POPULATION AND PUBLIC POLICY: ESSAYS IN HONOR OF PAUL DEMENY
CONTENTS

Preface ix

INTRODUCTION
Reflections on Post-Transition Demography 3
Geoffrey McNicoll

INTERGENERATIONAL RELATIONS
Intergenerational Transfers, the Biological Life Cycle, and Human Society 23
Ronald Lee

The Intergenerational Welfare State 36
Nancy Folbre
Douglas Wolf

LOW FERTILITY, POPULATION AGING, AND THE BODY POLITIC
Why Demographic Suicide? The Puzzles of European Fertility 55
Lant Pritchett
Martina Viarengo

Low Fertility in Historical Perspective 72
Massimo Livi Bacci

On Demographic and Democratic Transitions 83
Tim Dyson

A Demographic Perspective on Japan’s “Lost Decades” 103
Reiko Aoki
POLICY AND PROGRAM: CHINA


WANG FENG
YONG CAI
BAOCHANG GU

Patriarchal Demographics? China’s Sex Ratio Reconsidered 130

SUSAN GREENHALGH

POLICY AND PROGRAM: AFRICA

Fertility Transition: Is sub-Saharan Africa Different? 153

JOHN BONGAARTS
JOHN CASTERLINE

The Recent Fertility Transition in Rwanda 169

CHARLES F. WESTOFF

Fertility and Population Policy in Algeria: Discrepancies between Planning and Outcomes 179

ZAHIA OUADAH-BEDIDI
JACQUES VALLIN

Working Misunderstandings: Donors, Brokers, and Villagers in Africa’s AIDS Industry 197

SUSAN COTS WATKINS
ANN SWIDLER

ENVIRONMENT, TECHNOLOGY, AND WEALTH

Peak Farmland and the Prospect for Land Sparing 221

JESSE H. AUSUBEL
IDDO K. WERNICK
PAUL E. WAGGONER
National Wealth

PARThA DASGUPTA

The Last Eighty Years: Continuities and Change

VACLAV SMIL

POpULATIOn THEOry AND MEASUREMENT

Demographic Metabolism: A Predictive Theory of Socioeconomic Change

WOLFGANG LUTZ

Cross-Sections Are History

RICHARD A. EASTERLIN

Population Change among the Elderly: International Patterns

ANDREW STOKES

SAMUEL H. PRESTON

The French School of Demography: Contextualizing Demographic Analysis

VÉRONIQUE PETIT

YVES CHARBIT

The Twilight of the Census

DAVID COLEMAN

SELECTED PUBLICATIONS OF PAUL DEMENY

AUTHORS
Paul Demeny, founding editor of *Population and Development Review*, relinquishes his editorship with Volume 38 in 2012. This collection of essays marks the occasion and, in the year he turns 80, celebrates his scholarly career.

*PDR*’s main interests lie in the relationships between population and social, economic, and environmental change, and related issues of public policy. The journal from its inception has sought to extend the conventional boundaries of population studies and to avoid disciplinary constrictions. It encompasses problems of low- as much as high-fertility regimes, environmental futures, population and geopolitics, population history, and the history of population thought. And it endeavors to be accessible to a broad range of readers in the social sciences and public affairs. In accord with this last aim, *PDR* has always welcomed contributions written as essays—a time-honored and distinguished genre unduly neglected in standard-order academic articles in many journals.

The 21 essays assembled here were selected from a larger number considered. Each on its merits could have found its way into a regular issue of the journal (all underwent its standard process of peer review). While each is a free-standing contribution, a loose thematic link is offered by the volume’s title and subtitle: population and public policy, and the wide-ranging scholarly interests of Paul Demeny.

The collection begins with three essays, by Geoffrey McNicoll, Ronald Lee, and Nancy Folbre and Douglas Wolf, on aspects of population renewal and the management of intergenerational relations. McNicoll discusses the various ingredients of generational succession and their bearing on the societal future. How the fundamental continuity problems facing every society have been dealt with over the course of human history makes an intriguing story, set out by Lee. Its culmination in the sustainability issues confronting the modern welfare state is the subject of Folbre and Wolf.

The transition to low-mortality, low-fertility regimes is a staple of population research, whether as an expected and welcome part of social, economic, and even political development or, in many countries, as an object of explicit policy intervention. Low fertility and its various consequences are explored in four contributions. Lant Pritchett and Martina Viarengo consider the puzzles for theorists presented by ultra-low fertility and natural population decrease. Massimo Livi Bacci offers an historical perspective on the subject. Tim Dyson points to evidence of the relationship between the demographic transition and the process of democratization. And Reiko Aoki sketches the intractable problems for social policy in Japan created by that country’s low fertility and the extreme population aging it is giving rise to.

The role of public policy in lowering high fertility, an often controversial matter, receives particular attention. The cases considered are those of China and
various African countries. In the former, government intervention has achieved unprecedented success of a kind, but attained at high human cost and with unintended long-run ill effects—as recounted by Wang Feng, Yong Cai, and Baochang Gu and by Susan Greenhalgh. In Africa a skeptical story of government roles in the fertility transition emerges in Zahia Ouadah-Bedidi and Jacques Vallin’s essay on Algeria. However, John Bongaarts and John Casterline and Charles Westoff find generally promising recent developments in family planning program outcomes in sub-Saharan Africa. Susan Watkins and Ann Swidler give a striking depiction of the gulf between international donor intentions and on-the-ground realities in service delivery in Malawi, instancing HIV interventions but suggesting an analogy with birth control programs.

