Human Genealogy Reveals a Selective Advantage to Moderate Fecundity*

Oded Galor Marc Klemp

February 1, 2019

Abstract

This research presents the first evidence that moderate fecundity had maximized long-run reproductive success in the human population. Using a reconstructed genealogy for nearly half a million individuals in Quebec during the 1608–1800 period, we find that while a high fecundity was associated with a larger number of children, perhaps paradoxically, a moderate fecundity had maximized the number of descendants after several generations. Moreover, the finding suggests that the level of fecundity that maximized long-run reproductive success was above the population average, indicating that natural selection had decreased the level of fecundity in the population over this period. This evolutionary process may have contributed to the onset of the demographic transition and thus to the evolution of societies to an era of sustained economic growth.

Keywords Demography, Evolution, Human Capital Formation, Natural Selection, Fecundity, Quantity-Quality Trade-Off, Long-Run Reproductive Success, Economic Growth

JEL Classification Codes J10, N30, O10

---

[*] The data was collected and kindly provided by "Le Programme de Recherce en Démographie Historique" (PRDH) at the University of Montreal. The authors are grateful to Bertrand Desjardins for sharing the data. Marc Klemp is funded by the Carlsberg Foundation, the Danish Research Council, and the European Commission. Oded Galor (oded_galor@brown.edu): Department of Economics and Population Studies and Training Center, Brown University, 64 Waterman St., Providence, RI 02912, USA. Marc Klemp (marc.klemp@econ.ku.dk): Department of Economics and Population Studies and Training Center, Brown University, 64 Waterman St., Providence, RI 02912, USA, and Department of Economics, University of Copenhagen, Øster Farimagsgade 5, building 26, 1353 Copenhagen K, Denmark.

Introduction

The influential life-history theory suggests that observed fecundity of organisms reflects a trade-off in reproductive success between the quantity and quality of offspring. Central to the theory is the supposition that there exists an optimal level of fecundity beyond which fitness diminishes (Lack et al., 1954; Cody, 1966; Roff, 1992; Stearns, 1992). A negative association between the quantity and the quality of offspring has been documented in a wide variety of species, ranging from plants to humans. In particular, researchers uncovered an inverse relationship between the number of seeds and their size as well as between the quantity and quality of offspring within and across mammals (Salisbury et al., 1942; Harper et al., 1970; Roff, 2002; Charnov and Ernest, 2006; Walker et al., 2008). Moreover, a trade-off between fertility on the one hand and offspring survival probability and education on the other hand has been documented for pre-industrial human societies (Lee, 1993; Hill and Hurtado, 1996; Strassmann and Gillespie, 2002; Gillespie et al., 2008; Meij et al., 2009).

Nevertheless, as shown theoretically in supplementary materials section S1, the presence of a static trade-off between the quantity and quality of offspring is merely a necessary but not a sufficient condition for the presence of an adverse effect of fecundity on reproductive success in the long run. The extensive exploration of the static trade-off between quantity and quality of offspring, while confirming an important building block of the theory, has therefore not shed a direct light on the effect of the reduction in quantity, and the associated investment in offspring quality, on the long-run reproductive success. Moreover, few attempts to examine the effect of fertility on long-run reproductive success in pre-industrial societies have been largely inconclusive (Kaplan et al., 1995; Borgerhoff Mulder, 2000).

This research presents the first conclusive evidence about the trade-off between fertility and long-run reproductive success in the human population. Using a reconstructed genealogy for nearly half a million individuals based on the parish registers of the St. Lawrence Valley in Quebec during the 1608–1800 period, the study traces the number of descendants of early inhabitants of this Canadian province in the subsequent four generations. Using the protogenesic interval, i.e., the time interval between the date of marriage and the first live birth, as a proxy for fecundity over this period and thus as a source of variation in family size, the research establishes that while high fecundity was associated with a larger number of children, moderate fecundity had maximized the number of descendants after several generations. See supplementary materials section S2 for information the empirical strategy.

The research finds that the maximal reproductive success is attained by couples with a moderate protogenesic interval (i.e., those whose first delivery occurs 65 weeks after their marriage, in comparison to a sample median of 53 weeks). In particular, in comparison to highly fertile couples whose first child is born 38 weeks after the marriage, those individuals have on average 0.3 fewer children, but 0.4 more grandchildren, 8.4 additional great-grandchildren, and 15.7 added great-great-grandchildren. In light of the heritability of fecundity (see supplementary materials section S5 and (Christensen et al., 2003; Pettay et al., 2005; Ramlau-Hansen et al., 2008; Kosova et al., 2009)), the finding that the optimal protogenesic interval is above the population median suggests that in pre-industrial Quebec, the representation of individuals with lower levels of fecundity, and thus higher pre-disposition towards child quality, has gradually increased in the population.

The evidence from pre-industrial Quebec suggests that the forces of natural selection favored individuals characterized by moderate fecundity, increasing the population's predisposition towards investment in child quality. Interestingly, the conditions that were faced by the founder population of Quebec during this high-fertility time period may resemble the environment that anatomically modern humans confronted during their migration out of Africa, as they settled new territories where the carrying capacity of the new environment was an order of magnitude greater than the size of the founder population. Thus, the findings suggest that during the high-fertility segments of the Malthusian epoch in which evolutionary forces could have made a significant impact on the composition of the population (e.g., during the Neolithic transition and the formation of sedentary agricultural communities), natural selection favored individuals with a larger predisposition towards child quality, contributing to human capital formation, the onset of the demographic transition, and the evolution of societies from an epoch of stagnation to sustained economic growth (Galor and Moav, 2002; Galor, 2011).

The empirical strategy employed in this research takes advantage of the inherent uncertainty in the process of human reproduction. In light of the social norm observed in pre-industrial Quebec, in which marriage marked the intention to conceive a child (see supplementary materials), this research exploits the time interval between the date of first marriage and the first birth (the protogenesic interval, PI) as a source of variation in fertility, reflecting the probability of successful conception (i.e., fecundability). Since fecundability reflects genetic and socio-environmental factors, individuals' PI is affected by their genetic predisposition, their socio-environmental conditions, as well as the realization of random elements that affect conception. The latter reflects the randomness in the occurrence of a pregnancy in a given monthly cycle amongst couples that are actively engaged in an attempt to conceive, have the same genetic predisposition towards fecundity, and operate in similar socio-environmental conditions.

Accounting for a range of genetic and socio-environmental confounding factors that may affect the protogenesic interval, reproductive success, and the quality of offspring, the study attempts to isolate the random variations in the number of offspring across individuals. In particular, socio-environmental factors are accounted for by the concentration on a homogenous founder population in a single geographical location – Quebec – and by the further statistical control for parish fixed effects within the province and for year of birth fixed effects. Cultural, genetic and socio-economic factors that may affect lineage fecundity are accounted, i.e., by identifying the effect based on variations in PI among heads of dynasties who shared the same mother. Confounding socio-economic characteristics are further accounted for by controlling for the parental marriage age and literacy status. This comprehensive strategy permits the identification of the underlying causal effect of the number of offspring, as captured by the random component of PI, on long-run reproductive success.

The study exploits the demographic history of Quebec, using the reconstructed genealogy based on the entire parish registers of Quebec, covering 471,412 individuals from the beginning of the French colonization in the 17th century to the turn of the 19th century. More than 94% of these individuals were born and died in Quebec. The sample is restricted to the reproductive success of 3,798 heads of dynasties born in Quebec prior to the end of 1685 and died in the province thereafter, focusing on individuals whose PI is at least 38 weeks.

Results

The effect of PI on the number of descendants is depicted in Fig. 1. As expected, Fig. 1A shows a negative effect of PI on the number of children, reflecting the observation that *ceteris paribus*, a shorter protogenesic interval in the pre-demographic transition era increases the total number of children born in a family. Perhaps paradoxically, Fig. 1B depicts a hump-shaped effect of PI on the number of great-great-grandchildren. Heads of dynasties with an intermediate PI, and therefore an intermediate level of fertility, achieved the maximal number of great-great-grandchildren, despite having a smaller number of children relative to those with lowest PI.

The results of linear regression models are presented in Table 1, estimating the effect of PI on reproductive success while accounting for birth year and age of marriage (columns 1–4) as well as lineage fixed effects, literacy and gender (columns 5–8). The results can also be established using a GLM model with a negative binomial distribution and a logarithmic link function (supplementary materials Table S2). There exists a significantly negative effect of PI on the number of children (columns 1 and 5). An increase in PI by one year results in a reduction of approximately one child. The association is monotonically negative. In particular, including a quadratic transformation of PI generates a negative and insignificant coefficient estimate on the squared term in both specifications.

Meanwhile, there is a significant hump-shaped relationship between PI and the number of descendants in generation 2–4 (columns 2–4 and 6–8). In particular, the first-order effect of the quadratic expression is positive and significant and the second-order effect of the quadratic expression is negative and significant. Moreover, an additional test establishes a significant hump-shaped relationship (p=0.01) in the fourth generation accounting for lineage fixed effects, literacy and gender (column 8). The peak of the estimated quadratic equation occurs at PI of 1.25 years (i.e., approximately 65 weeks). Interestingly, the peak of the hump is located above the average and the median PI in the population (Table 1), suggesting that the forces of natural selection plausibly had a negative effect on average PI in the population over the time period under consideration. As established in the supplementary materials the results in Table 1 are robust to accounting for remarriages, spousal migration, birth and death parish, month of marriage, month of birth of the firstborn, firstborn status, within-lineage birth order, sample division based on gender, as well as a range of other re-specifications (supplementary materials sections S12–S13).

The research identifies several mechanisms that had contributed to the trade-off associated with higher fecundity and to the observed hump-shaped effect of fecundity on reproductive success in the long run. While individuals with lower fecundity had fewer children, the observed hump-shaped effect of fecundity on long-run reproductive success reflects the beneficial effect of lower fecundity on various measures of the quality of each child. In particular individuals with lower fecundity were more likely to have children that: (i) survived and got married, (ii) married at an earlier age, and (iii) were educated (in accordance with Klemp and Weisdorf, 2019). Thus, despite the positive effect of fecundity on the number of children, the adverse effect of fecundity on child quality and the reproductive success of each child, generated the observed hump shaped relationship between fecundity and long-run reproductive success.

The mechanisms are explored in Table 2. The fraction of children surviving to age 40 is indeed negatively related to PI (columns 1–2). In particular, accounting for lineage fixed effects, literacy and gender of the head of dynasty, there is a highly significant negative relationship between PI and the fraction of children surviving to the average marriage age. Furthermore, as established in column 3 and 4, the association between the average marriage age among offspring that got married and PI of heads of lineages is negative and highly statistically significant, accounting for the marriage age and birth year of heads of lineages, as well as for their signature status and gender. Finally, PI of heads of lineages has a highly significant positive association with the fraction of children who signed their marriage certificate (a proxy for literacy), accounting for the marriage age and birth year of heads of lineages, as well as for the confounding effects of the signature status, gender, and stoppage age of heads of lineages (columns 5–6). Moreover, literacy of heads of lineages has a highly significant positive effect on the literacy of their children. Hence, the positive effects of lower parental fecundity on the child quality is further amplified, generating higher reproductive success for subsequent generations. The results are robust to alternative specifications (supplementary materials section S4.5).

Conclusion

Evidence from preindustrial Quebec suggests that the forces of natural selection favored individuals characterized by moderate fecundity. While high fecundity was associated with a larger number of children, moderate fecundity maximized the number of descendants after several generations, reflecting the beneficial effect of lower fecundity on various measures of child quality.

Moreover, the analysis suggests that evolutionary forces decreased the level of fecundity in the population over this period. These findings are consistent with theories of the onset of the demographic transition and the transition of economies from stagnation to modern growth in which natural selection favored individuals with a larger predisposition towards child quality (Galor and Moav, 2002) and with the predictions of Unified Growth Theory (Galor, 2011).

References

Borgerhoff Mulder, M. (2000): "Optimizing offspring: the quantity–quality tradeoff in agropastoral Kipsigis," *Evolution and Human Behavior*, 21, 391–410.

Charnov, E. L. and S. K. M. Ernest (2006): "The offspring-size/clutch-size trade-off in mammals," *The American Naturalist*, 167, 578–582.

Christensen, K., H.-P. Kohler, O. Basso, J. Olsen, J. W. Vaupel, and J. L. Rodgers (2003): "The Correlation of Fecundability Among Twins: Evidence of a Genetic Effect on Fertility?" *Epidemiology*, 14, 60–64.

Cody, M. L. (1966): "A general theory of clutch size," *Evolution*, 174–184.

Galor, O. (2011): *Unified Growth Theory*, Princeton: Princeton University Press.

Galor, O. and O. Moav (2002): "Natural selection and the origin of economic growth," *Quarterly Journal of Economics*, 117, 1133–1191.

Gillespie, D., A. Russell, and V. Lummaa (2008): "When fecundity does not equal fitness: evidence of an offspring quantity versus quality trade-off in pre-industrial humans," *Proceedings of the Royal Society B: Biological Sciences*, 275, 713–722.

Harper, J., P. Lovell, and K. Moore (1970): "The shapes and sizes of seeds," *Annual review of ecology and systematics*, 1, 327–356.

Hill, K. and A. Hurtado (1996): *Ache life history: The ecology and demography of a foraging people*, Aldine de Gruyter.

Kaplan, H., J. Lancaster, S. Johnson, and J. Bock (1995): "Does observed fertility maximize fitness among New Mexican men?" *Human Nature*, 6, 325–360.

Klemp, M. and J. Weisdorf (2019): "Fecundity, Fertility and the Formation of Human Capital," *Economic Journal*, 129.

Kosova, G., M. Abney, and C. Ober (2009): "Heritability of reproductive fitness traits in a human population," *PNAS*, 107, 1772–1778.

Lack, D. et al. (1954): "The natural regulation of animal numbers." *The Natural Regulation of Animal Numbers.*

Lee, R. D. (1993): "Population dynamics: Equilibrium, disequilibrium, and consequences of fluctuations," in *Handbook of Population and Family Economics*, ed. by M. R. Rosenzweig and O. Stark, Elsevier, vol. 1, chap. 19, 1063–1115, 1 ed.

MEIJ, J., D. VAN BODEGOM, J. ZIEM, J. AMANKWA, A. POLDERMAN, T. KIRKWOOD, A. DE CRAEN, B. ZWAAN, AND R. WESTENDORP (2009): "Quality–quantity trade-off of human offspring under adverse environmental conditions," *Journal of evolutionary biology*, 22, 1014–1023.

PETTAY, J. E., L. E. B. KRUUK, J. JOKELA, AND V. LUMMAA (2005): "Heritability and genetic constraints of life-history trait evolution in preindustrial humans," *PNAS*, 102, 2838–2843.

RAMLAU-HANSEN, C. H., A. M. THULSTRUP, J. OLSEN, AND J. P. BONDE (2008): "Parental Subfecundity and Risk of Decreased Semen Quality in the Male Offspring: A Follow-up Study," *American Journal of Epidemiology*, 167, 1458–1464.

ROFF, D. (1992): *Evolution of life histories: theory and analysis*, Routledge, Chapman and Hall.

ROFF, D. A. (2002): *Life History Evolution*, Sunderland, MA: Sinauer Associates.

