The Historical Fertility Transition and Theories of Long-Run Growth: A Guide for Economists

Timothy W. Guinnane
Department of Economics
Yale University

Revised April 2010

10,000 words

Note: This paper is under revision for Journal of Economic Literature. Please do not cite this incomplete version. My research on this and related topics has been generously funded by the National Institutes of Health (US) and the Leverhulme Trust (UK) and the Economic and Social Research Council (UK). I have learned much about these issues from working with and talking to John Brown, Oded Galor, Carolyn Moehling, Thomas Mroz, Sheilagh Ogilvie, Barbara Okun, Cormac Ó Gráda, Jean-Laurent Rosenthal, Mark Rosenzweig, Paul Schultz, William Sundstrom, Nathan Sussman, James Trussell, Christopher Udry, and David R. Weir. My interest in this subject is due to Paul David, and any (good) ideas here probably reflect his influence more than is reflected in the citations. I thank the editor and referees for comments on an earlier version.
Abstract

Recent developments in growth theory place the historical fertility transition at the center of models that explain the transition from high-fertility, low-growth economies to the low-fertility, high-growth experience that characterizes most of the world today. This body of work, especially in its most recent “Unified Growth Theory” (UGT) form, marks an important advance in making demographic and economic behavior endogenous to each other. This literature also takes the historical record seriously, and is motivated by an attempt to explain actual experience. This paper discusses the larger literature on the historical fertility transitions and what it means for economic explanations of the historical fertility transition. I discuss several possible traps for the unwary economist and stress that we know less about the historical fertility transition than we would like to know.
The neoclassical model of economic growth attributed to Robert Solow treats population as exogenous to the economy. Early models of population (such as Malthus) treated economic change as exogenous to the demographic system. Thus neither model was equipped to elucidate one of the most important transformations in human history, the fertility transition. Prior to the fertility transition, which took place in most of Europe and North America between the late eighteenth and early twentieth centuries, fertility was regulated only indirectly, via age at marriage and whether women married. Married women could expect to have eight or more births. After the fertility transition, the link between marriage behavior and fertility became weak, if not non-existent. Economists have recently developed enhanced models to study the interaction between economic and demographic change, focusing on long-run growth. An important focus of the newer research has been to make fertility decisions endogenous to the economy, and thus to construct models in which economic change induces couples to have smaller families, and those families in turn affect economic dynamics in ways that shape long-run growth.

Early efforts in this area usually embed the microeconomic model of fertility decisions due to Gary Becker in a framework that allowed feedbacks from the economy to fertility decisions. These models have either a single equilibrium, or possibly multiple equilibria without explicit focus on how an economy would transition from one to another. A more recent body of research focuses explicitly on how economies transition from a “Malthusian” economy of high fertility and little growth in per-capita incomes to one in which fertility is much lower and per-capita incomes grow rapidly. For this reason the literature usually refers to it as “Unified Growth Theory” (hereafter UGT). This research, due to Oded Galor, David Weil, and others, integrates a microeconomic model of the demand for children with feedbacks from a model of growth, and as such accounts in a coherent way for a central issue in our understanding of long-run growth.

These discussions have provoked renewed interest in the empirics of the historical fertility transition: when and why it took place, and how rapidly it created the low fertility we
now see in most wealthy countries. Theorists are to be commended for trying to explain historical facts, and to highlight the importance of issues that were previously of interest primarily to economic historians and to others, such as demographers, outside the economics profession. Unfortunately, some recent discussions misunderstand the demographic literature. More importantly, the use of certain theoretical ideas has led others to believe that we know more about the historical fertility transition than is the case. This paper has two purposes. I provide an overview of the historical fertility transition, and then discuss the main hypotheses that are current in the economics and economic history literature. This discussion stresses what we know and do not know.

Throughout I set aside three other, issues. First, most current economics work naturally stresses the *implications* of demographic change for economic growth and development. That issue lies beyond the scope of this paper. Second, economists typically think of fertility decline as the reduction in the number of children born to a woman or to a couple. Demographers and others stress heterogeneity in the way fertility declines, for example, whether couples reduce the number of surviving offspring by spacing their child-bearing or ending child-bearing before that is biologically necessary. This paper only discusses this second issue where it becomes important to understanding the evidence on the fertility transition. Finally, the fertility transition in developing countries since World War II has been studied far more intensively than its historical counterpart. This paper focuses on the earlier, historical transition, which is most relevant to what theorists have in mind in modeling the industrial revolution.

### 1. The basic contours of the historical fertility transition

Figure 1 reports fertility experience for the period 1800-1970 for five major countries: France, England and Wales, Germany, the United States, and Italy. This paper focuses primarily on the first four countries; Italy is included in Figure 1 only to suggest the heterogeneity of
historical experience. Figure 2 focuses on the single case of Germany to show the relationship between fertility and mortality decline. Ignoring the heterogeneity in Figure 1 for the moment, we see a single, broad pattern of fertility declining starting in the eighteenth or nineteenth century, and in most cases accelerating in the second half of the nineteenth century. The two world wars produced dramatic, temporary reductions in fertility, and in most of these countries the post-World War II period saw some type of “Baby Boom.” By the 1970s most of western Europe saw the emergence of “low-low” fertility, fertility so low that, if sustained, the population would shrink over time. Today, few OECD countries have fertility rates high enough to sustain population growth through natural increase.\(^1\) The relationship between fertility and mortality declines differs across these countries, but the German experience detailed in Figure 2 is fairly typical. For most of the nineteenth century, birth rates exceeded death rates and the population grew (even with, as in the German case, extensive emigration). A long tradition assigned to an exogenous decline in mortality a causal role in the fertility transition. More recent research tempers or even rejects this line of argument, but the idea lives on in many accounts. Fertility and mortality began to decline at roughly the same time, and today mortality rates are so low that at the earlier, prevailing fertility levels population would grow rapidly. That this has not happened reflects the very low fertility prevailing today; as Figure 2 shows, for much of the late twentieth century, German rates of natural increase were negative.\(^2\)

**The timing of the fertility transition**

Fertility fell first in western Europe and North America, and began to decline later in eastern Europe. Most countries outside Europe and North America did not experience fertility declines until after World War II. When we focus on the timing within Europe and North

---

1. Kohler, Billari, and Ortega (2002) discuss the emergence of very low fertility in the 1990s; Goldstein, Sobotka, and Jasilione (2009) discuss a recent, partial reversal.
2. The measures reported in Figures 1 and 2 are the Crude Birth Rate (CBI) which is the number of births per thousand population. The Crude Death Rate (CDR) is defined analogously. The Crude Rate of Natural Increase (CNRI) equals CBR - CDR.
America, however, matters become less clear. Some economists have accepted the view that with two exceptions, the fertility transition took place essentially all at once across Europe.

