A Brief History of Equality

Geoffrey Brennan, Gordon Menzies and Michael Munger*

February 2016

Total Word Count: 11,400

**Corresponding author:* Michael Munger, Box 90204, Department of Political Science, Duke University, Durham, NC 27708-0204. Fax: 919-844-0954. Phone: 919-369-6453. Email: munger@duke.edu

Abstract: Trends in distribution of wealth and income have profound implications for economic and political stability in societies. We point out an "Iron Law" of distribution: : Future inequality of wealth is decreasing in the income elasticity of family size. It is an "Iron Law" not because it is always the most significant influence on the income/wealth distribution at any point in time—it may not be—but because its effects are relentless, systematic and cumulative. If parents care more about their own children than others, then family size switches from being a brake on concentration of wealth (if wealthy families have more children) to rapidly concentrating wealth (if wealthy families have fewer children). As we show, this consideration, in surprisingly accurate and modern form, was recognized by Smith (1776) in anticipating Malthus (1798), connecting current concerns with income inequality with classic debates over population growth.

JEL Codes: B12; B16; D31

Acknowledgements: The authors acknowledge the comments and assistance of the Economics Discipline Empirical Reading Group seminar at University of Technology—Sydney. In addition, Jonathan Anomaly, Bruce Caldwell, Michael Gillespie, Olena Stavrunova, and Georg Vanberg offered useful comments but deserve no blame for the errors that remain.

A Brief History of Equality

I Introduction

I.A. Overview

Distribution and inequality of wealth are perennial concerns, both for academic economists and for practical politics. But interest in the sources and consequences of inequality has recently burgeoned, perhaps because of (apparent) trends towards increasing inequality within, and across, countries[1]. The casual factors identified in the explanatory accounts of this phenomenon are varied: Piketty (2014; 2015) emphasizes the role of relative factor prices; Cowen (2014) identifies a major cause as lying in increased globalization and increased acquisition of human capital in developing countries.

There is, in our view, one explanation for long-term trends in inequality that is so fundamental, and so obvious that it seems (paradoxically) nearly to have escaped attention. We refer to this mechanism as the Iron Law of intergenerational transmission of dispersion (Iron Law for short).[2] The Iron Law is simply this: *Future inequality of wealth is decreasing in the income elasticity of family size.* It is an "Iron Law" not because it is always the most significant influence on the income/wealth distribution at any point in time—it may not be—but because its effects are relentless, systematic and cumulative. The overall tendency for inequality to increase, or decrease, over successive generations operates constantly according to the sign of that elasticity, and the speed of the effect is determined by the magnitude of that elasticity, whatever else is going in national economies. Like other Iron Laws in the social

---

[1] For a detailed examination of the phenomenon itself see Atkinson et al (2011); for a general introduction, see Blinder (1973).

[2] The first use of "iron law" in this context appears to have been from Goethe, in 1783, and then used by Malthus (1798). Karl Marx in Part II of his "Critique of the Gotha Program," (1875), referring to Lasalle's "Iron Law of wages," which Marx claimed was borrowed from Goethe's line from "DAS GÖTTLICHE" : "As great, eternal iron laws dictate, we must all complete the cycles of our existence." Lasalle was reformulating the Malthusian (Malthus, 1802) that the working population would be tightly constrained by the food supply, and that wages would adjust to ensure that the constraint was binding. Roberto Michels (1911) later advances what he calls "The Iron Law of oligarchy." We use the phrase not to be grandiose, but to emphasize that the "law" has mechanistic features and tightly constrains events.

sciences, this one operates 'invisibly' in Adam Smith's sense – the aggregate patterns it produces are in no way conscious intention of any of the participating agents.

The theory is explanatorily 'parsimonious' in the sense that its operation depends on nothing more than elementary logic and one (robust) empirical assumption about the human species. That assumption is simply that parents and testators care more about their own children than they do about others' children. The consequence is that parents make transfers to their own children during their (the parents') lives, and then leave their estates primarily to their own children.[3]

The Iron Law is derived, then, from parents' transfer to children of ordinary private goods, which by definition are 'rival in consumption'. This implies an economic analogue of 'sibling rivalry', which we shall term 'sibling rivalness' to distinguish it from its familiar psychological cousin. Sibling rivalry is contingent: it may be common but it is not logically necessary. Sibling rivalness is a logical necessity, because the magnitude of the transfers received by each offspring is *ceteris paribus* a negative function of family size.

It is easy to conflate the fact of inheritance—the passing of a fortune to one's children—with the particular mechanism we identify, which derives from inheritance interacting with family size as a function of wealth. For example, Wedgewood (1929:60-61) claimed:

> … Inheritance perpetuates and may intensify inequalities arising originally from other causes. In that sense, it is a secondary cause of inequality; but that is not, of course, to say that it is of secondary importance. The extent of its influence on distribution remains an open question, which cannot be decided merely by theoretical reasoning … but requires in addition something in the nature of a quantitative analysis of the relevant facts.[4]

[3] In the absence of such transfers, it is doubtful whether the human race could survive. In this respect, the human species is like others in which it takes a significant period before offspring are self-sufficient. In the human case, of course, what counts as "self-sufficiency" is itself historically dependent. And it is clear that the psychological factors of natural affection (and norms of parental duty) that cause parents to look after their children do not disappear once children are self-sufficient: transfers proceed up to and including death. For evolutionary purposes of course what matters is survival rate to age of reproduction. An interesting model of the impacts of different motivations for intergenerational transfers is Zilcha (2003).

[4] Quoted in Davies (1982).

Fortunes would not inherited if there were no inheritance, of course. But our focus is on a dynamic aspect of the problem: family size in one period is a (socially and culturally determined) function of the effect of family income. But then family size in the first period determines the dispersion of that income among children, and laters heirs, in the following period. To exemplify, suppose that the income elasticity of family size in a cultural epoch is positive – the richer the family, the more surviving offspring, *ceteris paribus*. Then the intra-family lifetime transfers (the sum of nutrition, education, inheritance, etc.) to offspring will be divided among a larger number of claimants in richer families. This implies that the ratio of the incomes of rich families compared to poor families in a period will be greater than the ratio of the incomes of their offspring, because the wealthy testators will have their wealth divided among more heirs. This effect will substantially mitigate, and might even reverse, tendencies in the economy that might otherwise concentrate wealth in fewer hands. So long as wealthy transferers have more recipients, the operation of the Iron Law is anodyne: Dispersion in the distribution of income over time is inherently self-limiting.

The problem is that the Iron Law also operates if income elasticity of family size has a negative sign. But then the effect is far from anodyne. In fact, if increasing income is associated with smaller family size, some of our basic assumptions about the structure of democratic societies may prove untenable. Concentrated wealth becomes more concentrated, and the variance of income increases rapidly. This effect is not self-limiting, but is relentlessly accelerated by the interaction of family size, intra-family transfer, and subsequent family size.

This logic suggests that the empirical sign of the income elasticity is of central importance. And the trend in different societies shows a striking—and for the reasons described above, sinister—pattern. Society after society has reached, at different times, but at approximately the same stage of development, a point where income increases come to be associated with fewer, rather than more, children. For example, Becker (1981) and Becker and Tomes (1994) have suggested that, in modern times, the higher opportunity cost of children for families with greater human capital will mean that richer families will tend to be smaller. Österberg (2000) and Lindahl (2008) offer some empirical estimates of the effects of this changing elasticity on "income mobility," and support the idea that the pattern is important, though worrisome.[5]

Our contribution is to nail down the underlying logic of formal relationship rather than relying on empirical estimates. Further, it turns out that there is precedent for our theoretical claims: Smith (1776) foreshadows Malthus (1798) in emphasizing the positive causal connection between income and population – which in Smith is driven by the rate at which "a great part of the children which … fruitful marriages produce" are "destroyed" (op. cit. p.98). So at the level of nations, it is a familiar observation that the rate of growth in income during the Malthusian era was associated with increasing population; but in the modern era, higher income is associated with lower population growth. Supposing that the national experience is mirrored at the individual level, the obvious conjecture is that in the Malthusian era richer families had larger numbers of surviving children and poorer families fewer surviving children because of the differential capacity to protect offspring against malnutrition and disease.

The importance of the dynamic process we adumbrate is hard to overstate. At its base is a key observation: in the modern era, as survival rates rose, family size became a

[5] Our claim that this effect, in its full force, has been overlooked is borne out by Atkinson (2015, pg. 159) who cites us (Brennan et al., 2014) when referring to the mechanism, though not the name, of the Iron Law. The idea is also mentioned briefly in the context of inherited wealth in a survey chapter, Davies and Shorrocks (2000, pg. 622), "… if the wealthy consistently have fewer children, inherited wealth can become continuously more unequal." However, we would argue that the Iron Law has a far wider jurisdiction than mere inheritance, operating as it does via in vivo transfers or via beneficial parental attention, which differentially enhances the human capital of children in small families.