Three contributions treat dimensions of human development beyond the narrowly demographic. Jesse Ausubel, Iddo Wernick, and Paul Waggoner examine human demands on the natural environment—usually a bleak account, but here much more hopeful, anticipating a future restoration of Nature. Partha Dasgupta probes the concept of national wealth and finds gross underestimation of the contributions of improved health and environmental services. And Vaclav Smil investigates the true nature of human advances over the past eighty years as a demonstration of the unpredictability of fundamental change.

The final group of essays is concerned with theory and data. Wolfgang Lutz proposes a model of social change as a cohort succession process. Richard Easterlin explicates the life expectancy–income relationship in cross-section and as it has evolved over time. Andrew Stokes and Samuel Preston show that the familiar pattern of death rate and birth rate declines constituting the demographic transition is repeated, decades later, in exit from and entry to the elderly population. Véronique Petit and Yves Charbit provide a perspective on recent advances in formal demography. And David Coleman looks at the possible demise of the centuries-old instrument of data collection that is the population census.

Appended to the collection is a selected bibliography of Paul Demeny’s writings.

Geoffrey McNicoll
John Bongaarts
Ethel P. Churchill
INTRODUCTION
Reflections on Post-Transition Demography

GEOFFREY MCNICOLL

The demographic regimes of modern affluent societies are characterized by long life, below-replacement fertility (sometimes, well below replacement), large and increasing proportions of elderly, and appreciable immigration from poorer areas. Their main population problems, gauged by intensity of academic scrutiny and public debate, are concerned with the last two of these: aging and immigration. In both cases, at the country level, heated contention over details disguises a fair amount of consensus on what should be done. Dealing with aging, the consensus would say, calls for raising retirement ages, somehow limiting the health care burden, and ensuring the viability of pension systems. Dealing with immigration calls for revising selection criteria, smoothing obstacles to assimilation, and controlling illicit entry. Both, it might be concluded, are fairly “easy” problems, technically if not politically.

Affluent societies also have population problems that by any measure are “difficult”: without obvious technical solution and lacking an agreed basis for policy debate. Most are connected to the fundamental demographic process of population renewal—that is, generational succession (to use the convenient fiction of non-overlapping generations). Population size is the most obviously problematic succession feature, at a time when some countries are facing the prospect of substantially falling numbers. (Others, of course, are still experiencing moderate population increase.) Lower numbers need not be a serious matter where change is gradual, but where it is not—where, for instance, successive generations seem set to fall by one-quarter or more, with ballooning proportions of elderly—governments are finding they have scant proven remedy at hand. The “quality” of the next generation—to draw on a term long out of favor in demography—is another characteristic of succession with major policy significance, taking this to encompass both human and social capital and socialization into good citizenship. A third policy-freighted characteristic of succession is the extent of intergenerational social mobility. And a fourth is the degree of preservation of societal identity over time—how
the balance is struck between acculturation of newcomers entering by birth or migration and acceptance of social change and diversity. These four aspects of generational succession—size, quality, mobility, identity—are not exhaustive but seem to warrant particular attention in considering the post-transition demographic landscape.

Assurance of the next generation’s well-being requires that its members have a claim on the social product. In a modern service- and transfer-based economy, the apportionment of such claims becomes more than a little arbitrary, increasingly an outcome of political bargaining and rent-seeking rather than reflecting the divisions decreed by economic theory. Moreover, the concept of social product itself grows elusive as the proportion of it consisting of readily grasped goods and services—in broad terms, “stuff”—diminishes. These considerations too necessarily bear on policy debate on succession issues.

And as yet another complication, while generational succession is mostly a domestic matter for a country, its various component processes are influenced in many ways by demographic realities elsewhere. Beyond actual international migration, an open trade regime, technology transfer, and large international financial flows create strong competitive pressures on the domestic labor force. More generally, any but the most isolated country is necessarily subject to the economic, environmental, and geopolitical effects of differential paces of demographic transition in the rest of the world, and of the diverse age structures and changes in relative population size that result.

In the following pages I will expand on these points to suggest some of the population policy difficulties ahead in the post-transition demographic era, likely to disturb any hoped-for placidity. Readers of Population and Development Review over its nearly four decades of existence will have encountered them already in various guises, not least in the journal’s Archives department, and I will draw on some of those materials. I rely on the essay form to excuse an extreme sketchiness of treatment.

Generational succession

Consider briefly how the four demographic ingredients of generational succession mentioned above—size, quality, mobility, identity—bear on the societal future.

Size

Once-conventional thinking about demographic transition held that a low-fertility–low-mortality regime, with near stationary population size, was the end of (demographic) history. The regime came, but not that outcome: many countries instead see impending or actual natural decrease. Whole regions of the world with fertility a half child below replacement—Eastern and Southern
Europe, Russia, East Asia—are facing a rapid dwindling in numbers. Indeed, global projections foresee the world’s population as a whole peaking during this century (or, in the UN’s newly conservative medium scenario, shortly after). Early accounts by demographers of the fertility decline observed in Western countries in the late nineteenth century had already envisioned it leading to such a shrinkage. Later theorists of the phenomenon eventually agreed—after calling into play a second transition, a response to what they discerned as the emerging mores of post-industrial societies.

