Selection on Fertility, and the Environmental Decline That Would Cancel It

A compositional mechanism in a 237-country cohort-component projection to 2150

Dylan Slagh* Claude Opus 5† ChatGPT 5.6 Sol‡

Version 1.2.2 • 22 August 2026

Abstract

Population at a 150-year horizon is governed principally by fertility after the demographic transition. Standard projections represent that process with a national average that tends toward an assumed long-run level. Such an average contains no information about the parental composition of successive cohorts.

That omission matters because women differ substantially in completed family size: across low-fertility countries with completed-cohort data, its standard deviation is more than half its mean. Completed family size is also transmitted across generations, with a parent–child correlation of about 0.15. The next generation is therefore drawn disproportionately from larger families, making fertility-increasing dispositions more prevalent. This selection raises fertility but cannot be represented by a model that tracks only the national average.

We add that mechanism to a cohort-component projection covering 237 countries, built on the United Nations' own age-specific inputs and reproducing its published projection to 0.001% of world population at 2100. Switching the mechanism off returns the original engine to machine precision, so whatever the mechanism changes is the mechanism and not different arithmetic. Its two parameters come from sources that have nothing to do with the projection they are used to alter.

Under the benchmark specification, selection is measurable and consequential but insufficient to restore population growth. It raises world population in 2150 from 8.54 to 10.36 billion — a shift of about 21% — by changing the level at which the population curve settles. The high-fertility religious groups prominent in public discussion contribute 2.5% as much as ordinary, unlabelled variation within the general population.

The countervailing uncertainty is whether conditions that make children costly continue to worsen. Because selection and the fertility environment both enter multiplicatively, the model can identify a break-even threshold without assigning a preferred value to future environmental change. A further uniform decline of 1.52% per decade after 2050 cancels selection's effect on the fertility rate; cancelling its effect on the 2150 *population* takes 2.45%, because the extra people selection has already produced are still alive. Across the measured uncertainty in the two parameters the first threshold ranges from 0.33% to 4.49%. Present evidence does not establish which side of this threshold the world occupies; reporting it expresses the unresolved question as an observable rate of change.

*Independent researcher, Atlanta, Georgia. Correspondence: dylanslagh@gmail.com.

†Anthropic.

‡OpenAI.

1 Introduction

At a 150-year horizon, almost nobody alive at the start is alive at the end. A projection to 2150 from a 2024 base year is therefore barely a statement about today's population at all. It is a statement about the process that produces births, run four or five times over, and everything else washes out. Freeze today's fertility rates and the arithmetic gives about 53 billion people; use the United Nations' medium assumptions and it gives about 8.78 billion. Because both calculations use the same base data and projection engine, the difference arises entirely from the assumed fertility path.

Conventional long-run projections make that assumption in a specific form. Total fertility is treated as a national time series that walks down through the demographic transition and then reverts probabilistically toward a long-run level [Alkema et al., 2011, Raftery et al., 2012, Ševčíková et al., 2016]. The United Nations [United Nations Population Division, 2024] and the Institute for Health Metrics and Evaluation [Vollset et al., 2020] disagree about the level but agree about the object being modelled: one number per country per year, evolving by its own history.

That object has no composition in it, and composition is not a detail here. Women differ a great deal in how many children they have. Across 19 low-fertility countries with completed cohort data the coefficient of variation of completed family size has a median of 0.570, so the standard deviation is more than half the mean. This variation persists across generations: adult children of larger families have somewhat larger families themselves, with reported parent–child correlations usually between 0.1 and 0.2 [Murphy, 1999, 2013].

Together, dispersion and intergenerational persistence imply a process that a national-rate model cannot express. The parents of the next generation are not a random sample of this one — they are sampled in proportion to how many children they have. Whatever inclined a woman toward four children rather than one is over-represented among the people raising the next cohort, and if it carries at all, the composition of the population moves toward it. This is selection in the sense used since Fisher [1930], and its direction on fertility is always upward, because fertility here is simultaneously the trait and the measure of reproductive success. A model that projects only the national mean does not estimate this force to be small. It has no way to see it.

1.1 What the existing arguments get right, and where they stop

That selection acts on fertility is not a new observation. Kolk et al. [2014] show that intergenerational fertility correlations can prevent low fertility from persisting; Burger and DeLong [2016] argue that treating the fertility decline as permanent is unjustifiable on evolutionary grounds; and Collins and Page [2019] apply the breeder's equation to national fertility and conclude that global population stabilisation this century is unlikely. There is also direct evidence that contemporary human populations are under measurable selection on fertility-linked traits [Byars et al., 2010, Beauchamp, 2016, Kong et al., 2017].

The sharpest objection to that work is also the most useful. Arenberg et al. [2022] point out that a model built from a transmission coefficient alone contains no absolute fertility level: it can describe which dispositions become more common without being able to say whether the population grows, and the empirical fact that higher-fertility subgroups are themselves drifting toward replacement has nowhere to enter. Their condition is clean — a high-fertility type shrinks

unless the expected number of same-type daughters each woman leaves exceeds one — and it simply cannot be evaluated in a model that carries no fertility schedules.

1.2 What this paper does

1. **Puts the mechanism inside a real projection.** Selection is added as an extra axis on a single-year, two-sex cohort-component engine for 237 countries carrying the United Nations' own age-specific fertility, survival and migration schedules (Section 4). Absolute levels are present at every step, so the Arenberg et al. [2022] condition is enforced by the arithmetic rather than assumed away. With every propensity set to one the extended engine reproduces the ordinary one to machine precision, which is what makes any difference in the results attributable to the mechanism rather than to a second implementation of the bookkeeping.
2. **Calibrates it from evidence about something other than the answer.** The two mainstream parameters are the dispersion of completed family size and the parent–child correlation in it, taken from independent sources (Section 3). Their combination reproduces an intergenerational covariance, which is a weaker and more defensible claim than a heritability, though Section 2.2 is explicit that the covariance is constructed from two datasets rather than measured in one.
3. **Reports a threshold instead of a preferred number.** How the fertility *environment* moves — how many children a person of fixed disposition has under prevailing conditions — cannot be measured in advance, and it is where a long-run projection can be made to say anything one likes. Because selection and the environment both multiply fertility, we can ask how much additional environmental decline would exactly offset measured selection, and report that (Section 6) rather than choosing a value.

1.3 What we find

Selection is real, measurable, and smaller than its advocates suggest: it shifts the level of the long-run curve without restoring growth, and it is overwhelmingly a property of ordinary variation rather than of the visible high-fertility minorities that dominate discussion of it. Cancelling it takes a different amount of environmental decline depending on whether the target is the fertility rate or the accumulated population. Section 5 gives the magnitudes and Section 6 the thresholds.

2 The mechanism

The model keeps two questions apart and answers them separately, because conflating them is what makes long-run fertility arguments irresolvable:

Who is becoming more common?
The composition of the population by fertility disposition. This is what selection moves.

How many children does a person of a given disposition have?
The fertility environment. This is what social and economic change moves.

Observed fertility is the product of the two. It is an output of the model, not an input to it, and that is the structural difference from a conventional projection, in which observed fertility is the modelled object itself.

2.1 Types, propensities and transmission

Each country's population is divided along an extra axis into K fertility *types*. A type is not an ethnic, religious or political label; it is an effective fertility propensity ϕ_k, a multiplier on the country's own base-year age-specific fertility schedule. A woman of propensity 1.4 has forty per cent more children than the national average did in the base year, at every age, and that multiplier does not change when conditions do. The propensities are normalised so that the base-year composition reproduces each country's published fertility exactly, which anchors the model to the present instead of fitting it to the future.

Writing ℓ for country, t for year, a for single year of age, k for type, $f_t[\ell, a]$ for the baseline age-specific fertility schedule and $E_t[\ell]$ for the environment multiplier, births are counted by the mother's type and then reassigned to the child's:

$$B_t[\ell, k] = \sum_a \bar{P}_t[\ell, k, \mathrm{F}, a]\ f_t[\ell, a]\ E_t[\ell]\ \phi[\ell, k], \tag{1}$$

$$\tilde{B}_t[\ell, j] = \sum_k B_t[\ell, k]\ M_t[\ell, k, j], \tag{2}$$

where $\bar{P}_t$ is mid-year exposure of women and M_t is the transition matrix from mother's type to child's type. Everything the mechanism claims is in those two lines. Survival, ageing, the open-ended age group and migration are untouched and carried identically for every type.