Economists accepting this conclusion probably would not admire the data and methods upon which it is based. In addition, the all-at-once view has important implications for causality. The observation of simultaneous fertility transitions is usually used to support a claim that economics has little to do with the fertility transition. In a famous comparison, some demographers have argued that the apparently simultaneous onset of the fertility transition in England and Hungary implies that economic development cannot have been the cause.

Studies based on micro-level evidence often conclude that the fertility transition started earlier than implied by the aggregate data such as reported in Figures 1 and 2. Knodel (1988) reports that couples married 1800-24 in five of the fourteen German villages he studied exhibited significant control of marital fertility; that is, fertility in the mid-nineteenth century already shows strong efforts to limit family size. The marriage cohort of 1850-1874 exhibits significant fertility control in his villages taken as a whole. One could worry that Knodel’s villages were exceptional in some way. They are: they were poorer and less developed than most of Germany. If anything, we would expect later transitions there.

There are more general reasons to use caution in talking about simultaneous fertility transitions. Figure 1 reminds us that there were two important exceptions. Most scholars agree that France’s fertility decline started in the early nineteenth century at the latest. France’s CBR was already lower than elsewhere, and declining, in the early nineteenth century (Figure 1). Using

---

3 Clark (2007, p. 225) claims “… the timing of the demographic transition in Europe and the United States places it circa 1890…” This assertion is patently false in the U.S. case and French cases. Clark does not give sources for his Figure 11.6, or explain why he limited his illustration to England and Sweden. Galor (2005, Footnote 33), accepts the view that the fertility transition began everywhere at once in Europe. He is citing the results of a large project undertaken at Princeton University in the 1960s and 1970s. The Princeton conclusion reflects problems with sources, measures, and econometrics. The Princeton index of marital fertility $I_p$ in particular does not perform as desired in Monte Carlo studies. (Guinnane, Okun and Trussell (1994); Brown and Guinnane (2007). Doepke (2004) focuses on the differential timing of fertility transitions. But his stress is on the broader variation in timing across the world, and not on variations within the first societies to experience this change.
more refined approaches Weir (1994, Table B3) and Mroz and Weir (1990) show that fertility in France had fallen at least ten-percent, by 1800, from its high in the early eighteenth century; and that in northern France this decline was especially pronounced. Birthrates in the United States were much higher than elsewhere, but already falling rapidly in the early nineteenth century. By the end of the nineteenth century U.S. fertility (measured this way) was lower than much of western Europe. At the beginning of the nineteenth century France was one of the largest countries in Europe, and a cultural and intellectual leader. By the end of the nineteenth century the U.S. was a major industrial producer. These are not two esoteric exceptions one can ignore.

The other concern reflects the CBR so often used in economics research. Not accounting for changes in either age-structure or marriage patterns, the CBR can increase when couples are having smaller families (if, for example, age-structure shifts so that the denominator includes fewer people of child-bearing age), or decline when couples are having more children (if, for example, the proportion married declines). During the period of the fertility transition, to take one example, in many societies more Europeans were getting married, and doing so earlier. Thus what looks like constant fertility (as measured by the CBR) reflects two offsetting trends, an increase in marriage and a decrease in fertility within marriage. Taking this concern into account, for example, the precocious fertility declines in France and the United States are even sharper than the CBR would suggest. Figure 3 reports the available CFR for the five countries in Figure 1. This measure requires age-specific fertility data, and thus are not available as far back as CBR. But Figure 3 clearly shows that the experience of specific cohorts born in the nineteenth century is not well-captured by CBR. Figure 6 shows the effect of changes in marriage patterns for France, where the proportions married rose by about twenty percent through the nineteenth

---

4 Weir’s estimate of Ig, a measure of marital fertility we discuss more below, falls by nearly half from 1820 to 1900. The decline in the Total Fertility Rate (TFR) for the white population in the US from 1800 to 1900 is from 7.04 to 3.56 (Haines 2000, Table 4.3).  
5 For the United States we have estimates of the Total Fertility Rate (the cross-section analogue of CFR) back to 1800. These start at 7.04 in 1800 and fall to 3.56 by 1900. These figures, like the CBR in Figure 1, are for the white population only (Haines (2000, Table 4.3)).
century. This marriage boom masked to some extent the sharpness of the decline in marital fertility.

One could quibble about the relevance of measures defined over nation-status as the basis for a study of the causes of the fertility transition. A more general concern for economists should be who started limiting family size, because who implies why: the correlations of fertility behavior with micro-level characteristics of couples are central to testing the competing hypotheses we discuss below.

2. Theories of fertility and the economy

Economists tend to use the term “Malthusian” as a synonym for poverty or low growth rates. In the literature at stake here, in contrast, the term refers to a specific model. The central stylized fact is that in a pre-transition population, marital fertility depends only on age at marriage and the proportions who marry. This in turn is a characterization of what is called the “European marriage pattern” in which young adults deferred marriage until well after puberty, often into their middle and late twenties, and as much as ten to twenty percent of some cohorts never married at all. Couples could not marry until they could support themselves and their offspring, implying that marriage decisions depended on the real wage in young adulthood. Thus economic conditions regulated fertility, but indirectly, through marriage. Cohorts that experienced low wages married later and thus had smaller completed family sizes. In Malthus’s model the elasticity of fertility with respect to the real wage is positive.6 Mortality in the Malthusian model

---

6 The Malthusian model implicitly assumes away illegitimacy. Most studies find a substantial proportion of pregnant brides, but children actually born outside of marriage accounted for more than 10 percent of all births. The exceptions, such as in southern Germany, reflect, at least in part, legal restrictions on the right to marry. These restrictions were unusual and do not account for the role of marriage in European demographic patterns. For the European marriage pattern, see Hajnal (1965, 1982). Most scholars accept Hajnal’s European Marriage Pattern as a stylized account, but it is not clear which parts of Europe it describes, outside of the very northwest corner. Some accounts of the Malthusian model describe it as assuming that the elasticity of births with respect to the real wage is zero, that is, that all adjustments take place via mortality. This amounts to a restriction on the model that is at variance with what we know from historical populations.
depends negatively on the real wage, as higher incomes reduce the incidence of malnutrition and disease. With some simple assumptions there is only one real wage at which the number of births equals the number of deaths. Standard classical assumptions on production close the model; with a given stock of land, capital, and technology, the marginal product of labor is a negative function of population size.