‘demand-side’ phenomenon (dominated by the birth rate) rather than a supply-side phenomenon (dominated by the survival rate). And Becker’s emphasis on demand-side considerations (in determining the choice between quantity and quality of offspring) becomes increasingly relevant.

On this basis, the Iron Law suggests a simple reduced-form “history of equality”. In the Malthusian era, the Iron Law tended to produce ever greater equality: it suppressed the incomes of the offspring of richer, larger families vis-a-vis the offspring of poorer, smaller ones. In that sense, the Iron Law was favourable to the emergence of an ever-expanding ‘middle class.”

In the modern era, by contrast, as wide-spread availability of contraceptive measures meant the predominance of demand-side considerations in the determination of family size, the iron law’s effects are reversed: the forces of differential family size across income classes involve ever greater inequality (whether measured in terms of wealth or income).

Our description of the iron law’s operation involves an independent role for income as a determinant of family size – in part because the economic literature[6] suggests independent reasons why income should play an independent causal role. However, it is worth noting that the iron law could in principle operate without any independent role for income at all.

To illustrate, suppose in a situation of identical incomes in period, there is some random variation in family size. Sibling rivalness effects will still be in operation and ceteris paribus children from smaller families will receive larger intra-family transfers than children from larger families. In the next generation, the same random effects will be in evidence: the

[6] Both Smith (1776) and Malthus (1798) offered speculative reasons why this might be true. The modern economic and demographic work on this question suggests two vectors of influence: capacity and taste. Capacity is related to the ability of a family to provide food and medical care, meaning that more offspring will survive, or that mothers may survive childbirth more often. Taste is the decision to have children, so that if economic factors matter at the margin the effects of price, income, and other variables will be in the expected directions. Becker’s method is important, because he ruled out routine resort to “changes in preferences” as an explanation for observed changes. That is not to say that tastes, culture, and societal norms are invariant. But explanations based on changes in observable parameters should be privileged in explanations, in this view.

proportion of rich individuals with a ‘small family history’ will tend to increase over time, whatever other factors in income determination are in play.

I.B. Caveats

Some clarifications and caveats should be inserted at the outset.

1. The mechanism we focus on makes an implicit assumption of perfect assortative mating. Mating patterns are clearly important and in the limit, systematic offsetting mating could undermine the entire mechanism, though we find this an implausible scenario.

2. The Iron Law involves no claim of any genetic connection between the income-earning capacities of parents and children. Indeed, any such would presumably operate as an intra-family ‘public good’ to all siblings; and hence not be subject to the sibling rivalness effect. In this sense, any claim concerning a putative relation between income elasticity of family size and the aggregate quality of the genetic pool is entirely alien to the spirit of this paper.

3. Although we have illustrated the central property of sibling rivalness via appeal to ‘economic goods’ narrowly construed, we do not think the phenomenon is so restricted. For example, we take the apparently robust empirical finding in the sociological literature that educational performance is negatively correlated with family size (other things equal)[7] to emphasize that parental attention and energy in relation to their children’s education is no less ‘rival’ in the economic sense than food or Princeton fees.

4. This is not an empirical paper. We make no attempt here to assess how significant the iron law effects are in the current situation. Our focus is on the mechanism itself, which operates

[7] See for example Blake (1989), Alexander and Cherlin (1990) and for the developing world, Dang and Rogers (2013).

always in the background. We consider that mechanism sufficiently interesting to be worth independent conceptual scrutiny.

5. Although bequests are one important mechanism of intergenerational transfer, they are not the most significant for most families. Over most of human history, because humans do not reach material independence for an extended period (which may itself be endogenous), much the most significant intra-family transfer receipts for most individuals are those that occur over the first decade or so of the individual's life. Even in the relatively affluent modern era, bequests are likely to represent a small part of total lifetime transfers from parents to children, except for the very richest, perhaps because human capital has become more important than relatively alienable claims on physical capital. Accordingly, although we use some bequest data later as the basis for measures of the income elasticity of family size, we want to emphasise that we do not see bequests as the main game in the operation of the Iron Law.

I.C. Outline

The lay-out of the discussion is as follows.

In section II, we illustrate the effects of the iron law by appeal to some simple simulations. Part of the objective here is to show in simple arithmetic terms what the effects on the Gini coefficient of the Iron Law's operation are, under various possible values of the key parameter – the income elasticity of family size. In developing those simulations we take it that the income level strictly determines family size, so that there is no independent variance of family size within income classes.

In section III, we offer a somewhat more general treatment by taking a Taylor series approximation of the relation between child income and parent income – and allowing explicitly for variance in the relation between family size and parental income.

In section IV we turn to the historical material, using secondary data to try to locate different periods of the value of the income elasticity of family size; and exploiting Malthusian period data to buttress our claims about the importance of income-mediated factors. The function of these empirical investigations are essentially to justify the particular parameter values used in the simulations in section II.

Section V draws together some brief conclusions.

II The Operation of the Iron Law

In this section we simulate the operation of the Iron Law to see how it might affect inequality, measured in a standard way. The setup is deliberately simple, and we focus on different values of a key parameter – the income elasticity of family size – and exclude all other influences on family size. In the next section we offer a somewhat more general treatment, modeling other influences on family size by an i.i.d. error.

There are many influences on income determination, of which the Iron Law is only one. Moreover, transfers to children and the acquisition of human capital might occur over an extended period, so at any point in time the operation of the Iron Law might be invisible to both econometricians and the actors alike. However, despite this, we demonstrate in this section that the cumulative effects can be sizable.

Our simulations examine the changes in inequality over three generation to confirm, firstly, that they are small over a timeframe of interest to short-sighted policymakers, secondly, that there are no reversals and all the compounding heads in one direction, and, finally, that the relentless compounding delivers a cumulated effect that is non-trivial.

To abstract from economic growth, which is not *per se* determinative of relative incomes, the simulations start from a state where the distribution of income is normalized with Uniform distribution over (0, 1). We assume that income in cohort i, Y_{it}, is divided equally among individuals in a family, N_{it}, and therefore that income available for the next generation is proportional to per person income:

$$Y_{i,\,t+1} = n\frac{Y_{i,t}}{N_{i,t}} \tag{1}$$

Where n is the average number of people in household across all income classes. We assume perfect assortative mating by men and women on the same income level, and, the pre-multiplication by average family size n is so that the income of successive generations does not decline on average.[8]

Agents live for two periods. Over the first period, children are born according to the income of their parents, as dictated by ($N_{it} - 2$) in (2), and acquire human capital and other *in vivo* transfers. At the end of the first period, their parents die and the sum total of *in vivo* transfers plus bequests is given by (1). In the next period, they become parents, and make transfers to the next generation. The number of children born in a period ($N_{it} - 2$) is determined only by income in this section, and therefore the average family size n in (1) is determined only by average income y.[9]

$$N_{i,t} = \alpha + \beta Y_{i,t} \qquad\qquad n = \alpha + \beta y. \tag{2}$$

The simulations loop through three iterations, for each new generation.

Step 1: The income for each cohort is specified.

Step 2: Family size is determined by (2), and then the income for the children who are coming of age in each cohort is determined by (1).

Step 3: Income inequality is measured for the new generation.

We then return to step 1 using the new distribution of income and loop through the three steps again.

[8] As Fernández Rogerson (2001) point out, assortative mating alone will have some impact on inequality.

[9] We will assume two parents throughout this model.

To operationalize and interpret the simulations an initial distribution of income is needed for the first time Step 1 is performed. Then, there needs to be a choice of α and β in (2) for Step 2 and, finally, a choice between measures of inequality in Step 3.

For any linear distribution of income, including *Y~Uniform (0,1)* which is the starting distribution for the simulations, the two most natural measures of inequality – the Gini coefficient and the standard deviation – are very close.[10] The importance of this result lies in the fact that it is much easier to work with standard deviations analytically than with Gini coefficients, and we shall exploit this in the next section. Accordingly, we only need to plot the Gini below, and the reader can be confident the standard deviations's are virtually the same.[11]

We now show the influence of β for the operation of the Iron Law in two contrasting cases. In what follows the elasticity of family size with respect to income evaluated at the means of y and n, namely $\beta y/n$, will be denoted η.

In the first case, which for reasons we will elaborate later we call the *Mathusian Era,* there is a positive income elasticity of family size (η is positive because β is positive). In our simulations we assume the poorest families had only two surviving children whereas the richest families, better able to protect their progeny from the vicissitudes of disease and starvation, had an average of between 3 and 4 surviving children. This corresponds to an intercept and slope pair (α, β) of *(4.0, 1.55)* in (2).