Selection is what happens when the rows of M differ and ϕ is not constant. Higher-propensity women appear more often in Equation (1) because they have more children; those children land in their mother's type more often than chance through Equation (2); so the composition drifts upward, the population-average propensity rises, and observed fertility departs from the environment path without anyone imposing a floor on it. The quantity we report throughout is that departure,

$$S_t[\ell] \ = \ \frac{\text{observed fertility}}{\text{environment alone}} \ = \ \bar{\phi}_t[\ell], \tag{3}$$

the propensity averaged over women of childbearing age. It equals one when selection is switched off, and its distance above one is the size of the mechanism's claim.

2.2 The response, and what it does and does not rest on

The mainstream transition rule is deliberately the simplest one that can carry the evidence. A child takes its mother's type with probability ρ and is otherwise drawn from the surrounding population of childbearing adults. Under that rule the one-generation response has a closed form. Mothers are sampled in proportion to their own fertility, so the birth-weighted distribution of types is $q_k \propto w_k \phi_k$ where w_k is the share of childbearing women of type k. The mean propensity of the newborn cohort is then

$$\rho \sum_k q_k \phi_k + (1-\rho) \sum_k w_k \phi_k \ = \ \bar{\phi} + \rho \, \frac{\sigma^2}{\bar{\phi}}, \tag{4}$$

so the relative gain per generation is $\rho\,\mathrm{CV}^2$: the transmission coefficient times the squared coefficient of variation of family size. This is the secondary theorem of natural selection in its covariance form [Robertson, 1966, Price, 1970] — the response equals the covariance of the trait with fitness

divided by mean fitness — specialised to the case where the trait and the fitness are the same quantity, so that the covariance is a variance.

Interpreting Equation (4) requires three qualifications.

The transmission probability and the reported correlation coincide only because the variances are equal. Under this rule a child is either identical to its mother in type or an independent draw from the same distribution, so the implied parent–offspring correlation in propensity is exactly ρ, and $\mathrm{Cov} = \rho\sigma^2$ rather than the general $\rho\sigma_X\sigma_Y$. That is a property of the rule, and it is why ρ can be read off a reported correlation. It also means the model assumes the dispersion of family size does not change between generations, which is a real restriction and not a harmless normalisation.

The covariance is constructed, not observed. The dispersion comes from one set of countries and cohorts and the correlation from another (Section 3). Multiplying them produces the covariance a single intergenerational sample would have if both applied to it; it does not reproduce a covariance measured in one. The right estimate would come from paired parent–child completed-fertility data with both moments taken from the same sample; no such dataset has the country coverage this model requires. This construction is the weakest link in the calibration and is reported as an assumption rather than as a direct measurement.

Matching one generation does not validate four. Many transition rules reproduce the same first-generation response and compound differently over 2150. What is checked is the first generation, which the implementation reproduces to nine decimal places; what is assumed is that the rule keeps holding. Section 8.3 states which alternative rules would change the answer and by how much.

At the central calibration ($\rho = 0.15$, CV $= 0.57$) the one-generation response is 4.87%. Compounded over the roughly four generations between 2024 and 2150, that closed form implies a selection multiplier near 1.22, against the 1.174 the full projection produces. The closed form runs slightly hot because it ignores the age structure that delays each generation's effect and the migration that mixes compositions between countries. That the two agree within a few per cent is a check that the simulation is doing what the theory says.

2.3 Named groups are transitions, not labels

Populations with both very high fertility and strong boundaries are the version of this mechanism that reaches public discussion, and they are modelled here as an additional type with two properties rather than one. A named group has its own fertility, and its own *retention*: the share of children raised in it who remain in it as adults. Those who leave are not deleted into an average; they land in the mainstream, weighted toward its high-propensity type, on the reasoning that leaving a community is not the same as abandoning everything it taught you. How much they carry is a scenario knob (Section 3), set at 60% and reported separately for that reason.

Retention is as load-bearing as fertility. A group at one per cent of a population with six children per woman and ninety per cent retention, against a mainstream at 1.4, reaches 63% of the population in 4 generations, while the same group with weak retention goes nowhere. The condition stated by Arenberg et al. [2022] uses single-sex units: the expected number of same-type daughters surviving to reproduce exceeds one, not that retention times the total fertility rate exceeds one. The cohort engine applies survival and the sex ratio at birth to every cohort, so the condition it

enforces is the single-sex one; the loose phrasing is only ever a shorthand.

Two limitations follow from the available data. No conversion *into* a named group is modelled, although such conversion is small rather than zero for these two groups. Every type starts with its host country's age structure, because nobody publishes one by type; since a population with six children per woman is far younger than its host, this understates group growth, and every group share reported below is a lower bound.

3 Calibration

Of the mechanism's 13 parameters, 8 are estimated from published evidence and 5 are not. The distinction is enforced in code rather than described in prose: the loader refuses a row with no provenance, a row marked as sourced that carries no evidence, and a scenario knob that claims to have been verified. The full table and its sources are in the supplement.

One rule governs the exercise:

> **The standing rule on parameters**
> No mechanism parameter may be fitted to the series it is meant to explain. A quantity that cannot be sourced independently of the fertility path the model produces is labelled a scenario knob, and the paper does not build a result on one. This restricts what the model is allowed to learn, and it is adopted because a mechanism tuned until it reproduces observed fertility has explained nothing about it.

3.1 The dispersion of completed family size

The first mainstream parameter is the coefficient of variation of completed family size, from the Cohort Fertility and Education database [Wittgenstein Centre, 2017], which reports parity distributions by birth cohort. We pin 43 country files and apply one declared rule everywhere: women rather than couples or mothers only; children ever born, including childless women; whole birth-cohort bins whose members were all aged 50–59 at observation; and one estimate per country, with education categories summed.

The database reports exact total children, which identifies the mean of the open highest-parity category. The primary estimate distributes that category geometrically, the maximum-entropy integer distribution with that mean. Substituting the minimum-variance integer tail instead moves the coefficient of variation by 0.005 at the median across all 43 countries and by at most 0.059.

Across the 19 countries with mean completed fertility at or below 2.2 children (29.1 million women), the country-median coefficient of variation is 0.570, with an observed range of 0.443 to 0.787. We take the unweighted country median rather than the women-weighted one because these are a set of plausible settings rather than a harmonised global sample, and weighting would let the single largest national extract dominate. The adopted value is 0.57 with a propagated range of 0.44 to 0.80.

As a check against a source of different construction, paired cohort fertility tables from the United States National Center for Health Statistics [National Center for Health Statistics, 2024] give mean

completed fertility 2.04 and a coefficient of variation of 0.695 for cohorts born 1947–1956 at exact age 50 — inside the declared range and above the cross-country median.

3.2 What that dispersion is, and is not

Hruschka and Burger [2016] show, using two hundred surveys, that a large share of completed-fertility variance can be produced by a counting process alone, with no stable differences between women. The objection is correct and it applies here. These parity marginals do not identify a latent trait with coefficient of variation 0.57; realised family size also contains timing, infertility, partnership history and measurement error.

The parameter is therefore used as an *effective phenotypic spread*, paired with the observed intergenerational correlation; together they reproduce a covariance (Section 2.2) without asserting a variance decomposition. Subtracting a Poisson variance to estimate a latent-trait coefficient would not resolve the problem, because most of these low-fertility distributions are *under*-dispersed relative to Poisson, consistent with targeted family sizes. The empirical range 0.44–0.80 is consequently propagated through every result rather than collapsed to a point.

3.3 Parent–child transmission

The second mainstream parameter is the parent–child correlation in completed family size. Murphy [1999] reports correlations mostly between 0.15 and 0.20 in the larger recent British and American samples reviewed there. Murphy [2013], covering 46 populations in 28 countries, reports unweighted regional means for respondents aged 40 and over of 0.09 in Northern America, 0.11 in Northern Europe, 0.15 in Western Europe, 0.17 in Southern Europe and 0.19 in Eastern Europe. We adopt 0.15, with a deliberately conservative cross-setting interval of 0.05 to 0.30.

This transmission assumption excludes evidence that is not numerically commensurable with a parent–child correlation. Twin studies report substantial genetic variance components for fertility behaviour [Kohler et al., 1999], and there is molecular evidence of contemporary selection on fertility-associated variants [Beauchamp, 2016, Kong et al., 2017]. None of it is used. A heritability is a variance decomposition and is not numerically interchangeable with a parent–child correlation; substituting one for the other would inflate the mechanism by a factor of two or more with no warrant. The parameter here is the observed continuity, whatever produces it.

3.4 How far the evidence is being stretched

Both parameters are applied to all 237 countries, despite the much narrower geographic coverage of the evidence.