The Malthusian model’s pessimistic reputation reflects the implication that demographic behavior determines the long-run equilibrium real wage. Increases in labor productivity brought about by capital investment, new land, or even technological improvement affect, in the long run, only population size. Figure 4 illustrates such a comparative static exercise. A shift in the marginal product of labor schedule initially generates a higher real wage and thus increases in fertility and decreases in mortality. Eventually however, natural increase (births minus deaths) increases the population size, bringing the system to a new equilibrium at the same real wage but with a larger population.

A large empirical literature tries to test various aspects of the Malthusian model. In the long run, the model has three equations and three endogenous variables; in addition, populations can exhibit “echoes” of past population shocks, posing modeling problems that have fascinated many. Figure 5 reproduces Wrigley and Schofield (1981)’s graphical illustration of the relationship between fertility and the real wage in England. A sympathetic observer can agree

---

7 There is also good evidence for the dependence of mortality on short-term shocks to the real wage. See the references cited in the previous note.

8 A large literature that will not be discussed at length here attempts to determine whether, in a particular time and place, adjustment to shocks took place via fertility or mortality. This amounts to estimating the elasticity of fertility and mortality with respect to the real wage. Clearly a society in which the adjustment was primarily via fertility would be more pleasant to live in and be devoting fewer resources to the clothing, feeding, and education of people who would die in the next shock. Thus, contrary to recent assertions to the contrary, there are important variations across societies characterized by this kind of population system.

9 Figure 1 reflects such an echo effect: population growth rates increased in the early nineteenth century, creating a large cohort of women of child-bearing years in the mid nineteenth century. Lee and Anderson (2002) is only one of many recent papers proposing methods to estimate long-run versions of the model.

10 The gross rate of reproduction (GRR) is the average number of daughters that would be born to a woman (or a group of women) if she survived at least to the age of 45 and experienced the age-specific fertility rate of a given year.
with Wrigley and Schofield, on the basis of this figure, that the changes in wages drove changes in fertility. Others have their doubts. A different estimation strategy relies on the fact that in the short run population is approximately fixed, making the model more tractable and generating a large literature estimating the elasticity of fertility and mortality with respect to shocks to the marginal product schedule. These papers often employ variation in grain prices as proxies for real-wage shocks. This approach is attractive given the wide availability of grain price series, and is justified by the dominance of agricultural output in the pre-industrial economy. Estimates of the price elasticity of marriages and deaths vary widely, as one would expect, but most studies confirm a negative marriage elasticity with respect to prices, and a positive mortality elasticity. This literature generally finds that prior to the nineteenth century, births \textit{per se} did not respond to shocks to prices, thus confirming the Malthusian model’s key assumption.

The sustained economic growth experienced by many countries from the nineteenth century clearly cannot be explained within Malthus’s model. Economic historians emphasize that the Industrial Revolution in Britain and elsewhere reflected increases in TFP more than capital accumulation, but neither force can explain the fertility transition in any simple way. If we take the Malthusian model seriously, the resulting increases in labor productivity would have to be more rapid than the ability of net reproduction to offset it. More importantly, the Malthusian model implies that most societies today would consist of couples who married young and had large families. Obviously that has not happened, which led economists such as Becker to ask why the elasticity of births with respect to incomes is now, apparently, negative.

\textit{The demand for children}

\footnotesize

\begin{itemize}
  \item[11]*** Cite Weisdorf and other recent papers
  \item[13]In the Solow model there is no paradox; because the labor force grows at a rate determined exogenously to the model, higher incomes due to technical advance are not offset by increases in fertility and decreases in mortality.
\end{itemize}
Virtually all micro-economic analysis of fertility today is based at least in part on Becker’s models of the demand for children. Becker’s basic insight was that children can be viewed as consumer durables, and the demand for children analyzed using the tools of consumer choice. The model yields important insights. For example, observers have long noted that fertility tends to be negatively correlated with income in the cross-section, and, since the beginnings of the fertility transition, over time. Becker’s model implies that this is a standard substitution effect, that children are not inferior goods: wealthier couples have higher opportunity costs of time, and time is a major cost of child-rearing. We sketch Becker’s model here because later we will organize our discussion around its implications. We start with a household utility function \( U = U(n, Z) \), where \( n \) is the number of children and \( Z \) is a vector of all other commodities. The household maximizes this utility subject to a standard budget constraint. Anything that increases the costs of children (say, an increase in the costs of food) will induce substitution away from children and towards other commodities. A pure income effect would increase the demand for children, as we expect. But if the income comes from rising wages, then that increased wage may show up as an opportunity cost of having children, and once again induce a reduction in the demand for children via the substitution effect.

Later interest in the Becker model focuses on the possible trade-off between the number of children and their quality. We now work with a household utility function of the form \( U(n, q, Z) \), where \( q \) is the quality of each child. Becker (1981, pp.107-108) divides child costs into three categories. Some costs depend only on the number of children: an example of this \( p_n \) would be the costs associated with the mother’s pregnancy and delivery. Another cost is related to child quality, but does not depend on the number of children, as it goes to purchase household

---

14 The important paper references are Becker (1960) and Becker and Lewis (1973). Becker (1981, Chapter 5) is a more elegant and expansive exposition.

