exponents stayed in a narrow band: α ranged from 1.593 to 1.764 for girls and from 1.548 to 1.721 for boys, with minima around 1957–1962 and maxima in 1994 (girls) and 2004 (boys). These values are similar to Hahn and Bentley's (2003) decade estimates (about 1.6–1.9).

The formal comparison does not support a power law in particular. In all 290 year-by-sex fits the log-likelihood ratio favoured the lognormal in sign ($R < 0$); in 45 fits (35 for girls, 10 for boys) the lognormal was significantly better, and in the other 245 the two could not be distinguished. A power law was never significantly preferred. With x_min fixed at 5 the lognormal was significantly better in 118 fits. Lognormal preference was concentrated in recent girls' names (19 of the 25 years from 2000 onward). The x_min selected by the KS criterion had a median of 7, so the fitted tail typically covered about three-quarters (median 75.5%) of listed names.

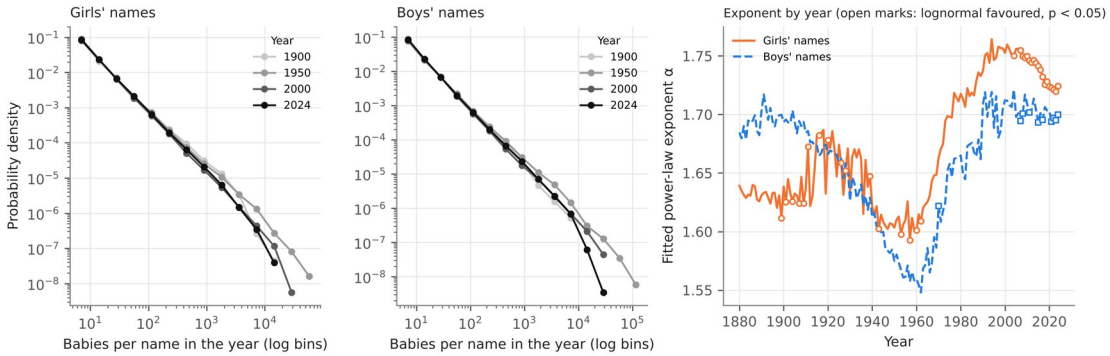



*Figure 1. Name-frequency distributions. Left and centre: log-binned probability density of the number of babies per name for 1900, 1950, 2000 and 2024 (SSA files, names with at least five babies). Right: fitted discrete power-law exponent α by year and sex; open markers show years in which a lognormal fitted significantly better ($p < 0.05$).*

## 3.3 Signature 2: a turnover exponent that hovers near 0.86

Turnover rose over the period. The mean number of names entering the top 100 each year was 3.24 for girls and 2.86 for boys in 1891–1960, and 5.47 and 3.66 in 1961–2024. On the primary grid (top 10 to top 100), the decade exponent b fluctuated with no obvious long-run trend for girls and drifted downward for boys in the 2000s (Table 1). The 95% bootstrap interval included 0.86 in 10 of 14 decades for each sex. For girls, the four exceptions were all below 0.86 (1901–1910, 1921–1930, 1951–1960, 1961–1970). For boys, three decades were above 0.86 (1881–1890, 1921–1930, 1961–1970) and one below (2011–2020).

The interval included b = 1 in 11 of 14 decades for boys but only 3 of 14 for girls. Annual exponents were noisy (mean 0.782, SD 0.338 for girls; 0.941, SD 0.415 for boys).

Table 1. Turnover exponent b by decade and list-size grid (Poisson GLM; 95% cluster-bootstrap interval for the primary grid).

| Window | Girls, y 10–100 (95% CI) | Girls, y 1–30 | Girls, y 50–1000 | Boys, y 10–100 (95% CI) | Boys, y 1–30 | Boys, y 50–1000 |
|---|---|---|---|---|---|---|
| 1881–1890 | 0.72 (0.52–0.97) | 0.75 | 1.86 | 1.11 (0.93–1.29) | 1.51 | 1.72 |
| 1891–1900 | 0.64 (0.41–0.91) | 0.65 | 1.78 | 0.86 (0.66–1.04) | 2.00 | 1.75 |
| 1901–1910 | 0.42 (0.25–0.61) | 1.09 | 1.75 | 0.96 (0.84–1.07) | 1.24 | 1.71 |
| 1911–1920 | 0.86 (0.67–1.09) | 1.20 | 1.55 | 0.97 (0.78–1.19) | 0.95 | 1.56 |
| 1921–1930 | 0.62 (0.44–0.83) | 1.01 | 1.40 | 1.08 (0.91–1.30) | 1.02 | 1.60 |
| 1931–1940 | 0.79 (0.62–0.96) | 0.78 | 1.27 | 0.85 (0.55–1.18) | 1.30 | 1.59 |
| 1941–1950 | 0.96 (0.79–1.13) | 0.72 | 1.24 | 0.76 (0.57–0.96) | 1.35 | 1.46 |
| 1951–1960 | 0.71 (0.57–0.85) | 0.67 | 1.14 | 0.92 (0.60–1.21) | 0.60 | 1.42 |
| 1961–1970 | 0.60 (0.41–0.83) | 0.57 | 1.11 | 1.22 (0.91–1.55) | 0.53 | 1.34 |
| 1971–1980 | 0.71 (0.56–0.91) | 0.61 | 1.08 | 0.83 (0.62–1.08) | 0.70 | 1.23 |
| 1981–1990 | 0.97 (0.82–1.11) | 0.70 | 1.04 | 0.79 (0.53–1.14) | 0.68 | 1.26 |
| 1991–2000 | 0.82 (0.66–0.98) | 0.66 | 1.06 | 0.94 (0.72–1.18) | 0.46 | 1.18 |
| 2001–2010 | 0.84 (0.70–0.98) | 0.53 | 0.95 | 0.70 (0.56–0.89) | 0.57 | 1.14 |
| 2011–2020 | 0.85 (0.71–0.99) | 1.11 | 0.96 | 0.58 (0.47–0.69) | 0.43 | 1.02 |
| 2021–2024 | 0.81 (0.56–1.12) | 1.22 | 0.93 | 0.62 (0.42–0.81) | 0.94 | 1.04 |

