The Causes and Consequences of the Demographic Transition

David Canning

July 2011

PGDA Working Paper No. 79
http://www.hsph.harvard.edu/pgda/working.htm

The views expressed in this paper are those of the author(s) and not necessarily those of the Harvard Initiative for Global Health. The Program on the Global Demography of Aging receives funding from the National Institute on Aging, Grant No. 1 P30 AG024409-06.
The Causes and Consequences of the Demographic Transition

David Canning

Harvard School of Public Health

July 2011

Abstract:

The causes and consequences of the demographic transition are considered in light of the recent book by Dyson (2010) on demography and development. In the last 50 years the world has seen an exogenous decline in mortality that generated a decline in fertility and an increase in urbanization that has had profound economic, social and political consequences. However, historically, declines in mortality and fertility, and escape from the Malthusian trap, have required countries to have already undergone considerable economic and political development. We therefore argue for two way causality between the demographic transition and economic and political outcomes.

JEL Numbers: J11, I15, O10.
**Introduction**

Tim Dyson (Dyson (2010)) has given us a short book on a big subject. The last 300 years have seen three remarkable changes in human society. One is the demographic transition, substantial reductions in mortality rates set off a population explosion, followed by reductions in fertility that are leading to stable, and in some cases declining, population numbers. A second is economic growth, the emergence for the first time in history of sustained increases in income per capita. The third is an increase in social and political equality, particularly between men and women, with the adoption of democracy and universal adult franchise. This book asks what causes the demographic transition and what have been its consequences for social, political, and economic development which is essentially the question of how we explain these three fundamental changes and the links between them.

Most academic work, both by individuals and within disciplines, is narrow, focusing on particular mechanisms and there is a tendency to shy away from statements about the big questions. To its credit, the book does not do this but rather addresses these questions directly by integrating insights from different disciplines to produce a coherent view of population and development.

It is easy to critique the book in terms of the individual mechanisms that lie behind the theory. Each of these has its own literature, with established and dissenting views, and in every case more could be said than in this short book. I do talk about a few of these mechanisms, where I feel there is evidence for larger effects than are reflected in the book. However, the key issue is whether the “big picture” of the demographic transition and its relationship to development put forward by the book is correct. This means tackling the central thesis of the book which is that the demographic transition is a largely self-contained process that proceeds...
independently of social and economic development. The argument is that the mortality transition occurs exogenously and is the cause of a subsequent decline in fertility and rise in urbanization. This exogenous demographic process plays out over time, and has effects on economic, social, and political development.

The case for the complete independence of demographic processes is absurd. There are clear examples of cases, such as civil wars, when economic, social, or political circumstances have affected population numbers. While acknowledging this, and discussing the complex mechanisms that affect demographic processes, the thesis of the book is that by and large demographic processes are relatively independent and that as a first approximation they are exogenous and lead to changes in development. One key issue is, therefore, whether there are important mechanisms that flow the other way, from development to demography that might have substantial effects and would undermine this view.

A second question is how important the demographic transition has been for the economic and political changes we have seen. The book makes the case that the demographic transition made a substantial contribution to these processes but does not address the question of whether the transition was necessary – could the economic and political changes have taken place without the transition? I deal with these two central questions in detail below.

The book takes population as defined by fertility, mortality, sex, age structure, and urbanization. This focus on the number of people, and where they live, is the standard approach in demography but it misses a key element in terms of the quality of people. There is discussion in the book that reductions in mortality are probably accompanied by improvements in population health leading to social and economic benefits but this quality mechanism is not investigated in the same way as the quantity mechanisms.
Dyson’s view of population and development can be contrasted with the alternative "unified growth theory" as expounded by Galor (2011). The major theme of that theory is that there are two epochs in development, the first characterized by a Malthusian trap where any improvement in living standards inevitably leads to pressure on agriculture and starvation, followed by an epoch of sustained economic growth following escape from the trap. Dyson essentially focuses on the modern epoch after we have escaped the trap, but the issue of how we did this is a fundamental question in development.

Causes of the Transition

Mortality

The book documents the large decreases in mortality over the last 300 years and argues that they have mainly been due to advances in health knowledge that lead to the prevention and treatment of infectious diseases. The key argument here is the Preston Curve (Preston (1975)) which shows that while life expectancy is higher in high income countries, most health gains come from the curve moving up over time with gains in life expectancy at all income levels. The contribution of rising incomes to improved health, while real (Pritchett and Summers (1996)), is modest in comparison. This picture is true both across countries and within developed countries (Canning and Bowser (2010)) though in developed countries there is likely to be a contribution from the treatment of non-infectious disease (Cutler, Deaton, and Lleras-Muney (2006)).