[10]The Gini coefficient is based on the so called Lorenz curve, which measures cumulative income against cumulative population. A Gini of zero indicates no inequality (see Dorfman, 1979). Let $f_Y=1-(b/2)+bY$ be the linear income pdf over *(0, 1)* with slope b, constructed to integrate to unity. The standard deviation of Y is $\frac{1}{3}\sqrt{\frac{3}{4}-\frac{b^2}{6}}$ and we use Dorfman (1979) $G=1-\frac{1}{E(y)}\int(1-F_y)^2 dy=\frac{1}{3}\left(\frac{1-b^2/20}{1+b/6}\right)$ to obtain the Gini where F_y is the cdf. Now in order to guarantee a positive pdf it must be the case that $-2 \le b \le 2$ so these expressions will not be too far numerically from *1/3* for many admissible values of b. The standard deviation and the Gini are exactly $1/\sqrt{12}$ and *1/3* when f_Y is uniform, since $b=0$ in that case.

[11] The simulations that follow can be replicated in an online appendix and the reader can confirm that the Ginis and standard deviations are indeed very close to each other.

In the second case, called the *Modern Era*, there is a negative income elastic of family size (η is negative because β is negative). Here, the demand for children is far more important for family size than survival technology. We assume the richest families have between one and two on average, while the poorer families have two. We choose the parameters *(4.0, -0.35)* to generate this behavior.

Later we will justify our choice of the slope and intercept parameters used in the two cases. Briefly, the (α, β) pair from the Malthusian era come from regressions using estates in England prior to the industrial revolution (Clark and Cummins, 2010), and the values for the Modern era are derived from US data in the second half of the twentieth century (Jones and Tertilt, 2008).

The simulations are shown in Figure 1 and are available in detail in an online appendix. In Table 1 we then provide the compounded change in the standard deviation over three generations. We have checked that the Gini coefficients in our spreadsheets are virtually identical to the standard deviations, so we will freely interchange these two terms as we discuss the results. The per cent change in the Gini over one year in the second row of Table 1 is calculated assuming that each generation comprises one third of a century, and so we take $[G_{end}/G_{start}]^{1/100}-1$ as our estimate for the annual proportional change.

Figure 1: The Compounding of the Iron Law: Malthus vs. Modern

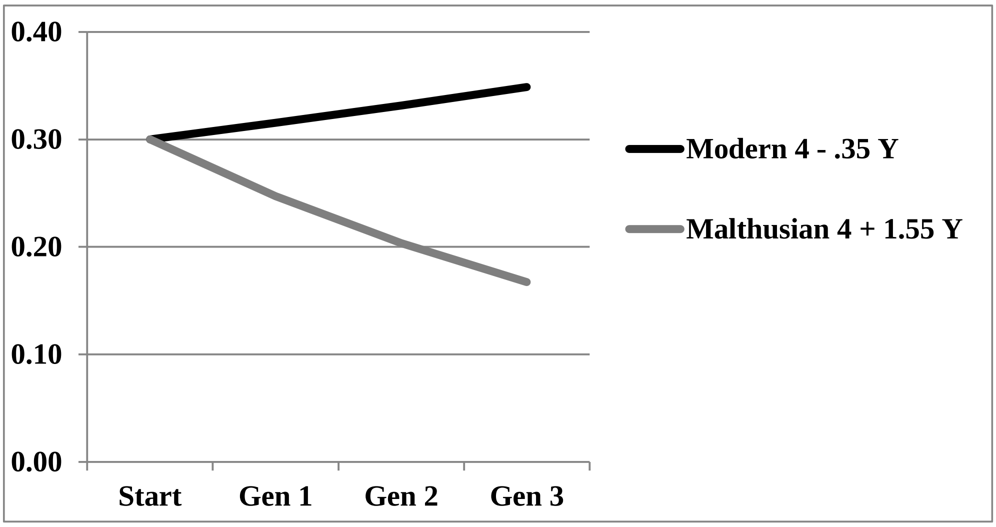


Figure 1 has several interesting features. Measured inequality exhibits a sizable change over the period. Using the Gini equivalence we may compare the numbers in Figure 1 with some contemporary Gini estimates (Corak, 2012). The starting Gini value of just under 0.3 is close to current estimates for Germany. The endpoint for the modern era, just under 0.4, places inequality at the upper end of inequality in developed countries, towards China and the US (both just over 0.4). The decline in the bottom axis is very large, considering that the Scandinavian countries, which are known for their low inequality, have Ginis between 0.2 and 0.3.

Table 1: Inequality Effects of the Iron Law

	Malthusian	Modern
% Δ over three Generation	-44%	16%
% Δ over one year (3 Gen = 100 yrs)	-0.58%	0.15%

These sizable effects – nearly 50 per cent lower in the Malthusian era and 16 per cent higher in the Modern era – are possible because there are no reversals in the compounding of the Iron Law. The effects on each generation are relentlessly one-directional for all simulations, as flagged in the introduction. The contrast between the numbers in the two rows of Table 1 is startling, and the reconciliation between the two hinges on the timescale of the bottom axis of Figure 1. Three generations amounts to one century, and so while it is hardly surprising that the annual changes are small, it amounts to a significant policy-relevant fact. These kinds of changes would be very difficult to detect econometrically, with estimated coefficients susceptible to attenuation bias.[12] Crucially, with an impact horizon extending far beyond the political cycle, they would scarcely attract the interest of a short-sighted policymaker.

To reinforce the message of both Figure 1 and Table 1, we recap the basic Iron Law mechanism. For countries in the Modern era, any dynasty with high income has fewer children amongst which to divide their wealth. These children will therefore find themselves in the higher end of the income distribution in their adult lives and, since the only thing affecting income in these simulations is the number of children and starting income, these children will themselves have fewer children who will in turn gain a greater transfer from their parents, and so on.

In the Malthusian era, rich parents have more surviving children than their poor counterparts and so a dynasty that starts out rich will find its wealth diluted as it is progressively spread out over relatively large families in its dynasty. In subsequent generations, the same effect will repeat itself, compounding towards greater equality.

Actually, the effects for the Malthusian era in Figure 1 are so sizeable that it is worth reiterating that this is a stylized ceteris paribus exercise. There are many factors that acted to hold up inequality in the Malthusian era, not least primogenitor conventions which we sidestep by assuming in (1) that income is divided equally.[13] Furthermore, according to Adam Smith

[12] As is well known, a mismeasured RHS variable in a regression will cause the estimator of the coefficient to be biased towards zero – so called attenuation bias. Small or gradually evolving effects are therefore notoriously difficult to capture in econometric analysis, because in the presence of measurement error estimates close to zero run the risk of being deleted on a significance criterion.

[13] As Menchik (1980) notes, primogeniture could perpetuate inequality, once it has already been established for some other reason.

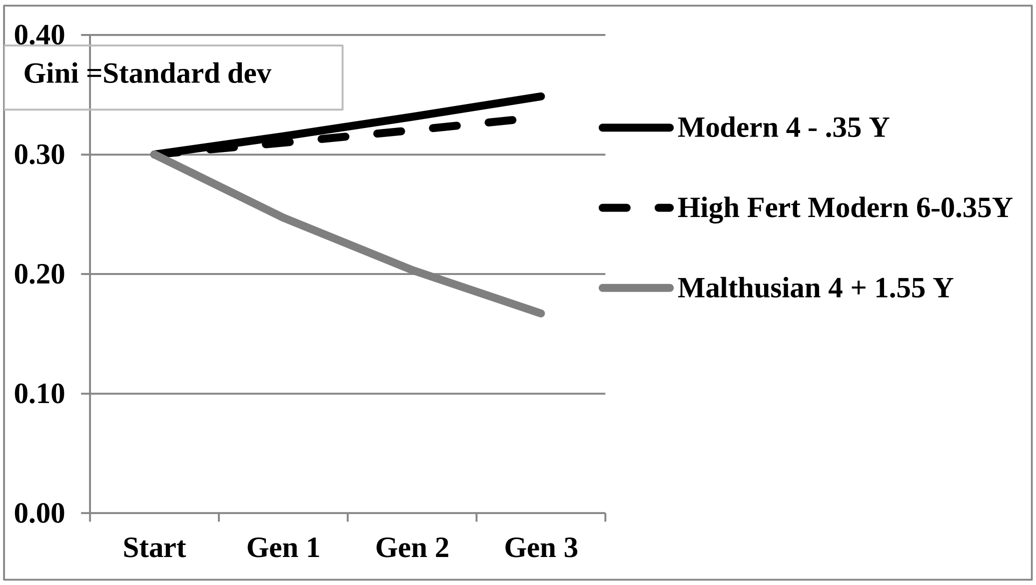
(quoted later) the tragedy of child bereavement was far more common among the poorer classes, so that the income in (1) for poor people was depleted by children who did not survive. That said, the impact of primogenitor was probably muted outside the upper echelons of society in the Malthusian era, since poorer people may have lacked significant bequests, and there is an offsetting factor that holds up Malthusian inequality within the Iron Law framework itself, as we shall see in the next section.

We have so far operated on η the elasticity of family size with respect to income, by positing changes to β. However, there is another dimension along which to engineer changes in this elasticity. If we now consider a high-fertility modern country, such as those found in the developing world, we note that a higher value of α will result in a lower elasticity for a higher mean household size n (engineered by increasing α in (2)). From (1) this in turn will have less impact upon per person income in subsequent periods, so the compounding of the Iron Law will be weaker.