SALISBURY, E. ET AL. (1942): "The reproductive capacity of plants. Studies in quantitative biology." *The reproductive capacity of plants. Studies in quantitative biology.*

STEARNS, S. (1992): *The evolution of life histories*, vol. 248, Oxford University Press Oxford.

STRASSMANN, B. AND B. GILLESPIE (2002): "Life–history theory, fertility and reproductive success in humans," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269, 553–562.

WALKER, R. M., M. GURVEN, O. BURGER, AND M. J. HAMILTON (2008): "The trade-off between number and size of offspring in humans and other promates," *Proceedings of the Royal Society*, 275, 827–833.

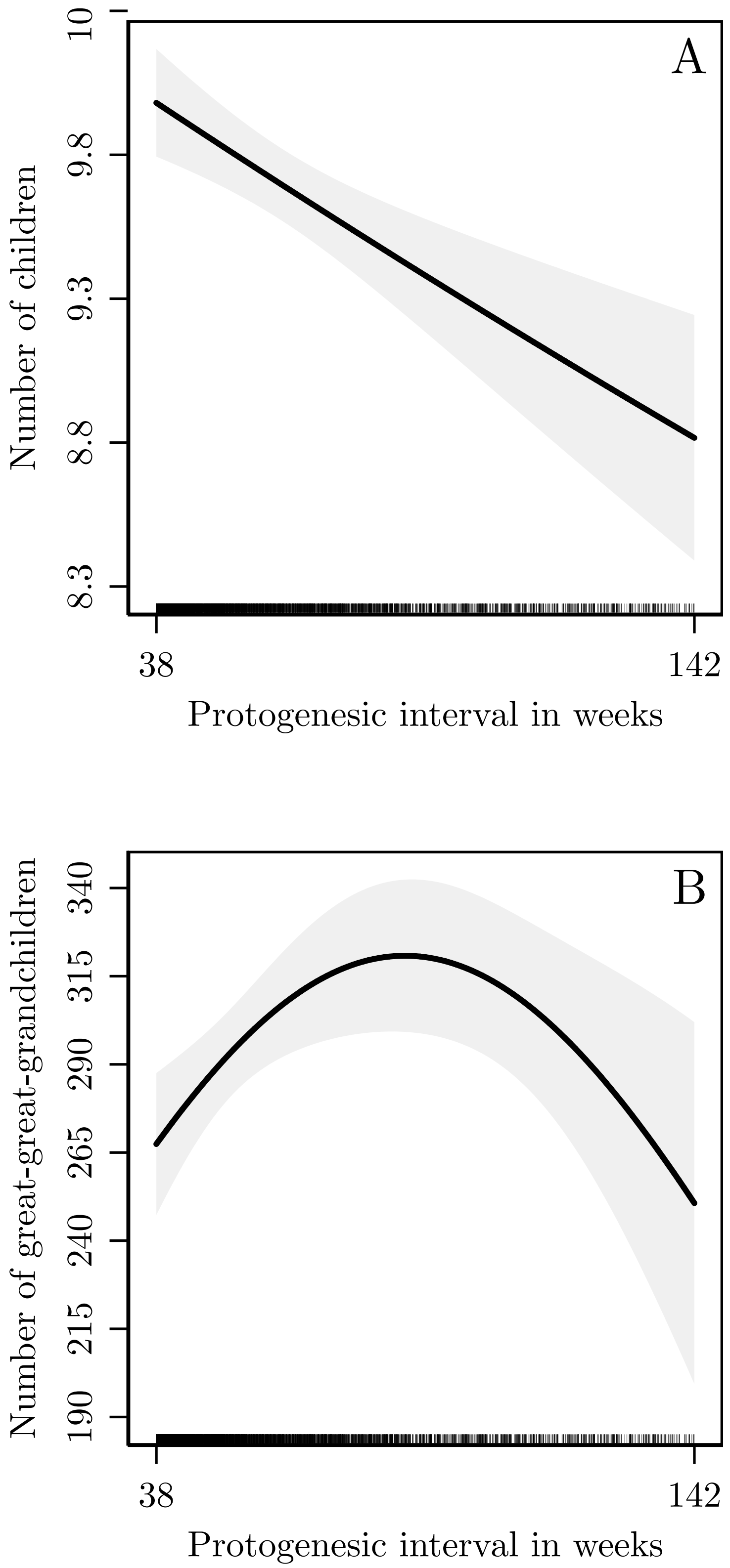


Fig. 1: Predicted number of children and great-great-grandchildren of 3,798 individuals as a function of the protogenesic interval, corresponding to columns 5 and 8 of Table S3. (**A**) Number of children. (**B**) Number of great-great-grandchildren. Rugs at bottom of panels represent the distribution of observations.

Table 1: The association between the PI and number of descendants for heads of lineages born prior to the end of 1685

	Log number of descendants in:							
	Gen. 1	Gen. 2	Gen. 3	Gen. 4	Gen. 1	Gen. 2	Gen. 3	Gen. 4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Protogenesic interval	-.080***	.137	.509**	.830***	-.077***	.208	.535***	.810***
	(.024)	(.180)	(.232)	(.311)	(.011)	(.130)	(.181)	(.258)
Protogenesic interval squared		-.054	-.174**	-.290***		-.089**	-.210***	-.325***
		(.058)	(.075)	(.099)		(.042)	(.059)	(.084)
Proper signature					-.027*	.044	.125***	.109*
					(.014)	(.032)	(.046)	(.066)
Male					-.028*	.025	.085*	.036
					(.015)	(.031)	(.043)	(.063)
Maternal Founder fixed effects	No	No	No	No	Yes	Yes	Yes	Yes
Stoppage age fixed effects	No	No	No	No	Yes	Yes	Yes	Yes
Number of observations	3,798	3,798	3,798	3,798	3,798	3,798	3,798	3,798
Adjusted R^2	.018	.013	.026	.301	.799	.442	.296	.355
Joint sign.-level of PI & PI^2	.001	.439	.055	.007	.000	.002	.000	.000
Maximizing PI		1.256	1.46	1.431		1.163	1.272	1.247

Table 1: This table presents the results of a series of OLS regressions of the number of descendants in generation t on the protogenesic interval measured in years, i.e. PI and PI^2 for heads of lineages born prior to the end of 1685. Birth year and marriage age dummy variables are included as controls. The specifications in column 5–8 account for lineage fixed effects, meaning that the estimation is based on variation across siblings, as opposed to across the population as a whole. Furthermore, signature status, sex, and stoppage age dummy variables are included in columns 5–8. A dummy variable indicating unknown signature status is also included in the regressions underlying column 5–8. Supplementary materials section 2 provide further information. Standard errors clustered at the level of the firstborn are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: The effect of the PI on offspring quality

	Fraction of children surviving to age 40 who got married		Average marriage age of children		Fraction of children with proper signature	
	(1)	(2)	(3)	(4)	(5)	(6)
Protogenesic interval	.299***	.232**	-.430***	-.339**	.401***	.337***
	(.113)	(.110)	(.005)	(.023)	(.090)	(.091)
Proper signature		.779***		-.705***		1.305***
		(.109)		(.000)		(.095)
Male		.365***		-.720***		.407***
		(.120)		(.000)		(.098)
Stoppage age fixed effects	No	Yes	No	Yes	No	Yes
Number of observations	3,727	3,727	3,727	3,727	3,727	3,727

Table 2: Column 1–2 presents the results of fractional logit regressions of the fraction of children, i.e., individuals in the first generation, surviving to age 40 that got married on the protogenesic interval measured in years, i.e., PI, for heads of lineages with at least one child surviving to age 40. Column 3–4 presents the results of OLS regressions of the average marriage age of children, i.e., individuals in the first generation, on the protogenesic interval measured in years, i.e., PI. Column 3–4 presents the results of fractional logit regressions of the share of children, i.e., individuals in the first generation, who signed their marriage certificate with a proper signature (a proxy for literacy) on the protogenesic interval measured in years, i.e., PI, for heads of lineages with at least one surviving child with observed signature status. Birth year and marriage age dummy variable are included as controls in all specifications. Furthermore, dummy variables indicating signature status, gender, and stoppage age (in years) are included in column 2, 4, and 6. A dummy variable indicating unknown signature type is included in the regressions underlying column 2, 4, and 6. Supplementary materials section 2 establish that the results are robust to alternative specifications and provide further information. Standard errors clustered at the level of the firstborn are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Supplementary Information

Human Genealogical Data Suggests a Selective Advantage to Moderate Fecundity

Contents

S1 Theoretical Model

This section presents a theoretical model that explores the effect of the level of fecundity of members of a founding generation (i.e., individuals who settled new territories where the carrying capacity of the new environment was an order of magnitude greater than the size of the founder population) on their long-run reproductive success.

S1.1 The Basic Structure of the Model

Consider an overlapping-generations society originated in a newly acquired (sparsely populated) territory in which the economic environment could support fertility rates at the maximal feasible physiological rate. Individuals live for two periods: childhood and adulthood. In the first period of life, childhood, individuals are passive economic agents, and their consumption needs are provided by their single parent. In the second period of life, adulthood, individuals work and allocate their income between their own consumption and child rearing.

Preferences of a member i of generation t are represented by a utility function, u_t^i, defined over the individual's consumption, c_t^i, the individual's number of children, n_t^i, and the children's level of human capital, h_{t+1}^i (i.e., when they are engaged in production in the subsequent period as adults).[1] Specifically,

$$u_t^i = (1-\gamma)\ln c_t^i + \gamma[\ln n_t^i + \beta \ln h_{t+1}^i], \tag{1}$$

where $\beta \in (0,1)$ captures the individual's preference for quality, and $\gamma \in (0,1)$ captures the individual's preference for for children relative to consumption.

The income of a member i of generation t, y_t^i, reflects the individual's human capital, h_t^i, and the productivity, w, of each efficiency unit of labor.[2] In particular,

$$y_t^i = wh_t^i. \tag{2}$$

The income of a member i of generation t, y_t^i, is divided between consumption, c_t^i, and expenditure on children. The household's budget constraint is

$$n_t^i(\tau + q_t^i) + c_t^i \leq y_t, \tag{3}$$

where τ is the minimum amount of parental resources required to raise a child, and $\tau + q_t^i$ is the cost of raising a child with quality q_t^i.

---

[1] For simplicity, parents derive utility from the expected number of surviving offspring, and the parental cost of child rearing is associated only with surviving children. The introduction of costs associated with non-surviving children, or risk aversion, would not affect the qualitative features of the theory.

[2] In line with the evolution of wages over the sample period, wages are assumed to be constant over time. The introduction of a rising productivity would not alter the qualitative results.

The level of human capital of a child of a member i of generation t, h^i_{t+1}, is an increasing and strictly concave function of the parental investment in the education of the child, q^i_t :

$$h^i_{t+1} = h(q^i_t), \tag{4}$$

where $h'(q^i_t) > 0$, $h''(q^i_t) < 0, \lim_{q_t \to 0} h'(q^i_t) < \infty$, $\lim_{q_t \to \infty} h'(q^i_t) = 0$, $h''(q^i_t)q^i_t + h'(q^i_t) < 0$, and $h(0) = 1$ (i.e., in the absence of investment in quality, each individual has a one efficiency unit of labor).[3]

S1.2 Optimization

A member i of generation t allocates resources between consumption and child rearing in order to maximize utility.

S1.2.1 Unconstrained Optimization

In the absence of additional constriants, given the homotheticity of preferences, a member i of generation t allocates a fraction $(1-\gamma)$ of income to consumption and a fraction γ to to child rearing. Hence,

$$\begin{aligned} c^i_t &= (1-\gamma) y^i_t; \\ (\tau + q^i_t) n^i_t &= \gamma y^i_t. \end{aligned} \tag{5}$$

Moreover, applying the Kuhn-Tucker conditions, the level of quality of the children of a member i of generation t, $q^i_t \geq 0$, will be determined such that

$$\beta h'(q^i_t)(\tau + q^i_t) - h(q^i_t) \quad \begin{cases} \leq 0 & \text{if} \quad q^i_t = 0 \\ = 0 & \text{if} \quad q^i_t > 0. \end{cases} \tag{6}$$

The maximal number of children, $[n^i_t]^{\max}$, that a member i of generation t can raise is, therefore,

$$[n^i_t]^{\max} = \frac{\gamma y^i_t}{\tau} \equiv n^{\max}(y^i_t). \tag{7}$$

This maximum will be reached only if the individual does not invest in the quality of children (i.e., $q^i_t = 0$).

S1.2.2 Constrained Optimization

Given income, y^i_t, each member i of generation t will determine an age of marriage and, therefore, an onset of the reproductive period so as to achieve a desired fertility level, n^i_t, that would not exceed $n^{\max}(y^i_t)$. Suppose, however, that the level of fertility of a member i of generation t is affected by the

---

[3] These conditions are satisfied if, for instance, $h(q^i_t) = (1 + q^i_t)^{\alpha}$, for $\alpha \in (0,1)$.

randomness associated with conception, and that the actual fertility of the individual therefore may deviate from the desired level. In particular, suppose that, in the absence of reliable contraceptive methods or fertility enhancing drugs, nature dictates to each individual the number of children that will be born within the feasibility constraint, i.e.,

$$n_t^i \in (0, n^{\max}(y_t^i)], \tag{8}$$

where the number of children is assumed to be strictly positive.[4]

Since preferences are homothetic, the allocation of the individual's resources between consumption and children will not be altered.[5] However, the division of resources between the quantity and quality of each child will be affected by this random assignment of the number of children. A lower number of children will generate higher child quality.

S1.3 The Founder Population

Members of the founder population (generation 0) are assumed to be identical in their level of human capital, h_0, and therefore in their income, y_0. Their desired level of fertility, therefore, is identical as well. Nevertheless, due to the randomness associated with conception, the number of children may differ across members of generation 0. In particular, the number of children of a member i of generation 0 is randomly drawn from the interval $(0, n^{\max}(y_0)]$, i.e.,

$$n_0^i \in (0, n^{\max}(y_0)]. \tag{9}$$

Consider a member i of generation 0 that draws n_0^i children. In light of the budget constraint, the optimization of individual i, given by (5), implies that the quality of individual i's children is given by

$$q_0^i = \frac{\gamma y_0}{n_0^i} - \tau, \tag{10}$$

where $\partial q_0^i / \partial n_0^i = -\gamma y_0 / \left[n_0^i\right]^2 < 0$. Thus, the quality of the children of member i of generation 0 is declining in the number of children.

S1.4 Subsequent Generations

Members i of generation $t = 1, 2, 3, \ldots, T$ (i.e., the descendants of member i of generation 0) generate an income, y_t^i,

$$y_t^i = wh(q_{t-1}^i), \tag{11}$$

---

[4] In light of the possibility that individuals engaged in within-marriage fertility control, the source of exogenous variation in the number of children in the empirical analysis is based on the random component of the variation in the time from the onset of unprotected intercourse as marked by the date of marriage to the first birth. Thus, the source of exogenous variation in the number of children exploited in the empirical analysis is entirely related to the random component of the variation in the timing of first birth and not to subsequent birth intervals.