15 This discussion follows Becker (1981) and uses his notation.
public goods: examples of $p_q$ would include books that children could share. A final cost $p_c$ is the
cost of augmenting the quality of any child.\(^\text{16}\) The household’s budget constraint is then:

$$p_n n + p_q q + p_c n q + \pi_z Z = I$$

(1)

where $I$ is household income and $\pi_z$ is the cost of the $Z$s. The marginal rate of substitution
between quantity and quality is

$$\frac{MU_n}{MU_q} = \frac{q}{n} \frac{(1 + r_n)}{(1 + r_q + \varepsilon_{pq})}$$

(2)

Where $r_n$ and $r_q$ are the ratios of fixed to variable costs for quantity and equality, respectively, and
$1 + \varepsilon_{pq}$ is the ratio of the marginal variable cost to the average variable cost of quality.\(^\text{17}\) The
substitution effects between quantity and quality are more complicated than in the model without
the Q-Q tradeoff. Consider an increase in $p_n$. The household will substitute away from a large
family to both child quality and other commodities, as one would expect. But because of that
interaction term $p_c n q$, the shadow cost of child numbers depends on the level of child quality,
and the reduction in child numbers raises the shadow cost of numbers even more, inducing more
substitution of quality for quantity. Thus the interactions between child quantity and quality in the
budget constraint can produce substantial effects.

This model is now a workhorse of micro-economic studies as well as UGT and other
models of the role of fertility in long-run economic growth. Below we will argue that it has not
really be meaningfully tested in historical contexts. This claim reflects to sorts of worries. First,
Becker assumes that $q$ is the same for all children. Thus parents cannot react to changes in the
price of quality by, for example, educating some of their children but not all. We know this is at
variance with the facts: for example, in most societies boys receive more education than girls.
This simplification could be avoided by elaborating the model to have different kinds of children,
and to assume that $q$ is the same within a “child-class.” A two-class model would account for

---

\(^{16}\) Becker allows $p_c$ to vary with $q$.

\(^{17}\) That is, $r_n = \frac{p_n}{p_c n}$; $r_q = \frac{p_q}{p_c q}$.
differential treatment of boys and girls, for example. The Q-Q idea becomes less useful, unfortunately, if parents pursue more complicated strategies, such as having one high-quality child and N-1 low-quality children. One could imagine a set of returns to education and financing constraints that would lead a household to finance one (high-quality) child’s education with the earnings of the other (low-quality) children. A more general concern reflects a lack of care about the difference between the “Q-Q” model and what Rosenzweig and Wolpin (1980) call the q^1 model, which amounts to Becker’s model with \( p_e nq \) equal to zero.\(^{18}\)

Many recent growth models embed the Q-Q approach within a growth model intended to understand the causes and implications of the transition to low fertility. In Galor and Weil (2000), for example, the transformation rests on two features of the model. First, there is sustained technological change, which raises the rate of return on human capital. Parents shift from having a lot of low-quality children to a small number of high-quality children. This is, in a nutshell, the fertility transition. Second, Galor and Weil assume that the rate of technological progress depends on the aggregate stock of human capital. The economy develops a virtuous circle in which the demographic behavior (smaller families) leads to changes that in turn make smaller families more attractive.

3. Explanations and evidence

Discussions of the reasons for declining fertility in the nineteenth and early twentieth century have entertained many explanations, some of which do not deserve to the taken

\(^{18}\) In a number of articles often called collectively the “Easterlin synthesis,” Richard Easterlin integrated Becker’s approach with a better appreciation of the costs of fertility control as well as the biological limits on reproduction. See especially Easterlin (1978). Easterlin’s insights have probably not received the attention they deserved. The “wealth flows” model attributed to the demographer John Caldwell claims that “… the fundamental issue in demographic transition, is the \textit{direction and magnitude of intergenerational wealth flows} or the net balance of the two flows – one from parents to children and the other from children to parents – over the period from when people became parents until they die.” (Caldwell 1976, p.344; emphasis in original). It is possible to see in this a rough version of Becker’s model, if we interpret “wealth flows” in terms of costs, and allow parents to react on the margin to changes in magnitudes, not just signs. Caldwell appears to believe his is an alternative to an economic model and does not cite microeconomic works in his discussions. See also Caldwell (2005).
seriously. The primary explanations of interest to economist can be grouped under six headings, which we consider in turn. The first is an exogenous decline in infant and child mortality, as in the “demographic transition” story. The second turns on innovations in the technology of contraception, or more widespread availability of contraceptive devices. The third looks for increases in the direct cost of childbearing. The fourth explanation assumes increased opportunity costs of child-bearing. Within the Q-Q framework, both of these accounts imply a shift to quality as a by-product of reductions in child numbers. The fifth looks for a net increase in returns to quality directly. The sixth argument assumes that children were an important way to ensure against risk and to provide for old age, and that the rise of state social insurance as well as private insurance and savings vehicles led households to substitute out of children. We consider these explanations in ceteris paribus fashion. An acute problem for empirical research is that many of these relevant changes occurred at roughly the same time.

“Demographic transition theory” and the role of mortality decline

A long tradition assigned to mortality decline a causal role in the fertility transition. The central idea, famously represented in the so-called “demographic transition theory” associated with the demographer Frank Notestein (1945), is that couples in high-mortality societies have a lot of births because that is necessary to ensure a surviving brood. An exogenous mortality decline would induce couples to have fewer children because they would not need to have so many to achieve a given number of survivors. Notestein’s account was motivated by the experience of developing countries after World War II, where public-health interventions created...

---

19 One discredited view is “racial degeneration.” Another, less absurd argument rests on the alleged deterioration of health caused by urban residence or factory work. I know of no convincing work that shows this can account for a significant part of the transition. The same goes for the spread of venereal disease: some such diseases do cause premature sterility, and in some cases they did become more widespread in the nineteenth century, but most scholars doubt this played an important role in what we observe.

20 Textbooks for economic development often invoke this explanation. Even Ray (1998, pp. 302-303), which is otherwise unusually sophisticated in its treatment of population issues, makes this argument.
dramatic reductions in infant and child mortality in a short period. The mortality decline in Europe was not as abrupt. Historians and others still debate the causes of the historical mortality decline, but most scholars stress improvements in public health systems (such as clean water supplies and food-safety measures) combined with modest results from medical interventions (such as vaccines against smallpox). Some of these developments reflect local decisions about public-good investments, but it is plausible to view them as largely exogenous to any couple’s decision-making.

The empirical weakness in this explanation of fertility decline is timing. Fertility in the United States declined for decades before any noticeable decline in mortality. In other places, as in Germany (Figure 2), the fertility and mortality declines took place at roughly the same time. This does not rule out a role for exogenous changes in mortality as a causal force, of course, but it suggests that Notestein-style accounts will not get us far. We should also note that mortality decline alone could not get us to the low levels of fertility we observe today. Recall that the total fertility rate in the U.S. in the early nineteenth century was about seven. Even if thirty percent of children then died in infancy or childhood (a high estimate), this implies that households wanted a surviving brood of four or five. By the end of the nineteenth century, in contrast, white, urban women in the U.S. were increasingly having just two children (David and Sanderson (1987)).