We pursue this thought in Figure 2, where we define a *Modern Era (High Fertility*[14]*)* regime for countries that exhibit a negative correlation between income and family size, but where all families have more children on average. We assume the poorest families have four children but richer families have between three and four, so that the intercept and slope pair is *(6.0, -0.35)*.

[14] We use the term *fertility* like demographers, who reserve the term *fecundity* for the ability to bear children.

Figure 2: The Compounding of the Iron Law: Low vs. High Fertility



The movement of the high fertility country from 0.3 to somewhere around 0.35 it is rather like a move between Germany and the UK, where the UK has somewhat more inequality than many other countries in Europe. As expected, the compounding exhibited in the dashed line is less than the Modern Era simulation.

We have shown in this section firstly that the Iron Law gains its strength from its relentless compounding power. We take it as fact that rivalness of sibling consumption is such a robust claim about the human species that the Iron Law will indeed be permitted to operate over the kinds of timescales that make this compounding nontrivial. Yet, somewhat paradoxically, we have also shown that the Iron Law, by virtue of its underwhelming annual progression, may be invisible to those with the most to profit from knowing about it, namely policymakers.

We have also shown the impact of the key drivers of the elasticity of family size with respect income, namely β and n, on the compounding of the Iron Law. The compounding is negative for a positive β and positive for a negative β. However, the Iron Law is attenuated for high fertility countries, because changes in income have a smaller *proportional* impact on N_{it} in (2).

III Compounding, Entropy and the Iron Law

In this section we derive an analytic representation for evolution of income inequality due to the Iron Law, measured as the square of the standard deviance (the variance). Our first main finding is that modern societies with low numbers of children are particularly susceptible to the compounding of the Iron Law towards inequality. When we introduce other uncorrelated factors that bear on the numbers of children – what we call entropy – we find that this acts to increase inequality in any era, which is our second main finding. In the Mathusian era entropy puts a brake on the large Gini declines seen in Figure 1, which is a welcome fillip to the realism of the model.

Our point of departure from section II is to rewrite (2) with other factors uncorrelated with income on the RHS. We suppress subscript for dynasty *i* and think of *Y* and *N* as random variables distributed across individuals.

$$N_t = \alpha + \beta Y_t + u_t \qquad (3)$$

Equation (3) obeys some standard assumptions, namely:

$$Cov\ (N_t, Y_t) = \beta V(Y_t) \qquad (4)$$
$$V(N_t) = \beta^2 V(Y_t) + \sigma^2, \qquad V(u) = \sigma^2. \qquad (5)$$

We take a Taylor series linearization of income in period *t+1* and work out its variance. It turns out to depend on the variance in period t, and the covariance between income and family size in period *t*.

$$Y_{t+1} = n\left(\frac{Y_t}{N_t}\right) \approx n\frac{y}{n} + n\frac{1}{n}(Y_t - y) + n\frac{-y}{n^2}(N_t - n) = y + Y_t - \frac{y}{n}N_t. \qquad (6)$$

We then take the variance of this linearization.[15]

[15] A subtlety here is that the variance in t+1 uses the distribution of income in that period. The distribution is altered by different numbers of children populating different income cohorts. It turns out that the pre-multiplication by *n* elegantly addresses this issue, as proved in Appendix 1.

$$V\left[n\left(\frac{Y_t}{N_t}\right)\right] = V[Y_{t+1}] \quad \approx V[Y_t] + \left(\frac{y}{n}\right)^2 V[N_t] - 2\frac{y}{n}Cov(Y_t, N_t) \qquad (7)$$

The appearance of a negative covariance term is obvious upon reflection. If the numerator and denominator of the ratio of two random variables move together (positive covariance) a high value in the numerator will be offset, on average, by a high value in the denominator, However, if the denominator tends to fall, on average, when the numerator rises (negative covariance) some very large ratio draws will occur. We can now substitute the above expressions for $Cov(N_t, Y_t)$ and $V(N_t)$ into (7) to obtain the time series process for the variance.

$$V_{t+1} = \left(1 - \frac{\beta y}{n}\right)^2 V_t \ + \left(\frac{y\sigma}{n}\right)^2 = \qquad (1-\eta)^2 V_t \ + \left(\frac{y\sigma}{n}\right)^2 \qquad (8)$$

Crucially, the autoregressive process for inequality in (8) only converges when the elasticity is positive, in the Malthusian era. In this case, the steady state variance is given by (9).

$$V = \frac{(y\sigma/n)^2}{1-(1-\eta)^2} \qquad (9)$$

Simple comparative statics for the Malthusian era can be inferred from (9). Inequality in the steady state is increasing in income (y) and entropy (σ) and decreasing in fertility (n) and the income elasticity of family size (η). Outside the Malthusian era, we have to rely on the growth rate in the autoregression, which can be inferred from the growth rate of V_t in (8).

$$\frac{V_{t+1}}{V_t} - 1 = (1-\eta)^2 - 1 \ + \ \frac{\left(\frac{y\sigma}{n}\right)^2}{V_t} \qquad (10)$$

If η is negative this is always positive, which is another way to say that inequality is explosive *ceteris paribus* in the modern era. However, we can confirm our comparison of profiles in the modern era in Figure 2. The lower n is (for given β) the faster will be the growth in inequality.

If on the other hand η is positive we can distinguish two cases: First, if V_t is growing it does so at a decreasing rate. The RHS of (10) will fall to zero as the last term diminishes and stops there (because V_t will no longer change). Second, if V_t is falling it must do so at an increasing because as V_t falls the last term rises the RHS of (10) will rise to zero and stop there (because V_t will no longer change).

We conclude from (9) and (10) that entropy *($\sigma > 0$)* must increase the growth rate of inequality (from (10)) in either the Malthusian or Modern era and the level of inequality in any existing steady Malthusian state (from (9); and, that the property of the relentlessness of the Iron Law is unperturbed by entropy, meaning that any positive/negative growth in V_t continues monotonically. All these points are illustrated in Figure 3.

Figure 3: Entropy and Inequality (dashed with entropy)

Taking the left panel first, there is no steady state in the modern era, but we know that for an arbitrary starting value (the black dot) the growth rate must be greater when entropy exists. That is, independently of any causal relationship between income and family size volatility in family size creates extra inequality because of rivalness.

In the right panel, we note from (9) that a steady state must exist in the Malthusian era, that it is independent of the starting value, and that the steady state is increasing in the level of entropy.[16] Thus we show the fixed points as the closed dot for the no-entropy case and the open dot for the entropy case. Would inequality have increased or decreased due to the operation of the Iron Law in the Mathusian era? Clearly, from the panel, it depends on the extent of entropy and the starting value.

Three cases can be distinguished. If inequality starts very low (below *a* on the axis) it will increase to one or other of the steady states shown by the dots. If inequality starts below the entropy steady state and above the no-entropy steady state (between *a* and *b*), it will rise if there is entropy and fall if there isn't. Finally, if inequality starts very high (due, say, to a rigid social hierarchy) then the operation of the Iron Law will be to reduce inequality, with or without entropy.

We conclude this section with a number of observations about this model. First, from equation (11), which is (10) with $\sigma=0$, we can forge a link to section II. In the absence of entropy (the world of section II), we have that the growth rate of the standard deviation is the negative of the income elasticity. (And we recall that the standard deviation is approximately the Gini).

$$\sqrt{V_{t+1}} = (1-\eta)\sqrt{V_t} \tag{11}$$

Over three generations we then have that the proportional change in the standard deviation should be roughly three times the standard deviation, if η is small. This insight, unpacked in Table 2, provides a useful cross-check on our simulations from in section II. In the last two rows of Table 2 we compare the compounding from the simulations with the theoretical compounding implied by equation (11), both over three generations. They are indeed very close.

[16] There is a case where there is explosive growth in inequality in the Malthusian era. If the income elasticity of family size is greater than one, relatively rich families have relatively poor children, since the numbers of children rise so much that the ratio in (1) falls. Beyond an elasticity of two, this effect is so strong that it leads to explosive inequality.

Table 2: No Entropy Comparison

	Malthusian	Modern	Modern (High Fertility)
α	4	4	6
β	1.55	-0.35	-0.35
y	0.5	0.5	0.5
n	4.775	3.825	5.825
$\beta y/n=\eta$	0.162	-0.046	-0.030
3rd Generation ***(11)***	-41%	14%	9%
3rd Generation ***Section II***	-44%	16%	10%

Armed with an analytic solution, we are now in a position to explain the differential Modern era compounding between columns two and three, which was a brute fact in section II. Noting that the denominator of $\beta y/n$ is itself $\alpha+\beta y$ it is immediately obvious that if α rises for a given β, as it does in moving from column two to three, then $|\beta y/n|$ falls and the compounding is reduced.