[5] Given income, the individual determines the marriage age so as to assure that, given the gestation period of a child, the number of children that could be born during the remaining reproductive period could not exceed $n^{\max}(y_t^i)$. The qualitative results will remain unchanged if alternatively individuals are subjected to a non-binding subsistence consumption constraint, $\tilde{c}$, and the the value of $n^{\max}(y_0)$ could be $[y_0 - \tilde{c}]/\tau$, independent of preferences.

reflecting the parental investment in their human capital, q^i_{t-1}.

Suppose, for simplicity, that a member i of generation t can obtain their desired number of children for all $t \geq 1$.[6] The household's optimization implies, therefore, that parental consumption $c^i_t = (1-\gamma)y^i_t$, the optimal level of children, n^i_t, and their quality, q^i_t, are obtained by the unconstrained optimization (5) and (6).

Lemma S1.1. *The level of investment in child quality,* q^i_t, *and a level of fertility,* n^i_t, *of members* i *of generation* $t = 1, 2, 3, \dots, T$ *is such that*

$$\begin{aligned} q^i_t &= \begin{cases} 0 & \text{if} \quad h'(0) \leq 1/\beta\tau \\ q^* > 0 & \text{if} \quad h'(0) > 1/\beta\tau; \end{cases} \\ n^i_t &= \begin{cases} \frac{\gamma y^i_t}{\tau} & \text{if} \quad h'(0) \leq 1/\beta\tau \\ \frac{\gamma y^i_t}{\tau + q^*} & \text{if} \quad h'(0) > 1/\beta\tau, \end{cases} \end{aligned} \tag{12}$$

where $\frac{\partial q^*}{\partial \beta} > 0$, $\frac{\partial q^*}{\partial \tau} > 0$, $\frac{\partial n^i_t}{\partial \beta} < 0$, $\frac{\partial n^i_t}{\partial \tau} < 0$, *and* $\frac{\partial n^i_t}{\partial y} > 0$.

Proof. As follows from (6), the household's optimization with respect to $q^i_t \geq$ is given by the implicit function

$$G(q^i_t) \equiv \beta h'(q^i_t)(\tau + q^i_t) - h(q^i_t) \quad \begin{cases} \leq 0 & \text{if} \quad q^i_t = 0 \\ \\ = 0 & \text{if} \quad q^i_t > 0, \end{cases} \tag{13}$$

where, as follows from the strict concavity of $h(q^i_t)$.

$$\frac{\partial G(q^i_t)}{\partial q^i_t} = \beta h''(q^i_t)(\tau + q^i_t) - (1-\beta)h'(q_t) < 0. \tag{14}$$

Moreover, since $h(0) = 1$, $\beta \in (0,1)$, and $h(q^i_t) - h'(q^i_t)q^i_t > 0$,

$$\begin{aligned} &\lim_{q^i_t \to 0} G(q^i_t) = \tau\beta h'(q^i_t) - 1 \begin{cases} \leq 0 & \text{if} \quad h'(0) \leq 1/\beta\tau \\ > 0 & \text{if} \quad h'(0) > 1/\beta\tau; \end{cases} \\ &\lim_{q^i_t \to \infty} G(q^i_t) < 0. \end{aligned} \tag{15}$$

Hence, it follows from the *Intermediate Value Theorem* that, if $h'(0) > 1/\beta\tau$, there exists a unique, interior value of $q^i_t = q^*$ such that $h(q^*) = \beta h'(q^*)(\tau + q^*)$ and a boundary solution $q^i_t = 0$ if $h'(0) \leq 1/\beta\tau$. The level of fertility, n^i_t, is given by (5). □

The interior unconstrained optimal level of investment in human capital, q^*, reflects the costs and the benefits of this investment and is, therefore, independent of the parental income. Hence,

---

[6] This assumption permits the analysis to focus on the effect of fecundity of the founder generation on long-run reproductive success. Moreover, this is a good approximation for the actual fertility if the offspring generation is sufficiently large, as is the case in our sample where the average fertility rate in nearly 10. In particular, if there is a continuum of children and the realization at the individual level is drawn from a normal distribution with a mean at the desirable level of fertility, the law of large numbers will assure that on average the desired level of fertility is realized in the offspring generation. Clearly, a distribution of fertility may affect the level of average human capital but this is a second-order effect.

as long as parental income is constant across generations and there are no changes in the cost and the benefits associated with this investment, this level remains constant across generations. In particular, as follow from Lemma 1, the unconstrained level of investment in human capital of a member i of generation t is independent of the individual level of income for all $t \geq 1$.[7]

Hence, the level of human capital, h_t^i, of each member i of generation $t \geq 2$ is equal within and across generations:

$$h_t^i = h(q_{t-1}^i) = h(q^*), \tag{16}$$

and the level of income of each member i of generation $t \geq 2$ is equal within and across generations:

$$y_t^i = wh(q_{t-1}^i) = wh(q^*) = y^*, \qquad \text{for all } t \geq 2.$$

Corollary S1.2.

- *The fertility level of a member i of generation 1 is*

$$n_1^i = \frac{\gamma y_1^i}{\tau + q^*} = \frac{\gamma wh(q_0^i)}{\tau + q^*} \equiv n_1(n_0^i),$$

  where $q_0^i = \frac{\gamma y_0}{n_0^i} - \tau$.

- *The fertility level of member i of generation $t \geq 2$ is*

$$n_t^i = \frac{\gamma y_t^i}{\tau + q^*} = \frac{\gamma y^*}{\tau + q^*} = n^*,$$

  where $y^* = wh(q^*)$.

As established in the subsection "Extension of Model: Intergenerational Transmission of Human Capital" below, if parental human capital has a direct positive spillover on the level of the human capital of the child, the level of investment in human capital will increase monotonically and will gradually converge to a steady-state, $\bar{q}$, as long as the complementarity between parental investment in human capital and the spillover exceeds the direct effect of the spillover.

S1.5 Long-run Reproductive Success

The long-run reproductive success of a member i of generation 0 is affected by the the initial fertility draw, n_0^i, and the implied level of investment in the quality of each child, q_0^i. In particular, the number of descendants that individual i will have in generation T, D_T^i, is

$$D_T^i = n_0^i n_1^i n_2^i n_3^i ... n_T^i \equiv \prod_{j=0}^{T} n_j^i, \tag{17}$$

---

[7] Since generation 0 faces a constrained optimization, the level of investment in human capital of this generation is affected by income, as established in (10).

where n_j^i is the number of children that descendants of member of generation i are expected to have in period $j = 1, 2, \ldots, T$. Thus, as follows from Corollary 1,

$$D_T^i = (n^*)^{T-2} n_1^i n_0^i. \tag{18}$$

Nature will give an evolutionary advantage to a member i of generation 0 whose level of fecundity maximizes long-run reproductive success, D_T^i. Hence, the level of fecundity, n_0^i, that would be favored by the forces of natural selection is such that

$$\begin{aligned} n_0^i &= \arg\max D_T^i \qquad (19)\\ s.t. \quad n_0^i &\in \left[0, \left[n_t^i\right]^{\max}\right]. \qquad (20) \end{aligned}$$

As follows from (18),

$$\frac{\partial D_T^i}{\partial n_0^i} = (n^*)^{T-2} \left[n_1^i + n_0^i \frac{\partial n_1^i}{\partial n_0^i} \right] = (n^*)^{T-2} n_1^i [1 - \eta_{n_1,n_0}], \tag{21}$$

where the elasticity of fertility in period 1 with respect to fertility in period 0, η_{n_1,n_0}, is

$$\eta_{n_1,n_0} \equiv -\frac{\partial n_1^i}{\partial n_0^i} \frac{n_0^i}{n_1^i} = \frac{(\tau + q_0^i) h'\left[q_0^i\right]}{h\left[q_0^i\right]}. \tag{22}$$

Moreover, given the properties of $h\left[q_0^i\right]$ imposed in (4) and the negativity of $\partial q_0^i / \partial n_0^i$ established in (10),

$$\frac{\partial \eta_{n_1,n_0}}{\partial n_0^i} = \frac{\left((\tau + q_0^i) h''\left[q_0^i\right] + h'\left[q_0^i\right]\right) h\left[q_0^i\right] - (\tau + q_0^i)\left(h'\left[q_0^i\right]\right)^2}{\left(h\left[q_0^i\right]\right)^2} \frac{\partial q_0^i}{\partial n_0^i} > 0. \tag{23}$$

Hence, as follows from (21), $\partial^2 D_T^i / \partial \left[n_0^i\right]^2 < 0$, and D_T^i is strictly concave in n_0^i.

The optimal level of fecundity in period 0, n_0^*, is therefore[8]

$$\begin{array}{lclcl} n_0^* \in \left(0, \left[n_0^i\right]^{\max}\right) & \iff & \frac{\partial D_T^i}{\partial n_0^i}(n_0^*) = 0 & \iff & \eta_{n_1,n_0}(n_0^*) = 1; \\ n_0^* = [n_0]^{\max} & \iff & \frac{\partial D_T^i}{\partial n_0^i}\left([n_0]^{\max}\right) \geq 0 & \iff & \eta_{n_1,n_0}\left([n_0]^{\max}\right) \leq 1. \end{array} \tag{24}$$

Hence, if over some of the feasible range of fertility in period 0, $\left[0, [n_0]^{\max}\right)$, the elasticity of fertility in period 1 with respect to fertility in period 0 is greater than 1 in absolute value (i.e., if a one percent increase in fertility in period 0 reduces the quality, and therefore the fertility, of the descendant in period 1 by more than one percent), an intermediate level of fecundity $n_0^* \in \left(0, [n_0]^{\max}\right)$ will maximize long-run reproductive success. However, if the elasticity is smaller than

---

[8] As established in the subsection "Extension of Model: Intergenerational Transmission of Human Capital. Suppose" below, if parental human capital has a direct positive spillover on the level of the human capital of the child, the optimization would depend on the effect of the fecundity of the first generation on all subsequent generation, rather than the next generation only, but the qualitative results would remain intact.

1 for the largest feasible level of fertility, $[n_0]^{\max}$, the level of fecundity that maximizes reproductive success will be the highest feasible one, $[n_0]^{\max}$.

In particular, it follows from (22) and (24) that the highest level of fecundity will maximize long-run reproductive success if and only if $\eta_{n_1,n_0}\big([n_0]^{\max}\big) \leq 1$, which holds if and only if

$$h'[\tilde{q}] \leq \frac{h[\tilde{q}]}{(\tau + \tilde{q})}, \tag{25}$$

where $\tilde{q} = q(([n_0]^{\max})$.

Noting that $q_0^i = 0$ when $n_0^i = [n_0]^{\max}$ and, therefore, $h[\tilde{q}] = h(0) = 1$, it follows that the highest level of fecundity will maximize long-run reproductive success if and only if

$$h'[0] \leq \frac{1}{\tau}. \tag{26}$$

Proposition S1.3. *The level of fecundity of the founding generation that maximizes reproductive success is*

- *The maximal level,*
  $$[n_0]^{\max} = \frac{\gamma y_0}{\tau},$$
  if the elasticity of fertility in period 1 *with respect to fertility in period* 0 *evaluated at the highest feasible fertility,* $\eta_{n_1,n_0}\big([n_0]^{\max}\big)$, *is weakly less than* 1 *in absolute value, i.e., if*
  $$h'[0] \leq \frac{1}{\tau}.$$
- *An intermediate level,*
  $$n_0^* \in \big(0, [n_0^i]^{\max}\big),$$
  if the elasticity of fertility in period 1 *with respect to fertility in period* 0 *evaluated at the highest feasible fertility,* $\eta_{n_1,n_0}\big([n_0]^{\max}\big)$, *is greater than* 1 *in absolute value, i.e., if*
  $$h'[0] > \frac{1}{\tau}.$$

Proof. The proposition is a corollary of (24) and (26). □

S1.6 Theoretical Insights

The theory establishes that, in contrast to a prevailing perception, the presence of a static trade-off between the quantity and quality of offspring is merely a necessary but not a sufficient condition for the presence of a negative effect of fecundity on reproductive success in the long run. In particular, it demonstrates that individuals with the highest level of fecundity could have had the largest reproductive success if, for any feasible number of children, an additional child would have contributed directly to the number of offspring in the long-run more than the adverse indirect effect of this additional child on the quality of his siblings and hence their long-run reproductive success.

The theory suggests that if investment in child quality is sufficiently productive in terms of its effect on the potential income and thus the reproductive success of the child (i.e., if the elasticity of fertility of an offspring with respect to the fertility of the founding parent, evaluated at the highest feasible level of fertility) is greater than one in absolute value, then an intermediate level of fecundity generates the highest reproductive success and will therefore be favored by natural selection.

S1.7 Extension of Model: Intergenerational Transmission of Human Capital

Suppose that parental human capital has a direct positive spillover on the level of the human capital of the child. In particular, the level of human capital of a child of a member i of generation t, h^i_{t+1}, is an increasing strictly concave function of both the parental investment in the education of the child, q^i_t, as well as the parental level of human capital.

$$h^i_{t+1} = h(q^i_t, h^i_t), \tag{27}$$

where for $j = 1, 2$, $h_j(q^i_t, h^i_t) > 0$, $h_{jj}(q^i_t, h^i_t) < 0, \lim_{q_t \to 0} h_1(q^i_t, h^i_t) < \infty$, $\lim_{q_t \to \infty} h_j(q^i_t, h^i_t) = 0$, $h_{11}(q^i_t, h^i_t)q^i_t + h_1(q^i_t, h^i_t) < 0$, and $h(0, 1) = 1$.

Suppose that $h_{12}(q^i_t, h^i_t) > 0$, (i.e., parental human capital increases the productivity of parental investment in the human capital of each child), and suppose that this complementarity exceed the direct spillover from parental human capital to the human capital of the child, i.e.,

$$h_{12}(q^i_t, h^i_t) > h_2(q^i_t, h^i_t). \tag{28}$$

Modifying (13) accordingly, the interior optimal level of investment, q^i_t, is determined by

$$G(q^i_t, h^i_t) \equiv \beta h_1(q^i_t, h^i_t)(\tau + q^i_t) - h(q^i_t, h^i_t) = 0, \tag{29}$$

where $\partial G(q^i_t, h^i_t)/\partial q^i_t = (1 - \beta)h'(q^i_t) - \beta h''(q^i_t)(\tau + q^i_t) < 0$. Hence, it follows from the Implicit Function Theorem that there exist a single-valued function $q^i_t = \xi(h^i_t)$, such that[9]

$$\xi'(h^i_t) = -\frac{\beta h_{12}(q^i_t, h^i_t)(\tau + q^i_t) - h_1(q^i_t, h^i_t)}{\partial G(q^i_t, h^i_t)/\partial q^i_t} > 0. \tag{30}$$

The evolution of human capital within lineage i, is given by the the sequence $\{h^i_t\}_{t=0}^{\infty}$, such that

$$h^i_{t+1} = h(q^i_t, h^i_t) = h\big(\xi(h^i_t), h^i_t\big) \equiv \phi(h^i_t), \tag{31}$$

where $\phi(0) > 0$ and $\phi(h^i_t)$ is strictly concave in h^i_t. Hence, the level of human capital increases monotonically and it converges to a steady-state equilibrium,

$$\bar{h} = h(\bar{q}, \bar{h}), \tag{32}$$

---

[9] It is further assumed that $\xi''(h^i_t) < 0$ and $\xi'(h^i_t) > -2h_{12}(q^i_t, h^i_t)/h_{22}(q^i_t, h^i_t)$.

in which investment in human capital, $\bar{q}$, coupled with parental externality, $\bar{h}$, generate a level of human capital $\bar{h}$.