Even the poorest developing countries today have seen rapid improvements in health over the last 50 years. The caveat to this, of course, is the HIV/AIDS epidemic which has led to sharp reductions in life expectancy in many countries in Sub-Saharan Africa. While this picture of health improvements being largely exogenous seems true today, this may not have been the case
for health improvements in the 18th and 19th centuries. This is linked to the debate about the relative importance of nutrition, which is endogenous and income dependent, and public health measures due to exogenous advances in health knowledge in historical mortality improvements. It is clear that the mortality transition first started in richer countries and spread to poorer countries over time. It is also the case that the implementation of basic public health measures requires public health to be a policy goal and the existence of at least rudimentary infrastructure for the delivery of public health services. Sen (1991) has argued that democracy plays an important role in ensuring the provision of basic nutrition and public health services.

This suggests that historically the timing of the start of the mortality transition in each country was dependent on its level of political development. However, while this may have been true before 1950, since then all countries have started the mortality transition. To some extent this is simply a process of technological diffusion but it has been aided by the fact that expertise and funding is available from international organizations to support these measures. This supports Dyson’s argument, that at least in the modern era, the mortality transition is largely exogenous.

While the mortality transition has had important consequences for economic and social development it is important to realize that longer healthy life spans are themselves a gain in human welfare. If we measure the value of these health gains relative to gains in income we find that the two have been roughly equal as sources of welfare gains over the last century (Becker, Philipson, and Soares (2005)).
Fertility

A key idea in the book is that while fertility is influenced by government policy, contraceptive availability, education, ideation, and culture, the central underlying cause of the fertility transition is the prior mortality transition. The strongest evidence for this is the fact that the fertility decline always follows the decline in mortality. Against this, the initial response to the start of the mortality decline is a rise in fertility and the very rapid fertility transition in some countries in recent years relative to historical episodes means the relationship may not be automatic. There is undoubtedly an effect of lower child mortality on desired fertility. The increase in child survival rates reduces fertility required to achieve a desired number of surviving children. An alternative explanation is the economic model which emphasizes fertility decisions as being the result of a quantity quality tradeoff. A key is the returns to education and human capital. If technological progress raises the returns to schooling, families may decide to have fewer children but invest more in each child. A third idea is that advances in contraceptive technologies have played a key role in allowing the reduction in fertility.

It is difficult to decide between these theories using household level data. Replacement fertility, in direct response to a child death is fairly low, but the real issue is how important future child mortality risk is in fertility. Similarly with education, the issue on returns is about expectation of earnings over the child’s future life time. Neither the mortality risks, nor earning expectations, are usually measured directly in household surveys. However there is clear evidence that while household level characteristics do affect fertility, most of the fertility change we see during the fertility transition is not due to changes in these characteristics. Rather all women move towards lower fertility without large changes in household level socio-economic variables. Changes in fertility behavior is more like a change in a social norm than changing
independent individual decisions characteristics, and is likely because of common, community-wide factors, such as mortality expectations or earnings expectations. The book also makes the case that while the mortality transition may be the underlying cause of the fertility transition it may work through intermediate factors, making causality even more difficult to determine. For example, reductions in infant mortality may lead to large family sizes, with negative consequences, and it is the observation of these large families in their communities that may lead to women deciding to have fewer children without explicitly considering mortality rates.

The availability of contraception and abortion clearly affect fertility rates, but may not be decisive in allowing fertility to decline. In Ireland despite contraception and abortion being illegal the total fertility rate had declined to four by 1979 which was very high by European standards of the time. After 1979 and the legalization of abortion, fertility fell quite quickly to replacement levels with consequent effects on female labor supply, age structure, and economic growth. My view is that it is very difficult to see how a country could achieve replacement fertility without access to contraceptives or abortion. However since all countries now seem to allow access to some form of contraception\textsuperscript{1} in practice this is no longer a key issue.

In many countries the fertility transition has occurred after mortality decline but before substantial gains in income per capita, for example, both India and China saw large fertility declines before each of their growth surges. However, the continuing decline in fertility to replacement and below usually occurs in conjunction with rapid growth making it difficult to establish causality. While the aggregate level data is consistent with the idea that the mortality transition drives the fertility transition, my own view is that the question is still open on the importance of the returns to human capital and the quality quantity tradeoff. Today this is a

\textsuperscript{1}The mix of different contraceptive methods, sterilization, and abortion used to control fertility varies enormously across countries suggesting that access to at least one method has a large effect while the availability of some methods and not others may have a smaller impact.
central question in Sub-Saharan Africa, will the substantial declines in child mortality we have seen be sufficient to induce large declines in fertility without economic growth or will fertility decline stall?