“Demographic transition theory” also has two the theoretical flaws. First, at least part of the decline in infant and child mortality is endogenous to the fertility decline. There are several lines of argument here, all of which assume that parents can assert some influence on their children’s mortality risks by providing health-enhancing resources. In an historical context these resources include breast-feeding (which isolates an infant from possibly contaminated water and

---

21 The discussion is sometimes called “the McKeown debate.” Thomas McKeown (1976) famously argued that prior to about 1900, medical science had done little increase human longevity, and concluded that the observed mortality declines to that point reflected direct and indirect effects of better nutrition brought about by higher incomes and better food supplies. Fogel (xx)’s more nuanced account also stresses the role of nutrition. Some reactions to McKeown and Fogel pose the introduction of public-health systems as an alternative explanation to “economics.” Brown (1988) shows that the introduction of public-health measures reflected local incomes, as one would expect with a public good.
food supplies); other nutrition; and protection from danger such as hearth fires. Most historical
studies that consider the issue seriously find, indeed, that in a regression framework, reasonable
instruments for infant and child mortality reduce mortality’s impact on fertility. Brown and
Guinnane (2002) are fairly typical in finding that while mortality has a strong, positive effect on
fertility in an OLS model, in the counterpart IV model the effect is zero. Becker’s model implies
that reduced infant and child mortality could arise from changes in other costs. For example,
improved contraceptive technology (discussed next) could allow parents to more tightly control
the link between actual and desired fertility. Parents might have a smaller number of children and
care for them more intensively, in effect not relying any more on high mortality to cull their
brood to its desired size.22

The second theoretical problem with the “demographic transition” account is that even a
fully exogenous reduction in infant mortality would have two, countervailing effects. An
exogenous mortality decline would reduce pn and thus make child numbers cheaper relative to
both child quality and Zs. An exogenous mortality decline could actually raise fertility.

As an aside, it is worth noting the implications of the historical age-patterns of mortality
change for some UGT-style accounts. Clearly the return to investment in human capital depends
in part on how long an individual lives to employ the human capital. Thus it is correct to link
improving mortality with an increased return to human capital, and, within the Q-Q model, falling
n. But we have to be careful about the age-patterns of mortality decline. The decline in death rates
in the nineteenth and twentieth century reflects mortality improvements concentrated in the early

---

22 If this claim seems extreme, consider the practice of wet-nursing, which was extensive in France into the
twentieth century. Legislation passed in 1874 sought to regulate the practice of sending children elsewhere,
sometimes long distances, to be nursed. The remaining wet-nurses still exercised Zola, who included a
nasty portrayal in his novel Fécondité. In his account, a single woman would take several urban babies
soon after their birth, and take them by train to the location where they would be cared for. Some would die
en route, from cold or hunger, and others would die from neglect at their destination. Martin-Fugier (1978,
pp. 26-27) quotes a thirty percent mortality rate for wet-nursed infants in the Paris region. This is roughly
twice the infant mortality rate for France as a whole at the time. Other estimates were higher. Even if, as
she suggests, these figures reflect hysteria more than fact, it is clear that sending a child out to wet-nurse
did not reflect much anxiety about its survival.
years of life. In Germany, in the decade 1871/80 235 infants per thousand died before their first birthday. This figure declined to 162 by 1910 and 61 in 1939 (Marschalck 1984, Table 3.17). In a striking graph, Wrigley et al show the decline of death rates at different ages in England from the late 17th through the end of the 20th century. In percentage terms the declines under age 5 dwarf those for older people.\(^{23}\)

What are the implications of these figures for the expectation of life? We usually lack life tables before the 19th century, and so cannot calculate expectation of life from a cohort’s complete experience. Most studies combine estimates from infant or child mortality with model life tables to estimate later experience. In this spirit, Knodel’s estimates for the pre-1800 period imply a Coale-Demeny model life table “North” level 7. A population fitting this life table would have an expectation of life at birth of about 35, but those who survived the dangerous early years would fare much better; at age 15, the expectation of life is about 45 years, implying a mean lifespan of 60 years. Put differently, about 40 percent of those who survive to age 15 live to age 65.\(^{24}\) This picture emerges clearly from the period when we have life tables. A German woman born in the 1870s had an expectation of life of 38.5 years. If she had already reached aged 15 in that decade, she could expect to live a further 44.2 years. Forty years late, infant girls had gained almost ten years extra life expectation, while a 15 year-old German woman had gained not quite five years (Marschalck 1984, Table 3.15).

\(\text{Innovations in contraceptive methods}\)

Michael and Willis (1976) first integrated the costs of averting unwanted births into the microeconomic model. Their model assumes that couples can effect control over their actual number of births, which is a random variable, using methods that imply various utility and money

---

\(^{23}\) This is not the case everywhere; in France, for example, it appears that a larger part of the mortality decline in this period was due to improvements for children over age 5.

\(^{24}\) Reconstitution studies cannot track individuals who leave the place where they are born, meaning that mortality estimates derived from these sources are unreliable past about age 10. The example given in the text uses the Coale-Demeny model life table “North” level 7 for females.
costs. The couple’s optimization decision takes into account both the costs of contraception and the utility costs of having “too few” or “too many” children. (Thus the Q-Q tradeoff enters their thinking indirectly; a couple with too many children may, because of the budget constraint, be forced to choose a lower level of quality than it would have preferred). The costs of any contraception method are a function of the difference between the number of children the couple would have with no contraception, and the mean number of children expected using a given method. Their model implies that any contraceptive method implies both a fixed costs (which must be “paid” to use the method at all) and a marginal cost (which is proportional to the number of births averted). Their model implies (Figure 3) that the lifetime cost of using any contraceptive strategy would depend on the number of births the couple seeks to avert, in addition to the actual fixed and marginal costs. A couple that wanted to avert three of an expected eight births, for example, would be happier with a relatively high marginal cost approach than would a couple that wanted to avert all but two of eight expected births.25