Taken at face value, the primary grid looks like support for drift. The sensitivity grids show why that reading is fragile (Figures 2 and 3). With the small grid (top 1 to top 30), boys' names had convex profiles in the early decades (b = 1.508, 2.000 and 1.239 for 1881–1910) and concave profiles from the 1950s to 2010s (0.433 to 0.699), which reproduces the qualitative pattern Acerbi and Bentley (2014) reported; girls' names on the small grid swung from 0.531 (2001–2010) to 1.106 (2011–2020) and 1.216 (2021–2024). With the large grid (top 50 to top 1,000), b was well above 1 early in the record (1.861 for

girls and 1.724 for boys in the 1880s) and declined smoothly to about 0.93–0.96 for girls and 1.02–1.04 for boys in the last two windows. The log–log turnover profile is therefore curved, not straight, so “the” exponent depends on which segment of it is measured. Part of the early large-grid steepness is mechanical: in the 1880s each sex had only about 1,000–1,250 listed names, so top-1,000 lists reach down into names tied at five and are reshuffled by small count changes and the alphabetical tie rule.

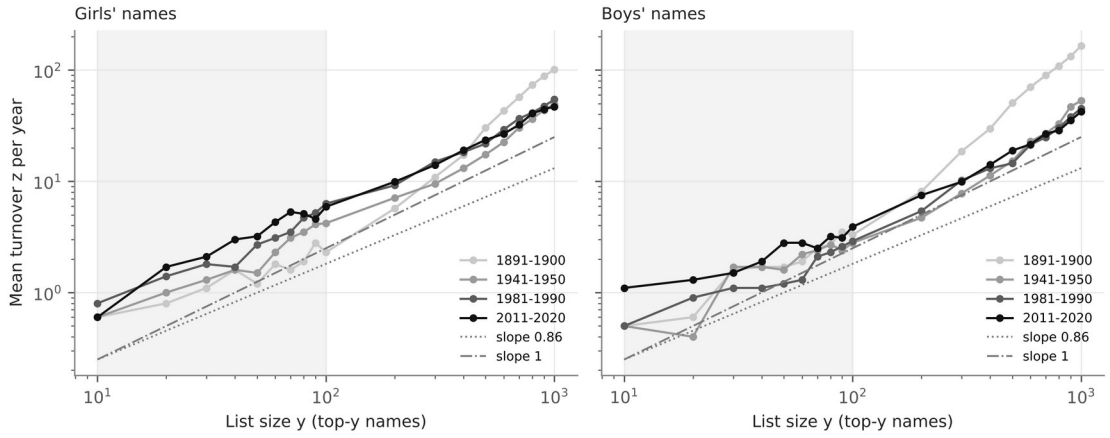



*Figure 2. Mean annual turnover z against list size y (log–log) for four decades. The shaded band marks the primary grid (y = 10–100). Reference lines have slopes 0.86 and 1 (intercepts arbitrary).*

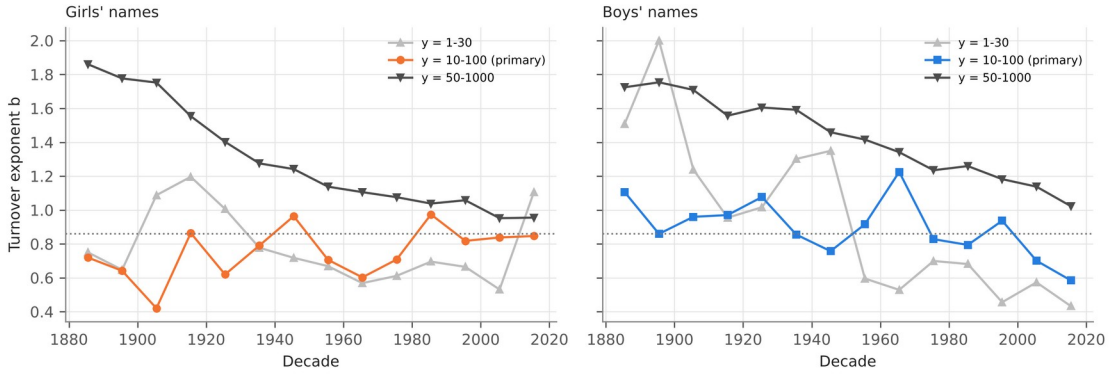



*Figure 3. Decade turnover exponent b on three list-size grids. Dotted line: 0.86.*