The level of fertility of a member i of generation t is, accordingly, along the transition to this steady-state equilibrium, therefore

$$n_t^i = \frac{\gamma y_t^i}{\tau + q_t} = \frac{\gamma w h_t^i}{\tau + \xi'(h_t^i)} = \psi\big(h_0^i(n_0^i)\big), \tag{33}$$

where $h_t^i = \phi(h_t^i) = \phi^{\{T\}}(h_0^i)$

Hence, the level of fecundity of the founding generation, n_0^*, that would maximize reproductive success in period T will be based on the the effect of n_0^* on the level of fertility in each of the subsequent T generations. Moreover, if $h'[0]$ is sufficiently small, the maximal level of fecundity will be favored by the forces of natural selection.

S2 Supplementary Text

The analysis focuses on an important heritable character determining fertility namely fecundity.[10] In order to investigate the hypothesis that an intermediate level of fecundity maximizes long-run reproductive success, the research isolates random variation in the time interval between the date of first marriage and the first birth, i,e, the protogenesic interval (PI), to infer the general effect of PI, and therefore fecundity, on long-run reproductive success.[11]

S2.1 Obstacles in Identifying Intergenerational Trade-offs in Reproductive Success

Two major obstacles affect the identification of intergenerational trade-offs in reproductive success in existing studies. First, omitted correlates of offspring quality may also be correlated with offspring quantity, obscuring the presence or the absence of a quantity-quality trade-off. For instance, if variations in the access to resources across individuals, enables individuals to produce more offspring of higher quality, failing to account for the effect of resources will obscure the presence of a trade-off between quantity and quality of offspring, and an observed positive relationship between the number of children and that of grandchildren may misleadingly be interpreted as indicative of the lack of a trade-off in reproductive success. Similarly, if variations in education across individuals affect parental earning capacity, enabling individuals to produce more offspring of higher quality, failing to account for parental education will obscure the presence of a trade-off between quantity and quality of offspring.

Second, reverse causality from offspring's quality to the aggregate quantity of offspring may obscure the presence or the absence of a quantity-quality trade-off. For instance, some form of reverse causality may generate a negative relationship between quantity and quality of offspring that may be falsely interpreted as indicative of the presence of an intergenerational trade-off between quantity and quality. In particular, the adverse effect of low offspring quality on offspring survival rate may contribute to the total number of offspring born (via the child replacement channel), generating a negative correlation between the quantity and the quality of offspring that has no bearing on the presence or the absence of an intertemporal trade-off in reproductive success.

This research mitigates these major hurdles by focusing on the effect of fecundity, rather than fertility, on reproductive success. Furthermore, it designs an empirical strategy that exploits the inherent uncertainty in the process of human reproduction to identify the effect of fecundity on reproductive success. In particular, in light of the social norm observed in pre-industrial Quebec, in which marriage marked the intention to conceive, the research exploits variation in the random

---

[10] The findings presented here could generalize to other factors co-determining intergenerational transmission of fertility, such as social transmission of demographic behavior (see e.g., Austerlitz and Heyer (1998); Gagnon and Heyer (2001)).

[11] The PI was correlated across generations, although insignificantly, in an earlier analysis of a subset of the present data (see Desjardins et al. (1991)). As is established in section S5, using both animal models as well as parent-offspring regressions, PI is heritable in the present population. Furthermore, see Christensen et al. (2003); Pettay et al. (2005); Ramlau-Hansen et al. (2008); Kosova et al. (2009). It should also be noted that in contrast to the PI, inter-birth intervals are largely determined by deliberate behavior, rendering inter-birth intervals unsuitable as sources of variation in family size in the present analysis Cinirella et al. (2017).

component of the time interval between the date of first marriage and the first birth to capture the effect of fecundity on long-run reproductive success.[12]

S2.2 Protogenesic Interval

As depicted in Figure S1, a marriage over this period signaled a deliberate attempt to conceive. A sharp spike in birth rates occurs starting in the 35th week after marriage and nearly a third of births occurs within the 36–44 weeks time interval. Furthermore, premarital conception is insignificant, reflecting possibly an adherence to the social and religious norms existing at the time. In particular, only 7.9 percent of the births over this period occurred within 35 weeks of marriage, and the incidence of premature births suggests that even this small fraction overstates the share of babies conceived prior to marriage.[13] These proportions conforms to those from other studies, finding that just around five percent of women married in Canada were pregnant before marriage, a very low proportion and comparable to that of France over the same time period Charbonneau (1993); Charbonneau et al. (2000). Furthermore, the existence of births reported within 35 weeks of marriage reassuringly rules out the possibility that social taboos prevented the recording of pre-nuptially conceived children.

Since fecundity reflects genetic and socio-environmental factors, the protogenesic interval is affected by genetic predisposition, socio-environmental conditions, as well as the realization of random elements that affect conception. Accounting for a range of genetic and socio-environmental confounding factors that may affect PI, reproductive success, and the quality of offspring, the study attempts to isolate the effect of the random variations in PI across individuals. In particular, genetic, as well as cultural and socio-economic factors that may affect fecundity are accounted for by the inclusion of Maternal Founder fixed effects. Namely, the effect of fecundity on reproductive success is identified based on variations in reproductive success among siblings, capturing the similarities in the genetic predisposition of these genetically linked individuals, as well as their cultural and socio-economic proximity.

Additional confounding variations between siblings are accounted for by the inclusion of indicator (dummy) variables for their marriage age, birth year, gender, and a proxy for literacy. Furthermore, additional control variables include the geographic location at birth and death, birth order, month of birth of the head of lineage and their firstborn, number of marriages, and immigration and emigration status of the spouse.

---

[12] In addition, while, in accordance with the hypothesis, the analysis is conducted using data from a pre-industrial and pre-demographic transition era where the Malthusian constraint is binding, the study population is characterized by being a founder population. This has two major advantages. First, for the majority of the existence of anatomically modern humans and their migration out of Africa, humans have been continuously conquering and settling new territory. Thus, the present study population yields a high level of external validity for extrapolation of the findings to the conditions of human populations throughout human existence. Second, the constraint of available land, important for aggregate production, is presumably less binding for this founder population than for populations that has had time to reach a final Malthusian steady state. This will tend to bias our analysis *against* finding evidence of an inter-generational reproductive trade-off compared to a similar analysis conducted on a population subject to a more binding Malthusian constraint.

[13] In the sample of all 59,238 mothers, 3.2 percent of births occurred prior to the marriage date, 5.5 percent of the births occurred after two years and 38 weeks of marriage (i.e., two years after first conception), and 1.6 percent of births occurred within one week of marriage (reflecting possibly a tendency of mothers who gave birth before marriage to baptize their firstborn at or shortly after their wedding date).

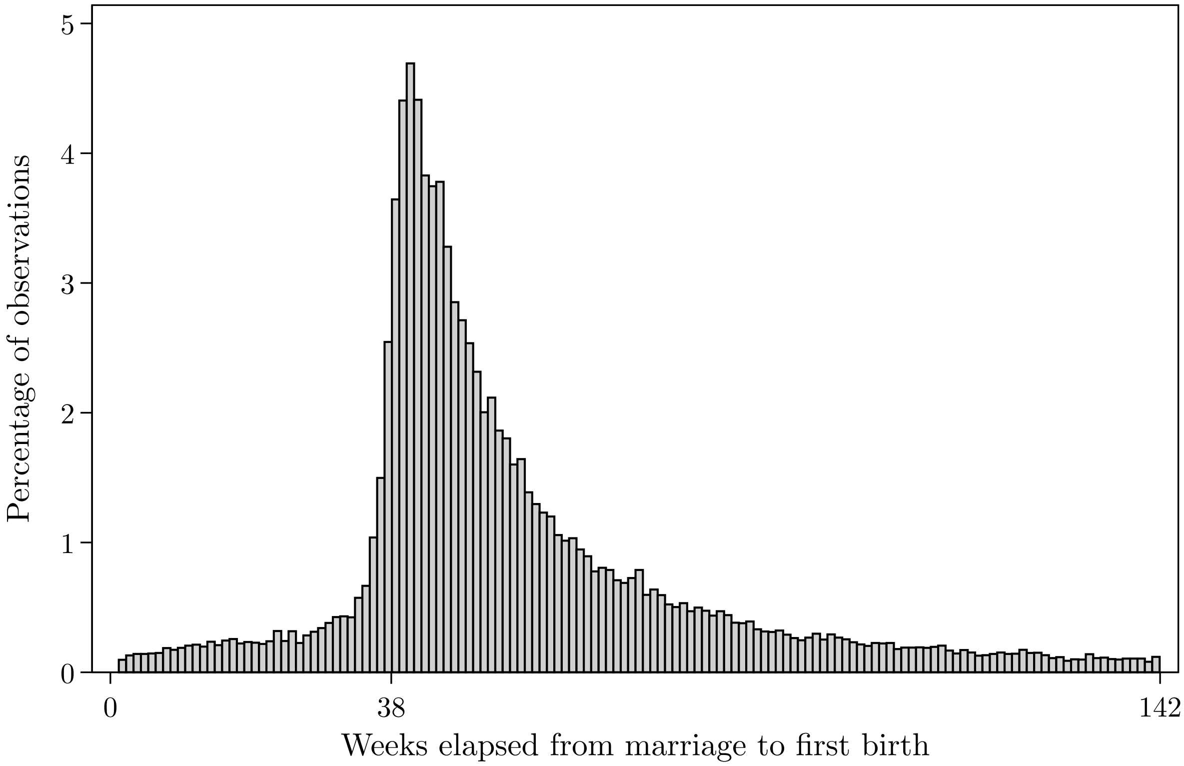


Fig. S1: PI (in weeks) of 53,154 mothers in Quebec in the 17th–18th century who gave birth after the 7th day of their marriage date.

S2.3 Determinants of the Protogenesic Interval

We here summarize the underlying characteristics and determinants of the key independent variable: the length of the protogenesic interval. We establish (i) that the protogenesic interval is a meaningful proxy for parental fecundity; (ii) that, according to existing evidence from historical data, the protogenesic interval is unaffected by potentially confounding factors such as parental socio-economic characteristics including skills and wealth; (iii) that studies of modern data also suggest that the protogenesic interval is not affected by unobserved confounding factors.

S2.3.1 The Distribution of PI and its Relation to Fecundity

Imagine a population consisting of equally-fecund couples who have unprotected sexual intercourse on a regular and identical basis. Despite identical levels of fecundity, these couples will have widely different times to conception. This means that the process of conception is randomized with a large variation. Indeed, in a homogeneous population, the protogenesic intervals follow a geometric distribution. Fecundity is measured by *fecundability* – the probability of achieving conception in a given monthly cycle Gini (1924). If $p \in (0, 1)$ denotes fecundability, then, in light of the geometric distribution, the coefficient of variation of the time to conception will be $p^{-1/2} > 1$. With a monthly probability of conception of 10%, the standard deviation of the time to conception in a homogeneous population will therefore be more than 3.16 times the average time to conception.

In a heterogeneous population, on the other hand, the proportion of the variance that can be attributed to differences in fecundity depends on the variation in fecundability between more and less fecund individuals. With small differences in fecundability, only a small proportion of the variance can be attributed to differences in fecundity. If half the population has a fecundability of 17% per monthly cycle and the other half has a fecundability of 15% per cycle, then simulations show that less than 0.5% of the variance can be attributed to the difference in fecundity between the two groups. However, if the fecundability of the less fecund group drops to 5% per cycle, then almost 20% of the variance in time to conception can plausibly be attributed to differences in fecundity.

S2.3.2 Determinants of Time to Pregnancy in Modern Populations

Studies using data for modern populations have found that time to pregnancy is primarily determined by purely biological factors (e.g., age, menstrual cycle length, parity, or the degree of oxidative damage to sperm DNA).[14] Moreover, those factors that correlate with fecundity have very little predictive power for determining the time to pregnancy. Axmon et al. Axmon et al. (2006) had information on a large variety of factors and concluded that their multivariate model "explained only a small fraction of the variation in the observed time to pregnancies" (p. 1279). This is consistent with the literature review in Ecochard (2006), which states that "most of the biological heterogeneity" in fecundity "remains unexplained" (p. 142). At any rate, biological determinants of variations in fecundity are of no concern with respect to our identification strategy because our interest regards the effect of the inherent biological nature of fecundity on offspring quality as well as its potential effect via fertility.

Other factors sometimes found to influence fecundity include exposure to chemicals. The most robust findings relate to pesticides,[15] but even these findings are ambiguous. Curtis et al. (Curtis et al., 1999, p. 112) argue that pesticides may potentially affect fecundity. However, they still found "no strong or consistent pattern of associations of pesticide exposure with time to pregnancy." Other environmental exposures are even less likely to influence the time to pregnancy. For example, Joffe et al. (2003) found no effect of lead exposure on the time to pregnancy. Regardless, chemical factors are not a concern here, because humans were not usually highly exposed to these chemicals in the time period of our study.

Lifestyle-related and socio-economic factors are generally found not to determine the time to pregnancy. Juul et al. Juul et al. (1999) found that smoking, body mass index, age, and parity did not explain regional differences in fecundity in their data, and Juhl et al. Juhl et al. (2001) found no association between moderate alcohol intake and fecundity. Joffe et al. Joffe and Barnes (2000) concluded that a wide range of socio-economic characteristics did not affect the time to pregnancy of the first birth among the subsequent generation. Characteristics included body-mass index, height, smoking habits, and social class. The lack of intergenerational effects is interesting

---

[14] See Juul et al. (1999); Loft et al. (2003); Ecochard (2006); Axmon et al. (2006); Wise et al. (2011); Schwartz and Mayaux (1982).

[15] See de Cock et al. (1994); Curtis et al. (1999); Cohn et al. (2003); Axmon et al. (2006).

here because it suggests that grand-parental socio-economic characteristics are unrelated to time to pregnancy of parents.[16]

The medical literature generally concludes that fecundity is determined by biological components and much less so by lifestyle or socio-economic circumstances. Axmon et al. (Axmon et al., 2006, p. 1279) concluded that "female biological factors seemed more important predictors of [time to pregnancy] than lifestyle factors." Consistent with these findings, Klemp and Weisdorf (2019) show that the observed socio-economic variables in their data have no statistical explanatory power over the length of their protogenesic intervals. Thus, while the possibility of omitted variables can never be entirely ruled out, the limited scope for lifestyle-related and other potentially omitted variables in explaining parental fecundity is highly reassuring.

S2.3.3 The Protogenesic Interval and Observed Socio-economic Variables

One other paper has studied the association between the protogenesic interval and socio-economic factors Klemp and Weisdorf (2019). It found a complete absence of significant correlations between the length of the protogenesic interval and observed socio-economic characteristics of the parents, except the maternal age. The lack of a significant correlation is interesting, because it suggests that the protogenesic interval is not correlated with key parental socio-economic characteristics that could potentially co-determine offspring human capital, including parental literacy, skills, and wealth.