We usually lack direct evidence on the use of specific contraceptive methods until the twentieth century. What we know about contraception in the past often comes from sources condemning contraception and discussing efforts to prevent its use. Indirect evidence supports the view that until the second half of the nineteenth century, most couples who sought to control their fertility did so with a combination of withdrawal (coitus interruptus) and abstinence from sexual relations.26 As Santow (1993) shows, coitus interruptus remained widely-used well into the twentieth century, even in countries where alternatives were available. In the Michael-Willis framework, withdrawal and abstinence count as low fixed-cost, high marginal-cost approaches. Their population may surprise, but we must bear in mind the objectives and the data. Couples today associate birth control with the effort to prevent conception entirely, or to time births within narrow time-windows that reflect professional and other concerns. Much contraception in our

25 See also Bailey (2010)’s useful graphical exposition.
26 Santow (1995) provides the best recent evidence on this issue.
society is used by couples with little or no commitment to a stable, long-term union. The French couples responsible for that late eighteenth-century fertility decline, on the other hand, simply wanted to avoid having six or seven births they could otherwise expect within a long marriage.\(^{27}\)

The nineteenth century also witnessed the spread of “marriage manuals,” a euphemism for guides to sexuality, sexual health, and contraception. The usefulness of such guides in limiting fertility of course depends on their information being more accurate than what couples already knew from other sources, and in some cases that is doubtful. Some, for example, stressed the dangers *coitus interruptus* posed to male health. Still, the popularity of such guides as well as efforts to limit their dissemination suggests that couples viewed them as better than alternative sources of information.

The first “modern” methods appeared in the second half of the nineteenth century. These techniques relied partly on advances in medical understanding, but also on the invention of vulcanized rubber (in 1844). Applied to the production of condoms (in 1855), this new industrial process allowed couples to replace expensive and relatively unreliable condoms made from natural materials. (Most studies, in fact, conclude that prior to vulcanization, condoms were used to prevent the spread of disease, not pregnancy). Rubber condoms were at first expensive, but more technical improvements brought their price within the budgets of even working-class people. Vulcanized rubber was also the basis for the introduction of the diaphragm and similar intrauterine barriers in the later nineteenth century. The latter required the attention of a trained medical professional and thus was more expensive to use. The contraceptive pill as used today dates from the 1950s, with its first widespread use taking place in the 1960s.

\(^{27}\) Bailey (2010) find that relaxations on the availability of contraceptives in the United States, especially the oral contraceptive or “pill,” had significant impacts on several outcomes, including fertility. But this was in a context where U.S. fertility was well below its historical maximum or biological limit; as she argues, the spread of the oral contraceptive reduced the marginal cost of averting births. Goldin and Katz (xx) find that the pill gave women the additional control over their fertility necessary to integrate into the paid labor force.
In the nineteenth and early twentieth century, the U.S. and many European countries made concerted efforts to limit the spread of contraceptive knowledge and technologies. The focus and practical enforcement of such laws were uneven. In the United States, the “Comstock Laws,” a collection of state and federal statutes made it illegal to disseminate both marriage manuals and contraceptive devices such as condoms. Similar measures were enacted in England, Germany, and many other European countries. Police interest in enforcing the regulations also waxed and waned. The policies themselves could also be self-contradictory. Germany’s Lex Heinze (1900) made it illegal to advertise or sell condoms and other contraceptive devices at the same time that the German army distributed free condoms, with instructions for use, to all its recruits. Less important for our purposes, some efforts to regulate the availability of contraceptive devices reflected the medical profession’s efforts to create demand for their services. The diaphragm required a doctor’s services; condoms did not.

The effects of these restrictions on the fertility transition are not really known. Demographers today tend to argue that the availability of contraceptives and contraceptive information is the most important barrier to fertility decline in developing countries. Economists are more skeptical. Bailey (2010) is one of the few careful empirical studies of the issue, but deals with a period well the U.S. fertility transition. She concludes (p.122) concludes from her study of the Comstock Laws that in 1965 that, without the bans in place, marital fertility in the affected states would have been eight percent lower.

We know enough to understand that lacking modern contraceptive methods could easily have reduced their fertility voluntarily in ways consistent with the data we observe in the nineteenth and early twentieth centuries. The early nineteenth-century French did limit their fertility and they did not have access to condoms, the pill, etc. In an unfortunately neglected study, David and Sanderson (1986) show that even withdrawal could produce the fertility reductions we observe in the nineteenth century. They develop a model of a couple’s lifetime
fertility as a renewal process. Using parameter values drawn mostly from studies of married American women after World War II, they derive estimates of the number of live births a couple would experience in a twenty-year marriage under various assumptions about coital frequency and contraceptive failure rates. Their baseline couple (no contraception) would have about nine births if they had sexual intercourse, on average, five times per 24-day cycle. If this couple used a method with a 12.5 percent failure rate, and failed to use it about 10 percent of the time, they would have only three births in twenty years. This “method” amounts to what we know about the use of coitus interruptus in modern populations. Conscientious use of condoms would get the couple below one birth.28

Before leaving this subject we should stress one implication of the ubiquity of coitus interruptus. Some scholars argue that even the high, pre-transition fertility reflects some fertility control, largely achieved with withdrawal (Santow (1995)). This argument implies that what we call the fertility transition is not a shift from fertility at a corner solution to some interior solution, but an increase in fertility control itself. More directly, it underlines the endogeneity of contraception’s use. The French achieved their fertility transition with methods that had been known for centuries. The English presumable understood the same basic biological facts, but chose not to reduce fertility in the early nineteenth century. Changes in contraceptive technology can only get us so far.29

Increases in the direct costs of children

Two types of costs go into Becker’s $p_n$, direct of out-of-pocket costs, and opportunity costs that are invariant to child quality. In the Q-Q model, increases in $p_n$ induce substitution

---

28 Their model is fairly standard in the demographic literature, and is based on Sheps and Menken (1973). They present a range of simulations; the one discussed here is unrealistic only in making these methods seem less effective than they actually are. Michael and Willis (1976, Table 2) report a similar exercise using different parameter values. In their model, a couple using no contraception would have 11.42 births. If they used condoms, they could expect 2.33 births (with a variance of 1.64). Withdrawal, in their results, would produce 2.74 births (with a variance of 1.79).