### 3.4 Random copying from the actual cohort is far too slow

The one-step forecast (§2.6) asks how many names random copying from last year's actual cohort would push into this year's top lists. The answer, in most years, is much fewer than were observed (Figure 4, left). Over the 134 year-pairs from 1891 to 2024, observed top-100 turnover was 4.77 times the forecast mean for girls and 3.50 times for boys, and it exceeded the forecast's 97.5th percentile in 82.1% of years for girls and 61.2% for boys. The gap is smaller for shorter lists (top 50: 68.7% and 51.5% of years above the forecast range; top 10: 35.8% and 20.9%) but in the same direction. In no year was observed top-100 turnover below the forecast's 2.5th percentile. Names enter and leave the top 100 several times faster than binomial sampling noise around last year's frequencies can produce, given the number of babies actually named.

### 3.5 How small would the copying pool have to be?

Random copying can match the observed level of top-100 turnover if the effective pool being copied is much smaller than the birth cohort (§2.7; Figure 4, right). The required pool was 21.4–33.9% of recorded births for girls and about 45% for boys in the 1890s and 1900s (when recorded births were small and incomplete), 4.1–8.2% for girls and 6.8–12.2% for boys in 1911–1940, and 1.42–2.73% of births for girls (24,640–47,654 "effective parents" per year) and 1.51–4.78% for boys (30,047–89,476) in each decade from 1941 to 2020. At those pool sizes, drift also had to reproduce the shape of the turnover profile. It did so in 5 of 13 decades for girls and 7 of 13 for boys, in the sense that the forecast exponent b fell inside the observed decade's 95% interval. In every other decade the forecast exponent was above the observed one (girls: 1901–1990 except 1911–1920; boys: 1891–1900, 1931–1960 and 1971–1990): once drift is made strong enough to match how many names enter the top 100, it makes the very top of the list too stable relative to its lower part.

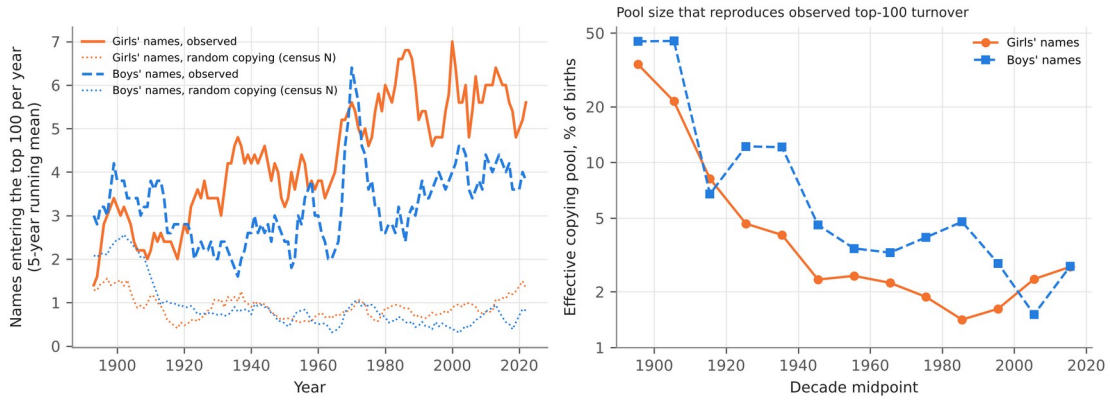



*Figure 4. Left: names entering the top 100 each year, observed (solid or dashed) and forecast by random copying from the previous year's actual cohort at census size (dotted); 5-year running means. Right: the effective copying pool, as a percentage of recorded births, at which random copying reproduces each decade's observed top-100 turnover.*

## 3.6 Neutral simulations at census scale

The planned census-scale simulations (540 runs; burn-in 985 to 19,326 generations) test the neutral model on its own terms rather than against a fixed benchmark (Figure 5). Calibration A's innovation rates ($8.13 \times 10^{-4}$ to $4.06 \times 10^{-3}$ for girls, $2.56 \times 10^{-4}$ to $3.96 \times 10^{-3}$ for boys) produced 0.82 to 4.19 times as many listed names as observed; calibration B's lower rates ($3.48 \times 10^{-4}$ to $1.43 \times 10^{-3}$ and $2.07 \times 10^{-4}$ to $1.41 \times 10^{-3}$) matched the observed number of names to within 9.6% in every decade.

On the turnover exponent, the neutral model at census scale is hard to reject. The observed decade b fell inside the simulated 2.5–97.5% range in 7 of 13 decades for girls and 10 of 13 for boys under calibration A, and in 11 of 14 for each sex under calibration B. Where it fell outside, the observed exponent was almost always below the simulated range (concave relative to neutral); the exceptions were boys in 1881–1890 (calibration B) and 1961–1970 (both calibrations), where it was above. The level of top-100 turnover matched less often: inside the simulated range in 5 (girls) and 4 (boys) decades under calibration

A, where observed turnover was mostly lower than simulated, and in 3 and 7 decades under calibration B, where it was mostly higher (above the range in 10 girls' and 6 boys' decades).