S2.4 Main Statistical Model

To assess the negative relationship between PI and the number of children and the hump-shaped relationship between PI of heads of dynasties and long-run reproductive success formally, the study estimate a series of quadratic regression models of the form

$$\ln(1 + D_{i,1}) = \beta_{0,1} + \beta_{1,1} PI_i + Z_i \beta_{3,1} + \varepsilon_{i,1}$$

and

$$\ln(1 + D_{i,t}) = \beta_{0,t} + \beta_{1,t} PI_i + \beta_{1,t} PI_i^2 + Z_i \beta_{3,t} + \varepsilon_{i,t}$$

where $D_{i,1}$ is the number of children born to heads of dynasties in generation 0; $D_{i,t}$ is the number of descendants that the head of dynasty i, in generation 0, has in the subsequent three generations t, $t = 2, 3, 4$; PI_i is the time from the first marriage to the first birth of the head of dynasty i; $Z_{i,t}$ is a vector containing the control variables; and $\varepsilon_{i,t}$ is an error term clustered by individuals sharing the same firstborn. The addition of one ensures that the transformation is defined for extinct dynasties. Table S2 demonstrates that the results are robust to the use of a GLM model with a negative binomial distribution and a logarithmic link function.

The theory predicts that $\beta_{i,1} < 0$, i.e., that PI negatively affects the number of children. (In contrast, the quadratic specification of the relationship between PI and the number of children is

---

[16] This view is consistent with Aguero and Marks (2008, 2011) who show that infertility is largely unrelated to parental characteristics, besides age, concluding that according to their analysis, "infertility is not correlated with 'predetermined' or background characteristics of women" (Aguero and Marks, 2011, p. 806).

insignificant statistically.) Furthermore, the theory predicts that $\beta_{1,t} > 0$ and $\beta_{2,t} < 0$ for $g \geq 2$, i.e., PI has a hump-shaped effect on the number of grandchildren, great-grandchildren and great-great-grandchildren. Estimates on the original scale of the number of descendants presented in the paper, when based upon a linear regression model with log-transformed outcomes, are corrected for retransformation bias in accordance with Duan (1983) and the test for a presence of a hump-shape is based on Lind and Mehlum (2010). We test the significance of the regression coefficients using t-tests and hypothesis tests used throughout the analysis are two-sided.

In addition to the remarks on Table 1 stated in the main part of the paper, it should be noted that column 5 indicates that literacy is negatively associated with the number of children, and positively associated with long-run reproductive success, suggesting a possible quality-bias of literate parents. This is consistent with the finding reported in Table 2, that literacy is positively associated with the number of surviving children, thus indicating that the quality strategy is associated with higher short-run reproductive success.

S3 Additional Data Description

S3.1 Descriptive Statistics and Basic Relationships

This subsection explores demographic patterns in Quebec between the 16th and the 18th century that could shed light on the empirical exploration of the effect of fecundity of long-run reproductive success. In particular, it examines: the distribution of the number of children and grandchildren, the distribution of the age at first marriage across individuals, and the relationship between the protogenesic interval and the number of children.

The high-fertility regime in Quebec over this historical period is evident from supplementary materials Fig. S2. The histogram of the fraction of individuals with a given number of children takes on a predominantly mound-shaped form. The density of individuals with a given number of children increases at larger number of children, for less than 10 children, and drops significantly thereafter with each successive birth, with less than 0.5% of individuals (predominantly male) with more than 20 children. In particular, the average fertility of heads of lineages in the main sample, consisting of the early settlers of Quebec, is 9.70 children (supplementary materials Table S1). However, as a consequence of a high mortality rate among offspring of early settlers, heads of lineages produce on average 4.63 children surviving to age 40 (supplementary Table S1).

Supplementary materials Fig. S2 also depicts the density of individuals with a given number of grandchildren. Interestingly, despite the predominantly mound-shaped distribution of the number of children, the distribution of the number of grandchildren is J-shaped, reflecting the compound nature of long run reproductive success and that the potential role of child quality in generating dispersion in the number of grandchildren. The histograms of the density of individuals with a given number of great-grandchildren and great-great-grandchildren, also shown in supplementary materials Fig. S2, display a similar pattern and establish that the range in reproductive success increases dramatically as the number of generations over which it is recorded increases.

The age distribution at first marriage of individuals during this time period is depicted in supplementary materials Fig. S3. The mean age at first marriage of 23.8 over this period is substantially lower than the age of 30.1 observed in Canada in the year 2008.[17] Since marriage over the study period marked the intention to conceive, this early marriage age combined with a late age at last birth (also supplementary materials Fig. S3), contributed to a high-fertility regime in historical Quebec.

Supplementary Fig. S4 depicts the relationship between PI and the number of children nonparametrically, dividing PI into 15 equal-size groups (15-quintiles) and plotting the mean value of the number of children in each bin.[18] This binned scatter plot provides a nonparametric representation of the conditional expectation function without showing the underlying variance in the individual-level data. The regression coefficient and standard error, corresponding to the overlaid regression line (and reported in the Fig. caption), is estimated on the micro-data (not the binned

---

[17] See Indicators of Well-being in Canada developed by *Human Resources and Skills Development Canada* at www4.hrsdc.gc.ca/.3ndic.1t.4r@-eng.jsp?iid=78.

[18] This flexible nonparametric method is also employed, and further described, by Chetty et al. (2014), investigating relationships found in large datasets.

averages). The Fig. and the associated linear regression establishes that PI has a statistically highly significant negative effect on the number of children over this time period.

S3.2 Data and Analysis Sample

The data is based on the demographic history of Quebec, using the reconstructed genealogy of the parish registers of the St. Lawrence Valley in Quebec, covering 471,412 individuals born between 1572 and 1799, spanning the French settlement and colonization of Quebec Légaré (1988). The data exhaustively covers the Catholic population of European descent who lived in the St. Lawrence River valley from 1608 to 1799 Lacroix and Desjardins (2012).[19] Nearly all individuals in the data (i.e., more than 99.96%) were born after the founding of Quebec City in 1608.[20] Indeed, more than 94% of these individuals were born and died in Quebec.

The heads of lineages immigrating to Quebec before 1680 are responsible for the majority of the genetic makeup of today's population in Quebec Charbonneau et al. (2000). The analysis focuses on the reproductive success of individuals who were born in Quebec prior to the end of 1685, died in the province, and had at least one great-great grandchild.[21] The summary statistics for this sample can be found in supplementary materials Table S1. The focus on the reproductive success of heads of lineages born prior to the end of 1685 permits the tracing of most descendants of these individuals in the subsequent four generations, while accounting for the lineage fixed fixed effects. Furthermore, since the early period of the data relates to a very small founder population (i.e., 19 births satisfying the sample restrictions in the period 1620–1639, 68 births in 1640–1649, and 287 births in 1650–1659), the focus on reproductive success of heads of lineages born prior to the end of 1685 allows the inclusion of years in which births rates were sufficiently large to permit a precise estimate of the regression coefficients while accounting for a large set of fixed effects.

The demographic conditions in the St. Lawrence Valley before the 1670s were characterized by population growth via immigration, including the arrival of refugees from other parts of New France following the end of the French and Indian War in 1663. Furthermore, with the arrival of vast majority of the so-called "King's daughters" by 1671 and the prohibition of permanent migration out of the colony and into France in the following year, the population started growing mainly via natural increase in the 1670s Charbonneau (1993). We therefore also performed the analysis on a sample restricted to cohorts born after 1670 while keeping the other sample restrictions. In particular, as established in supplementary materials Table S4, Panel C–E, the main results are

---

[19] The data is provided by *Le Programme de recherche en démographie historique* at the University of Montreal.

[20] The data covers all parishes of Quebec, and thus in light of the negligible inter-provincial migration over this period, intra-provincial migration does not prevent the tracking of reproductive success of individuals over several generations.

[21] All children of these heads of lineages, virtually all grandchildren, and most of the descendants in the third and the fourth generations are observed. Furthermore, the fraction of extinct lineages from generation to generation in the regression sample is rather constant, implying that the end date of the entire data need not generate concerns regarding the effect of potential censoring. Indeed, as can be derived from the numbers of observations in columns 1–4 in supplementary materials Table S4, Panel A, of the 4,240 observations satisfying all the sample restrictions while also including heads of lineages with no observed great-great-grandchildren, 5.6% of lineages produced no observed grandchild, and of the remaining lineages, 1.7% produced no observed great-grandchild, and of these remaining lineages 3.4% produced no observed great-great-grandchild. Furthermore, as established in supplementary materials Table S4, Panel A and B, the results are robust to the inclusion of extinct lineages. Potential systematic associations between the birth year of the head of the lineage and the number of unobserved descendants in the third and fourth generations are accounted for by the inclusion dummy variables for the birth year of the heads of lineages.

qualitatively unchanged if the analysis focuses on the reproductive success of heads of lineages born: (i) in 1660–1685, (ii) in 1670–1685, or (iii) over the entire sample period. Additional data description can be found in supplementary materials 1.

We restrict the sample to heads of lineages with the highest quality grade of data on the first marriage date as well as the birth date of the first child. Furthermore, in addition to the focus explained in the main text, the study focuses on individuals whose time from first marriage to first birth is at least 38 weeks.[22] Moreover, the study follow the convention in the literature, e.g., Milot et al. (2011), and restricts the sample to individuals whose time to first conception is less than 2 years (i.e., PI not shorter than 38 weeks and not longer than 2 years and 38 weeks), excluding 10.9% of this subset.[23] This further restriction is designed to mitigate the effect of extreme protogenesic intervals which may reflect measurement errors or underlying biological conditions that may directly affect long-run reproductive success. Thus the analysis focuses on the reproductive success of 3,798 heads of lineages in the pre-1685 period (consisting of individuals born in the period 1620–1685) that satisfy the entire sample restrictions.

The sample includes men and women for two reasons. First, due to the occurrence of remarriages, PI can differ between men and women, whereby the inclusion of both provides a complete picture of the variation in fertility in households due to variation in PI.[24] Second, by including men and women, the analysis accounts for the differential effects of paternal and maternal characteristics such as birth year, marriage age and a proxy for literacy. It should be noted that the analysis accounts for the correlation in reproductive success between heads of lineages that produced the same firstborn and thus shared their PI by clustering the standard errors on the level of the firstborn.[25]

Furthermore, immigrants among heads of lineages are excluded from the sample for two reasons. First they may differ systematically from natives, reflecting the circumstances that led to their decision to immigrate as well as the effects of immigration on their socio-economic status and thus their reproductive success. Second, reproduction of immigrants prior to their arrival to Quebec is unknown. Similarly, emigrants are excluded as well since they may possess unique attributes and their subsequent marriages and births outside of Quebec are not observed.

S3.3 Analysis Variables

The analysis makes use of the following variables.

---

[22] For the 8.2 percent of the individuals in the sample of non-migrating heads of lineages whose firstborn's date of birth is unknown, it is estimated to be one week prior to the date of baptism.

[23] The use of alternative cut-offs would not affect the qualitative results. In particular, focusing on individuals whose time from first marriage to first birth is at least 36 weeks, the hump-shaped relationship remains highly significant. Likewise, focusing on individuals whose time from first marriage to first birth is at least 40 weeks, the hump-shaped relationship remains significant. Furthermore, exploiting alternative estimation methods that do not rely on a quadratic specification reveals qualitatively similar patterns in the presence of observations with PI exceeding 2 years and 38 weeks.

[24] Supplementary Table S5, Panel A, establish that the results are robust to accounting for remarriages.

[25] The results are robust to the use of the sample of men or women only. In particular, supplementary materials Table 4, Panel G, establishes that the baseline results are robust to the focus on women only.

S3.3.1 Main Dependent and Independent Variables

In the main analysis that explores the effect of the time interval, measured in years, between the first marriage date of the head of a lineage and the birth date of the individual's first child (the protogenesic interval, PI) on reproductive success, the dependent variable is the number of offspring of each head of lineage in the subsequent four generations (i.e., children, grandchildren, great-grandchildren and great-great-grandchildren). In the additional analysis that examines the mechanism through which PI affects long-run reproductive success, the dependent variables are the fraction of ever-married children among children observed to have survived to age 40, the average marriage age of children with observed marriage age, and the fraction of literate children among children with observed literacy status (recorded at marriage). The independent variable throughout the analysis is the protogenesic interval of the head of a lineage and the birth date of the individual's first child (PI).[26]

S3.3.2 Lineage Fixed Effects

The effect of fecundity on reproductive success may be affected by variation in genetic predisposition among genetically distinct individuals, as well as variation in cultural and socio-economic background. Hence, similarities in the genetic, cultural, and socio-economic characteristics across siblings within each household are exploited to isolate the effect of random variation in PI on reproductive success. Accounting for lineage fixed effects, as well as the confounding factors underlined below, the analysis explores the effect of random variation in PI on long-run reproductive success within lineages headed by siblings, as opposed to across all heads of lineages.[27]

S3.3.3 Control Variables

The analysis accounts for the confounding effects of the marriage age, birth year, literacy, and the maternal identity, for each head of lineage. Furthermore, additional control variables include the geographic location at birth and death, birth order, month of birth, month of birth of the firstborn, number of marriages, and immigration and emigration status of the spouse, for each head of lineage.

The confounding associations between the marriage age of heads of lineages and their affluence, fecundity, and reproduction is accounted for by the inclusion of dummy variables indicating the

---

[26] In couples where neither spouse remarried, PI is identical for the husband and the wife. Nevertheless, given that the frequency of remarriage over this period is substantial, reflecting in part a considerable mortality rate, PI and the number of offspring of each spouse often differ. The correlation in reproductive success between parents sharing the same firstborn, and therefore the same PI, is accounted for by clustering the standard errors for heads of lineages sharing the same firstborn.

[27] Accounting for the family characteristics of the spouse of the head of dynasties would restrict the sample for siblings who married siblings from a different household. Accounting for these fixed effects would reduce the sample size by more than 90%, and more importantly, would introduce biases associated with the unique characteristics of these families and their matching process. Reassuringly however, the lineage fixed effects do account for the characteristics of the spouses of heads of households in the likely scenario of non-random (assortative) mating. Moreover, since the regression analysis is nearly unaffected by the introduction of lineage fixed effects, the potential effect of random mating does not compromise the analysis and further control for fixed factors in the family of the spouse of the head of lineage are unlikely to alter the results.

marriage age of heads of lineages. The marriage age is associated with reproductive success through three channels. First, fecundity is affected by age Baird et al. (2005). Second, the marriage age affects the length of the reproductive period of the couple. Third, in the pre-demographic transition era that corresponds to our sample, the marriage age was inversely related to the affluence of individuals, and marriage age and its potential association with affluence could have had an independent effect on long-run reproductive success.

The time-path of socioeconomic and demographic factors may differentially affect fecundity and reproductive success across cohorts of heads of lineages. In particular, the affluence, fecundity, and reproductive success of heads of lineages may be affected by the socioeconomic and demographic conditions during their lifetime, as partly captured by their birth year. These confounding factors are accounted for by the inclusion of dummy variables indicating the birth year of heads of lineages.[28]

The human capital attainment that may reflect the socioeconomic status of heads of lineages may affect their PI and reproductive success. This confounding factor is partly accounted for by the inclusion of the literacy status of heads of lineages, inferred from the existence of a proper signature on the marriage certificate. Additional confounding variations between heads of lineages are accounted for by the inclusion of dummy variables capturing gender, geographic location at birth and death, birth order, month of birth, month of birth of the firstborn, number of marriages, and immigration and emigration status of the spouse.