*Urbanization*

Prior to the demographic transition high mortality in cities tended to give them an excess of deaths over births and they were sustained only by rural to urban migration. As mortality declined the death rates in cities fell below the death rate in rural areas. The urban population became self sustaining though urbanization, meaning the increasing proportion of people in urban areas, requires continuing rural to urban migration.

The economic theory of urbanization is that migration equalizes net advantages of living in different settings; these net advantages take into account all factors such as earnings, living space, health, and social contact (Glaeser (2011)). Cities tend to have higher wages on average than rural areas but housing prices are higher and crowding may make cities more or less attractive than the countryside. The rise in rural population that follows mortality decline means more workers per acre of land and lower agricultural wages due to diminishing returns, the classical Malthusian argument. Production in the manufacturing and service sector in cities do not have a fixed factor and do not suffer from decreasing returns, in fact we see higher productivity in larger and denser cities.

From the point of view of a single country with international trade, population pressure will therefore tend to produce urbanization. However on a worldwide basis, the larger population still needs to be fed. In a Malthusian setting this requires an ever increasing labor inputs in farming with diminishing gains in output. In fact, in the modern era we see increased
productivity in agriculture attributable to both technology and increased capital intensity. Technological progress and increased capital tends to reduce the demand for labor in agriculture and increase the demand for labor in industry. A striking feature of the modern era has been the reduction of the absolute numbers of workers in agriculture and the reduction in land area devoted to agricultural production – technological progress and the mechanization of farming have essentially removed the Malthusian check. A similar process is now happening in the second transition where technological advances and increasing capital intensity are reducing the demand for labor in industry and are being accompanied by a shift of labor into services.

My view until quite recently was that urbanization was a combination of push factors from the countryside and pull factors from the city, with excess labor in the rural area producing a push and industrialization and high wages in the cities providing a strong pull. The problem with this view is that, as pointed out by Dyson, urbanization in Africa over the last 50 years has proceeded apace with that of Asia (Bloom, Canning, and Fink (2008)). Given the more rapid productivity gains in agriculture and industry in Asia, this is difficult to explain. Dyson’s conclusion is that urbanization is the inevitable result of the demographic transition. The excess population in the rural areas tends to move to the cities largely independently of economic considerations, partly because cities are more attractive places to live for non-economic reasons. For example, mortality rates are now often lower in urban than in rural areas, even among people of low socioeconomic status, and cities may have cultural and social advantages not found in rural settings.

My view is that large scale urbanization requires technological progress and increased capital intensity in agriculture to allow sufficient food production for the growing population with a declining agricultural workforce. However, the similarity of the urbanization rates in Asia
and Africa means that the pull factors of industrialization and high urban wages may not be of central importance.

*The Quality of People*

A remarkable feature of the modern era is the increase in the “quality” as well as quantity of people. In economics, quality or human capital has been taken to be synonymous with education but we should think of the quality of people in a broader sense. Over the last 100 years we have seen substantial increases in the physical and cognitive development of children, as evidence by gains in adult heights (Fogel and Costa (1997)) and IQ scores (Neisser (1998)), even in the component not associated with education. These gains have been due to improvements in nutrition and health in the first few years of life (Akachi and Canning (2007)), the mortality transition that sets off the demographic transition is also usually a health transition that is associated with better nutrition and a lower burden of disease. In addition, the onset of disability and physical and cognitive decline in old age is occurring later in life – the compression of morbidity (Mor (2005)). These improvements in physical and cognitive ability, in the quality of people, have played a large role in economic and social development. I would argue that including measures of these improvements in population quality, in addition to population quantity, in demography provides a richer basis for thinking about the link between population and development. My view is that the economic and social effect of these quality changes has been fundamental in promoting economic growth. One reason for a higher return to education is the increased healthy life span people now expect, which increases the time horizon for the returns to education.
Consequences of the Transition

Economic Consequences

Surprisingly, the original Malthusian argument that larger populations are associated with lower incomes, because of population pressure on agriculture and food leading to impoverishment, is not discussed in the book, perhaps because it has already been discussed in Dyson (1996). I agree with the argument there that productivity gains and mechanization have meant that food production has not proved fundamental barrier to economic growth though the case for the Malthusian argument in the African context has been undergoing something of a revival (Young (2005)). Newer versions of the argument are that large population numbers will lead to impoverishment in the future because pressure on other scarce resources such as energy, or that it will contribute to global warming. My view is that these possible negative consequences can be addressed through incentive mechanisms, such a pricing carbon emissions, and do not fundamentally require population control, but it would be helpful to have at least some discussion of these issues.