The simulations fail decisively on the structure of the distribution. In every decade, for both sexes and both calibrations, the observed share of babies given a top-100 name (33.6% to 79.1% across decades) was above the largest value in any replicate; simulated shares ranged from 2.9% to 51.0%. Correspondingly, the power-law exponents fitted to simulated end states (medians 3.00 to 5.51) were far steeper than the observed decade means (1.57 to 1.75). A neutral population of one to two million births per year with enough innovation to sustain the observed number of names is much less concentrated than American naming has ever been: the heads of real name distributions are too heavy for census-scale drift.

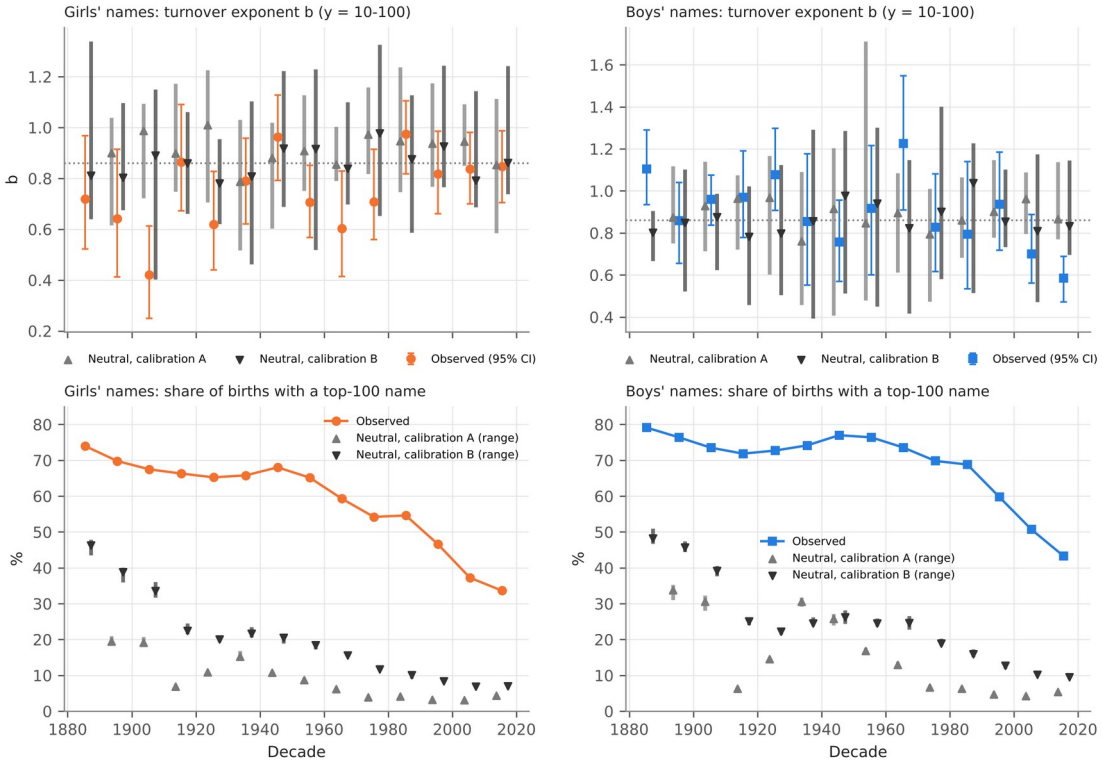



*Figure 5. Observed decade statistics against census-scale equilibrium neutral simulations (10 replicates per decade, sex and calibration). Top: turnover exponent b (primary grid), observed with 95% bootstrap*

*interval and simulated 2.5–97.5% range with median. Bottom: share of births with a top-100 name, observed decade mean and simulated range with median.*

### 3.7 Did the drift fit break? Pre-specified tests

For the primary outcome, the annual turnover exponent b_t (Figure 6), no candidate year showed a break for girls (all Holm-adjusted p = 1). For boys, the 2000 break was significant after correction (Wald = 13.36, p = 0.0013, Holm-adjusted p = 0.0075): the level shift at 2000 was small and uncertain (d_0 = −0.155, SE 0.123), but the trend after 2000 was steeper downward than before (slope change d_1 = −0.0123 per year, SE 0.0047). In decade terms, boys' b fell from 0.937 (1991–2000) to 0.701 (2001–2010) and 0.585 (2011–2020); 2011–2020 was the only full boys' decade whose interval lay entirely below 0.86, and the four-year 2021–2024 window (0.62, interval 0.42–0.81) was also entirely below it. The 1965 break for boys was nominally significant (p = 0.0118) but not after correction (Holm-adjusted p = 0.0591); 1946 was not (p = 0.4987).

Secondary outcomes (Table 2). The annual top-100 turnover rate showed breaks for girls at 1965 (Holm-adjusted p = 0.0333; a level rise of 1.1 percentage points) and for boys at 1946 and 1965 (both Holm-adjusted p < 0.001), but not at 2000 for either sex. The power-law exponent α_t "broke" at all six candidates (all Holm-adjusted p < 0.0001). I do not read much into that last result: α_t is a smooth, strongly autocorrelated series (Figure 1, right) with a long U-shape, and a broken-trend model at almost any year improves on a single line. The exploratory scans placed the largest Wald statistic for b_t at 1977 for girls (7.39) and 1956 for boys (25.49); for top-100 turnover at 1922 (girls) and 1897 (boys); and for α_t at 2005 (girls) and 1968 (boys). Because these maxima are chosen after looking, they carry no p-value; they mainly show that the boys' b series has its largest single shift in the 1950s, not at a pre-specified date.