The intuition is that N appears on the denominator of nY/N and so as the mean of N drops nY/N is exposed to low draws of N which, in view of this inverse relationship, will boost the children's incomes a lot. To put the point straightforwardly, the sibling rivalness implications of a family having two children versus one child are very much more substantial than the rivalness implications of having five children in a family that already has four.[17] It is mere conjecture, but our suspicion that this is rarely a topic of conversation among parents

[17] A nice line of mathematical reasoning shows that the Iron Law greatly accelerates the removal of inequality in the Mathusian case, to the point where everyone attains the mean in a single generation (with the Gini being zero in this case). We can take the limit of nY/N as α approaches zero in the Malthusian case (the limit does not exist for the modern era, because of the assumption that n and N must always be positive) and we obtain $nY/N \rightarrow (\beta y)\, Y\,/\,(\beta Y) \rightarrow y$.

contemplating an extra child is what led us to float the idea, in the introduction, that the Iron Law is invisible in the Smithian sense – that is, invisible to the actors themselves.

The combination of compounding and entropy in the general case (equation (8)) helps explain a puzzle from the last section.

We noted there that the Iron Law predicted inequality in the Malthusian period should have withered quickly, leaving the impression that the poverty stricken and hierarchical Middle Ages was a more-than-Scandinavian paradise of equality. However, entropy, which we shall show in section IV abounded during the pre-modern era, puts a break on the fall in the Gini, and implies a non-zero steady state for the general equation (9). We will quantify these effects in the next section, and show that Mathusian entropy was quite high (a standard deviation of 2.23 children).

We conclude this section by reiterating that low levels of fertility in the modern era strengthen the operation of the Iron Law. We have already commented on the difference between columns two and three of Table 2. We now make the additional observation that entropy – which was not a feature of Table 2 at all – always heads in the direction of greater inequality. In considering the Malthusian era, this was a matter of comfort, at least to us who are promoting the Iron Law as an explanatory tool, since it rescued the Iron Law from making a perilously optimistic prediction about the Malthusian world. However, when we consider the modern world, the second term in (8) can only fan the flames of inequality, making any policymaker who wishes otherwise busier, and hotter, than ever.

In particular, a policy maker concerned about inequality can be both encouraged and discouraged by the general equation (8) with compounding and entropy.[18] The good news is that future inequality drives off current inequality, and so all the standard tools of addressing

[18] The last part of this section presumes without critique that one wishes to alter inequality.

inequality now through the tax and transfer system will have an intergenerational legacy.[19] Furthermore, the Iron Law operation via compounding gives *extra policy teeth* in the Modern Era (where β is negative and so $1-\eta > 1$) because this enhances the transmission from one generation to the other. Pulling families in from rich and poor extremes towards the centre of the distribution in the current generation reduces the variation in family size across the income distribution *ceteris paribus* attenuating the differential impact of sibling rivalness. However, the existence of entropy in the equation also is cause for pessimism. Even if redistributive policy could create a perfectly equal society now, the operation of the Iron Law via entropy sows the seeds of inequality afresh every generation.

Consideration of the Iron Law puts two standard policies front and centre in any debate about intergenerational inequality. Public education, which tries to take one of the most important forms of a child's capital accumulation out of the reach of the Iron Law, has an obvious appeal. It is not a panacea, though. Parental attention or any other family factors subject to sibling rivalness are complementary inputs into the development of a child's human capital (Blake 1989), even if they are combined in 'public school production function'.

Inheritance taxes, too, are worth considering (Davies, 1982) as a means to diluting the pool of resources subject to sibling rivalness. Perhaps inheritance taxes would gain greater appeal if people understood the current implicit system of 'taxation' derived from the Iron Law, where the share lost to siblings depends on the number of siblings present, and the prospective number of siblings.

[19] There is substantial evidence, of the sort produced and reviewed by Roine and Waldenström (2008), and Roine, Vlachos, and Waldenström (2009), that such policies can change both behavior and levels of inequality. Our goal is to identify the underlying tendency in the absence of such institutions.

Table 3: Implicit Inheritance 'Taxes' (Decline in Personal Inheritance assuming two parents[20])

Existing siblings / extra siblings	1	2	3
1	-25%	-40%	-50%
2	-20%	-33%	-43%
3	-17%	-29%	-38%

IV Learning from the Past: Changes in the Income Elasticity of Family Size

In section II, we conducted a set of simulations to illustrate the effects of the iron law – and to make those effects apparent, we contrasted a case where the income elasticity of family size (ß) is positive from a case in which it is negative. We termed the first case "modern" and the second "Malthusian". In those simulations, we used particular values for α and β in equation (2) to indicate the effects over specific periods.

Our task in this section is to justify the values so selected for the two contrasting periods; and to indicate the time spans that these 'periods' represent.

The Modern Era:

Although for many countries, data on the connection between income and fertility is not easy to come by, we illustrate in Table 4, the relation between household wealth (not income) and fertility for a range of countries within the contemporary period (last fifty years say). On this basis, it seems reasonable to conclude that the income elasticity of family size is negative (or at best zero as for the Ukraine) across a range of countries for that period. Reliable data *is* available for the US on the relation between fertility and income within the US back through

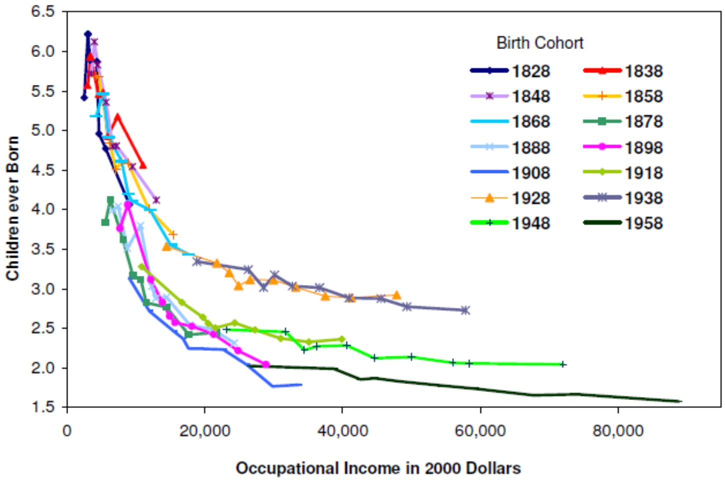
[20] Calculation: Consider the top right cell. We are going from a division 3 ways (two parents plus one child implies a third each) to a division four ways (two parents plus two children implies a quarter each). The percentage change between one-third and one-quarter is -25%.

the nineteenth century [Jones et al (2008), Jones and Tertilt (2008)]: and the relation between income and fertility for various periods is shown in Figure 4. For the US, at least, it seems as if the income elasticity of family size has been negative throughout this period. In other words, in the US, the “modern period”, as we characterize it, extends back at least to 1830.

The elasticity for the most recent cohort in the US (the line commencing in 1958) is reported to be -0.22 in Jones and Tertilt (2008). We therefore linearized N = $(N_{Y=1})\ Y^{-.22} = (1.6)Y^{-.22}$ around a (normalized) income of unity to get:

$$N = 1.952 - .352.Y \qquad 0 < Y < 1. \qquad (13)$$

Figure 4: Fertility by Occupational Income in 2000 Dollars



Source: Jones and Tertilt (2008)

Table 4: Fertility by Wealth Quintiles

Fertility rates: Total fertility rate						
Country	Survey Year	Household wealth index				
		Lowest	Second	Middle	Fourth	Highest
Sth. Africa	1998	4.8	3.6	2.7	2.2	1.9
Peru	1991-2	7	4.8	3.3	2.5	1.5
Brazil	1996	4.8	2.7	2.1	1.9	1.7
Pakistan	2006-7	5.8	4.5	4.1	3.4	3.0
Ukraine	2007	1.7	1.3	1.3	0.9	1

Source: (ICFI, 2012) http://statcompiler.com/?share=91398D8C3B

Given the near-linear relationship evident in the figure, equation (13) tracks the actual data quite well even as *Y* approaches its minimum value (normalized to zero). Recall that in section II, N refers to family size, not numbers of children, so we need to add two (parents) to the constant term in (13). We referred to this as $N = 4 - 0.35\ Y$ and this is what we use in our simulation.