The dispersion comes from 19 low-fertility countries, almost all European, and the transmission literature is concentrated in Europe and North America. Neither source covers sub-Saharan Africa, where most of the century's births will happen, or South and East Asia, where most of the departure from replacement now is. Applying a European parent–child correlation to Nigeria is therefore an unsupported geographic extrapolation, represented transparently as a global constant rather than as a locally estimated parameter.

Using one value avoids introducing regional parameters without evidence. The stated range is

propagated instead, which is why 0.44 to 0.80 and 0.05 to 0.30 propagate through every result and why Section 6 reports a threshold across that whole grid rather than at the central pair. A reader who thinks transmission is weaker outside the sampled countries can read the answer off Figure 3 directly.

3.5 The named groups

Two populations have enough published demography to pin fertility, retention and population share separately, which is the threshold for modelling a group by name rather than inventing one.

Haredi Jews in Israel: fertility 6.40 children per woman, population share 13.9%, and retention 86.7%, implied by an estimated 13.3% of adults raised Haredi having left [Regev and Gordon, 2020, Israel Democracy Institute, 2024]. The retention interval is wide because the same study documents leaving rates from 5.4% to 26.4% across factions, and that heterogeneity is real rather than sampling noise.

The Amish in the United States: fertility 6.10 children per woman and retention 84.5% at ages 40 and over, from a population-wide analysis of more than 50,000 households [Anderson and Thiehoff, 2025], with a population share of 0.115% from a specialist census of settlements [Young Center, 2024].

Both fertility figures are period rates, used only to initialise a relative propensity at the base year. They are not scoring targets, since the model's claim concerns cohort fertility and period rates are contaminated by timing [Bongaarts and Feeney, 1998].

3.6 The parameters that are not measured

5 parameters cannot be sourced independently and are labelled scenario knobs. The number of mainstream types (3) is a discretisation choice. The share of group leavers landing in the mainstream's high-propensity type (60%) has no direct measurement. The rate at which a named group's fertility advantage decays as the group stops being unusual has documented instances but no established general value. And two describe the fertility environment: the additional proportional decline per decade, and the floor beneath it.

The environmental path is the consequential unmeasured quantity. It cannot be observed in advance and must not be fitted, because fitting it to observed fertility would consume exactly the evidence the mechanism is supposed to explain. Section 6 is the response: rather than choosing a value and reporting the population it implies, we report the value at which the model's answer changes sign.

4 What is validated, and what is not

A mechanism is interpretable only if its effect can be distinguished from projection arithmetic. The accounting layer is therefore validated against published outputs, and the mechanism is implemented as an additional axis on the same engine rather than as a separate model.

It also sets a distinction the rest of the paper depends on. Three different things are involved here and only one of them is validated:

The accounting is validated.
The engine reproduces the United Nations' published projection from the United Nations' published inputs, to 0.001% of world population at 2100.

The post-2100 continuation is ours.
The source stops at 2100. What happens after it is this project's emulator of the source's own late-horizon behaviour, and Section 7 measures what that choice is worth rather than asserting it is small.

The mechanism is not validated at all.
It is calibrated, from evidence independent of the projection, and its first-generation response is checked against theory. Nothing here tests it against an outcome, because the outcome it predicts has not happened yet. Section 8.4 says when it could.

4.1 The accounting layer and its test

The engine is a standard cohort-component projection with no theory in it: single year of age from 0 to 100 with an open-ended final group, two sexes, 237 countries and territories, annual steps, populations dated 1 January, and rates in force during the calendar year. It takes fertility, survival and net migration as given and does the bookkeeping. Its inputs are World Population Prospects 2024 [United Nations Population Division, 2024].

The test that matters is the zero-migration variant, because every input to it is published and nothing in it is tuned, so any discrepancy is our arithmetic being wrong. Running from 1 January 2024 to 1 January 2100 on the UN's own inputs, world population lands within 0.001%, the worst of 227 countries above ten thousand people within 0.13%, and the worst five-year age group below 100 within 0.006%. Freezing the base year's rates instead gives about 53 billion people by 2150, within 0.05% of the UN's own constant-fertility variant at 2100, which is the absurdity check at the other end of the range. The full table, and the conventions that produce plausible wrong answers if guessed, are in the supplement.

4.2 Why the mechanism's effect is attributable

The typed engine of Section 2 is the same code with a composition axis. Setting every propensity to one leaves the composition free to move but unable to change fertility, and in that configuration the typed engine reproduces the ordinary engine to a relative difference of 3×10^{-16} — machine precision.

This identity establishes attribution: the difference between a run with selection and a run without it cannot arise from a second implementation of the arithmetic, different handling of the open age group, or a changed exposure convention.

4.3 Migration, and what it is allowed to support

The United Nations does not publish net migrants by single year of age. For the scenario comparisons in Section 5 the migration input is backed out of the UN's own medium path as a residual, which makes it a usable forward-model input and no evidence at all: it absorbs every difference between the UN's procedure and ours. Runs using it are labelled diagnostics.

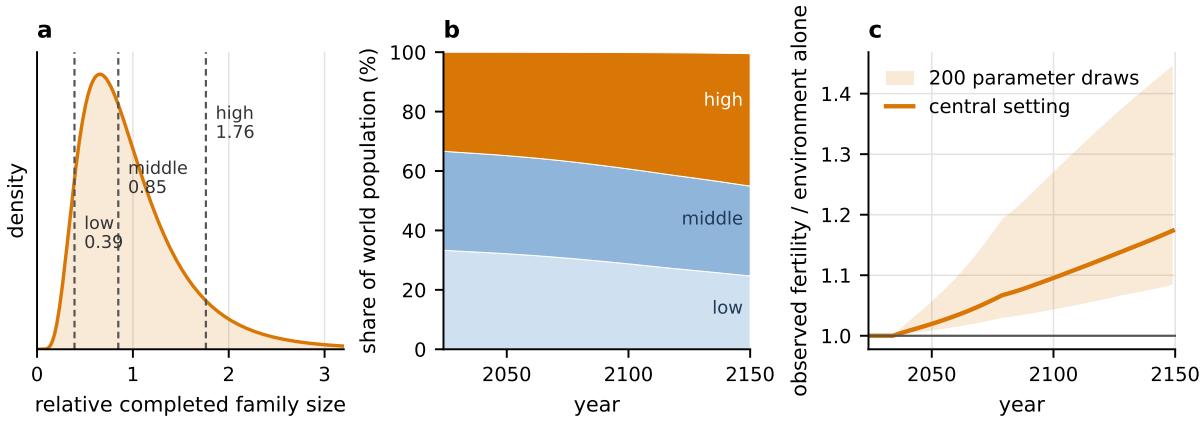

Figure 1: The mechanism in three steps. **(a)** Completed family size is right-skewed with a coefficient of variation of 0.57; the model stands 3 effective types in for that distribution, at the conditional means of its equal-probability thirds, rescaled to reproduce the dispersion exactly. **(b)** World composition drifts because higher-propensity women contribute more of the next cohort: the high type rises from 33.3% of the world population at 2024 to 44.6% at 2150. **(c)** The resulting selection multiplier, observed fertility divided by the environment path alone. The band spans 200 draws with all 13 mechanism parameters sampled uniformly from their stated ranges; it is a sensitivity envelope over judgemental intervals, not a posterior.

Where migration uncertainty is the object of study — Section 7, and the robustness check in Section 6 — a stronger construction is used: the published Bayesian migration trajectories of Azose and Raftery [2015], with rates applied to each path's own evolving population and every draw-year balanced so that world net migration is exactly zero. The public trajectory archive is balanced in expectation but not draw by draw. Without draw-year balancing, individual paths can create or delete millions of people globally even though migration must sum to zero. Enforcing that accounting constraint yields a migration contribution of 0.34 billion to the world uncertainty width.

5 How large is selection?

5.1 The size of the effect, step by step

The benchmark is deliberately not the UN's own path, because that path already contains an assumed fertility recovery whose justification is exactly what is in dispute. Instead it holds that today's fertility already embodies today's conditions: the country's fertility follows the UN's projected path down and then stops at its running minimum, so the decline the source projects is kept and the recovery it assumes is removed. Formally the environment multiplier is $E_t[\ell] = \min_{s \le t} f_s[\ell]/f_t[\ell]$, where f_t is the UN's own projected total fertility. Selection is then switched on against that background, and the components are added one at a time (Figure 2).