Table 2. Pre-specified break tests (Wald test of d_0 = d_1 = 0, Newey–West HAC covariance, Holm correction across 3 years × 2 sexes within each outcome).

| Outcome | Sex | Year | d0 (SE) | d1 per year (SE) | Wald | p | Holm p |
|---|---|---|---|---|---|---|---|
| Turnover exponent b_t | Girls | 1946 | -0.115 (0.098) | -0.0017 (0.0031) | 1.39 | 0.4994 | 1.0000 |
| Turnover exponent b_t | Girls | 1965 | -0.090 (0.107) | 0.0008 (0.0027) | 0.74 | 0.6900 | 1.0000 |
| Turnover exponent b_t | Girls | 2000 | 0.019 (0.079) | -0.0036 (0.0036) | 1.01 | 0.6050 | 1.0000 |
| Turnover exponent b_t | Boys | 1946 | 0.113 (0.165) | -0.0037 (0.0033) | 1.39 | 0.4987 | 1.0000 |
| Turnover exponent b_t | Boys | 1965 | 0.203 (0.160) | -0.0093 (0.0032) | 8.88 | 0.0118 | 0.0591 |
| Turnover exponent b_t | Boys | 2000 | -0.155 (0.123) | -0.0123 (0.0047) | 13.36 | 0.0013 | 0.0075 |
| Top-100 turnover rate | Girls | 1946 | -0.001 (0.004) | -0.0000 (0.0001) | 0.16 | 0.9253 | 0.9253 |
| Top-100 turnover rate | Girls | 1965 | 0.011 (0.004) | -0.0001 (0.0001) | 9.58 | 0.0083 | 0.0333 |
| Top-100 turnover rate | Girls | 2000 | 0.002 (0.005) | -0.0005 (0.0003) | 2.86 | 0.2392 | 0.4784 |
| Top-100 turnover rate | Boys | 1946 | 0.009 (0.005) | 0.0005 (0.0001) | 27.95 | < 0.0001 | < 0.0001 |
| Top-100 turnover rate | Boys | 1965 | 0.016 (0.007) | 0.0003 (0.0002) | 17.89 | 0.0001 | 0.0007 |
| Top-100 turnover rate | Boys | 2000 | 0.012 (0.005) | -0.0001 (0.0002) | 5.42 | 0.0665 | 0.1994 |
| Power-law exponent α_t | Girls | 1946 | -0.040 (0.017) | 0.0020 (0.0004) | 24.88 | < 0.0001 | < 0.0001 |
| Power-law exponent α_t | Girls | 1965 | 0.062 (0.018) | 0.0016 (0.0005) | 58.33 | < 0.0001 | < 0.0001 |
| Power-law exponent α_t | Girls | 2000 | 0.059 (0.018) | -0.0024 (0.0002) | 204.99 | < 0.0001 | < 0.0001 |
| Power-law exponent α_t | Boys | 1946 | -0.059 (0.015) | 0.0037 (0.0003) | 121.11 | < 0.0001 | < 0.0001 |
| Power-law exponent α_t | Boys | 1965 | 0.058 (0.017) | 0.0037 (0.0005) | 159.79 | < 0.0001 | < 0.0001 |
| Power-law exponent α_t | Boys | 2000 | 0.089 (0.024) | -0.0001 (0.0003) | 71.08 | < 0.0001 | < 0.0001 |