Of course there are other developed regions—much of Western and Northern Europe and the United States—where a so-far resilient, if not fully realized, two-child norm and tolerance of significant immigration point to an at least medium-term future of continued population increase. The US is exceptional in this company in the scale of the anticipated increase and in viewing such an outcome with equanimity. Forty years ago, in the ebbing years of the baby boom, a government-appointed commission could find no benefit to the US from further population increase. Yet America’s population now is 100 million above its level then, and the prospect of another 100 million being added by 2050, as the Census Bureau projects, arouses scant public interest, let alone protest—at least, qua numbers alone: views on migration are another matter.

In countries long accustomed to immigration-fueled population growth, commercial interests can be a potent lobby for its continuation, often against wide though inchoate public opposition. A modest rate of increase also presents age-transfer advantages, in Ponzi-scheme fashion, that may offset adverse effects of increased scale, although that cannot last indefinitely. (Migration sufficient to halt the rise in old-age dependency, as the UN’s “replacement migration” calculations have dramatized, requires vast and ever-growing inflows.) Contraction, however, will be the more common situation, and here any gains in amenity that may result from smaller numbers are balanced and likely outweighed by the age-distribution disadvantages of an aging labor force and the health and welfare burden of rising proportions of elderly. Over the long term, zero growth thus should be an attractive demographic objective: there is always a lot for any society to be worried about, but at least concerns about overall scale—too many people or too few—can then be set aside.

Having an objective does not bring with it the ability to achieve it, or even perhaps much willingness to expend effort to do so. In the case of migration, countries routinely declare a national interest in who and how many are admitted, and enact policies to pursue that interest. For admission through birth, however, allowing any collective deliberation on future population size and make-up is widely seen to be denying the individual rights of citizens. This position is so familiar that its sociological distinctiveness is not noted. It is formalized in a succession of international covenants, from the Universal
Declaration of Human Rights onward, which uphold sole parental sovereignty in determining family size—only slightly qualified by asking for parents’ decisions to be “responsible.”

Yet unless we subscribe to the supposed Thatcherite dictum that there is no such thing as society, “society” can be ascribed rights too. And these plausibly extend into the population sphere. The alternative, after all, is not some freer expression of individuality but for outcomes to be left by default to entrenched behavioral norms and sanctions from an earlier time. And indeed with a sufficient departure from a zero growth trajectory, societal rights are likely to find expression in some form. Thus, in dealing with high fertility they were a background presence—in fact, justifying much of the original concern with the matter—even if rarely acknowledged in policy and program rhetoric, and often, post-Cairo, explicitly disavowed. And in practice, family planning programs have typically acted as if a societal interest was their rationale. (That usually mild assertion of interest does not, of course, compare to the radical usurpation of control over birth numbers at the individual level claimed to be in the national interest by China’s government.)

Over two centuries ago, the Enlightenment thinker Francis Hutcheson, teacher of Adam Smith, took for granted that a society—“mankind as a system,” he called it—had a fully valid right to influence its members’ childbearing, provided the right was exercised with restraint. The society’s members had an obligation to help in “continuing the human race.” He added (Hutcheson, 1755: v.2, p.106):

This duty however, must in a great measure be left to the prudence of individuals: it would scarce be wise in any society to compel them to have offspring, whether they desired it or not; tho’ there may be strong reasons for making celibacy burthensome and dishonourable.

The last two epithets, it can be remarked, neatly encompass most of the instruments available for modern-day social policy to wield.

In early-twentieth-century America, a time when fertility had been falling for several decades and was not much above replacement level (then around 3), the sociologist Edward Alsworth Ross (1907: 615) asked, only half in jest, “Shall we live to see the mother of more than three regarded as a public benefactor and placed on the pay-roll of the state?” (Soviet hero-mothers had not yet been envisaged.) In a subsequent comment, the demographer Walter F. Willcox (1907: 632) preferred recourse to an institutional response to boost fertility: an “ethical and economic” adjustment in societal arrangements “whereby the present and future birth-rate of the entire community or of the classes of pre-eminent social worth may be controlled less exclusively by the interests of the individual or the family, and more by the general interests of society, or whereby society may gradually modify the interests of
the former class into closer agreement with its own.” (The passing intimation of a eugenic aim was a standard proviso for the time.)

A century later, thinking along those lines is acquiring a renewed urgency.

**Quality**

In Hutcheson’s view the rights of society in the population sphere extended well beyond an interest in its future size. Society, he said, also had the right to influence parental roles: specifically, to compel parents “to educate and habituate their children, that they may be able to support themselves” and, more generally, “to form [their children] into proper members of society” (p.109). John Stuart Mill (1859: 189) took a similar view:

> [T]o bring a child into existence without a fair prospect of being able, not only to provide food for its body, but instruction and training for its mind, is a moral crime, both against the unfortunate offspring and against society:…if the parent does not fulfil this obligation, the State ought to see it fulfilled, at the charge, as far as possible, of the parent.