Measured mainstream selection — the two independently sourced parameters, with no named groups and no assumed change in conditions — raises world population in 2150 from 8.54 billion to 10.36 billion. That is 1.82 billion, or 21%, from a mechanism absent from conventional projections. Observed world fertility ends 17.4% above the environment path that produced it, and rises

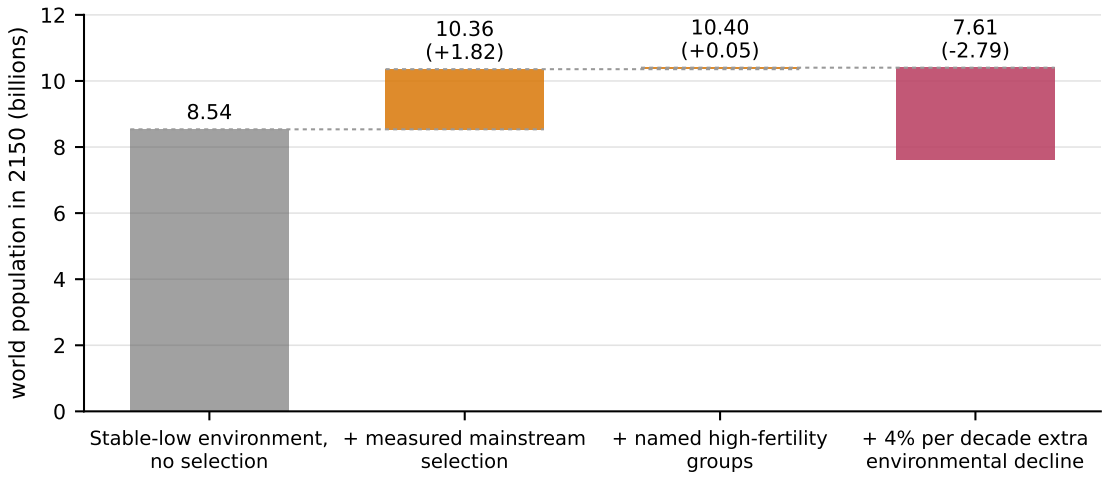

Figure 2: World population in 2150 as each component is added. Grey is the benchmark environment with no selection. Orange steps are the mechanism: measured mainstream selection first, then the two modelled high-fertility groups. Red is an illustrative stress test, a further 4% per decade decline in the fertility environment after 2050, which is a scenario knob and not a forecast. That last step is larger than the mechanism itself, which is why the paper's primary result is the threshold in Section 6 rather than any single bar on this chart.

measurably above it from about 2042 (Figure 1c).

The mechanism produces a level shift rather than a return to growth. Under the benchmark environment with mainstream selection the world still peaks and then runs close to flat; selection changes where the curve settles, not its shape. That is the distinction Arenberg et al. [2022] press against models built from a transmission coefficient alone, and it survives here because the projection carries absolute age-specific fertility.

The 1.82 billion estimate is conditional on more than the two measured calibration parameters. It also depends on the benchmark environment defined above, on children regressing toward the current population rather than a fixed baseline, on 3 types standing in for a continuous distribution, on every type sharing one age profile of fertility, on one parameter pair applied to all countries (Section 3.4), and on the post-2100 treatment in Section 7. It is a result under a stated specification, and Section 8.3 gives the size of the ones that have been measured.

5.2 The modelled groups are a small part of it

Adding the two named high-fertility groups — with their measured fertility, their measured retention, and one scenario knob routing their leavers — moves the world result by 0.05 billion. The world composition at 2150 is 0.5% named-group.

These results differ from the emphasis in much public discussion. The model assigns both groups their measured fertility, allows high retention to compound for four generations, and starts them with their host country's age structure, which understates their growth. Even under these assumptions they contribute 2.5% of what unlabelled variation contributes. The reason is arithmetic rather than sociological: a group at a tenth of a per cent of a national population has four generations to compound before it is large enough to matter to a world total, whereas the

Country	Selection multiplier at 2150	Composition of the effect
Israel	1.81	mainstream plus a large named group
United States	1.30	mainstream plus a small named group
Nigeria	1.16	mainstream only
Japan	1.15	mainstream only
Germany	1.15	mainstream only
Korea	1.14	mainstream only

Table 1: Observed fertility divided by the environment path alone, at 2150, in the run including both selection and the 4% stress test. The multiplier is a statement about composition and is very nearly unaffected by the level of the environment path, for the reason given in Section 6.1.

mainstream's dispersion acts on everybody at once from the first year.

This result applies to the two modelled groups rather than to high-fertility minorities in general. Only Haredi Jews in Israel and the Amish in the United States have sufficiently documented fertility, retention and population shares for separate treatment here. In these two cases, the projected contribution to the world total remains two orders of magnitude below that of mainstream variation; the same arithmetic applies to any group beginning from a comparably small population share.

Where a group is already large it dominates locally. In Israel, whose Haredi population is 13.9% of the country at the base year, the group reaches 58% of the population by 2150 and the country's selection multiplier reaches 1.81. In the United States the Amish reach 4.6%, contributing to a national multiplier of 1.30. Neither is a world-scale effect.

5.3 Cross-national variation

The same run produces very different multipliers by country (Table 1). The ordering follows from Equation (4): the response depends on the *relative* dispersion of family size, so a country at 0.8 children per woman has as much proportional room for recomposition as one at 4, and the multiplier is largest where a named group is large.

5.4 The mechanism's own uncertainty

Drawing all 13 mechanism parameters independently and uniformly from their stated ranges, 200 times, gives a selection multiplier at 2150 between 1.084 and 1.446 for the central 90% of draws, and a stress-test world population between 5.84 and 11.56 billion.

These are sensitivity ranges, not probability intervals. The intervals they are drawn from are judgemental, the uniform shape is a choice, and the parameters are sampled independently when several of them plainly are not. Reading a 90% figure off them as a credible interval would be reading a decision of ours as evidence.

The spread identifies family-size dispersion as the principal source: holding everything else fixed, the endpoints of the measured range, 0.44 and 0.80, move the stress-test world result from 7.10 to 8.92 billion. That is a genuine empirical uncertainty in a quantity that is measurable, unlike the environment path, and narrowing it is tractable work: it needs completed-cohort parity distributions from more than the small set of countries that publish them, and from more than one

continent.

The direction is more robust than the magnitude. In every draw the selection multiplier ends above one — which follows from the mechanism's structure rather than from the evidence, since a positive dispersion and a non-negative transmission coefficient cannot produce a downward response. What the draws vary is how large it gets, not whether the sign holds, and the sign should not be read as an empirical finding.

6 The boundary

Figure 2 shows that the unmeasured fertility environment can dominate the measured mechanism. The mechanism, built from two measured parameters, moves the 2150 world result by 1.82 billion. A single unmeasured scenario knob — how fast the fertility environment keeps falling — moves it by 2.79 billion in the opposite direction. Assigning a preferred value to that knob would allow an unsupported assumption to determine the headline result. A structural property of the model instead permits the knob to be reported as a break-even threshold.

6.1 Selection and the environment are commensurable

The environment multiplier $E_t[\ell]$ in Equation (1) multiplies every type equally. It therefore cancels from the relative weights with which types contribute to the next generation: whether conditions are generous or harsh, the ratio of a high-propensity woman's births to a low-propensity woman's births is unchanged, so the composition follows almost the same path either way.

The cancellation is exact in the birth weights and very nearly exact in the outcome. Running the same selection setting under two environments that end 2.79 billion people apart gives selection multipliers of 1.176 and 1.174, a difference of 0.22%. The residue is a real second-order effect rather than rounding: a harsher environment produces a different age structure, so the childbearing-age women over whom the propensity is averaged are drawn differently, and a fixed number of migrants mixes into a smaller population with proportionally more force.

That near-separability is a modelling assumption — Section 8.3 takes it seriously as one — but it is what makes the two forces directly comparable. Selection multiplies fertility by S; a uniform environmental decline of rate r per decade over d decades multiplies it by $(1-r)^d$. The two cancel when

$$S \cdot (1-r)^d = 1 \qquad \Longleftrightarrow \qquad r^\star = 1 - S^{-1/d}. \tag{5}$$

6.2 Two thresholds, because there are two questions

Equation (5) cancels the fertility *rate* in the final year. At the central calibration, measured mainstream selection reaches $S = 1.164$ in the final fertility-rate year (2149), giving 1.53% per decade of additional decline beginning in 2050. Country selection effects are aggregated using births from the no-selection counterfactual, so a scenario cannot choose the weights by which it is judged.