29 STOPPING SPACING STUFF/ABORTION
towards both child quality and the Z goods directly, and towards quality through the interaction
effect. One logical possibility to explain the fertility transition is that the direct costs of child-
bearing changed in ways that induced couples to have smaller families. The problem is that most
costs did not change, over the relevant period, in ways that would produce the observed fertility
decline. Most households in this period devoted the vast majority of their expenditure to food,
clothing, and housing. The real price of clothing dropped dramatically following the
technological innovations of the Industrial Revolution, many of which were in textiles. Food
prices varied over time and place, and protective tariffs on agricultural goods could raise the price
of food in one country well above its counterpart in others. But in general, food prices declined.

The only significant difference in direct costs took place because of urbanization. Most
European countries as well as the United States experienced rapid urbanization during the
nineteenth century. About six percent of the U.S. population lived in an urban place in 1800; in
1900 that was nearly forty percent (Haines 2000, Table 4.2). England was already very urban in
1800 (28 percent), and became even moreso over the nineteenth century. France started out the
period less urban, and while its cities did grow, remained less urban than England or Germany.
Urbanization in Germany was especially rapid in late nineteenth century. Germans living in
places with fewer than 2000 people fell from 64 to 40 percent of the population between 1871
and 1910. Equally striking, in 1910 about 21 percent of all Germans lived in a city of 100
thousand people or more, truly enormous for the period (Wehler 1995, Table 71). In urban areas
housing costs significantly exceeded costs in urban areas, but of course the decision to live in a
city was up to a couple. Most studies in fact find that urban fertility was lower than rural fertility
in the nineteenth century, although the precise causation has not been established. Once the
fertility transition began, fertility usually fell first in urban areas, with rural areas then catching
up, although there are exceptions: in the United States urban and rural fertility fell in step.
One literature is based on the idea that direct costs might include the costs of establishing a child in an independent life. Richard Easterlin’s (1976) famous explanation of the decline of fertility in rural American rests on the rising costs of farmland as an area is settled. Suppose a farm couple wanted to establish each child on a farm similar to their own. As the price of local farmland rises, parents either have to send their children further west, where land was cheaper, or have fewer children. Easterlin argues that parents prefer to have fewer children and be able to settle them locally. This very influential argument has the virtue of focusing attention on a set of facts that have as yet escaped the attention of growth theorists. Fertility in the rural U.S. began to decline long before the late nineteenth century. Easterlin dates the beginning of the decline in New York state to 1805 and even Iowa, much further west, to 1835. Later research has challenged the demographic details in his study, but most subsequent discussion focuses on the assumption that parents want to give each child a fixed bequest. Sundstrom and David (1988), for example, motivate their regression analysis using a bargaining model that presumes that a primary motivation for child-rearing is support in old age. In equilibrium, children can drive a harder bargain with their parents if they can point to better, off-farm opportunities. Cross-sectional regressions for U.S. states in 1840 show that fertility is strongly and negatively correlated with measures of non-farm labor-market opportunities. Once such proxies are introduced, land prices have no influence on fertility.

Child labor raises another source of variation in direct costs. In many societies children offset some of the direct costs to their parents by working either directly in parental income-generation activities (such as a farm) or by working in the labor market. Any change in children’s earnings would clearly alter the net costs to their parents. There are two important trends during the period in question. Many accounts argue that industrialization at first increased income-

---

30 His fertility measure is the child-woman ratio, or the number of children age 0-9 per thousand women 16-44. This measure is easy to criticize; it is sensitive to in- and outmigration of both children and adults, as well as to variations in infant and child mortality. He uses this measure because it is all one can do with census-based information.
earning opportunities for children. Nardinelli (1990) argues that the appearance of children in the workforce in industrial Britain reflects mostly their shift from other activities, including agriculture. Family income increased, and children remained at home longer than they had in an earlier pattern of “life-cycle service,” that is, a period of working as agricultural laborers on one-year contracts. By the 1830s, when the practice was controversial, large minorities of children were working. Nardinelli (1990, Table 4.2) reports that in most English counties, at least one-quarter of children aged 10-14 were reported in the workforce. Some parts of the textile sector depended heavily on children. One parliamentary inquiry reported that in cotton textiles, half of all workers where under 18, and 6.8 percent were under 10 (Nardinelli 1990, Table xx).

By the late nineteenth century, on the other hand, many countries had imposed restrictions on the use of child labor, especially outside of agriculture. The use of children in industry especially became quite controversial, and governments imposed age restrictions and other measures that dramatically reduce the earnings possibilities of children, especially in industry.31 The British “Factory Acts,” starting in 1833, imposed increasing restrictions on the ages of children who could work, and how many hours they could work. But they started at modest level; the 1833 Act restricted children aged 9-12 to forty-eight hour weeks. Many governments coupled restrictions on child labor with an education requirement. The British Factory Acts required that child workers also be in school. In some cases, the factory had to set up its own school to continue employing children (Nardinelli 1990, pp. 106-7). In Massachusetts, as of 1837 manufacturers could not employ anyone under the age of 15 who had not attended school at least three months in the previous year (Moehling 1999, p. 74). There is no careful study

---

31 There are two styles of explanation for this development. One is that a combination of social-welfare concerns, along with representatives of labor concerned about competition with adult males, overwhelmed industrialists’ opposition to child-labor restrictions. The other explanation is that these measures passed when industry no longer opposed them; either it had become easy to substitute capital and other sorts of labor for child labor, or the workforce had already changed in ways that the new laws where not a binding constraint when passed. Moehling (1999) argues that latter for the United States and gives references to the debate.
of the effect such measures had on the demand for children, but at least in this case the timing of
the changes are broadly consistent with the timing of the fertility decline.

*Increases in the opportunity costs of child-bearing*

Building a family also entails opportunity costs. Many studies have argued that
industrialization created new income-earning opportunities for women, and that to take advantage
of these opportunities, women had to curtail their fertility. The argument is not that women did
not “work” before. Rather, industrialization offered better-paying work that could not be
combined with child-minding; thus a woman spinning yarn at home could also care for children,
while a woman working in a textile factory could not.

Many studies find that local employment opportunities for women lower fertility. The
best is Crafts (1911). The English census conducted a special fertility survey in that year, which
Crafts uses to examine the relationship between local employment opportunities and the fertility
of marriage. Crafts (1911) finds a consistent, negative correlation between women’s local labor-
force opportunities and marital fertility, with elasticities ranging from -.13 to -.34. A number of
studies of the German fertility transition report similar results, although the data are not as
suitable as Crafts’.