Finally, variations in socioeconomic and physiological factors across heads of lineages may generate variation in the length of the reproductive period that may obscure the effect of PI on reproduction. In particular, conditional on the marriage age, the age at last delivery determines the length of the reproductive period. Hence, to account for the potential effect of the stoppage age, the analysis is shown to be robust to the inclusion of dummy variables indicating the stoppage age of heads of lineages, in addition to the marriage age, are introduced to account for this confounding factors, permitting the study to capture the effects of PI on fertility, for a given length of reproductive period.[29]

---

[28] In addition, the inclusion of birth year dummy variables mitigates the potential systematic associations between the birth year and the number of unobserved descendants after three or four generations.

[29] As depicted in supplementary materials Fig. S3, the stoppage age over this period marked the decline in fecundity and onset of sterility associated with age-related infertility and onset of menopause, with a modal stoppage age of 41.

S4 Additional Analysis Information

S4.1 Statistical Model

The negative relationship between PI of heads of lineages and the number of children, as well as the hump-shaped relationship between PI of heads of lineages and long-run reproductive success, is further assessed by estimating a series of regression models.

First, the effect of PI of the head of lineage on the number of children born to the head of lineage is estimated using the OLS regression model:

$$\ln D_{i,1} = \beta_{0,1} + \beta_{1,1} PI_i + Z_i \beta_{3,1} + \varepsilon_{i,1},$$

where $D_{i,1}$ is the number of children (i.e., offspring in generation 1) born to head of lineage i; PI_i is the protogenesic interval of the head of lineage i; Z_i is a vector of control variables capturing the characteristics of the head of lineage i; and $\varepsilon_{i,1}$ is an error term clustered at the level of heads of lineages sharing the same firstborn.[30] The coefficient of interest is $\beta_{1,1}$ and it is predicted to be negative, i.e., PI of heads of lineages negatively affects the number of children.

Second, the effect PI of heads of lineages and long-run reproductive success is estimated using the OLS regression model:

$$\ln D_{i,t} = \beta_{0,t} + \beta_{1,t} PI_i + \beta_{2,t} PI_i^2 + Z_i \beta_{3,t} + \varepsilon_{i,t},$$

where $D_{i,t}$ is the number of descendants that the head of household i, has in the subsequent generations t, $t = 2, 3, 4$.[31] The coefficients of interest are $\beta_{1,t}$ and $\beta_{2,t}$. The prediction is that $\beta_{1,t} > 0$ and $\beta_{2,t} < 0$, i.e., PI has a hump-shaped effect on the number of grandchildren, great-grandchildren and great-great-grandchildren.

This main regression methodology conforms to the Lande-Arnold model of selection on correlated characters, using a transformation of the number of descendants as the outcome variable Lande and Arnold (1983).

S4.2 Estimation Based on Variation within Heads of Lineages

In the baseline sample of 3,798 heads of households, 437 descended from mothers with no other offspring in the sample. Combined with the existence of unique dummy variables observation groups, the identification of the trade-off in the fixed-effects regressions is based on variation and 3,348 observations in the regressions including Maternal Origins fixed effects.

---

[30] The control variables (marriage age, birth year, sex, signature status (a crude proxy for literacy, see the discussion in Bouchard and Larouche (1989)), and stoppage age) are presented in supplementary materials Section 1.

[31] The results are robust to alternative regression methods suitable for count data. In particular, supplementary materials Table S2 demonstrates that the results are robust to the use of a GLM model with a negative binomial distribution and a logarithmic link function.

S4.3 Conversion of Estimates from Logarithmic to Level Scale

Throughout the analysis, in linear regression models involving outcome variables that are logarithmically transformed, estimates on the original scale of numbers of descendants reported in the main text are corrected for re-transformation bias in accordance with Duan (1983).

S4.4 Robustness and Other Potential Mechanisms

We establish the robustness of the qualitative results to (i) alternative sample periods and restrictions, such as restricting the sample to urban parishes, restricting the sample to parishes on the northern shore of the St. Lawrence River, including extinct lineages, excluding heads of lineages with just one child, excluding heads of lineages marrying before age 16, and more (supplementary materials Table S4); (ii) controlling for a wide range of potentially confounding factors or potential alternative mechanisms, including the number of marriages, spousal migration, birth parish,[32] death parish, month of marriage, month of birth of the firstborn, firstborn status of the head of lineage, birth order, parilineal or matrilineal migration (see Gagnon et al. (2006)), and patrilocal or matrilocal marriage (Table S5); and (iii) alternative estimation method, i.e., GLM (supplementary materials Table S2).[33]

S4.5 Robustness of Mechanism Results

The results in Table 2 are robust to alternative specifications as explained below.

S4.5.1 Fraction of Children Surviving to Age 40 Who Got Married

Due to the fact that the outcome variable is a fraction, the estimates presented in Table 2, columns 1 and 2, are based on the fractional logit model Papke and Wooldridge (1996). The results are robust to the inclusion of marriage age, birth year and stoppage age as continuous variables. In particular, the coefficient on PI in the first, second, third and fourth column would be 0.392 ($p < 0.001$), 0.334 ($p < 0.001$), 0.314 ($p < 0.001$), and 0.311 ($p < 0.001$), respectively. Furthermore, the results are robust to the use of a linear probability model. In particular, the coefficient on PI in the first, second, third and fourth column would be 0.071 ($p < 0.001$), 0.060 ($p < 0.001$), 0.054 ($p < 0.001$), and 0.055 ($p < 0.001$), respectively. Since the lineage fixed effects constitute an unbalanced panel, the fractional logit model does not account for these fixed effects Papke and Wooldridge (2008). Accounting for the fixed effects in a fixed-effects linear probability model on the baseline sample also yields positive estimates of the coefficient on PI, although the estimates are insignificant. However,

---

[32] The robustness to controlling for these parish-specific fixed effects is reassuring in light of the fact that while individuals located on the Canadian pioneer fronts produced more descendants (see Moreau et al. (2011)), the registration of births might have been deficient in these regions.

[33] Furthermore, since we focus on the reproductive success of early settlers, our analysis minimizes the potentially confounding effects of inbreeding Robert et al. (2009). Nevertheless, we also investigated the robustness of the results with respect to controlling for the degree of inbreeding in the extended sample underlying supplementary materials Table S5, Panel D, and found that the analysis is robust to accounting for this variable (results available upon request).

focusing on heads of lineages born prior to the end of 1660, the positive effect of PI on the average literacy of offspring is statistically significant. In particular, the estimate in column 4 would be 0.137 ($p = 0.002$). Thus, the statistical evidence of the positive effect of lower PI based on variation within siblings sharing the same mother on offspring literacy is strongest for the first part of the baseline period.

S4.5.2 Average Marriage Age of Children

In fixed-effects regressions accounting for lineage fixed effects, otherwise corresponding to Table 2, columns 3 and 4, the coefficient estimate drops somewhat and the standard error increases resulting in decreased statistical significance. However, the finding remains statistically significant at the 10% significance level in the entire sample period, i.e., the regression sample underlying supplementary materials Table S4, Panel D. In particular, in the regression corresponding to that underlying column 4, controlling for lineage fixed effects in the entire sample period, the effect of PI on the average marriage age of offspring is -0.181 ($p = 0.082$).

S4.5.3 Fraction of Literate Children

Due to the fact that the outcome variable is a fraction, the estimates presented in Table 2, columns 5 and 6, are based on the fractional logit model Papke and Wooldridge (1996). The results are robust to the inclusion of marriage age, birth year and stoppage age as continuous variables. In particular, the coefficient on PI in the first, second, third and fourth column would all increase in size to 0.316 ($p = 0.005$), 0.268 ($p = 0.015$), 0.255 ($p = 0.020$), and 0.256 ($p = 0.020$), respectively. Furthermore, the results are robust to the use of a linear probability model. In particular, the coefficient on PI in the first, second, third and fourth column would be 0.018 ($p < 0.005$), 0.015 ($p < 0.016$), 0.014 ($p = 0.029$), and 0.013 ($p = 0.035$), respectively. Since the lineage fixed effects constitute an unbalanced panel, the fractional logit model does not account for these fixed effects Papke and Wooldridge (2008). However, accounting for these fixed effects in a fixed-effects linear probability model yields similar results, although less precisely estimated. In particular, the estimate in column 4 would be 0.10 ($p = 0.127$). Under an alternative age criteria of 30 years instead of 40 years the estimate the estimate would be 0.011 ($p = 0.087$).

S5 Heritability of the Protogenesic Interval

S5.1 Estimation of Heritability using the Animal Model

We use the Animal Model to estimate the heritability of the protogenesic interval. We used the MCMCglmm R library to estimate the Bayesian implementation of the Animal Model. We used univariate models with various weakly informative priors, i.e., $V = 1$, $\nu = 0.002$ (see Milot et al. (2011)), and took samples from the posterior distributions of 1,500,000 iterations at every 1,000 iteration of the Markov chain after an initial burn-in of 500,000 iterations, for a total of 1,000 samples. We include a constant term.

We find that heritability was $h^2 = 0.054$ (95% credible interval: 0.027 to 0.082). More complex models that account for lineage fixed effects by use of random effects on the maternal level, as well as marriage age and a time trend as fixed factors, also yielded similarly tight estimates of heritability of the same order of magnitude (results available upon request). Thus, our Animal Model analysis establishes that the time from marriage to first birth is a heritable character.

S5.2 Estimation of Heritability using Parent-Offspring Regressions

We also estimate heritability of the protogenesic interval via parent-offspring regressions (see Falconer and Mackay (1996); Roff (2012)). In particular, we estimated a series of linear regression models of the years from marriage to the first birth for individual's on the average time from marriage to the first birth of those individual's parents, controlling for a range of factors. The analysis establish that the estimate of the heritability of the protogenesic interval is between 0.033 ($N = 82,911$) and 0.039 ($N = 84,263$), always significant at leat at the 5% level, depending on number of factors that we control for (marriage age, birth year fixed effects, stoppage age, signature status, and sex). Thus, our parent-offspring regression analysis also establishes that the time from marriage to first birth is a heritable character.

S6 Distributions of Numbers of Descendants

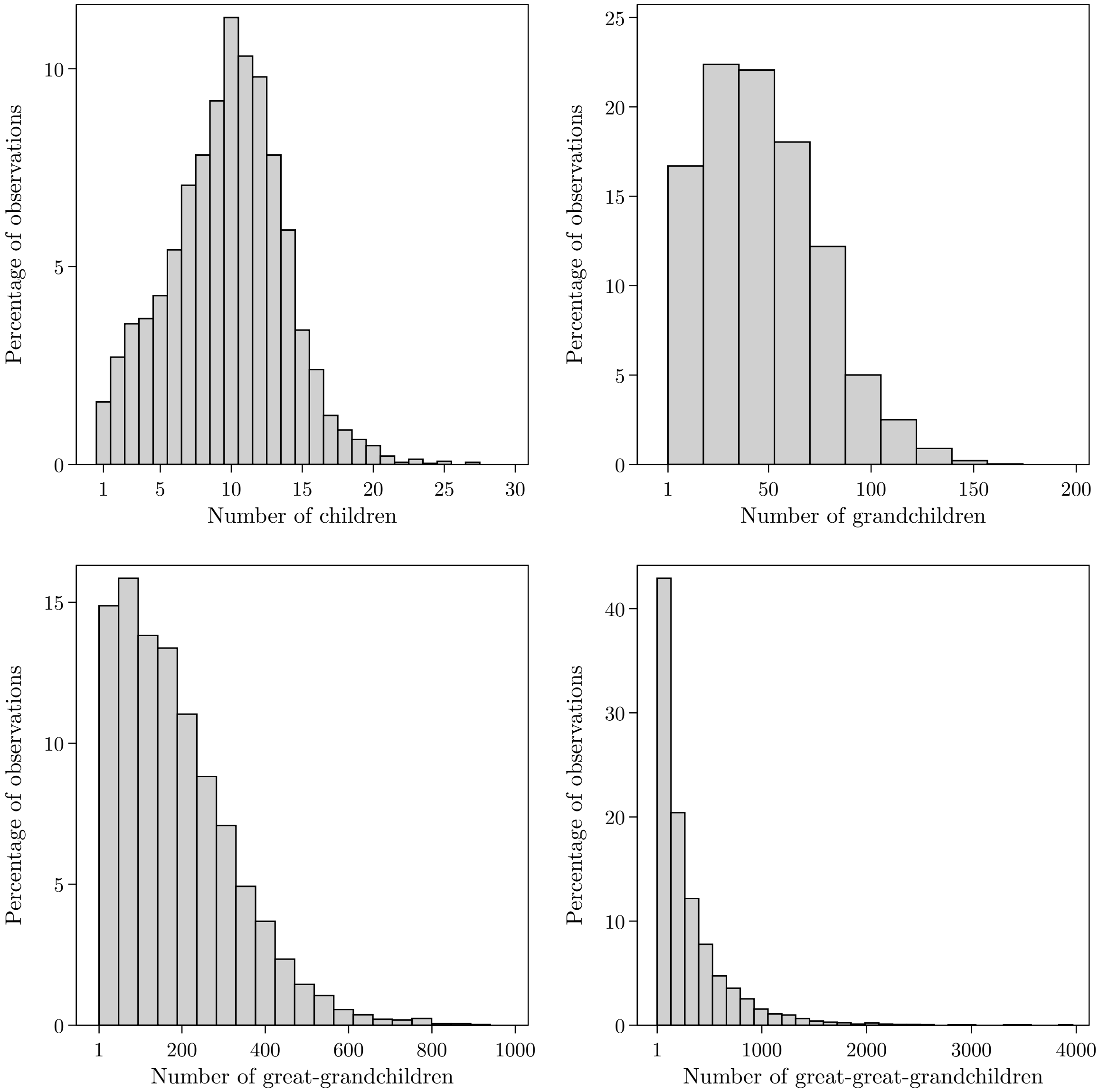
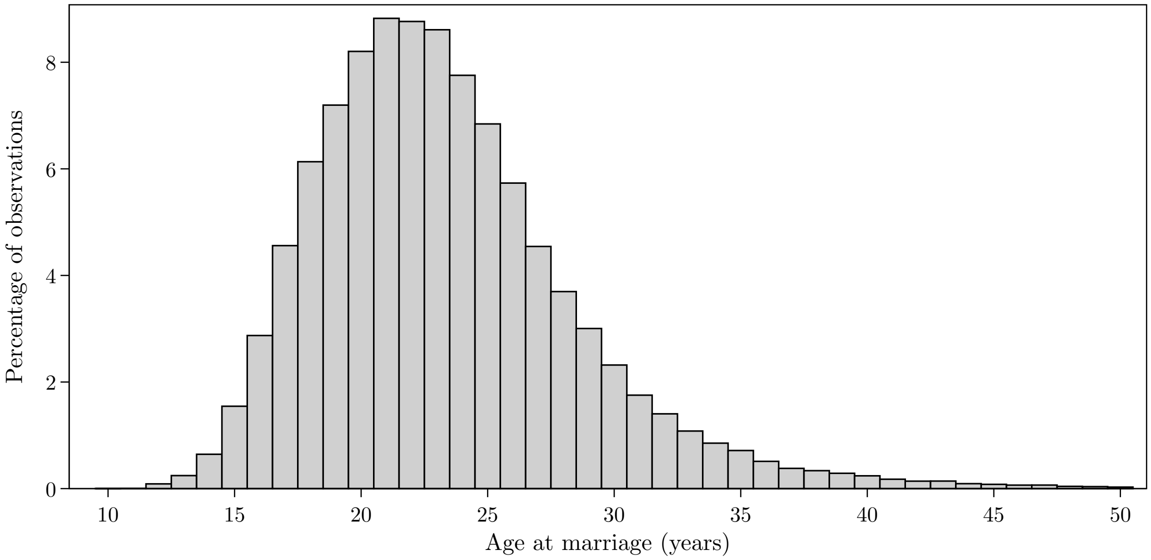


Fig. S2: Histograms of the number of children, grandchildren, great-grandchildren, and great-great-grandchildren of the 3,798 individuals in the baseline regression sample of Table 1. Panel A depicts the number of children, Panel B depicts the number of grandchildren, Panel C depicts the number of great-grandchildren, and Panel D depicts the number of great-great-grandchildren.