The Malthusian Era:

Although there is plenty of evidence about the macro-relations between income levels and population growth rates at the level of international comparisons, until recently it was thought by many economic historians that that macro-relation had no analogue at the level of individual families.[21] We referred earlier to Adam Smith's observations about the importance of survival of infants born; a more extensive reference here may be useful. As Smith puts it, about his own time and place:

> In some places, one half of the children die before they are four years of age; in many places before they are seven; and in almost all places before they are nine or ten. This great mortality will everywhere be found chiefly among the children of the common people who cannot afford to tend them with the same care as those of better station. [Smith, 1776; WN I.viii.38 p 97]

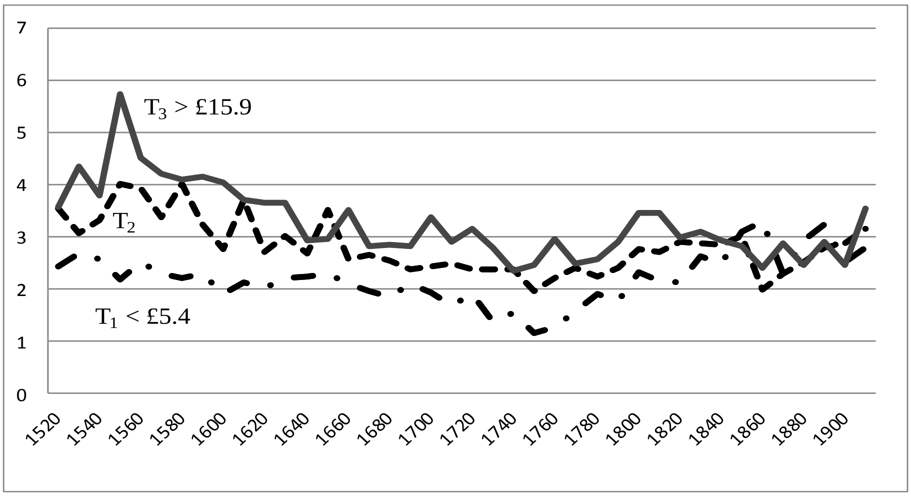
Thus it is the survival rate of children born rather than the birth rate that drives Smith's version of the Malthusian aggregate relation between rate of income growth and rate of population growth. But until recently, no direct empirical support for a positive relation between income and (surviving) family size has been available. Recently, however, Clark and Cummins (2010) have used wills from England over the 16th to 19th centuries to make a compelling case that there was a time when the within-economy relationship between income and the numbers of children was positive. They were able to back out estimates of lifetime earnings and numbers of surviving children for over 10,000 wills during a period when real incomes were broadly

[21] With regards to England, "The limited and contradictory earlier evidence on the relationship between wealth and fertility in pre-industrial England, and the fact that marriage ages and nuptuality were seemingly similar in 1850 to their earlier levels of many decades, created a false impression that the fertility regime of the mid nineteenth century [the poor having at least as many children as the rich] represented the entire pre-industrial period." pg. 2, Clark and Cummins (2010). See also their footnote 2.

stable. Figure 3 below, which is close to Figure 7 in Clark and Cummins (2010), shows some striking trends.[22]

Clark and Cummins defined income terciles in the figure as: T_1 below £5.4, T_2 between £5.4 and £15.9, and T_3 above £15.9[23]. They then divided the period 1520 to 1910 into 10-year sub periods and averaged the number of surviving children (at the time of the testator's death) for each income tercile. The figure suggests that from 1520 up to and including the 1780s, the number of children in the top income tercile was significantly higher than in lower terciles.[24] As we do, they call this the Malthusian era, but note that the relationship breaks down as England moved into the industrial revolution (which they date in the late 1700s). From that time, all income terciles show roughly the same numbers of children.

Figure 5: Surviving Children by Income Tercile



Source: Clark and Cummins (2010)

[22] We are grateful for the scholarly and generous way in which they provided, and explained, their data to us.
[23] It is sensible to define the terciles with the same figures over the whole period because income is stable. Only in the late C19th does this assumption start to weaken. The reader is referred to Clark and Cummins (2010) for a detailed discussion of their dataset. Although the sample size is very large (>10,000) it does not satisfy usual standards for random selection, given what can nowadays be accomplished by central statistical offices. It is, nevertheless, a remarkable and informative dataset.
[24] For the 1780s switch date, see Figure 8 of Clark and Cummins (2010).

Since this is an important result, in its own right and as indirect evidence of a causal relationship between income and numbers of children, we have wanted to establish the econometric credentials of their conjectures from their original data. The relevant hypothesis is that the average number of children in each tercile differs in the Mathusian era in the manner suggested by Figure 5, but not afterwards. To that end we added a Malthus dummy *D* to Clark and Cummin's tercile dummies T_1, T_2 and T_3. The Malthus dummy is unity up to the 1780s and is zero from 1790. The full results are in Appendix 2. We confirm formally that there was indeed a period of history where income and numbers of surviving children were positively related. The appendix also estimates a pooled regression over the Malthusian period, corresponding to the Malthus dummy above and shown here as (12), to justify the simulated relationship $N = 4 + 1.55\ Y$ in section II. Recall that two parents must be added to the constant term, and we rounded all coefficients to the nearest *0.05*.

$$N_t = 1.9947 + \underset{(t=14.34)}{1.5323}\ Y_t \qquad Y_t \in (0.1) \qquad R^2 = 0.0288,\ \ \sigma = 2.23 \qquad (12)$$

Of course, this positive relationship reflects the inclusion of what we have called entropy effects (which the simulations in section II abstract from). The rather large regression error standard deviation *(2.23)* suggests that Smith somewhat overstates the capacity of higher income households to protect their children; and the considerable variance in survival rates of children in households of identical size will serve to create greater dispersion of income in the next generation, effects that the iron law itself will have to overcome.

We do not of course have access to data that would allow our testing of the operation of the iron law directly in the Malthusian period. Nor have we attempted here to test the operation of the iron law in more modern times (for which reliable data on the time path of inequality is more readily available). That latter exercise will have to await a further occasion. All we have done here is to provide some empirical support for the parameters that are deployed in the simulations in section II. There is, we think, some reason to think that those simulations provide a picture that is unlikely to be too far from the mark for the relevant periods. But we have also provided some suggestion as to what the relevant periods might represent in terms of historical spans – 1520 to 1780 for the Malthusian, and perhaps 1900 to the present for the "modern".

V Summary and Conclusions

There was a time when there was abroad, in the social world, an "invisible force" that made for greater equality in the distribution of income. In the two hundred and fifty years between 1530 and the time of Adam Smith (a period of perhaps ten generations) that force would have essentially removed inequality altogether, had it been the only influence in operation. Allowing for the other factors that must have been at play, section III establishes that inequality has a stable steady state, provided the income elasticity of family size is positive (the characteristic feature of the Malthusian era). If, as is likely, social and cultural forces biased our forebears towards unequal societies above these steady states, the invisible force would have held inequality on a short leash.

In this paper, we try to uncover that force, to explain the logic of its operation and to provide some sense of its long-term magnitude. The force in question is the connection between the income elasticity of family size in the current period and dispersion in the income distribution in the next. We describe this connection as an "iron law" because it makes appeal to what seem to us fairly robust assumptions the most significant of which is that a primary mechanism of income determination for individuals lies in the life-time intra-family transfers each individual receives. Those lifetime transfers include not just bequests (if any) but rather resources of sustenance, human capital acquisition, and protection from disease and death through the years of child dependence (which for most individuals over most of human history are far more significant than bequests received). But whether bequests or gifts inter vivos, in larger families the transfers received by each individual are smaller because of the larger the number of children among whom total family resources have to be shared. If family size is positively correlated with income/wealth, then the effects of this process across the whole economy is dispersion reducing. And if wealth plays a causal role in family size then disposition to have large families is effectively inheritable and so the process is inexorable, working its egalitarian magic relentlessly from generation to generation.

Things are not now as they were in the days of Adam Smith. Those same inexorable forces that over the centuries prior to Smith produced greater equality are now directed at producing greater inequality: the crucial parameter – the income elasticity of family size – has changed sign. The larger transfers received in upper income families are systematically divided among a smaller number of siblings (than in lower income families). And again the effects of the process are cumulative, inexorable and over the long term very significant in magnitude.

The implications of the iron law in the contemporary world for egalitarian policy are significant. Beyond the conventional array of egalitarian policies, attention should be given to mechanisms that block or moderate the effective size of intra-family transfers – such as gift and bequest taxation; and publicly funded education provision on an equal per capita basis. However, the most distinctive suggestion that emerges here relates to the directly demographic aspects: greater consideration should be given to measures that increase the arithmetic value of the income elasticity of family size. The significance of this parameter for egalitarian policy has, we think, been obscured for too long.

Bio Note: Geoffrey Brennan is at Australian National University and Duke University; Gordon Menzies is at University of Technology--Sydney; and Michael Munger is at Duke University.

Compliance with Ethical Standards: The authors declare that they have no conflict of interest, and received no outside funding to support the preparation of this study.

References

Alexander, K. and A. Cherlin. "'Are Few Better than Many?' review of *Family Size and Achievement* by Judith Blake," *Contemporary Sociology*, 19(3)(1990): 343-346.

Atkinson, Anthony B. *Inequality: What Can Be Done?* (Cambridge, MA: Harvard University Press, 2015).

Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. "Top Incomes in the Long Run of History." *Journal of Economic Literature* 49(2011):1, 3–71.

Atkinson, Anthony B., and François Bourguignon. *Handbook of Income Distribution* (Volume II; Elsevier, 2015).