Society’s interest thus set out is in the education and socialization of the next generation.

As an object of a present-day societal interest, and thus potentially of public policy, the quality of the next generation in these respects ranks at least as high as its number. The salience of quality is seen in the weight attached to it in many cross-country comparisons. In the contemporary US, for example, growing public concern with the country’s perceived slippage in competitive standing in the world is buttressed by reference to indicators such as school dropout rates, international rankings of educational systems, and cohort proportions in college, in all of which US performance is seen as lagging. Rich countries also compete in attracting talented immigrants, and here too a once major US advantage is considered to be eroding.

State interest in population quality, beyond its general expression in public health measures, first found generational specificity in child labor laws and compulsory education. Looking at the trends around him in the late nineteenth century, the British historian Charles Pearson (1893: 261) characterized these interventions as a reluctant state takeover of formerly parental responsibilities:

> The State has interposed in the last resort, and not without many misgivings, because the interests of its future men and women—their health and mental equipment—were bound to be more important to it than the maintenance of parental authority.
James S. Coleman took up this theme in his 1992 presidential address to the American Sociological Association. The family, one of the primordial social institutions, was visibly weakening—part of a general decline in social capital. By the time of adolescence, many children were in effect abandoned by their parents, psychologically if not socially. But schools—he was a renowned and expert critic of the US public educational system—were not an adequate replacement: they stuck to a narrow training function, only a small part of “all the things entailed in becoming a mature adult.” He continued: “There is, however, one actor with strong interests in maximizing a child’s value to society or minimizing its cost. This is the state” (J. S. Coleman 1993: 12, 13).

Coleman promoted educational reform but did not envisage the state attempting to do “all the things” required in childraising. Instead, he proposed an elaborate incentive scheme to prod parents to do better by “restoring to them, in effect, property rights over a portion of their children’s productivity” (p.13)—rights that the state had usurped. Paul Demeny (1987) earlier proposed a very similar scheme—compensating parents in their later years for successful childraising—as a pronatalist measure with a quality angle. There are obvious difficulties in further complicating an already intricate array of transfers that has accreted in the contemporary welfare state, but the potential attraction of a policy that meets the joint societal interest in the next generation’s size and quality should be evident. It is unlikely that the last word has been heard on the subject.

Low-skilled immigration is widely seen as another quality issue. Nineteenth-century industrialists in the US, free trade doctrine notwithstanding, sought protectionist import barriers against foreign competition. An emerging labor movement fought for similar barriers to defend domestic wage gains against immigrant labor—epitomized by the 1882 Chinese Exclusion Act. E. A. Ross again was a strong voice on the matter, expressing a common sentiment of his time (and ours) when he wrote that until “backward mankind” had advanced out of poverty and ignorance, “the sunny spots created by scientific industry coupled with prudent parentage will be menaced by an influx, peaceful or armed, from the crowded areas, and the bristling frontiers between peoples and races will have to remain” (1907: 617). The frontiers around the sunny spots continued, and continue, to bristle, keeping low-skilled migrant numbers, and thus the effects of direct immigrant competition on base wages, at relatively modest proportions.

The modern formalization of this exclusion of the ill-educated is found in points systems for immigrant selection used in Canada and Australia, spreading to the UK, and often advocated for the US. In such systems, education and skills, youthfulness, and language competence typically count toward admission, with scoring in each area summed to determine the decision. (Awarding residence to in-country foreign students receiving higher degrees is an analogous procedure.) In contrast, US migration policy has long given
precedence to close family members of existing legal residents, irrespective of age and education—easing assimilation, it can be argued, but contributing less than proportionately in terms of human capital. (Seasonal migrant farm workers, a sector where education may count for little, are a different category.)

**Mobility**

An essential component of modernity is the breaking down of ascribed hierarchies, affording greater equality of opportunity and thereby greater social mobility. The US has always prided itself on this characteristic, celebrating what F. W. Taussig (1921: v.2, p. 248), in the classic economics textbook of a century ago, called its ability “to lessen, possibly to destroy, the caste-like character of social classes” of the Old World. Equality of opportunity offers a justification for acceptance of inequality of outcome. Such inequality peaked before the Great Depression and has again reached high levels in the decades since the 1970s. While the disparities are most striking in income, they are seen also in social class variation in marriage rates and family stability, and even in accentuated class differences in health and life expectancy.

Cross-sectional inequality tends to impair intergenerational mobility. Wealthy families have an obvious interest in maintaining their advantages and passing these benefits to their children. Politically, they can push for curtailment of redistributionist measures like inheritance taxes and generally for smaller public-sector transfers. Socially, they can marry within their class and ensure their children a high-quality education. In the US, parental investment in maximizing the career prospects of offspring begins with the competition for places in elite private pre-schools. Should the technology exist, it would no doubt start even earlier.

Some trends in the broader society also hinder social mobility. There have been decades of wage stagnation in the lower and middle percentiles of the US income distribution, attributable to hollowing out of labor demand through labor-saving technological change and competition from low-wage economies in manufactures and, increasingly, in services. The same trends have weakened and shrunk labor unions that would once have offered resistance. Additional downward pressure on wages, it has been argued, has come from the continuing large-scale shift of women into paid employment, a major step toward gender equality but achieved partly at the cost of institutionalizing the two-income family as a middle-class norm—in effect making economically unviable for most families the formerly dominant single-breadwinner model (Warren and Tyagi 2003). Wage stagnation and worsened job security for parents, abetted by residential segregation by income, in turn lessen the chances of upward mobility for children. Reinforcing that outcome may be some redirecting of public expenditures from child
support to programs benefiting the elderly, reflecting the political influence of an aging electorate.