That is not the same as cancelling selection's effect on the population, and the difference is not small. Bringing the rate back to parity in 2149 leaves every extra birth selection produced in the

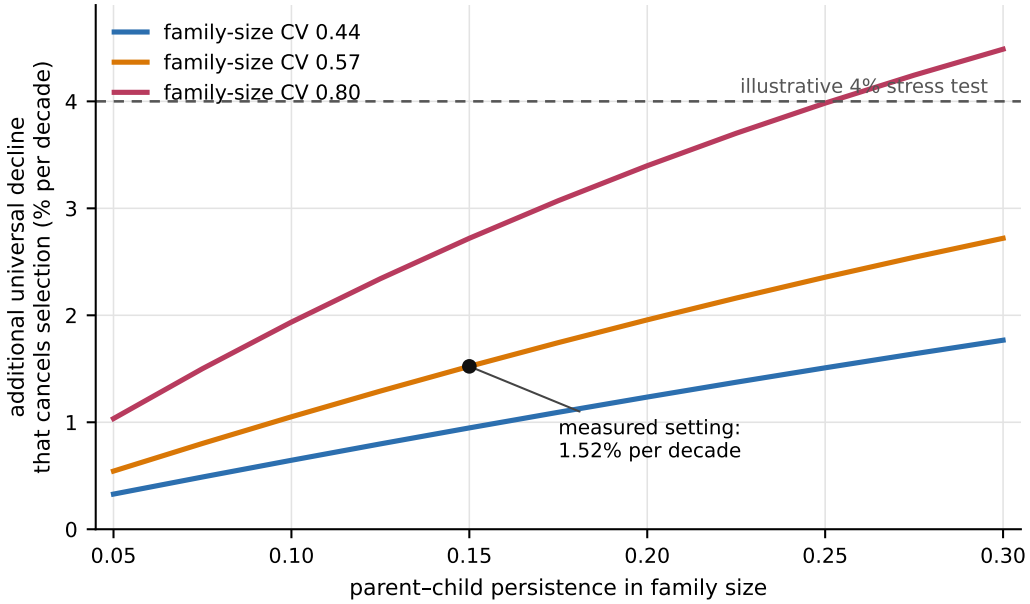

Figure 3: The additional uniform decline in the fertility environment, per decade after 2050, that cancels mainstream selection's effect on the fertility rate by 2150. Above a curve the environment wins; below it selection does. The two axes are the measured parameters: the dispersion of completed family size (three curves, spanning the observed low-fertility country range) and the parent–child correlation in it. Shaded bands are the 5th to 95th percentile across 50 paired stochastic migration paths, and are narrow enough to be nearly invisible, which is itself the result reported in Section 6.3. The dashed line marks the illustrative 4% stress test of Figure 2. Cancelling the effect on the 2150 *population* rather than on the rate requires 2.45% at the central setting, above the top of this figure's central curve.

preceding century, and those people and their descendants are alive in 2150. Solving directly for the decline that returns the 2150 world population to the no-selection benchmark of 8.54 billion gives 2.45% per decade — sixty per cent larger. Both are reported because they answer distinct inferential targets; the terminal-rate threshold alone understates the decline required to offset accumulated population effects.

Figure 3 shows how the rate threshold moves across the empirical range of both measured parameters, over a 33-cell grid. It runs from 0.33% per decade at the low corner — weak dispersion and weak transmission — to 4.49% per decade at the high corner. The threshold is close to linear in the transmission coefficient and roughly quadratic in the dispersion, as Equation (4) implies.

Read as a statement about the world, this says something more useful than a projected total. If the conditions that determine how many children a given person has continue to deteriorate at less than roughly one and a half per cent per decade after 2050, the compositional force wins on the fertility rate; below about two and a half per cent it still wins on the population at the horizon. Above those, the environment wins. Current evidence cannot determine which regime will prevail. The threshold expresses the unresolved question as a rate that can, in principle, be observed.

For scale, the illustrative 4% per-decade stress test is well above both thresholds. Sustained for a century, it removes a third of the fertility environment before its floor binds. It is included to

show the model's response to a strong environmental decline, not as a forecast.

6.3 The boundary survives stochastic migration

A threshold computed on one migration path would be worth little if migration could move it. It is checked against 50 stochastic paths drawn from the published Bayesian migration trajectories [Azose and Raftery, 2015], with one design choice that matters: the selection run and its no-selection counterfactual share each source path and each post-2100 innovation, so the comparison is paired and a migration draw cannot masquerade as a selection effect. Every draw-year is balanced to exactly zero world net migration.

The central rate threshold is 1.52% per decade with a 90% range of 1.51% to 1.53%. Migration moves it by about one part in seventy-five, against a factor of thirteen between the corners of the measured-parameter grid. That is what one should expect — migration sums to zero globally by construction and can affect a world fertility aggregate only by relocating people between countries with different schedules — but expecting it is not the same as having checked it.

7 The horizon, and whose numbers are whose

A projection running to 2150 on inputs that stop at 2100 has a problem of attribution rather than of arithmetic. Everything past the last year of published assumptions is the modeller's own extrapolation, and drawn as one continuous line the reader cannot see where the source ended.

7.1 Three separate objects

The UN reproduction.
World Population Prospects 2024's medium path, reproduced by this engine, through 2100 and no further. No value after that year is labelled a UN projection.

The project extension.
Begins from the published population on 1 January 2100 and runs to 2150 with stochastic continuations of the published WPP 2024 fertility and mortality trajectories, together with stochastic migration. Because the public archives do not include the fitted posterior state, these are emulators of the source trajectories rather than an official United Nations projection.

The selection model.
A separate object forking in 2024, so that selection acts over the whole projection rather than switching on after the UN's assumptions have already fixed the 2100 population.

7.2 What happens after the source stops

Holding final rates constant after 2100 assumes that fertility ceases to vary and longevity ceases to improve exactly at the source boundary. Because this assumption governs fifty of the projection's 126 years, it is evaluated against a continuation of the source's late-horizon behaviour. Fitted over 2070–2100, the two components exhibit different dynamics, which the continuation treats separately.

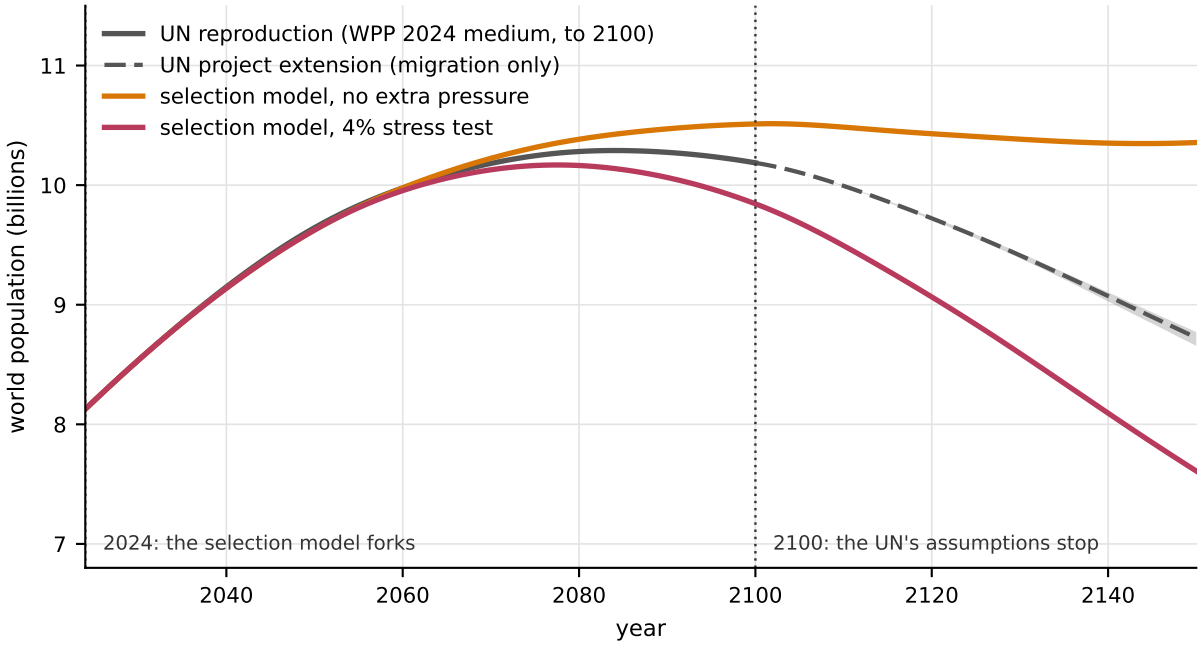

Figure 4: Three objects, drawn so it is visible which is whose. The solid grey line is the reproduced UN medium path and stops where the UN's assumptions stop. The dashed line and shaded fan isolate the migration component of this project's extension, with migration varying across 1000 paths; the fan is narrow because migration cancels globally by construction. Fertility and mortality are held fixed in this figure so that migration can be read on its own. The full project extension continues all three demographic components, as described in Section 7.2 and shown for the comparator ensemble in Figure 5. The two coloured lines are the selection model, which forks in 2024 and is not a UN projection at any point.