The fundamental point however is that in practice we have escaped the Malthusian trap. The demographic transition leads to four to ten fold increase in population numbers since the mortality decline precedes the fertility decline. In the Malthusian world this would put pressure on agriculture, leading to starvation, and the death rate would rise to return the population to its equilibrium size. We have avoided this outcome by massive gains in agricultural productivity. These agricultural productivity gains are a necessary condition for the demographic transition to occur- without them any mortality gains would be short term and would be reversed by starvation. This undermines the argument for the pure exogeneity of the mortality transition - it requires technological advances in agriculture to be available for the transition to occur,
Even when we have escaped the agricultural Malthusian trap, the Solow growth model predicts a negative association between population growth and economic growth as the growth in the work force dilutes the capital stock per worker in industry. There is some discussion of this negative relationship using recent data in the book but my view is that studying this relationship can be misleading. In practice the economic effects for population growth depend fundamentally on whether it is due to a high birth rate or a low death rate (Bloom and Freeman (1988), Kelley and Schmidt (1995)). Adding these two components together, and expecting the same results on growth independent of the source of the population growth is not feasible. If population growth has negative consequences for economic growth, we would expect to see a negative association between the birth rate and economic growth. Figure 1 shows just this expected negative relationship between the birth rate in 1980 and growth in real GDP per capita (purchasing power parity adjusted) over the period 1980 to 2000 in 127 countries using data from the World Bank (2011). On the other hand we expect to see a positive relationship between the death rate and economic growth as high death rates ease population pressure. However the relationship between death rates on economic growth shown in Figure 2 for the same sample is negative, indicating that countries with higher death rates have lower rates of economic growth. The effects of birth and death rates are not equal and opposite as would be required for population growth to be a meaningful summary statistic.

The book does discuss some reasons why birth and death rates have effects over and above their impact on population growth. The first is that they have different effects on age structure which can give rise to a demographic dividend, a rise in the working age share of the population, when the baby boom, brought about by falling infant mortality, is followed by a
decline in fertility. More importantly the decline in death rates leads to a longer life span which can increase the incentive to invest in human capital. Declining mortality is usually accompanied by declining morbidity. There is increasing evidence that improvements in early childhood health and nutrition lead to improved educational outcomes and labor market productivity.

Fewer children allows greater female labor market participation (Bloom, Canning, Fink, and Finlay (2009)). In addition, smaller numbers of children can allow a quantity –quality tradeoff with each child receiving a larger investment in health and education. While the book makes all these points there is a difficulty in establishing the magnitude of these effects since the associations seen in the data may not be causal. However there is recent evidence of large causal effects on education and earnings of eradicating malaria and improving childhood nutrition (Bleakley (2010), Hoddinott, Maluccio, Behrman, Flores et al. (2008)). There is also emerging evidence of causal effects of reduction in fertility on investments in children (Schultz (2009)). While everyone seems to agree on the negative effects of high fertility, there is still a debate about whether the quality effects of improvements in health outweigh the negative effects of additional population numbers when mortality declines (Acemoglu and Johnson (2007), Ashraf, Lester, and Weil (2007), Bloom, Canning, and Sevilla (2004)) .

**Social and Political Consequences**

The case for the social and political consequences of the demographic transition is perhaps the most difficult; though there is undoubtedly an influence, its magnitude relative to other forces is difficult to determine. In the Malthusian world we have high returns to land and excess labor at subsistence wages, which favors a political system under the control of those who own the land. The very high death rates, and decline in population, due to the Black Death in 14th century
Europe appear to have caused a shortage of labor, leading to a rise in wages and the breakdown of the feudal labor system (Herlihy (1997)), but this was due to a temporary rise in mortality, not the declines we see as part of the demographic transition.

The demographic transition and population growth would lead us to expect a world of labor surplus with high economic rents and political power for the owners of land and capital. This has not happened, rather the wages and the political power of workers have risen relative to other classes. The key force behind the rise in labor productivity and wages was the industrial revolution and the rise of industry. The new means of production increased the power of workers, particularly skilled workers. My view is that it is this rising economic power, rather than the demographic transition that lead to the progressive expansion of the franchise among men in several countries during the 18th and 19th centuries.