Becker, Gary S. "An Economic Analysis of Fertility." *Demographic and Economic Change in Developed Countries*. (Princeton: Princeton University Press, 1960).

_____. "A Theory of Social Interactions." *Journal of Political Economy* 82(6)(1974): 1063–93.

_____. *A Treatise on the Family*. (University of Chicago Press, 1981).

_____. "Fertility and the Economy." *Journal of Population Economics* 5(3)(1992): 185–201.

Becker, Gary, and Nigel Tomes. "Human Capital and the Rise and Fall of Families." In Gary Becker (editor), *Human Capital: A Theoretical and Empirical Analysis*. 3rd Edition. (Chicago: University of Chicago Press, 1994). pp. 257-298.

Blake, J. *Family Size and Achievement*. (Berkeley: University of California Press, 1989).

Blinder, A.S. "A model of inherited wealth," *Quarterly Journal of Economics* 87(1973): 608-626.

Clark, G. and N. Cummins. "Malthus to Modernity: England's First Fertility Transition, 1760-1800," *MPRA Paper* No. 25465, posted 28 September 2010.

Corak, M. "Inequality from Generation to Generation: the United States in Comparison," *University of Ottawa mimeo*. Cited in Alan B Krueger in 'The Rise and Consequences of Inequaltiy in the United States' Chairman Council of Economic Advisors, Jan 12 2012. See http://www.americanprogress.org/wpcontent/uploads/events/2012/01/pdf/krueger.pdf

Cowen, Tyler. *Average Is Over: Powering America Beyond the Age of the Great Stagnation*. (Plume Press, 2014).

Dang, Hai-Anh, and Halsey Rogers. The Decision to Invest in Child Quality over Quantity Household Size and Household Investment in Education in Vietnam World Bank Policy Research Working Paper #6487 (2013). https://openknowledge.worldbank.org/bitstream/handle/10986/15846/WPS6487.pdf

Davies, James. B. "The Relative Impact of Inheritance and Other Factors on Economic Inequality." *Quarterly Journal of Economics*. 97(3)(1982): 471-498.

Davies, James, and Anthony Shorrocks, "The distribution of wealth," *Handbook of Income Distribution*, Elsevier, 2000, Volume 1, Pages 605-675

Dorfman, R. 'A Formula for the Gini Coefficient', *Review of Economics and Statistics*, 61(1), (February, 1979): 146-149.

Fernández, Raquel, and Richard Rogerson. "Sorting and Long-Run Inequality." *Quarterly Journal of Economics*. 116(4P)(2001): 1305-1341.

ICF International, *MEASURE DHS STATcompiler*, at http://www.statcompiler.com , (2012), accessed August 30 2013.

Jones, L., Schoonbroodt, A. and M. Tertilt. "Fertility Theories: Can they Explain the Negative Fertility-Income Relationship?", *NBER working paper* 14266, August 2008.

Jones, L. and M. Tertilt. "An Economic History of Fertility in the U.S.: 1826-1960," in *Frontiers of Family Economics*, ed. by P. Rupert, vol. 1. (Emerald Press, 2008).

Lindahl, Lena. "Do birth order and family size matter for intergenerational income mobility? Evidence from Sweden." *Applied Economics*. 40(2008): 2239–2257.

Malthus, Thomas Robert. *An Essay on the Principle of Population, or a View of Its Past and Present Effects on Human Happiness, with An Enquiry into Our Prospects Respecting the*

Future Removal or Mitigation of the Evils Which It Occasions (2 Volumes, Fourth ed.), (London: J. Johnson, 1798/1807).
Marx, Karl. "Critique of the Gotha Program." In *Marx/Engels Selected Works*, Volume Three, p. 13-30; (Moscow: Progress Publishers, Moscow, 1875/1970).
Menchik, Paul L. "Primogeniture, Equal Sharing, and the U.S. Distribution of Wealth." *Quarterly Journal of Economics*. 94(2)(1980): 299-316.
Michels, Robert. *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy*. Translated by Eden and Cedar Paul. Kitchener, Ontario: Batoche Books, 1915/2001).
Österberg, Torun. "Intergenerational income mobility in Sweden: what do tax-data show?" *Review of Income and Wealth*, 46(4)(2000): 421–36
Piketty, Tomas. *Capital in the 21st Century*. (Belknap Press, 2014).
Piketty, Tomas. *The Economics of Inequality*. Belknap Press, 2015).
Roine, Jesper, Daniel Waldenström. "The evolution of top incomes in an egalitarian society: Sweden, 1903–2004." *Journal of Public Economics*, 92(1–2, 2008): 366-387.
Roine, Jesper, Jonas Vlachos, and Daniel Waldenström. "The long-run determinants of inequality: What can we learn from top income data?" *Journal of Public Economics*, 93(7–8, 2009): 974-988.
Smith, Adam. *An Inquiry Into the Nature and Causes of the Wealth of Nations*. (WN in text). General Editors: R.H. Campbell and A.S. Skinner. (Indianapolis: Liberty Fund, 1776/1981).
Wedgewood, Josiah. *The Economics of Inheritance*. (Nabu Press, 1929/2011).
Zilcha, Itzhak. "Intergenerational transfers, production and income distribution." *Journal of Public Economics*, 87(3–4, 2003): 489-513.

Appendix 1: Scaling Distributions by Family Size

We begin with the observation of Blake (1989) that in a random sample of people, those from large families will be over-represented. For example, suppose that the pdf of family size is uniform over the integer support N=0 to 4. In generation 2 the distribution of adults classified by family-of-origin size will be 0.0, 0.1, 0.2, 0.3, 0.4 corresponding to N=0 to 4. This is obtained by multiplying each N by 1/5 and then dividing by the mean of the original pdf, namely 2 children, to ensure the probabilities sum to unity.

The question is to what extent that distorts the relationship between $V(Y_1)$ worked out with a generation *1* distribution of income and $V(n\, Y_1 / N)$ worked out with a generation *2* distribution, over-represented by children from large families. It turns out that the multiplication of Y_1 by n/N within the bracket exactly compensates for the scaling of the distribution.

Notation:

- i generations $i=1, 2$
- f pdf
- Y 1st generation income (mean=y)
- N 1st generation children (mean=n)
- E_i the expectation based on the pdf in generation i ($i=1, 2$)
- V_i the variance based on the pdf in generation i ($i=1,2$)

We intuit a relationship between $f_{N|Yu,\,2}$ and the generation 1 density and then find the mean and variance of the new distribution. Whatever value of N is drawn for a family in generation 1, we assume the pdf will be scaled up by that value. Then, division by n is required so that the integral of the pdf is unity. This was the intuition above for Blake's observation that you are more likely to meet someone from a large family than a small one.

$$f_{NY,\,2} = \frac{N\, f_{NY,\,1}}{n}. \qquad (A1.1)$$

To simplify the algebra, we ignore ϕ in the text by setting it to unity. We are interested in $V_2(nY/N)$ – the variance of children's incomes calculated using the second generation distribution.

$$V_2\left(\frac{n\,Y}{N}\right) = E_2\left(\frac{n\,Y}{N}\right)^2 - \left(E_2\left(\frac{n\,Y}{N}\right)\right)^2 \qquad (A1.2)$$

The expectation of $n\,Y / N$ using the second period distribution is just the original expectation of Y.

$$E_2\left(\frac{nY}{N}\right) = \int_N\int_Y \frac{n\,Y}{N}\frac{N\,f_{NY}}{n}\,dY\,dN = y \qquad (A1.3)$$

Equation (A1.3) shows how the multiplication by n/N addresses the problem of the scaled distribution. When evaluating $E_2\left(\frac{nY}{N}\right)^2$ in (A1.2) we first place it into the expectations integral, to find how it relates to an expectation based on a generation-1 distribution.

$$E_2\left(\frac{n\,Y}{N}\right)^2 = \int_N\int_Y \frac{n^2\,Y^2}{N^2}\frac{N\,f_{NY}}{n}\,dY\,dN = E_1\left(\frac{nY^2}{N}\right) \qquad (A1.4)$$

To evaluate the latter, we first do a Taylor series expansion of the term in the expectations.

$$\frac{nY^2}{N} \approx y^2 + 2y(Y-y) - \frac{y^2}{n}(N-n) + \frac{1}{2!}\left\{2(Y-y)^2 + \frac{2y^2}{n^2}(N-n)^2 - \frac{4y}{n}(Y-y)(N-n)\right\} \qquad (A1.5)$$

Upon taking generation-1 expectations, the second and third terms fall out.

$$E_1\left(\frac{nY^2}{N}\right) = y^2 + \frac{1}{2!}\left\{2E_1(Y-y)^2 + \frac{2y^2}{n^2}E_1(N-n)^2 - \frac{4y}{n}E_1(Y-y)(N-n)\right\} \qquad (A1.6)$$

When (A1.3) and (A1.6) are substituted into (A1.2) we obtain the equivalent expression to the one in the text.

$$\begin{aligned}
V_2\left(\frac{n\,Y}{N}\right) &= E_2\left(\frac{n\,Y}{N}\right)^2 - \left(E_2\left(\frac{n\,Y}{N}\right)\right)^2 \\
&= E_1\left(\frac{nY^2}{N}\right) - y^2 \\
&= y^2 + \frac{1}{2!}\left\{2\,E_1(Y-y)^2 + \frac{2y^2}{n^2}E_1(N-n)^2 - \frac{4\,y}{n}E_1(Y-y)(N-n)\right\} - y^2 \\
&= V_1(Y) + \left(\frac{y}{n}\right)^2 V_1(N) - 2\left(\frac{y}{n}\right)C_1(Y,N)
\end{aligned}$$

This confirms (7) as the basis for the difference equation linking inequality across generations.