Fertility is stationary within a trajectory. Each published path fluctuates around its own level with an autocorrelation near 0.89, and a stationary spread of about 0.12 children. The spread *between* trajectories at 2100 is roughly 0.40 — three times larger. That is the published model's structure: the long-run level is itself uncertain, and most of the fertility uncertainty at 2150 is already present at the boundary. Each path is therefore continued as an autoregression around *its own* 2100 value. Pooling across trajectories would instead fit one level per country and draw every trajectory toward it, substantially reducing the posterior spread.

Longevity is still trending. The annual gain in female life expectancy over 2070–2100 averages 0.114 years and is strikingly uniform across countries. Freezing it asserts that improvement halts at the boundary. Each path is continued as a random walk with its country's own fitted drift plus autoregressive noise on the gain, with the male series derived from the female series and a separately continued sex gap so the two cannot cross.

Two limits are stated rather than buried. Fertility is still drifting downward slightly at the boundary, and centring each path on its 2100 value drops that residual drift rather than extrapolating it; extrapolating would subtract roughly 0.15 children by 2150, so this choice biases the result *upward*. And the continuation does not reproduce the published mortality model's deceleration, in which gains slow as life expectancy rises, so it slightly overstates longevity in the same direction. Both are emulators of a model's output, not continuations of the published model, for the same

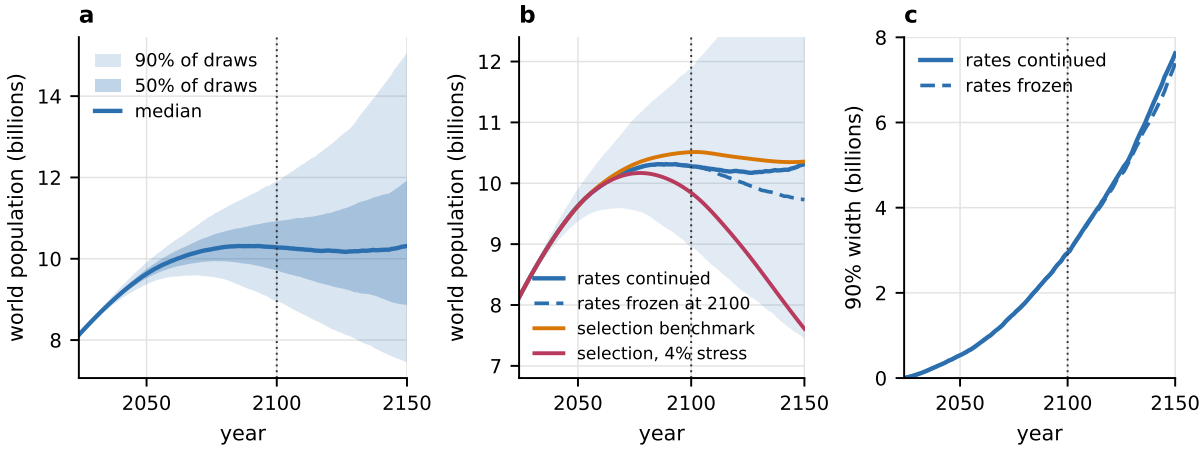

Figure 5: Uncertainty accumulating. **(a)** World population under the conventional mean-reverting posterior propagated through this engine, 1000 draws, with fertility and longevity continued past 2100 rather than frozen. Shaded bands are the central 50% and 90% of draws. The fan widens throughout because small differences in long-run fertility compound through both cohort size and later births, and it keeps widening after 2100 rather than stopping at an artificial boundary. **(b)** Medians only, against the version holding rates constant after 2100, with the selection model's benchmark and stress-test paths for scale. **(c)** The width of the 90% band year by year. The two are nearly indistinguishable, which is the finding: the frozen-rate assumption moved the centre of the distribution without much affecting how fast it opened.

reason as the migration continuation: the public archive omits the fitted posterior state. The age *patterns* — the shape of fertility across ages, and the mortality standard the levels are scaled onto — are still held at their final published shape, which is a much weaker assumption than holding a level and is stated separately for that reason.

7.3 Sensitivity to post-2100 continuation

Holding rates constant after 2100 gives a 2150 world median of 9.73 billion with a 90% range of 6.97 to 14.36. Continuing the source's own behaviour gives 10.31 billion with a range of 7.45 to 15.07, a median shift of 0.58 billion and a change in width of 0.23 billion.

Continuing the source moves the median *up*, for the two reasons stated above — longevity keeps improving and the residual fertility drift is not extrapolated — while the extra fifty years of fertility innovation widen the band only modestly, because the published fertility process is stationary within a trajectory and most of the eventual spread is already there at 2100.

7.4 Which uncertainty dominates depends on where you look

Varying one source at a time across 200 draws (Figure 6a), the 90% width of world population in 2150 is 7.26 billion for long-run fertility, 5.72 billion for the selection mechanism's parameters, 0.52 billion for mortality and 0.34 billion for migration. The mechanism entry is not directly comparable with the others: it varies judgemental ranges against the UN's medium demographic rates rather than propagating a posterior, and Section 5 explains why its band is a sensitivity

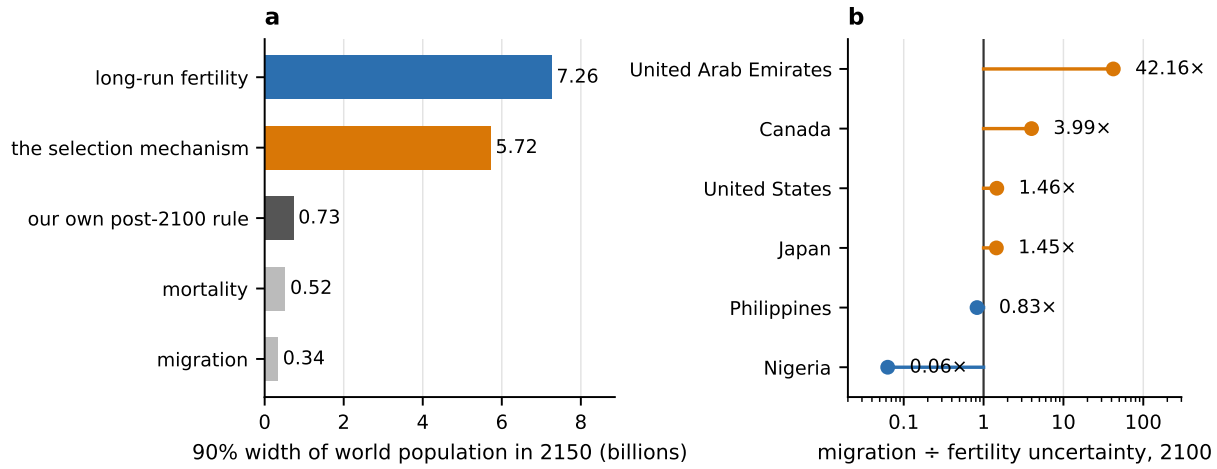

Figure 6: **(a)** One-at-a-time sensitivity: the 90% width of world population in 2150 when one source is varied across its own draws and the others are held at a median trajectory. These are not commensurable in the way a variance decomposition would be — fertility, mortality and migration are posterior draws, while the mechanism entry is a uniform sweep over judgemental ranges — so the comparison is of magnitude, not of attributable share. **(b)** The same comparison inside individual countries at 2100, where the ordering reverses.

envelope. It is included to place uncertainty in the mechanism's parameters on the same scale as the demographic inputs.

The country panel (Figure 6b) reverses the ordering in the world panel. At 2100, migration uncertainty is 42.16 times fertility uncertainty for the United Arab Emirates and 3.99 times for Canada, but only 0.06 times for Nigeria. In the project extension the Emirates' 2150 population runs from 1.0 million to 267.3 million around a median of 24.6 million — a factor of more than two hundred, for a component contributing almost nothing to the world total. Any single summary of "the" uncertainty in a population projection is answering a question about the world that most users of projections are not asking.

7.5 A conventional comparator

The published Bayesian posterior for fertility and mortality [Alkema et al., 2011, Raftery et al., 2012, Ševčíková and Raftery, 2016] was propagated through this same engine one draw at a time: 1000 draws, 236 countries, to 2150. This is the conventional mean-reverting view rather than the compositional specification. It is stored as an immutable dated record and marked as a comparator rather than as a claim of this paper.