Where the book may be on much stronger ground is in the decline in gender specific roles in society, the decline in marriage, and the economic and political empowerment of women being linked to the decline in fertility. Part of this political empowerment has been the extension of the franchise to women. The emphasis of psycho-social as opposed to purely economic mechanisms for these changes also seems plausible. As always, however, we have the problem of reverse causality, that these economic and social changes may determine fertility as well as being determined by it.

**Unified Growth Theory**

In unified growth theory we begin in a Malthusian world in which improvements in technology can raise living standard temporarily, reducing death rates. Population growth then puts pressure on agriculture which leads to starvation until equilibrium is reestablished at subsistence level of
wages but with a larger population. Improvements in health technology and reductions in mortality similarly raise population numbers and lead to starvation raising the death rate back to its old level. In such a world economic growth in per capita terms is impossible. However, technological progress does raise agricultural productivity and allows for a larger population to be sustained - all technological progress goes into population numbers not income per capita.

As population numbers increase the rate of technological progress increases; a large population produces more new innovations than a small population. Technological progress becomes sustained and cumulative; it becomes geometric rather than arithmetic. Improvements in agricultural techniques and the mechanization of agriculture make the fixed factor, land, less important and less of a constraint. More importantly this technological progress leads to new methods of production, and new research methods for producing new technologies, that increase the returns to education. This leads to a quality quantity trade off where families choose to have fewer children in order to enable investments in education that will make these children better off. This reduction in fertility means population growth is limited to a rate that can be covered by technological advances, and allows sustained economic growth. Economic growth is also assisted by improvements in health.

In Dyson’s view the central driving force is improvements in mortality that induce lower fertility. Unified growth theory gives pride of place to technological improvements in production and increased returns to education that produce the quantity quality tradeoff and induce fertility decline.
Conclusion

Dyson puts forward a coherent theory in which a largely exogenous demographic transition, fundamentally driven by reductions in death rates due to advances in public health has had major effects on economic, social, and political development. When I first read this book I was surprised by the argument for the exogeneity of the demographic transition. My own view is that the mortality transition is initiated by public health advances, but for it to be sustained we require productivity gains in agriculture that allow the larger population to be fed and to prevent the mortality gains from being reversed through starvation. I would also put more emphasis on the argument that a major underlying cause of fertility decline is technological progress that raises the return to education and induces a reduction in the number of children and greater investments in each child’s schooling - the quality quantity tradeoff. I would also emphasize the role of economic forces in determining urbanization, rather than seeing it as an inevitable consequence of the demographic transition.

While I disagree with Dyson, he has an arguable case that while these other forces exist, they have not proven decisive and in practice fertility decline and urbanization have inevitably followed the mortality transition. His view is essentially demography can be left to demographers - whereas I think it is a more organic part of the modern development that has been in part a consequence as well as a cause of the process. It may be that the issues are more about timing that about eventual outcomes. The process described by Dyson – the demographic transition and associated economic and political changes – is inevitable but the speed at which this process occurs may differ greatly depending on feedbacks to mortality and fertility. While Saharan Africa will likely follow a similar process of development to that already seen in other
countries this process may take 50 years, or it may take 200 years; the difference between these in terms of human welfare is enormous.

While the book makes the case that the demographic transition has contributed to economic growth and political change, a key question is whether it is necessary and sufficient for modern development. Unified growth theory places the fundamental source of growth as technological progress in production with a contribution from improved population health as mortality declines. However, in the absence of fertility reduction the theory predicts that the economy is in a Malthusian Trap; technological gains lead to growth in population numbers rather than income levels. While technological gains are the driving force, in order to turn these into higher income, and escape the Malthusian Trap, we need reductions in fertility. My view is that these reductions in fertility have been central in allowing us to escape this trap.

However, technological advances in health, and the improvement in health human capital, have played a major role in spurring economic growth. I would go further than Dyson in emphasizing that health improvements that reduced mortality and increased population numbers also improved population quality in the form of health human capital and these quality gains have helped fuel economic growth.

The story of the modern world is often told as a story of the industrial revolution driven by technological advances in manufacturing. To this we should add technological advances in public health that have allowed enormous gains in life expectancy. The modern world has seen both an industrial revolution and a health revolution. In addition to both of these we have the remarkable achievement of having escaped the Malthusian trap of population pressure leading to food shortages. This, in part, has been driven by technological advances and mechanization in the agricultural sector, but mainly by a decline in fertility.
Based on data for 127 countries from the World Bank's World Development Indicators 2011. The GDP per capita is measured at purchasing power parity and the growth rate is the annual average over the period.
Based on data for 127 countries from the World Bank's World Development Indicators 2011. The GDP per capita is measured at purchasing power parity and the growth rate is the annual average over the period.
References