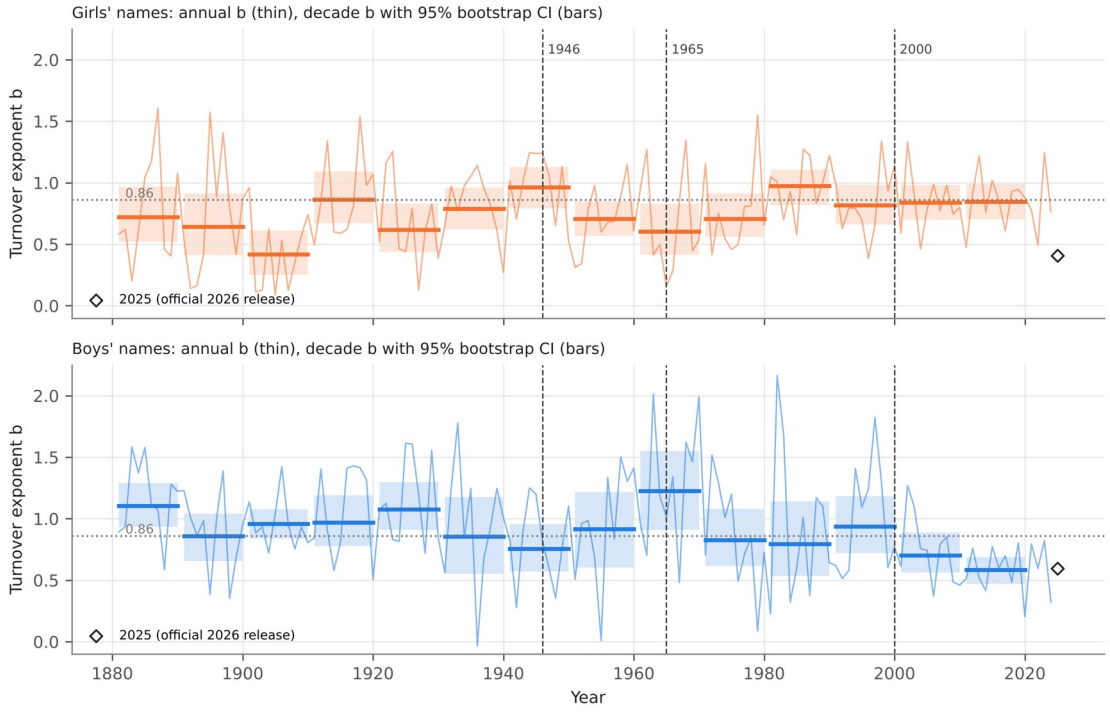



*Figure 6. Annual turnover exponent b_t (thin lines) and decade estimates with 95% cluster-bootstrap intervals (bars and bands) on the primary grid. Dotted line: 0.86. Dashed vertical lines: pre-specified candidate break years. Diamond: 2025, computed from the current SSA release.*

### 3.8 The current release and 2025

On the current SSA release, the primary-grid decade exponents differed from the mirror-based values by at most 0.02. The boys' 2000 break remained significant after correction (Holm-adjusted p = 0.0108) and the boys' 1965 break remained non-significant after correction (0.0656); no girls' break was significant. In 2025, 5 new girls' names and 6 new boys' names entered the top 100; the 2021–2025 window exponents were 0.709 (girls) and 0.625 (boys). The newest year does not change any conclusion.

## 4. Discussion

### 4.1 What survives of the drift account

Two of the classic neutral signatures are present in the data, in a loose sense. Name frequencies are heavy-tailed with a stable exponent near 1.6–1.8 for 145 years, and the turnover exponent over the top 10–100 sits near 0.86 in most decades. If the question is whether a random-copying null can be rejected using these two summary signatures alone, the answer is mostly no, and in that limited sense Hahn and Bentley's (2003) conclusion survives a century of extra data.

But both signatures are weak tests. Heavy tails arise from many processes, and the data do not single out a power law: the lognormal is never worse and often better, in line with the general warning of Clauset et al. (2009) that power laws are easy to see and hard to establish. The turnover exponent, in turn, depends strongly on the list sizes over which it is measured, because the empirical turnover profile is curved. Over the top 1–30, boys' names look conformist early and anti-conformist late (as in Acerbi & Bentley, 2014); over the top 10–100 they look neutral in most decades; over the top 50–1,000 everything looks convex early and nearly linear late. A diagnostic whose verdict can be moved by the choice of y range should be reported over several ranges, as here, rather than as a single exponent.

## 4.2 Where drift fails

The decisive results are the direct ones. Random copying from the cohort that was actually born the year before, at the actual number of births, cannot produce the year-to-year churn of the top 100 (§3.4). Turnover was three to five times higher than drift allows, and above drift's 95% range in most years. This agrees with the main conclusion that Newberry and Plotkin (2022) reached with a different method: common names decline and rare names rise faster than sampling noise permits. They also found that some names (Biblical names) carry a genuine selective advantage, so the departure from drift is not only frequency-dependent. In population-genetic terms, drift could only account for the churn if the effective population of name-givers being copied were about 1.4–4.8% of the birth cohort in every decade since 1941 (and up to 12% for boys in 1921–1940; §3.5). There is no obvious reason why a small fraction of parents should be the sole models for everyone else. A more natural reading is that name choice is not unbiased copying: the effective-size number stands in for systematic forces such as fashion cycles and the avoidance of names that have become too common. Such mechanisms have been documented directly: names that rose quickly in popularity were abandoned more quickly (Berger & Le Mens, 2009), the share of American babies given common names has fallen since 1950, after falling from 1880 to 1919 and rising from 1920 to 1949 (Twenge et al., 2010), and in recent decades parents' choices have tracked a name's momentum as well as its frequency, so that rising names keep rising, whereas in the early 20th century rising names tended later to decline (Gureckis & Goldstone, 2009). Even granting a small effective pool, drift tends to predict a steeper turnover profile than observed (§3.5): once it matches how many names enter the top 100, it makes the very top too stable. That is the concave, anti-conformist-like direction that Acerbi and Bentley (2014) described for recent names, here found across most of the 20th century for girls.