S7 Distributions of Marriage and Stoppage Age



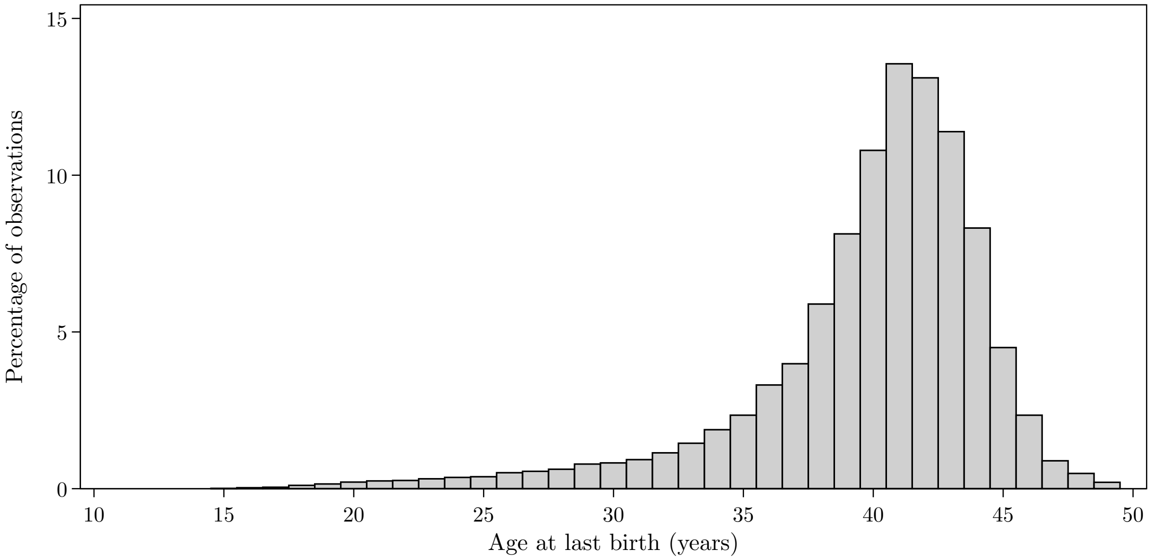


Fig. S3: The top histogram depicts age at first marriage (in years) of 107,887 non-migrant, ever-married women in Quebec between the 16th and the end of the 18th century, and whose marriage age was no later than age 50. The bottom histogram depicts the age at last delivery of 13,400 non-migrant, ever-married mothers in Quebec born before 1749 (and after 1624) who survived to age 50 and whose husband survived to age 50, and whose last delivery was no later than age 50.

S8 PI and Number of Children – Conditional Relationship

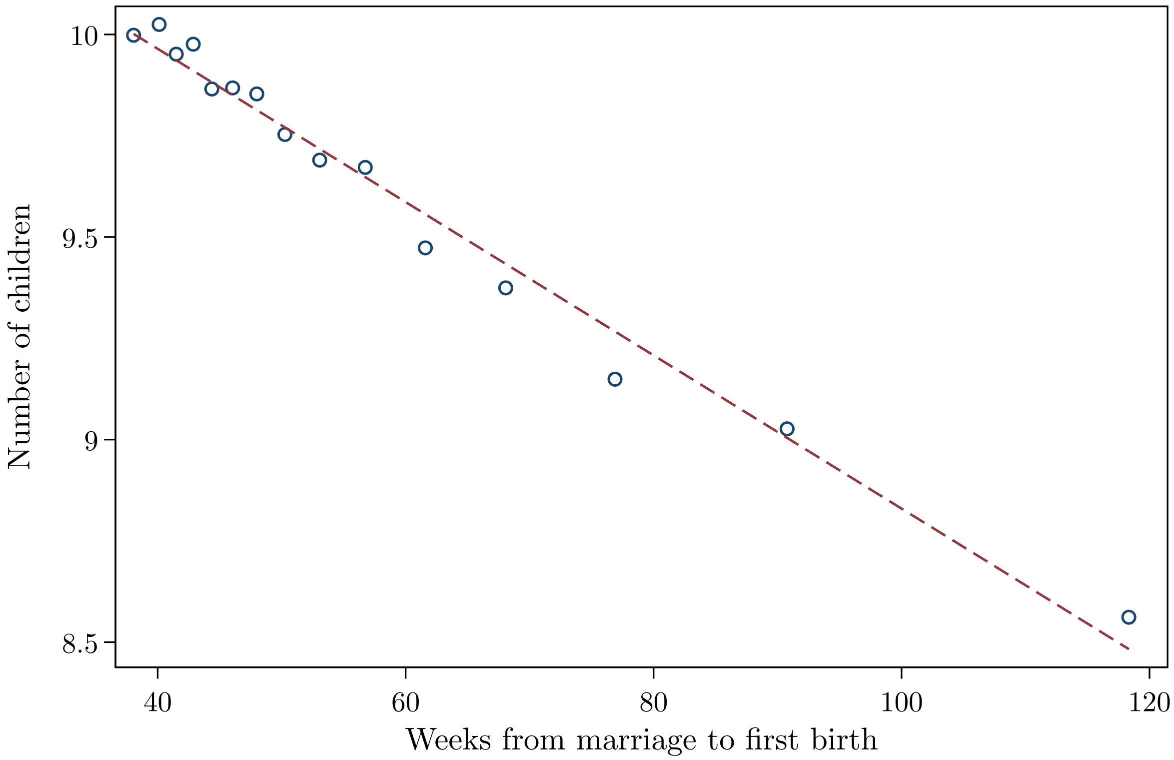


Fig. S4: This figure depicts a binned scatterplot of number of children versus PI for 38,498 non-migrating individuals in Quebec between the 16th and the end of the 18th century with known birth year, at least one observed grandchild and whose first conception occurred within marriage and in less than 2 years (i.e., with a PI not shorter than 38 weeks and not longer than 2 years and 38 weeks). This binned scatter plot is constructed by first residualizing the the number-of-children variable with respect to marriage age, stoppage age and birth year. Second, PI is divided into 15 equal-sized groups (15-quintiles) and the means of the number of children residuals within each bin is plotted against the mean value of PI within each bin. The line shows the best linear fit estimated on the underlying micro data using OLS. The estimated slope of the best-fit line, is -0.019 (i.e., one additional week from marriage to first birth results in 0.019 fewer children) and highly significant with a standard error clustered at the level of the firstborn of less than 0.001 ($p < 0.001$).

S9 Summary Statistics for Heads of Lineages

Table S1: Summary statistics of heads of lineages born prior to the end of 1685

	Mean (1)	Median (2)	S.D. (3)	Count (4)
Children	9.70	10	3.99	3,798
Surviving children[a]	4.63	4	2.59	3,798
Grandchildren	47.35	44	28.07	3,798
Great-grandchildren	187.59	159	140.17	3,798
Great-great-grandchildren[b]	294.01	171	363.58	3,798
Years from marriage to first birth (PI)	1.20	1.02	0.47	3,798
Literate	0.66	1	0.47	2,222
Fraction of literate children	0.74	1	0.35	3,448
Fraction of surviving children[a]	0.49	0.50	0.21	3,798
Fraction of surviving children with known literacy[a]	0.76	0.67	0.56	3,727
Age at first marriage	22.67	22.2	5.46	3,798
Age at last delivery	41.95	42.1	8.61	3,798

Table S1 Note: This table presents the summary statistics of the sample. [c] Survival is recorded at age 40. [b] The moderate increase in the mean and median number of descendants from the third to the fourth generation (i.e. from great-grandchildren to great-great-grandchildren) reflects the fact that these cohorts are less fully observed. Furthermore, since men produce children at later ages than women, this effect is more pronounced among men.

S10 Robustness to GLM Regression

Table S2: Robustness to GLM Regression

	Number of descendants in:							
	Gen. 1	Gen. 2	Gen. 3	Gen. 4	Gen. 1	Gen. 2	Gen. 3	Gen. 4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PI	-.063***	.120	.371**	.616***	-.073***	.195*	.468***	.595***
	(.018)	(.148)	(.185)	(.230)	(.010)	(.118)	(.160)	(.216)
PI squared		-.047	-.133**	-.232***		-.087**	-.191***	-.256***
		(.048)	(.060)	(.074)		(.038)	(.052)	(.069)
Literate					-.020	.055*	.118***	.104*
					(.013)	(.029)	(.040)	(.055)
Male					-.015	.044	.083**	.072
					(.014)	(.029)	(.040)	(.055)
Maternal Founder fixed effects	No	No	No	No	Yes	Yes	Yes	Yes
Stoppage age fixed effects	No	No	No	No	Yes	Yes	Yes	Yes
Number of observations	3,798	3,798	3,798	3,798	3,798	3,798	3,798	3,798
Joint sign.-level of PI & PI squared	.000	.446	.059	.000	.000	.000	.000	.000
Maximizing PI		1.273	1.399	1.328		1.116	1.225	1.159

Table S2 Note: This table presents the results of a series of GLM regressions, with a negative binomial distribution and a logarithmic link function, of the number of descendants in generation t on time to first birth measured in years, i.e. PI and PI^2. Birth year and marriage age dummies are included as controls. Furthermore, stoppage age dummies are included in columns 5–8. A dummy indicating unknown literacy is included in the regressions underlying column 5–8. Standard errors clustered at the level of the firstborn are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

S11 Simple Specifications

The findings can be established using simple specifications that excludes birth year control variables. Table S3 shows the results of simple specifications estimated with OLS and GLM.

Table S3: Simple Specifications

	Number of descendants in:							
	Gen. 1	Gen. 2	Gen. 3	Gen. 4	Gen. 1	Gen. 2	Gen. 3	Gen. 4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PI	-.079***	.138	.506**	1.030***	-.062***	.128	.380**	.706**
	(.001)	(.439)	(.028)	(.002)	(.000)	(.380)	(.039)	(.012)
PI squared		-.053	-.169**	-.323***		-.048	-.129**	-.213**
		(.359)	(.022)	(.002)		(.311)	(.030)	(.016)
Number of observations	3798	3798	3798	3798	3798	3798	3798	3798
Joint sign.-level of TFB & TFB2		.501	.07	.008		.476	.09	.037
Maximizing TFB		1.299	1.494	1.592		1.323	1.469	1.654

Table S3 Note: This table presents the results of a series of OLS and GLM regressions with a negative binomial distribution and a logarithmic link function, of the number of descendants in generation t on time to first birth measured in years, i.e. PI and PI^2. Marriage age dummies are included as controls. Columns 1–4 are estimated using OLS. Columns 5–8 are estimated using GLM with a negative binomial distribution and a logarithmic link function. Standard errors clustered at the level of the firstborn are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

S12 Robustness to Alternative Samples

Table S4 establishes the robustness of the analysis to the use of various alternative samples. All robustness-check control variables are included in the most flexible way possible, namely as categorical variables, using a dummy variable for each group. Furthermore, all regressions in this table accounts for lineage fixed effects and include control variables for literacy and sex (the coefficients are omitted from the table).

To save space, the specifications represent a middle ground of the parsimonious specifications in the main text that does not control for literacy, gender and stoppage age, and the full specification that controls for all these factors. In particular, recognizing the possible endogeneity of stoppage age, the specifications control for literacy and sex, but not stoppage age fixed effects. All the results are robust to excluding all these control variables or, alternatively, to including all the control variables.

The panels make use of the following different samples. Panel A: the baseline sample is extended to include extinct lineages (i.e., heads of lineages with no great-great-grandchildren). Panel B: the sample from Panel A is used, and the outcome variable is defined as $\log(1 + D_i)$, where D_i is the number of descendants in generation i, ensuring that the outcome variable is also defined for extinct lineages. Panel C: the baseline sample is restricted to heads of lineages born 1660–1685. Panel D: the baseline sample is restricted to heads of lineages born 1670–1685. Panel E: the baseline sample is extended by removing the restrictions on the birth year of the heads of lineages. Panel F: the baseline sample is restricted to heads of lineages with two or more children. Panel G: the baseline sample is restricted to heads of lineages marrying after age 15. Panel H: the baseline sample is restricted to female heads of lineages. Panel I: the baseline sample is restricted to parishes on the northern side of the St. Lawrence River. Panel J: the baseline sample is restricted to urban parishes (i.e., Montreal, Quebec City, or Trois-Rivières).

Table S4: Robustness to Alternative Samples

Panel A: Including Extinct Lineages				
PI	-.046*	.151	.462**	.773***
	(.026)	(.172)	(.208)	(.264)
PI squared		-.071	-.180***	-.309***
		(.057)	(.068)	(.087)
N	4,240	4,002	3,933	3,798
Panel B: Including Extinct Lineages, $\log(1 + D_i)$				
PI	-.041*	.414*	.822***	1.169***
	(.021)	(.215)	(.294)	(.336)
PI squared		-.149**	-.293***	-.427***
		(.068)	(.093)	(.108)
N	4,240	4,240	4,240	4,240
Panel C: Sample Restricted to 1660–1685				
PI	-.075***	.193	.487**	.802***
	(.025)	(.174)	(.216)	(.282)
PI squared		-.086	-.199***	-.330***

The table is continued on the next page.

Table S4: Continued from Previous Page

		(.057)	(.071)	(.093)
N	3,376	3,376	3,376	3,376
Panel D: Sample Restricted to 1670–1685				
PI	-0.071**	0.223	0.494**	0.805**
	(0.029)	(0.208)	(0.250)	(0.319)
PI squared		-0.102	-0.202**	-0.338***
		(0.070)	(0.085)	(0.107)
N	2,537	2,537	2,537	2,537
Panel E: Sample Extended to Entire Period				
PI	-.071***	.031	.293**	.393**
	(.017)	(.119)	(.140)	(.199)
PI squared		-.033	-.126***	-.187***
		(.039)	(.046)	(.065)
N	7,664	7,664	7,664	7,664
Panel F: Two Children or More				
PI	-.053**	.204	.560***	.891***
	(.021)	(.157)	(.201)	(.263)
PI squared		-.082	-.211***	-.344***
		(.051)	(.066)	(.086)
N	3,738	3,738	3,738	3,738
Panel G: Marriage after Age 15				
PI	-.052**	.129	.461**	.779***
	(.025)	(.175)	(.218)	(.279)
PI squared		-.057	-.180**	-.310***
		(.058)	(.073)	(.093)
N	3,604	3,604	3,604	3,604
Panel H: Females Only				
PI	-.080***	.380*	.993***	1.581***
	(.030)	(.215)	(.278)	(.349)
PI squared		-.142**	-.349***	-.564***
		(.069)	(.088)	(.110)
N	2,058	2,058	2,058	2,058
Panel I: Sample Restricted to North Shore				
PI	-0.063***	0.262*	0.542**	1.027***
	(0.013)	(0.152)	(0.213)	(0.302)
PI squared		-0.106**	-0.215***	-0.402***
		(0.049)	(0.069)	(0.099)
N	2,929	2,929	2,929	2,929
Panel J: Sample Restricted to Urban Parishes				
PI	-0.063***	0.313	0.658*	1.371***
	(0.020)	(0.255)	(0.351)	(0.494)

The table is continued on the next page.