Pursuit of a societal interest in intergenerational mobility thus has to overcome some strong obstacles. The case falls under the general problem of failures of collective action. In that broader sphere, as Mancur Olson (1982) has argued, organizations representing narrow segments of the society grow in number and influence over time, eroding economic dynamism. Indeed, Olson found no limit to “the social cost that such an organization will find it expedient to impose on the society in the course of obtaining a larger share of the social output for itself” (p. 44). Mobility is a kind of public good, widely acknowledged and celebrated as such, but little match for the distributional coalitions that have emerged, clustered at the upper incomes.

Identity

My fourth demographic succession issue is identity. Almost by definition a society has a collective interest in preserving its image of itself over time. Tacit that interest may be, but if pressed on the matter the society’s members, or most of them, would likely see themselves as stakeholders in an intertemporal entity, looking to a future that is culturally and institutionally a recognizable extension of the present. Part of their well-being is tied to the thought that they are participants in such an ongoing enterprise. New members, through birth or migration, are expected to become broadly acculturated to the society’s values. Except at the nativist extremes, that expectation is not to deny change, even what can eventually amount to transformative change, but to moderate its pace.

The sociologist Ferdinand Tönnies, of Gemeinschaft und Gesellschaft fame, wrote on this topic. Beyond simple demographic renewal (maintaining numbers), he distinguished two kinds of transmission process making for societal continuity. One derives from the intangible connection to “place and region, air and climate” absorbed in the course of a person’s upbringing and day-to-day interactions. The other is the deliberate effort of a self-aware society to influence its future by conveying its vision of itself—its traditions (no doubt including “invented” traditions), its historical memory, and its aspirations—to the next generation (Tönnies 1905). Nation-states routinely do this, sometimes to absurdity, seeking to maintain their sense of what they are and to strengthen the commitment of their members in the face of demographic turnover. Less benignly, national symbols and selective interpretations of history can also be manipulated in Orwellian fashion (“who controls the past controls the future”) to instill a regime’s strategic vision.

Identities tend to be resilient, entrenched both culturally and institutionally. In an illuminating study contrasting the US and Canada, Seymour Martin Lipset (1990: 212) traced the differences to Canada’s having attracted
the more conservative British Protestants and Empire loyalists, the US the more radical and democratic emigrés from Catholic Ireland and Europe. The contrasts, as he saw them, were enduring:

America reflects the influence of its classically liberal, Whig, individualistic, antistatist, populist, ideological origins. Canada...can still be seen as Tory-mercantilist, group-oriented, statist, deferential to authority—‘a socialist monarchy’....

(Echoes of such a division recur in US domestic politics—in views on the size of government, for instance, a perennial topic of dispute: in the 2012 election, the European welfare state was held up by Republicans as a dire threatened future for the US under the Democratic Party’s “socialist” policy agenda.)

In the context of generational succession, national identity is mainly in question in political struggles over immigration and multiculturalism. Affluence mostly resolves those struggles, though they may persist in plural societies or in situations where cultural divisions are sustained by continuing high migration levels. “Settler societies” like the US, Canada, and Australia—those where immigration has long been institutionalized as an element of national identity—can accept substantial cultural diversity; countries unused to that influx have greater problems. The growing Muslim presence in many European countries is an evident case in point. Increasing in scale and combined, as it often is, with poverty and residential segregation, such migration has been resistant to the usual forces of assimilation and in turn is giving rise to a rootless, anomic second generation. The eventual demographic-cum-cultural outcome in prospect is what David Coleman (2006) has called a third demographic transition.

Even in the US, where assimilation of migrants (sometimes after an initial experience of xenophobia) has long been a source of pride, the rising proportions foreign-born over recent decades—back to levels last seen in the 1920s—have reignited conservative protest against incipient multiculturalism. Thus Arthur M. Schlesinger, Jr. regretted that “America increasingly sees itself as composed of groups more or less ineradicable in their ethnic character” (1991: 16); George F. Kennan (1993), in his memoir, wrote in a similar vein. And as a populist voice complementing those establishment figures, Patrick J. Buchanan (2002) decried high levels of Hispanic immigration, which he saw being encouraged on partisan grounds by Democrats and at least tacitly, on economic grounds, by elements on the Republican right (illegal entry aside).

Of course, the alternative to cultural diversity is not cultural stasis. Simple demographic renewal in a closed population also brings change, as successive cohorts are acculturated in distinctive surroundings—the basis of the generation gaps that so intrigued Margaret Mead (1970) and that supply endless fodder for the popular media. Today’s youth grow up in a world
increasingly drenched by the internet and social media, opening strong new intragenerational paths of cultural influence.