The census-scale simulations (§3.6) sharpen the point. A drift model with realistic population sizes and innovation can produce a turnover exponent near the observed one, but only while generating name distributions with far lighter heads than any observed decade: at most about half of babies with a top-100

name in the simulations, against one-third to four-fifths in the data. Hahn and Bentley (2003) obtained their best distributional fits with $\theta = 4$. In the standard infinite-alleles model $\theta = 4N\mu$ (Kimura & Crow, 1964); the text of their paper available to me prints the definition as "4N", apparently with the $\mu$ lost, so I read their best fit as $\theta = 4N\mu = 4$, about one innovation per generation in the copying population. At census scale the calibrations used here imply 317 to 7,487 innovations per year from 1911 on ($4N\mu$ of about 1,300 to 30,000), and the same model then gives the wrong distribution. Drift fits the cross-section only if the effective number of independent name-givers is small, which is the same conclusion that the one-step forecast reaches from the dynamics.

### 4.3 Did the fit break?

The pre-specified tests give a qualified answer. Under the plan's own verdict rule (§2.10), the neutral fit held for girls' names and broke for boys' names. For girls' names there was no detectable break in the turnover exponent at the baby boom, the 1965 immigration reform or the arrival of household internet access; the girls' turnover rate rose after 1965 but its profile did not change shape. For boys' names the exponent began a decline after 2000 that left 2011–2020 as the only full decade clearly below the neutral benchmark, with the short 2021–2024 window also below it. This is consistent with a growing anti-conformist component in boys' naming in the internet era, but the test identifies timing, not cause: the same years also saw continued growth in diversity, and the exploratory scan suggests the largest single shift in boys' b came in the 1950s, which none of the pre-specified candidates captures. The internet-era interpretation is therefore offered as a hypothesis, not as a finding.

### 4.4 For memetics

In the memetic sub-series, earlier papers asked whether memes copy faithfully (Kalafatis, 2026a, 2026b). Names are copied with perfect fidelity in the sense that matters here (a name is either reused or not), so they isolate a different question: given faithful copying, is the choice of what to copy blind to frequency? For American first names it clearly is not. The pool behaves less like a randomly drifting gene pool than

like a market with fashion: a heavy-tailed distribution that looks neutral in cross-section and in coarse turnover summaries, sitting on top of year-to-year dynamics that are several times too fast for drift. Neutral models remain the right null hypothesis; the lesson is that matching a static distribution or a single turnover exponent is not enough to accept it.

### 4.5 Limitations

(1) The primary data come from a GitHub mirror of an earlier SSA release, not from ssa.gov directly; the current official release was used only through in-browser aggregates, produced in a browser driven by the AI assistant and transcribed by hand from its tool output (§2.2). The two releases differ by at most 0.54% in any year's births and by one name in 28 of 2,880 turnover values, and every turnover conclusion was rechecked on the current release, but the frequency-distribution and simulation analyses rest on the earlier tabulation. (2) The floor of five truncates the rare-name tail and affects the frequency-distribution fits and the innovation calibrations; it barely affects top-100 turnover. (3) Recorded births before about 1937 are survivors who later applied for a card and are not a complete birth cohort; early-decade results, including the large effective-pool percentages of the 1890s and 1900s, should be read with that in mind. (4) The neutral model treats each sex as a closed population with annual, non-overlapping generations; in reality parents copy names they have encountered over decades, from both sexes and from other countries. These choices follow the literature being tested but make the census-scale model a stylised benchmark. (5) Calibration A's innovation rate is a crude proxy, and the effective-pool estimates come from a one-step forecast that conditions on each observed year; neither is a full likelihood-based inference of N_e or of frequency-dependent selection of the kind Newberry and Plotkin (2022) performed. (6) Name strings were not merged across spellings, which inflates both diversity and turnover relative to a phonetic definition of a name. (7) Candidate break years were chosen from history before looking at results, but three years and three outcomes are a small set;

other plausible breaks (for example, the 1950s shift suggested by the scan) were not tested inferentially.

## 5. Conclusion

Re-tested on 145 years of SSA records, American baby names pass the two classic surface tests of the neutral model in a loose sense: heavy-tailed frequency distributions and a turnover exponent near 0.86 over the top 10–100 in most decades. They fail the more direct test. Names move in and out of the top 100 three to five times faster than random copying from the previous year's actual cohort allows; drift can match that churn only if the copying pool is about 1.4–4.8% of births (every decade since 1941; larger before), and even then it predicts the wrong shape of the turnover profile in about half of the decades. Simulated at census scale, the neutral model reproduces the turnover exponent but never the heavy concentration of babies in the most popular names. The one pre-specified break that survived correction is a post-2000 decline in the turnover exponent for boys' names. Baby names are a meme pool with drift in it, not a pure-drift meme pool.

## Acknowledgements

The author thanks no external funders or institutions for this independent, self-directed work. This study used Claude (Anthropic; model Claude Opus 5.5, model id claude-opus-5-5) as described in §2.11; the author reviewed all code, data and text and takes full responsibility for the final manuscript.

## Author contributions

Peter Kalafatis conceived the study, directed the analysis, reviewed all code and results, and wrote the final manuscript (with AI assistance as described in §2.11).

## Financial support

This research received no specific grant from any funding agency, commercial, or not-for-profit sector.

## Conflicts of interest

The author declares no conflicts of interest. This study used a product (Claude) made by Anthropic as a research assistant (§2.11); no compensation, data-sharing arrangement, or other relationship with Anthropic beyond ordinary product use is relevant to the paper's conclusions.