As an external consistency check, the separate deterministic run on the UN's published assumptions peaks at 10.29 billion in 2084, against the ensemble median's 10.31 billion in 2093. Those two calculations share only the engine, and they agree on the peak to within a fifth of a billion people and on its date to within a decade.

8 Discussion

8.1 A middle position, and why it is a position

Two literatures reach opposite conclusions about what intergenerational transmission of fertility implies. Collins and Page [2019] apply the breeder's equation to national fertility and conclude that global stabilisation this century becomes very unlikely, with European and North American trajectories turning from decline to growth. Arenberg et al. [2022] reply that a model built from transmission alone contains no absolute fertility level, cannot represent higher-fertility subgroups converging downward, and therefore cannot establish that transmission produces growth at all.

Both are right about what they are arguing. The disagreement is not really about the mechanism, which nobody disputes; it is about whether a model carrying only a transmission coefficient can support a conclusion about population size. It cannot; the mechanism must instead be embedded in a projection that carries absolute levels.

Embedding the mechanism produces a result intermediate between these positions. The mechanism is real, its direction is unambiguous, and it is worth 1.82 billion people at 2150 — too large to leave out of a long-run projection. It is also not a return to growth. Under the benchmark environment the world still peaks and then runs close to flat; what selection changes is where the curve settles. The Arenberg condition is satisfied rather than violated, because absolute age-specific fertility and survival are carried at every step, and the outcome is a level shift of about 21%.

The second finding cuts against how the subject is usually discussed. The two best-documented high-fertility religious groups, modelled favourably, contribute 2.5% of what unlabelled variation contributes. The compositional force in long-run fertility is not mainly a story about visible minorities; it is a property of the ordinary dispersion of family size, acting on everybody at once. That is the better result to have, because ordinary dispersion is something national statistical offices already measure.

8.2 Why a threshold is the right output

The future fertility environment is unidentified. The literature contains coherent arguments in both directions: the low-fertility trap hypothesis [Lutz et al., 2006] describes self-reinforcing downward pressure through ideal family size and cohort experience, while the mean-reverting specifications in official projections [Alkema et al., 2011, Raftery et al., 2012] embed an eventual recovery. Neither is settled, and a 2150 world total is enormously sensitive to the difference.

A projection whose headline is controlled by an assigned value for that quantity conflates an assumption with a result. Equation (5) provides an alternative. Because both forces multiply fertility, the model can be asked for the environmental decline that exactly offsets the measured compositional gain, and those numbers — 1.52% per decade for the rate, 2.45% for the population — are determined by the two measured parameters and the arithmetic rather than by a preference.

For this class of problem, the threshold is a more informative output. It is falsifiable in the parameters that are measurable, explicit about the one that is not, and expresses disagreement about the future as disagreement about a rate. Alternative assumptions about the future fertility environment can therefore be located directly on Figure 3.

8.3 Limitations

The environment multiplies every type equally.
This is what makes the threshold a single clean number and it is the assumption most likely to be wrong. Rising costs of housing, childcare and foregone earnings plausibly fall hardest on exactly the people whose disposition would otherwise produce large families, in which case the environment would interact with selection rather than merely oppose it and no single threshold would exist. Modelling that requires evidence on how fertility declines differ across the family-size distribution within cohorts, for which comparable cross-national data are not currently available. This is the most valuable single extension available.

Two parameters estimated in Europe are applied to the world.
Section 3.4 states this at length. Nothing in the calibration covers sub-Saharan Africa, where most of the century's births will occur. The propagated ranges represent this limitation only imperfectly.

The covariance is constructed from two sources, not measured in one.
Section 2.2. A single paired parent–child sample with both moments would replace the weakest link in the calibration.

Children regress toward the current population.
The alternative is regression toward a fixed baseline, which produces an equilibrium instead of compounding and lowers the multiplier from 1.174 to 1.072 and the stress-test population from 7.61 to 7.13 billion. We report it as a lower bound rather than as a rival specification, because a baseline permanently fixed at 2024 says a society's fertility culture never moves regardless of who is in it, which over 126 years is not credible. The true behaviour is presumably between the two, and where between is an empirical question about cultural transmission that demography could answer and largely has not.

Types are a discretisation, not a discovery.
The propensity distribution is stood in for by 3 points reproducing its first two moments exactly. Higher moments and the shape of the upper tail are approximations, and selection acts disproportionately on that tail. The number of types is a declared scenario knob for this reason.

Transmission is maternal only.
A child's type comes from its mother, with no paternal transmission, assortative mating or couple formation, while the correlations used to calibrate it come from studies covering both. Assortative mating on family-size preference would raise the response; the absence of paternal transmission lowers it. We have not established which dominates.

Every type shares one age profile of fertility.
Types differ by a multiplier and not by timing. If higher-fertility dispositions also start earlier, more generations complete by 2150 and the response compounds further than reported.

Selection is on fertility only.
Survival is identical across types. High-fertility groups often have distinctive mortality, and modelling that without evidence would add a parameter and subtract credibility.

Named-group results cover two groups.
Not high-fertility minorities in general, and no conversion into a group is modelled.

8.4 What would show this to be wrong

The mechanism's claim resolves in cohort fertility, not period fertility, and the distinction is not pedantry: period rates are depressed by postponement even when completed family sizes are unchanged [Bongaarts and Feeney, 1998], so a mechanism predicting a rise in completed cohort fertility could be scored as failing for decades by a measure that is not measuring it.

The first genuinely informative test arrives when the cohorts born around 1990 complete their childbearing, in the late 2030s. What can be checked before then is narrower — whether the dispersion of completed family size behaves as Section 3.1 measures it in countries not yet in the sample, whether the parent–child correlation holds outside Europe and North America, and whether the composition drift in Figure 1b is visible in successive cohorts of the same country. Each of those evaluates an input rather than the projection itself.

To preserve a testable claim through 2038, the projections are written as immutable dated records with proper scoring rules attached [Gneiting and Raftery, 2007], and the storage layer raises rather than overwrites an existing record. This record-keeping framework is intended to preserve the model's evidentiary value as the relevant cohort outcomes become observable.

9 Conclusion

Long-run population projection turns on a single quantity nobody can measure: what fertility does after the demographic transition. Conventional practice handles this by modelling the national rate as a mean-reverting time series, which is a defensible statistical choice and an odd substantive one, because it assumes away the fact that the parents of each generation are sampled in proportion to how many children they have.

Adding that compositional force to a cohort-component projection shows it to be real, measurable from evidence independent of the series it explains, and consequential: 1.82 billion people at 2150, or 21%, from two parameters that statistical offices already have the data to estimate. It also shows it to be more modest than either its advocates or its critics would suggest. It shifts the level of the long-run curve, it does not restore growth, and it is almost entirely a property of ordinary variation rather than of the visible minorities that dominate discussion of the subject.

The countervailing force is unmeasured, and the paper's main contribution is a comparison that does not require assigning it a preferred value. Both act multiplicatively on fertility, so "what will the population be" can be replaced by "how fast would conditions have to keep deteriorating for the compositional force to lose" — 1.52% per decade to cancel its effect on the fertility rate, 2.45% to cancel its effect on the population at the horizon, and 0.33% to 4.49% for the first of those across the empirical uncertainty in the two measured parameters.

Present evidence does not determine which side of that threshold the world occupies. Reporting the threshold isolates the unresolved assumption and makes the source of any future projection error explicit in Figure 3.

Author contributions and responsibility

The byline follows a project-specific definition of authorship based on substantive intellectual and technical contribution. Including two language models is a deliberate departure from prevailing journal policies that reserve authorship for humans. The public record of the project's conversations, committed with the repository, permits the allocation below to be checked against the collaboration itself.

Dylan Slagh conceived the project and its 2150 horizon, proposed that long-run population be studied through evolutionary demography rather than a national fertility rate alone, and set the standard that the model should be mechanistically explicit, probabilistic and readable beyond a specialist audience. He determined the final scientific architecture: an official UN reproduction, a separately labelled project extension and a selection model operating from 2024; selected compositional selection as the paper's primary subject; and chose a universal environmental multiplier as a transparent boundary parameter rather than an economic forecast. He set the publication and transparency requirements, adjudicated modelling choices throughout the work and reviewed the resulting analysis and manuscript.