**Stuff and satiety**

How should outcomes of (and policies bearing on) generational succession be appraised? One plausible answer would be by gauging their *fairness*. In considering social organization at a given time, the now-familiar device to draw on is the Rawlsian veil of ignorance, requiring judgment on the matter by a notional member of the society without knowledge of where in the socioeconomic, sex, and age mix he or she is positioned. (The exercise denies a judging role for persons from outside the society—not to allow that exclusion would make for a wholly different judgment, a radically impartialist perspective advocated by the philosopher Peter Singer [2004].) When the problem is intertemporal, as it is here, an analogous Rawlsian appraisal would be more complicated: it would need to gauge distributional fairness across as well as within generations in the light of the succession processes discussed above. Moreover, fairness for the next generation would have to build in that generation’s regard for its own successor—albeit presumably well-discounted. (Rawls himself did not seek to extend his theory intertemporally, limiting his discussion to finding an optimal intergenerational savings rate.) At the same time, the individualist orientation of this approach would have to be balanced by a concern for *societal* continuity, in the manner of Hutcheson, Tönnies, or James Coleman.

Specifying such a thought experiment is enough to suggest a rough answer. Public policies might be urged to stabilize population size, enhance the next generation’s human capital, promote more equality of opportunity (and, thereby, intergenerational mobility), and balance tradition and reform. And distributional fairness as well as the public interest would call for some socializing of the costs of childraising. Such a degree of deference to societal well-being is not the outcome likely to be delivered by the political process. As noted earlier, individuals well-positioned in the society prefer not to subject their and their children’s futures to veiled choices, and many would ally themselves with others of similar views to pursue their own narrower interests. The rights of society receive rhetorical support but, in their demographic content at least, are too vaguely specified to generate commensurate backing amid the usual ruck of interest-group politicking.

But with today’s high productivity, the naïve observer might ask: doesn’t the economy produce so much that these distributional problems become second-order concerns? What should affluence have brought if not that? This is the question John Maynard Keynes posed in his essay “Economic possibilities for our grandchildren,” originally a paper read to a student society in Cambridge in 1928, on the eve of the Great Depression (Keynes 1931). Not
much noticed then, it has had a remarkable afterlife (see, for example, Pecchi and Piga 2008 and Skidelsky and Skidelsky 2012). Keynes sketched a future two generations ahead, by which time, he speculated, the economy would have yielded a satiety of stuff, allowing—in fact requiring—a radical drop in labor participation and yielding an abundance of leisure. The “economic problem,” as he called it, would have been solved.

Admittedly, many people enjoy their work, or have become habituated to it, and do not seek additional leisure. Perhaps their work focus has been so intense that non-work interests have atrophied. Many other people are in thrall to the consumer economy or to the relative status their consumption levels bring them, setting an effectively high opportunity cost of forgone income from a marginal addition to leisure. And in the present dire economic conditions there are large numbers, losers in the Great Recession, who are unable to find employment.

Others, however, if offered the choice, would take increased leisure. What would they do with it? Generalizing perhaps from his own circle—the agreeable life of London’s Bloomsbury district and his friends’ country retreat in Wiltshire—Keynes envisaged a nation of aesthetes, hobbyists, and idlers, attending the opera, holidaying in the alps, or pottering in their greenhouses. Leisured manual workers, not explicitly described, might find a parallel set of activities, though no doubt encouraged to devote time to improving the mind through lectures at mechanics institutes, an earnest Fabian project of the time. Also not mentioned, since this was then a routine expectation of an unpaid female population, was the time-intensive task of raising children. With economic constraints removed, that activity might become primarily a variety of consumption—for the society, a hazardous bet on future lifestyles—supplemented, it might be hoped, with more than a dash of altruism.

For Keynes the fact that the good life he sketched was built on a low-fertility population—Britain’s at the time—was incidental: substantial further growth was not anticipated, mounting fear of population decline still a few years away. A century earlier, in the throes of the industrial revolution, the depiction of the good life in James Mill’s Elements of Political Economy (1821: ch. 2) had required a stationary population. Mill frankly acknowledged a class-bound society, with the idle rich at one extreme, the laboring poor at the other, but his interest was in those in between—“the men of middling fortunes,...to whom society is generally indebted for its greatest improvements.” These were in today’s terms the knowledge workers, “legislators, judges, administrators, teachers, inventors in all the arts...” and his hopes (he was a protégé of Bentham) were to enlarge this group: “It is peculiarly desirable that a class of this description should form as large a proportion of each community as possible.” And for that to be accomplished, population growth must cease: “There is a certain density of population which is convenient, both for social intercourse, and for that combination of powers by
which the produce of labour is increased. When these advantages, however, are attained, there seems little reason to wish that population should proceed any further.” (A better-known elaboration of the case was to come from Mill in a famous chapter of his own Political Economy.)

Clearly Keynes’s future has not arrived. What has gone wrong? Human acquisitiveness and jostling for status, encouraged by the blandishments of the consumer economy, explain part of the predictive failure. Thus Richard Easterlin (1996: 153–4) envisions “a world in which ever growing abundance is matched by ever rising aspirations...in the constant race to achieve the good life of material plenty.” By any measure the material advances since Keynes wrote have been dramatic, but the aspirations remain. In recent decades, moreover, the race has gone to the swift, concentrating material plenty at the upper end of the income distribution. A swollen financial sector has contributed to that outcome, much of it through risk-shedding onto the public or in a process akin to rent-seeking—offering vast returns to a narrow echelon for activities that have little or no (and, sometimes, hugely negative) social value.