*Changes in the costs of and returns to child quality*

Many recent accounts of fertility and long-run growth focus on the idea that fertility is
driven largely by reductions in the cost of child quality or increases in the return to education –
that is, increases in the return to child quality. The ideas are in principle distinct. Consider first
education. Historians of education stress important distinctions in the types of economically

---

32 Other studies find a negative elasticity of fertility with respect to local employment. The problem is that
we often do not know the opportunities for women. Brown and Guinnane (2002) use employment in
textiles as a proxy for female employment opportunities. This rough approximate does yield a significant
but small negative effect on marital fertility.
useful education. One could acquire basic literacy and numeracy at home (if the parents were literate) or in primary school. More advanced education or training required secondary schools, formal apprentice, or less formal on-the-job training. Tertiary education was restricted to a small elite and while perhaps important for overall TFP growth, would not figure heavily in demographic decisions.

The growth of literacy and its primary cause, compulsory public elementary education, differs dramatically across the countries on which we focus. Prussia led the way with the 1763 requirement that all children aged five to thirteen attend primary schools. The schools were not free, but there was tuition assistance for the poor. Like many grand educational reforms, this measure’s implementation was resisted by various interests, and in any case Prussia lacked sufficient teachers for all the children in the territory (Melton 1988, pp.174-177). Several U.S. states introduced free public elementary education starting in the 1840s, and for most of the nineteenth century the U.S was an outlier in the number of students. Free, compulsory education came much later to Germany (1872), France (1882), and England and Wales (1893) (Bruland 2003, pp. 160-161). Easterlin (1981, Appendix Table 1) estimates that in 1850, there were 1800 children in primary schools per 10,000 total population in the U.S. compared to 1600 in Germany, 930 in France, and 1045 in the United Kingdom.

By the end of the nineteenth century, then all western European countries, along with the United States, had free, compulsory public education systems. But there remained the opportunity costs of the child’s wages, and these seem to have been the real limit to children’s school attendance in the mid-nineteenth century. Mitch (1992, p.156) quotes Horace Mann’s comment in the British 1851 census of education: “It is not for sake of saving a penny per week, but for the sake of gaining a shilling or eighteenpence a week that a child is transferred from the school to the factory.” The opportunity cost declined as legislation restricted child labor.
What were the returns to education? Historical sources almost always lack the ingredients needed to estimate the relationship between education and future incomes. Goldin and Katz (2000) exploit the unique Iowa census of 1915 to estimate that the return to an additional year of high school or college then was, for males, on the order of 11-12 percent (Goldin and Katz 2000)). Goldin and Katz’s estimates do not net out the costs of acquiring that education. Mitch (1992) reports a calculation more nearly that of parents considering the Q-Q tradedoff. He uses data on school fees, childrens’ earnings, and the wage premium in occupations that required literacy to estimate the present value of acquiring literacy in the Victorian era. The present value of the cost of acquiring literacy by attending 40 weeks of school as a youngster would be £4. At a wage premium of 5 shillings per week for literacy, the present value of the higher wages for a 35-year work life would be over £200.

*Social insurance and old-age support*

One particular return to child-rearing that receives considerable attention in the literature is children’s potential role in old-age support. The argument is that children are a form of life-cycle savings; parents invest when they are young and healthy, and then expect their children to care for them in infirmity or old age. (The Sundstrom-David criticize of Easterlin’s model, noted above, is one version of this argument). Two versions of this argument have been advanced to explain the fertility transition. One is that it became harder for parents to hold children to the intergenerational bargain. This “child default” argument implies that the developing industrial economy made it simpler for children to run off and leave their parents to fend for themselves, making children a less desirable vehicle for savings.33 A second version of the argument points to the development of substitute means of providing for old age, in the development of financial instruments and more important, social insurance and the welfare state.

33 This would also be consistent with Caldwell’s changes in “net intergenerational wealth flows.”
At an analytical level both versions of the argument suffer from the problem that we know that economic ties between parents and children varied dramatically across the societies in question. In some peasant regions of Europe, peasant households would draft formal documents that turned a farmstead over to the heir, and carefully specified the heir’s obligations to his parents (as well as to siblings to had not yet received an inheritance). At the other extreme, rural laborer’s children in England would, from at least the early-modern period, leave home for good in the early teens, and the best evidence suggests no further economic relationship with parents. We should also note that the “child default” version of the argument is similar to arguments about mortality (although in reverse): from the parents’ viewpoint, rising child default is like increased mortality. Parents’ might actually invest in more, lower-quality children to ensure that at least some remained faithful.

The potential effect of rising social insurance has not received the careful attention it deserves. Every society in question here had some form of provision for the poor and those unable to provide for themselves because of illness or age. But in most places it was not as organized and secure as Britain’s famous Poor Law, and even the British system worked with a severe asset test. Social insurance strictly understood allows individuals to accumulate assets without sacrificing claims on payments in case of illness or old-age, and thus changes the incentives to accumulate wealth for the future. The first broad social-insurance system dates to 1883, when Bismark introduced compulsory sickness insurance in Germany. His system at first did not cover all workers, and only gradually did the government add accident (1884) and much later old-age insurance (1896). The German systems were self-funded. Britain’s non-contributory Old Age Pensions Act dates to 1908. One can imagine that these two sets of measures had some effect. France… The United States did not have anything comparable until the Social Security system was created in the 1930s.
4. Conclusions

IN PROCESS!
References


Deaton, Fogel review


Fogel, escape


Crude birth rates, selected countries, 1820-1970

- France
- England
- Germany
- United States
- Italy
Figure 2

Fertility and mortality in Germany
(Number of events per thousand population)
Figure 3: Cohort fertility rates

Source: Festy

Note: The cohort fertility rate is the mean number of children born to women belonging to the birth cohorts on the horizontal axis. The overlapping years are in the source. The precise birth cohorts vary slightly across countries.
Figure 4: Malthusian comparative statics

- Births and Deaths
- Birth and Death Rates
- Real Wage
- Population

Graph showing the relationship between real wage, population, and birth/death rates.
Figure 5: Fertility, mortality, and the real wage in England

Source: Wrigley and Schofield (1981, Figure 10.6)