Claude Opus 5 co-developed the initial specification, including the 2024 fork and the distinction between a mechanistic model and a purely phenomenological projection. It designed and implemented the cohort-component engine, historical backtest, interactive map, Bayesian demographic baseline, compositional mechanism and principal uncertainty analyses; developed much of the model's testing, visualisation and reproducibility infrastructure; and built the stochastic post-2100 fertility and mortality continuation. It wrote the first complete version of the present manuscript and supplement and revised them against the scientific review.

ChatGPT 5.6 Sol co-developed the hierarchical specification by separating mainstream intergenerational persistence, family and cultural transmission, and explicit group retention, and by treating the fertility environment as a force that can move independently of population composition. It established the official posterior-data extraction workflow, researched and audited the mechanism parameters, derived the completed-family-size calibration, developed the selection-only benchmark and break-even boundary with Dylan Slagh, and implemented the balanced migration continuation, paired stochastic selection comparisons and boundary robustness analysis. It also conducted the independent manuscript review that distinguished terminal-rate from accumulated-population thresholds (Section 6) and clarified the constructed covariance (Section 2.2), then edited the final paper and supplement.

The conversation archive documents this division of labour but is not needed to interpret or reproduce the paper. Dylan Slagh accepts responsibility for the analysis and manuscript.

References

Leontine Alkema, Adrian E. Raftery, Patrick Gerland, Samuel J. Clark, François Pelletier, Thomas Büttner, and Gerhard K. Heilig. Probabilistic projections of the total fertility rate for all countries. *Demography*, 48(3):815–839, 2011. doi: 10.1007/s13524-011-0040-5.

Cory Anderson and Stephanie Thiehoff. The population structure of the Amish, a rapidly growing ethnic religion in North America. *Population Studies*, 2025. doi: 10.1080/00324728.2025.2592576.

Samuel Arenberg, Kevin Kuruc, Nathan Franz, Sangita Vyas, Nicholas Lawson, Melissa LoPalo, Mark Budolfson, Michael Geruso, and Dean Spears. Research note: Intergenerational transmission is not sufficient for positive long-term population growth. *Demography*, 59(6):2003–2012, 2022. doi: 10.1215/00703370-10290429.

Jonathan J. Azose and Adrian E. Raftery. Bayesian probabilistic projection of international migration. *Demography*, 52(5):1627–1650, 2015. doi: 10.1007/s13524-015-0415-0.

Jonathan P. Beauchamp. Genetic evidence for natural selection in humans in the contemporary United States. *Proceedings of the National Academy of Sciences*, 113(28):7774–7779, 2016. doi: 10.1073/pnas.1600398113.

John Bongaarts and Griffith Feeney. On the quantum and tempo of fertility. *Population and Development Review*, 24(2):271–291, 1998. doi: 10.2307/2807974.

Oskar Burger and John P. DeLong. What if fertility decline is not permanent? The need for an evolutionarily informed approach to understanding low fertility. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1692):20150157, 2016. doi: 10.1098/rstb.2015.0157.

Sean G. Byars, Douglas Ewbank, Diddahally R. Govindaraju, and Stephen C. Stearns. Natural selection in a contemporary human population. *Proceedings of the National Academy of Sciences*, 107(suppl. 1):1787–1792, 2010. doi: 10.1073/pnas.0906199106.

Jason Collins and Lionel Page. The heritability of fertility makes world population stabilization unlikely in the foreseeable future. *Evolution and Human Behavior*, 40(1):105–111, 2019. doi: 10.1016/j.evolhumbehav.2018.09.001.

Ronald A. Fisher. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford, 1930.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Daniel J. Hruschka and Oskar Burger. How much between-group variation in fertility comes from individual-level processes? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1692):20150155, 2016. doi: 10.1098/rstb.2015.0155.

Israel Democracy Institute. Annual statistical report on Haredi society in Israel 2024. Technical report, Israel Democracy Institute, Jerusalem, 2024. URL https://en.idi.org.il/media/27532/idi-annual-statistical-report-on-haredi-society-2024.pdf.

Hans-Peter Kohler, Joseph Lee Rodgers, and Kaare Christensen. Is fertility behavior in our genes? Findings from a Danish twin study. *Population and Development Review*, 25(2):253–288, 1999. doi: 10.1111/j.1728-4457.1999.00253.x.

Martin Kolk, Daniel Cownden, and Magnus Enquist. Correlations in fertility across generations: Can low fertility persist? *Proceedings of the Royal Society B: Biological Sciences*, 281(1779):20132561, 2014. doi: 10.1098/rspb.2013.2561.

Augustine Kong, Michael L. Frigge, Gudmar Thorleifsson, Hreinn Stefansson, Alexander I. Young, Florian Zink, Gudrun A. Jonsdottir, Aysu Okbay, Patrick Sulem, Gisli Masson, Daniel F. Gudbjartsson, Agnar Helgason, Gyda Bjornsdottir, Unnur Thorsteinsdottir, and Kari Stefansson. Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences*, 114(5):E727–E732, 2017. doi: 10.1073/pnas.1612113114.

Wolfgang Lutz, Vegard Skirbekk, and Maria Rita Testa. The low-fertility trap hypothesis: Forces that may lead to further postponement and fewer births in Europe. *Vienna Yearbook of Population Research*, 4:167–192, 2006. doi: 10.1553/populationyearbook2006s167.

Michael Murphy. Is the relationship between fertility of parents and children really weak? *Social Biology*, 46(1–2), 1999. doi: 10.1080/19485565.1999.9988991.

Michael Murphy. Cross-national patterns of intergenerational continuities in childbearing in developed countries. *Biodemography and Social Biology*, 59(2), 2013. doi: 10.1080/19485565.2013.833779.

National Center for Health Statistics. Cohort fertility tables, 2024. URL https://www.cdc.gov/nchs/nvss/cohort_fertility_tables.htm. Tables 2 and 3; accessed 13 August 2026.

George R. Price. Selection and covariance. *Nature*, 227:520–521, 1970. doi: 10.1038/227520a0.

Adrian E. Raftery, Nan Li, Hana Ševčíková, Patrick Gerland, and Gerhard K. Heilig. Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences*, 109(35):13915–13921, 2012. doi: 10.1073/pnas.1211452109.

Eitan Regev and Gabriel Gordon. Transitions between religious groups among Israeli Jews. Israel Democracy Institute, 2020. URL https://en.idi.org.il/articles/32775. Accessed 13 August 2026.

Alan Robertson. A mathematical model of the culling process in dairy cattle. *Animal Production*, 8(1):95–108, 1966. doi: 10.1017/S0003356100037752.

Hana Ševčíková and Adrian E. Raftery. bayesPop: Probabilistic population projections. *Journal of Statistical Software*, 75(5):1–29, 2016. doi: 10.18637/jss.v075.i05.

Hana Ševčíková, Nan Li, Vladimíra Kantorová, Patrick Gerland, and Adrian E. Raftery. Age-specific mortality and fertility rates for probabilistic population projections. In Robert Schoen, editor, *Dynamic Demographic Analysis*, pages 285–310. Springer International Publishing, Cham, 2016. doi: 10.1007/978-3-319-26603-9_15.

United Nations Population Division. World Population Prospects 2024: Methodology of the United Nations population estimates and projections. Technical report, United Nations, New York, 2024. URL https://population.un.org/wpp/Publications/Files/WPP2024_Methodology-Report_Final.pdf. Department of Economic and Social Affairs.

Stein Emil Vollset, Emily Goren, Chun-Wei Yuan, Jackie Cao, Amanda E. Smith, Thomas Hsiao, Catherine Bisignano, Gulrez S. Azhar, Emma Castro, Julian Chalek, Andrew J. Dolgert, Tahvi Frank, Kai Fukutaki, Simon I. Hay, Rafael Lozano, Ali H. Mokdad, Vishnu Nandakumar, Maxwell Pierce, Martin Pletcher, Toshana Robalik, Krista M. Steuben, Han Yong Wunrow, Bianca S. Zlavog, and Christopher J. L. Murray. Fertility, mortality, migration, and population

scenarios for 195 countries and territories from 2017 to 2100: A forecasting analysis for the Global Burden of Disease Study. *The Lancet*, 396(10258):1285–1306, 2020. doi: 10.1016/S0140-6736(20)30677-2.

Wittgenstein Centre. Cohort Fertility and Education database, 2017. URL https://www.eurrep.org/database/database/. Wittgenstein Centre for Demography and Global Human Capital; data accessed 13 August 2026 under the database's terms of use.

Young Center. Amish population, 2024, 2024. URL https://groups.etown.edu/amishstudies/population-2024/. Young Center for Anabaptist and Pietist Studies, Elizabethtown College; accessed 13 August 2026.