Admittedly, the resulting patterns of consumption have their own valuation problems. A large part of GDP in a modern economy seemingly exists in some extended reality, held aloft, as it were, merely by mutual belief or expectation and by conventions about how certain kinds of goods and services should be priced. The economist and statistician Robert Giffen (1885: 119) gave an illuminating account of this phenomenon, instancing the art market. High-priced products and services “are now exchanged in part between small groups in society, and such exchanges, counting very largely in the aggregate, go to swell the total; but to some extent the whole thing is merely nominal—it pleases those concerned to count them for so much, and that is all. There is no corresponding ‘production’ to be divided.” Yet there is a common numéraire. Earnings from derivatives trading, say, can purchase a work of Damien Hirst or Jeff Koons in an airy market for modern art; but they are also an influential force in the down-to-earth economy.

Another part of the answer to the failed prediction lies lower down the income ladder, in the decades of stagnation in median incomes—attributable, it is usually argued, to the combined effects of the computerization and “offshoring” (Blinder 2005) of manufacturing and service jobs. What these changes have wrought is the detachment of a significant fraction of the working-age population in the advanced countries from secure claims on the social product. Reestablishing and refashioning such claims is a large and difficult problem for economic and social policy. The reattachment problem, moreover, is not just for the ill-educated and low-skilled. Other countries, the “emerging” economies notably, are increasingly able to compete in the skills market as their educational performance ratchets up. Offshoring of R & D is already in evidence. The residual claim by the forerunners, the US in particular, of a higher education system that better fosters creativity and
imagination, even if true now, is likely not a permanent advantage (especially in a world of MOOCs). And at best that will deliver a small—and geographically mobile—entrepreneurial elite. One might reasonably ask: what about the rest of the potential workforce?

The elaborate transfer programs of the welfare state are the standard means of reattachment—in effect buying off the disadvantaged with some of the proceeds of globalization. In Europe such transfers have enabled countries to endure their routine high and current recessionary levels of unemployment—though with insistent doubts about sustainability. The US offers less: Alan Blinder’s bleak recommendation to deal with offshoring, for example, is “to repair and thicken the tattered safety net that supports workers who fall off the labor-market trapeze” (Blinder 2005: 24–5), but that has proven politically difficult even in a period of economic crisis, let alone as everyday practice. Work-sharing has not turned out to be a successful option—the intent of France’s short-lived experiment with a 35-hour week and a not wholly credible defense of lengthy vacations and early retirement.

More ingenious and imaginative have been the ideas of utopian thinkers, like Robert Owen in the early nineteenth century or—when writing jointly—Percival and Paul Goodman in the mid-twentieth, who would solve the distribution problem not by social insurance but by splitting the economy. The Goodmans scorned the unemployment insurance model: under it, economic freedom is “trapped in regulation and taxes. The union of government and economy becomes more and more total” (1947: 189). “The amazing indirectness of this procedure,” they remark, “is brilliantly exposed by the discovery of a new human ‘right’—as if the rights of man could be so easily amended. This is the ‘right to employment’” (ibid.). Their Owenite alternative is to separate out a broadly defined subsistence sector of the economy—producing “whatever is essential for life and security”—to which access by all is guaranteed. The rest of the economy—“providing wealth, power, luxury, emulation, convenience, interest, and variety”—would then be left to operate without any government interference or underwriting (p. 191). Practicable the scheme is not, and any potential bearing on fertility unexplored, but it is not hard to see its abstract attraction.

Altogether, satiety seems very far removed from today’s advanced countries—whether in the US, with its visibly decaying infrastructure, high child poverty, and precarious public finances (conditions much predating the current hard economic times) or in many EU states, with their own overburdened welfare systems and major unemployment. But these are reflections of political arrangements and institutional structures that are at least potentially open to change. To a distant observer it would seem on the face of it absurd that such conditions should even exist under present levels of overall wealth and productivity.
Demographic arcadia, deferred

The post-transition demographic future of the affluent countries is not determined wholly within those countries themselves. In the discussion thus far, the outside world has entered as a source of not-always-welcome immigrants and lower-priced offshored manufactures and services. In theory, either flow could be slowed or halted by policy choice: by more stringent migration criteria or by a return to greater national or regional autarky. Or even without intention: an era of globalization has ended suddenly before.

There is certainly no approaching world shortage of the underlying stock of human capital. Looking just at the population in the 20–40 age group in selected regions, the projected fall-off in China—its low fertility no longer much tied to the one-child dictate—is more than compensated in continued growth elsewhere (UN medium variant projections, millions):

<table>
<thead>
<tr>
<th>Region</th>
<th>2010</th>
<th>2025</th>
<th>2050</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe and North America</td>
<td>304</td>
<td>275</td>
<td>273</td>
</tr>
<tr>
<td>China</td>
<td>432</td>
<td>397</td>
<td>276</td>
</tr>
<tr>
<td>India and Bangladesh</td>
<td>447</td>
<td>529</td>
<td>528</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>250</td>
<td>371</td>
<td>615</td>
</tr>
</tbody>
</table>

From the standpoint of the West, the phenomenon was described in Demeny and McNicoll (2006) as demographic overhang.