## Research transparency and reproducibility

Data: SSA national baby-name files (public domain), official source https://www.ssa.gov/oact/babynames/limits.html (names.zip, release of 8 May 2026, SHA-256 cd78e975ed7bb358e018dd62fbe14ced89295e9581c49172ca4eedcb011b3724). Primary analysis files: GitHub repository aramshiva/names, commit ecc921d17155412d4ed593d5d2b1c4290184c821, path db/names/ (https://github.com/aramshiva/names); unmodified copies are kept in pkg/data/raw/aramshiva_names/ with their hashes. Code: pkg/code/00_hash_plan.py (plan and file hashes), 01_load.py, 02_turnover.py, 03_powerlaw.py, 04_simulate.py (neutral simulations), 05_breaks.py, 06_effective_size.py, 07_browser_replication.js (official-release check run in a browser) with 07_make_browser_js.py, 08_official_release.py, 09_figures.py and 10_numbers.py; the pre-specified plan is code/00_analysis_plan.txt. All intermediate and final results are in pkg/results/; results/manuscript_numbers.json is the single source for every number in this manuscript. Code and results will be available at the links given with the published version of this preprint and in the GitHub archive x9m8/research-papers.

## References

Acerbi, A., & Bentley, R. A. (2014). Biases in cultural transmission shape the turnover of popular traits. *Evolution and Human Behavior*, *35*(3), 228–236. https://doi.org/10.1016/j.evolhumbehav.2014.02.003

Alstott, J., Bullmore, E., & Plenz, D. (2014). powerlaw: A Python package for analysis of heavy-tailed distributions. *PLoS ONE*, *9*(1), e85777. https://doi.org/10.1371/journal.pone.0085777

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, *61*(4), 821–856. https://doi.org/10.2307/2951764

Bentley, R. A., Hahn, M. W., & Shennan, S. J. (2004). Random drift and culture change. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1547), 1443–1450. https://doi.org/10.1098/rspb.2004.2746

Bentley, R. A., Lipo, C. P., Herzog, H. A., & Hahn, M. W. (2007). Regular rates of popular culture change reflect random copying. *Evolution and Human Behavior*, *28*(3), 151–158. https://doi.org/10.1016/j.evolhumbehav.2006.10.002

Berger, J., & Le Mens, G. (2009). How adoption speed affects the abandonment of cultural tastes. *Proceedings of the National Academy of Sciences*, *106*(20), 8146–8150. https://doi.org/10.1073/pnas.0812647106

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, *28*(3), 591–605. https://doi.org/10.2307/1910133

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703. https://doi.org/10.1137/070710111

Evans, T. S., & Giometto, A. (2011). *Turnover rate of popularity charts in neutral models* (arXiv:1105.4044) [Preprint]. arXiv. https://arxiv.org/abs/1105.4044

Gureckis, T. M., & Goldstone, R. L. (2009). How you named your child: Understanding the relationship between individual decision making and collective outcomes. *Topics in Cognitive Science*, *1*(4), 651–674. https://doi.org/10.1111/j.1756-8765.2009.01046.x

Hahn, M. W., & Bentley, R. A. (2003). Drift as a mechanism for cultural change: An example from baby names. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(Suppl. 1), S120–S123. https://doi.org/10.1098/rsbl.2003.0045

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Kalafatis, P. (2026a). *Too leaky to replicate: An error threshold for memes in human memory* [Preprint]. x9m8.com. https://doi.org/10.5281/zenodo.23751533

Kalafatis, P. (2026b). *Rephrasing without replicating: Transmission fidelity in language-model telephone chains* [Preprint]. x9m8.com. https://doi.org/10.5281/zenodo.23751529

Kalafatis, P. (2026c). *How memes fade: Comparing decay curves of Wikipedia attention across internet phenomena* [Preprint]. x9m8.com. https://x9m8.com/meme-attention-decay-wikipedia-pageviews/

Kimura, M., & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, *49*(4), 725–738. https://doi.org/10.1093/genetics/49.4.725

National Telecommunications and Information Administration. (2002). *A nation online: How Americans are expanding their use of the Internet*. U.S. Department of Commerce. https://www.ntia.gov/report/2002/nation-online-internet-use-america

Neiman, F. D. (1995). Stylistic variation in evolutionary perspective: Inferences from decorative diversity and interassemblage distance in Illinois Woodland ceramic assemblages. *American Antiquity*, *60*(1), 7–36. https://doi.org/10.2307/282074

Newberry, M. G., & Plotkin, J. B. (2022). Measuring frequency-dependent selection in culture. *Nature Human Behaviour*, *6*(8), 1048–1055. https://doi.org/10.1038/s41562-022-01342-6

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708. https://doi.org/10.2307/1913610

Social Security Administration. (n.d.-a). *Popular baby names* [National data file names.zip, release of May 8, 2026]. Retrieved November 25, 2026, from https://www.ssa.gov/oact/babynames/limits.html

Social Security Administration. (n.d.-b). *Background information for popular names*. Retrieved November 25, 2026, from https://www.ssa.gov/oact/babynames/background.html

Twenge, J. M., Abebe, E. M., & Campbell, W. K. (2010). Fitting in or standing out: Trends in American parents' choices for children's names, 1880–2007. *Social Psychological and Personality Science*, *1*(1), 19–25. https://doi.org/10.1177/1948550609349515