Appendix 2: Confirmation of the Malthusian Era and Simulation Parameters

This appendix provides additional details that 1) demonstrate the existence of a Malthusian era and 2) provide a Malthusian era regression equation which was the basis for our simulations using income in section II. For the first task the basic regression (supressing the error term) is:

$$N=\gamma_1+\gamma_2\, T_2+\gamma_3\, T_3+\gamma_4\, D+\gamma_5\, D.T_2+\gamma_6\, D\, T_3. \qquad \text{(A2.1)}$$

This can be rearranged using the fact that $\gamma_1= \gamma_1\,(T_1+T_2+T_3)$ and $\gamma_4 D= \gamma_4\, D\,(T_1+T_2+T_3)$ to give:

$$N\ =\ T_1\,\{\gamma_1+\gamma_4\, D\}+\ T_2\,\{\gamma_1+\gamma_2+(\gamma_4+\gamma_5)\, D\}+\ T_3\,\{\gamma_1+\gamma_3+(\gamma_4+\gamma_6)\, D\}. \qquad \text{(A2.2)}$$

Table 5 shows a battery of tests applied to (A2.2). We run the regression both on the decade data of Figure 5 (cols. 2 and 3 below) and on the unit record data of every will (cols. 4 and 5). First, we test if the numbers of children in each tercile differ in the Malthusian era by setting $D=1$ in (A2.2) and taking the relevant paired differences (see table notes). The overwhelming conclusion in rows 1 to 3 of the hypothesis tests (see the bottom left of Table 2) is that the number of children in every income tercile differed significantly from the other terciles at the 1 per cent level, establishing a Malthusian (i.e. positive) covariance between Y and N up to and including the 1780s.

Table 5: Regression Results

(* and ** are two-sided significance for 5% and 1%. D=1 in Malthusian era. N_{Ti} is children in ith tercile)

		Decade Data		Individual Data		Notes
Var	Coeff	obs=120	(s.e.)	obs=12314	(s.e.)	Tercile dummy coefficient differences:
C	γ_1	**2.5862	0.15	**2.5382	0.08	$N_{T2}-N_{T1} = \gamma_2+\gamma_5\, D$
T_2	γ_2	0.1174	0.22	*0.2180	0.10	$N_{T3}-N_{T1} = \gamma_3+\gamma_6\, D$
T_3	γ_3	0.3603	0.22	**0.3580	0.10	$N_{T3}-N_{T2} =[\gamma_3-\gamma_2]+[\gamma_6-\gamma_5]\, D$
D	γ_4	** - 0.5746	0.19	** - 0.5348	0.09	The differences in the Malthusian era
D*T_2	γ_5	**0.7538	0.26	**0.4640	0.12	ompared to afterwards are respectively γ_5,
D*T_3	γ_6	**1.0485	0.26	**0.7904	0.12	γ_6 and $\gamma_6-\gamma_5$.
Hypothesis Tests: H_1 Coefficient Combinations for F-tests						Hypotheses
$\gamma_2+\gamma_5>0$		**0.8712	0.15	**0.6820	0.06	$N_{T2}>N_{T1}$ in Malthusian era
$\gamma_3+\gamma_6>0$		**1.4088	0.15	**1.1484	0.06	$N_{T3}>N_{T1}$ in Malthusian era
$[\gamma_3-\gamma_2]+[\gamma_6-\gamma_5]>0$		**0.5377	0.15	**0.4664	0.07	$N_{T3}>N_{T2}$ in Malthusian era
$\gamma_5>0$ as above		**0.7538	0.26	**0.4640	0.12	$N_{T2}-N_{T1}$ gap falls after Malthusian era
$\gamma_6>0$ as above		**1.0485	0.26	**0.7904	0.12	$N_{T3}-N_{T1}$ gap falls after Malthusian era
$\gamma_6-\gamma_5>0$		0.2947	0.26	**0.3264	0.11	$N_{T3}-N_{T2}$ gap falls after Malthusian era

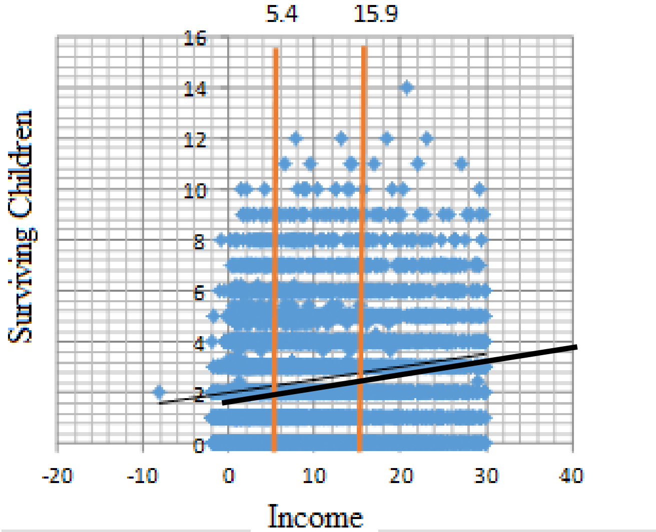
Notes: All tests are two sided, whereas the requirements of the different alternative hypotheses are one sided. However, when the estimated coefficients have the sign indicated under H_1, and H_0 is rejected with an F-test, these alternatives will certainly be rejected with a one sided t-test from an appropriately reparameterized regression, because the relevant p-value will halve.

Next, we tested whether this difference in children between the terciles was significantly greater in the Malthusian era compared with afterwards. We derived the tercile differences in the Malthusian era ($D=1$ in (A2.2)) and subtracted the same differences in the post-Malthusian era ($D=0$ in (A2.2)). We tested the significance of the relevant linear combinations of parameters. Here, the verdict from the individual data was that every tercile difference (N_{T2}-N_{T1}, N_{T3}-N_{T1} and N_{T3}-N_{T2}) fell significantly after the Malthusian era at the 1 per cent level. Apart from the difference N_{T3}-N_{T2}, the evidence from the decade data support the same conclusion.[25]

[25] Based on this dataset with Clark and Cummin's 1780s cut-off, there is weak evidence of a positive tercile difference in children numbers in the post-Malthusian era. That is, using the individual data and setting D=0 in Table 5, N_{T2}-N_{T1} (coefficient γ_2) is significant at the 5 per cent level, N_{T3}-N_{T1} (γ_3) is significant at the 1% level and N_{T3}-N_{T2} ($\gamma_3 - \gamma_2$, test not shown in Table 2) is not significant. However, the dataset contains only limited datapoints of the post-Malthusian era, and what significance there is almost certainly due to the ambiguity of

For our second task, we regressed the number of children on income, excluding income outliers. Naturally, the simulation regressions in the main text have two people added to these regressions to obtain family size.

Figure 6: A Malthusian Equation



The estimated equation in Table 6 has a highly significant slope coefficient, which is the marginal effect of income on family size. Since the maximum income (excluding the top 15% of income) was £30, we rescaled income to the interval [0, 1] and multiplied the coefficient by 30 to give the same marginal effect. The value of 1.55 says that the richest families (Y=1) had about one and a half more children than the poorest families (Y=0) during the Malthusian era.

Table 6: Malthusian Regression

Dependent Variable: N				
top 15% income excluded				
Sample (adjusted): 1 7060				
Variable	Coef	Std. Error	t-Statistic	Prob.
C	1.9947	0.0409	48.8281	0
Y (< £30)	0.0511	0.0036	14.2275	0
Rescaled Y ~ (0, 1)	1.5323	0.1077	14.2275	0
R-squared	0.0279	Mean dependent var		2.44
Adjusted R-squared	0.0277	S.D. dependent var		2.26
S.E. of regression	2.23			
F-statistic	205.5			
Prob(F-statistic)	0			

The R^2 of the regression for the Malthusian era is low, at around 2½ per cent, and the estimated error volatility σ is high, at just over two children. As noted earlier, the high error volatility, called entropy in the main text, is a welcome result because it saves the model from predicting vanishing inequality during the Malthusian era.

where to make the cutoff. Looking at Figure 5, it beggars belief that any differences would be significant after, say, the 1840s.