Table S4: Continued from Previous Page

PI squared		-0.110	-0.238**	-0.482***
		(0.082)	(0.111)	(0.160)
N	1,436	1,436	1,436	1,436

Table S4 Note: This table presents the results of a series of OLS regressions of the number of descendants in generation t on protogenesic interval measured in years, i.e. PI and PI^2 for heads of lineages born prior to the end of 1685. The models account for lineage fixed effects, meaning that the estimation is based on variation across siblings, as opposed to across the population as a whole. Birth year and marriage age, literacy, and gender dummy variables are included, as well as a dummy variable indicating unknown literacy. Standard errors clustered at the level of the firstborn are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

S13 Robustness to Additional Control Variables

Table S5 establish the robustness of the analysis to accounting for various alternative factors. All robustness-check control variables are included in the most flexible way possible, namely as categorical variables, using a dummy variable for each group. Furthermore, all regressions in this table accounts for lineage fixed effects and include control variables for literacy and sex, in addition to the control variables specified in the following (the coefficients are omitted from the table).

To save space, the specifications represent a middle ground of the parsimonious specifications in the main text that does not control for literacy, gender and stoppage age, and the full specification that controls for all these factors. In particular, recognizing the possible endogeneity of stoppage age, the specifications control for literacy and sex, but not stoppage age fixed effects. All the results are robust to excluding all these control variables or, alternatively, to including all the control variables.

The panels make use of the following different set of controls. Panel A: the model includes dummy variables for the total number of marriages experienced during the lifetime of the heads of lineages, as well as dummy variables for the total number of marriages experienced by the first spouses of the heads of lineages. Panel B: the model includes dummy variables indicating the immigration and emigration statuses of the first spouses of the heads of lineages. Panel C: the model includes dummy variables indicating the birth (or baptism) parish of the heads of lineages. Panel D: the model includes dummy variables indicating the death (or burial) parish of the heads of lineages. Panel E: the model includes dummy variables indicating the months of marriage of the heads of lineages; Panel F: the model includes dummy variables indicating the month of birth of the the firstborn of the heads of lineages. Panel G: the model includes includes a dummy that indicates if the heads of lineages are firstborn in their family. Panel H: the model includes dummy variables for the birth order the heads of lineages, thereby accounting for birth order fixed effects. Panel I: the model includes dummy variables for patrilocal marriage (i.e., marriage in the same parish as the birth of the husband), matrilocal marriage (likewise for the wife), and both (i.e., an interaction between the former two). Panel J: the model includes dummy variables for patrilineal non-migration (i.e., the marriage parish is the same as the birth parish of the husband), matrilineal non-migration (likewise for the wife), and both (i.e., an interaction between the former two).

Table S5: Robustness to Additional Control Variables

	Panel A: Number of Marriages			
PI	-.059***	.282*	.590***	.828***
	(.022)	(.157)	(.200)	(.263)
PI squared		-.109**	-.225***	-.328***
		(.051)	(.066)	(.087)
	Panel B: Spousal Migration			
PI	-.059**	.117	.423**	.720***
	(.024)	(.161)	(.204)	(.264)
PI squared		-.054	-.168**	-.289***
		(.052)	(.067)	(.087)

The table is continued on the next page.

Table S5: Continued from Previous Page

Panel C: Birth Parish				
PI	-0.062***	0.106	0.416**	0.695***
	(0.024)	(0.163)	(0.205)	(0.267)
PI squared		-0.052	-0.168**	-0.284***
		(0.053)	(0.067)	(0.088)
Panel D: Death Parish				
PI	-0.062***	0.032	0.391*	0.653**
	(0.023)	(0.161)	(0.202)	(0.255)
PI squared		-0.023	-0.154**	-0.262***
		(0.053)	(0.066)	(0.084)
Panel E: Month of Marriage				
PI	-0.061**	0.151	0.482**	0.798***
	(0.024)	(0.163)	(0.205)	(0.265)
PI squared		-0.067	-0.190***	-0.318***
		(0.053)	(0.067)	(0.087)
Panel F: Month of Birth of Firstborn				
PI	-0.050**	0.181	0.500**	0.787***
	(0.024)	(0.166)	(0.208)	(0.274)
PI squared		-0.073	-0.194***	-0.314***
		(0.054)	(0.068)	(0.089)
Panel G: Firstborn				
PI	-.062***	.139	.462**	.773***
	(.024)	(.162)	(.204)	(.264)
PI squared		-.062	-.183***	-.309***
		(.053)	(.067)	(.087)
Panel H: Birth Order				
PI	-.062***	.144	.456**	.752***
	(.024)	(.162)	(.204)	(.265)
PI squared		-.063	-.181***	-.303***
		(.053)	(.067)	(.087)
Panel I: Patrilocal Marriage, Matrilocal Marriage, or Both				
PI	-0.077***	0.208	0.537***	0.811***
	(0.011)	(0.130)	(0.182)	(0.258)
PI squared		-0.089**	-0.211***	-0.325***
		(0.042)	(0.059)	(0.084)
Panel J: Same Birth and Death Parish (Husband, Spouse, or Both)				
PI	-0.077***	0.210	0.547***	0.826***
	(0.011)	(0.129)	(0.180)	(0.256)
PI squared		-0.090**	-0.215***	-0.331***

The table is continued on the next page.

Table S5: Continued from Previous Page

		(0.042)	(0.059)	(0.084)
Number of Observations (Same for All Panels)				
N	3,798	3,798	3,798	3,798

Table S5 Note: This table presents the results of a series of OLS regressions of the number of descendants in generation t on protogenesic interval measured in years, i.e. PI and PI^2 for heads of lineages born prior to the end of 1685. The models account for lineage fixed effects, meaning that the estimation is based on variation across siblings, as opposed to across the population as a whole. Birth year and marriage age, literacy, and gender dummy variables are included, as well as a dummy variable indicating unknown literacy. Standard errors clustered at the level of the firstborn are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Supplementary Bibliography

Aguero, J. M. and M. S. Marks (2008): "Motherhood and Female Labor Force Participation: Evidence from Infertility Shocks," *American Economic Review*, 98, 500–504.

Aguero, J. M. and M. S. Marks (2011): "Motherhood and Female Labor Supply in the Developing World: Evidence from Infertility Shocks," *Journal of Human Resources*, 46, 800–826.

Austerlitz, F. and E. Heyer (1998): "Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population," *Proceedings of the National Academy of Sciences*, 95, 15140–15144.

Axmon, A., L. Rylander, M. Albin, and L. Hagmar (2006): "Factors affecting time to pregnancy," *Human reproduction*, 21, 1279–1284.

Baird, D. T., J. Collins, J. Egozcue, L. H. Evers, L. Gianaroli, H. Leridon, A. Sunde, A. Templeton, A. Van Steirteghem, J. Cohen, P. G. Crosignani, P. Devroey, K. Diedrich, B. C. Fauser, L. Fraser, A. Glasier, I. Liebaers, G. Mautone, G. Penney, B. Tarlatzis, and coauthors (2005): "Fertility and ageing," *Human Reproduction Update*, 11, 261–276.

Bouchard, G. and J. Larouche (1989): "Nouvelle mesure de l'alphabétisation à l'aide de la reconstitution automatique des familles," *Histoire sociale/Social History*, 22.

Charbonneau, H. (1993): *The First French Canadians: Pioneers in the St. Lawrence Valley*, University of Delaware Press.

Charbonneau, H., B. Desjardins, J. Légaré, and H. Denis (2000): "The population of the St-Lawrence Valley, 1608–1760," *A population history of North America*, 99–142.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014): "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104, 2593–2632.

Christensen, K., H.-P. Kohler, O. Basso, J. Olsen, J. W. Vaupel, and J. L. Rodgers (2003): "The Correlation of Fecundability Among Twins: Evidence of a Genetic Effect on Fertility?" *Epidemiology*, 14, 60–64.

Cinirella, F., M. Klemp, and J. Weisdorf (2017): "Malthus in the Bedroom: Birth Spacing as a Preventive Check Mechanism in Pre-Transition England," *Demography*, 54, 413–436.

Cohn, B. A., P. M. Cirillo, M. S. Wolff, P. J. Schwingl, R. D. Cohen, R. I. Sholtz, A. Ferrara, R. E. Christianson, B. J. van den Berg, and P. K. Siiteri (2003): "DDT and DDE exposure in mothers and time to pregnancy in daughters," *The Lancet*, 361, 2205–2206.

Curtis, K. M., D. A. Savitz, C. R. Weinberg, and T. E. Arbuckle (1999): "The effect of pesticide exposure on time to pregnancy." *Epidemiology*, 10, 112–117.

de Cock, J., K. Westveer, D. Heederik, E. te Velde, and R. van Kooij (1994): "Time to pregnancy and occupational exposure to pesticides in fruit growers in The Netherlands." *Occupational and environmental medicine*, 51, 693–699.

Desjardins, B., A. Bideau, É. Heyer, and G. Brunet (1991): "Intervals between marriage and first birth in mothers and daughters," *Journal of biosocial science*, 23, 49–54.

Duan, N. (1983): "Smearing estimate: a nonparametric retransformation method," *Journal of the American Statistical Association*, 78, 605–610.

Ecochard, R. (2006): "Heterogeneity in fecundability studies: issues and modelling," *Statistical methods in medical research*, 15, 141–160.

Falconer, D. S. and T. F. C. Mackay (1996): *Introduction to Quantitative Genetics*, Benjamin Cummings.

Gagnon, A. and E. Heyer (2001): "Intergenerational correlation of effective family size in early Quebec (Canada)," *American Journal of Human Biology*, 13, 645–659.

Gagnon, A., B. Toupance, M. Tremblay, J. Beise, and E. Heyer (2006): "Transmission of migration propensity increases genetic divergence between populations," *American journal of physical anthropology*, 129, 630–636.

Gini, C. (1924): "Prèmieres recherches sur la fécondabilité de la femme," in *Proceedings of the Int. Math. Congress*, Toronto: North-Holland, 889–892.

Joffe, M. and I. Barnes (2000): "Do parental factors affect male and female fertility?" *Epidemiology*, 11, 700–705.

Joffe, M., L. Bisanti, P. Apostoli, P. Kiss, A. Dale, N. Roeleveld, M.-L. Lindbohm, M. Sallmén, M. Vanhoorne, and J. P. Bonde (2003): "Time to pregnancy and occupational lead exposure," *Occupational and environmental medicine*, 60, 752–758.

Juhl, M., A.-M. N. Andersen, M. Grønbæk, and J. Olsen (2001): "Moderate alcohol consumption and waiting time to pregnancy," *Human Reproduction*, 16, 2705–2709.

Juul, S., W. Karmaus, J. Olsen, E. I. S. S. Group, et al. (1999): "Regional differences in waiting time to pregnancy: pregnancy-based surveys from Denmark, France, Germany, Italy and Sweden," *Human Reproduction*, 14, 1250–1254.

Klemp, M. and J. Weisdorf (2019): "Fecundity, Fertility and the Formation of Human Capital," *Economic Journal*, 129.

Kosova, G., M. Abney, and C. Ober (2009): "Heritability of reproductive fitness traits in a human population," *PNAS*, 107, 1772–1778.

Lacroix, C. and B. Desjardins (2012): "Adult mortality in preindustrial Quebec," *Canadian Studies in Population*, 39, 23–33.

LANDE, R. AND S. J. ARNOLD (1983): "The measurement of selection on correlated characters," *Evolution*, 1210–1226.

LÉGARÉ, J. (1988): "A population register for Canada under the French Regime: Context, scope, content and applications," *Canadian Studies in Population*, 15, 1–16.

LIND, J. T. AND H. MEHLUM (2010): "With or Without U? The Appropriate Test for a U-Shaped Relationship," *Oxford Bulletin of Economics and Statistics*, 72, 109–118.

LOFT, S., T. KOLD-JENSEN, N. H. HJOLLUND, A. GIWERCMAN, J. GYLLEMBORG, E. ERNST, J. OLSEN, T. SCHEIKE, H. E. POULSEN, AND J. P. BONDE (2003): "Oxidative DNA damage in human sperm influences time to pregnancy," *Human Reproduction*, 18, 1265–1272.

MILOT, E., F. M. MAYER, D. H. NUSSEY, M. BOISVERT, F. PELLETIER, AND D. RÉALE (2011): "Evidence for evolution in response to natural selection in a contemporary human population," *Proceedings of the National Academy of Sciences*, 108, 17040–17045.

MOREAU, C., C. BHÉRER, H. VÉZINA, M. JOMPHE, D. LABUDA, AND L. EXCOFFIER (2011): "Deep human genealogies reveal a selective advantage to be on an expanding wave front," *Science*, 334, 1148–1150.

PAPKE, L. E. AND J. M. WOOLDRIDGE (1996): "Econometric Methods for Fractional Response Variables with an Application to 401 (K) Plan Participation Rates," *Journal of Applied Econometrics*, 11, 619–32.

PAPKE, L. E. AND J. M. WOOLDRIDGE (2008): "Panel data methods for fractional response variables with an application to test pass rates," *Journal of Econometrics*, 145, 121–133.

PETTAY, J. E., L. E. B. KRUUK, J. JOKELA, AND V. LUMMAA (2005): "Heritability and genetic constraints of life-history trait evolution in preindustrial humans," *PNAS*, 102, 2838–2843.

RAMLAU-HANSEN, C. H., A. M. THULSTRUP, J. OLSEN, AND J. P. BONDE (2008): "Parental Subfecundity and Risk of Decreased Semen Quality in the Male Offspring: A Follow-up Study," *American Journal of Epidemiology*, 167, 1458–1464.

ROBERT, A., B. TOUPANCE, M. TREMBLAY, AND E. HEYER (2009): "Impact of inbreeding on fertility in a pre-industrial population," *European Journal of Human Genetics*, 17, 673–681.

ROFF, D. (2012): *Evolutionary quantitative genetics*, Springer Science & Business Media.

SCHWARTZ, D. AND M. MAYAUX (1982): "Female fecundity as a function of age: results of artificial insemination in 2193 nulliparous women with azoospermic husbands." *Obstetrical & Gynecological Survey*, 37, 548–550.

WISE, L. A., E. M. MIKKELSEN, K. J. ROTHMAN, A. H. RIIS, H. T. SØRENSEN, K. F. HUYBRECHTS, AND E. E. HATCH (2011): "A prospective cohort study of menstrual characteristics and time to pregnancy," *American journal of epidemiology*, 174, 701–709.