The forecast trajectory for sub-Saharan Africa warrants separate mention. The projections for that region assume fertility will drop from its current level of 5 to about 3 by mid-century. A faster decline might of course be achieved—especially with continuation of the relatively strong economic performance of the last decade—but a population aged 20–40 exceeding half a billion by 2050 is nearly assured. (The UN medium projections show the region’s total population, now 0.9 billion, reaching 3.3 billion—one in three people in the world—by 2100.) Based on the present quality of its infrastructure and education, the region’s ability to play a significant part in the global manufacturing and service economy in the medium term is far from assured. Admittedly, though, it is hard to forecast this: as late as 1981, Hong Kong businessmen I recall meeting at the Canton trade fair that year viewed the workers they employed in their first hesitant expansion on the mainland as wholly undisciplined, feckless, and lazy—a few years before these same workers would become the backbone of the powerhouse export industries of Shenzhen. And as illustrated by India’s service industries—not just call centers but a growing range of sophisticated back-office functions—infrastructure needs such as reliable electricity can be largely self-supplied.
Rest-of-the-world effects bearing on the demographic future are not just felt in migration and exports. There are other links. Still in the economic sphere, for example, the major shift underway from pay-as-you-go to funded pension schemes generates potentially large capital outflows from demographically older to younger countries, albeit diminished by the risk premiums often required (Börsch-Supan 2004). The US, still relatively “young,” has been a major beneficiary of such investment.

There are political links. Differential timing and pace of fertility decline around the world lead to successive regional “youth bulges” (see Jackson and Howe 2008), now widely assumed to be a factor behind the Arab Spring. The US National Intelligence Council is an avid scenario-builder in political demography based on those age distribution contrasts (see NIC 2012). One of its speculative reports predicts a future “grand bargain” between Europe and the Maghreb countries involving “European acceptance of regulated North African immigration in exchange for North African efforts to curb migration from further south” (“Effects of climate change…” 2010). Of course, the political drama may be overstated. Yahya Sadowski (1998) has disputed the empirical basis of what he scornfully termed “global chaos theory,” the population-linked political strife foreseen by writers such as Robert D. Kaplan (The Coming Anarchy) and Samuel P. Huntington (The Clash of Civilizations). That the approaching economic–demographic marginalization of the West will have profound geopolitical consequences, however, is not in doubt (see Demeny 2012).

And there are environmental links. Climate models point to an increasing rate of atmospheric warming, with ramifying, mostly deleterious, predicted effects on agricultural zones, disease patterns, and flood risks to low-lying areas, set to extend over many decades. Mitigation through control of carbon emissions on an international scale has thus far proven beyond the world’s institutional capabilities. The chief sticking point has been how to allocate claims on the atmospheric sink for greenhouse gases—and thus the burden of curtailment. One popular proposal would set a uniform per capita emission quota for each country (unchanged by population growth after a specified year). This “contraction and convergence” strategy would make the convergence burden fall on high emitters—justly so, many would claim. On the other hand, China has pointed to its own past achievement in birth control as unjustly penalizing it in the per capita scoring.

Even with advances in mitigation, substantial climate change seems a certainty. Adaptive responses will be more feasible for rich than for poor countries, but even there they may be inadequate to the task. It is quite conceivable that the resulting changed ecological circumstances will turn out to be a dominant factor in appraising future human well-being. (Happily for the rich, the prospect of masses of climate-caused environmental refugees besetting their lands—a recurring image of ecological alarmists—is generally disavowed by migration experts.)
The affluent countries make up a modest demographic segment of an expanding global community still 2 and perhaps 3 billion people short of its peak—people, moreover, who are rapidly acquiring middle-class consumption habits. There is no autarkic economic sanctuary to be had, scant protection against global environmental change, and no promise of geopolitical calm—though as yet little erosion of border sovereignty. It is hardly surprising, then, that declinists abound. Their message, of course, is not entirely new. Nearly a century ago Oswald Spengler, the most orotund among them, wrote of human history as “the endless uniform wave-train of the generations” flowing in a “stream without banks.” Over the water’s surface “the great Cultures accomplish their majestic wave-cycles. They appear suddenly, swell in splendid lines, lessen again and vanish, and the face of the waters is once more a sleeping waste” (1926 [1918]).

Yet there are environmentalists who discern a “Great Restoration” of the world’s forests (Waggoner and Ausubel 2001); there are political sociologists who foresee a reconstitution of the social order (Fukuyama 1999); and there are students of happiness—often economists, strangely—who tell us that incomes beyond a certain level of comfort gain us little. Demographers may contribute their own part to a defense against declinism. As Paul Demeny (1986: 487) wrote in his notable presidential address to the Population Association of America, they should lift their ambition above the humdrum of their discipline to ask: “What kind of society would we like to be part of? What kind of arrangements should that society have concerning demographic matters?” What, in short, should be “the desirable demographic constitution of contemporary societies”?

References


INTERGENERATIONAL RELATIONS
Intergenerational Transfers, the Biological Life Cycle, and Human Society

RONALD LEE

Humans are highly social, living and sharing in families and in larger social groups, not autarkic individuals living in self-reliant isolation. Intergenerational transfers are an economic expression of the links among individuals of different ages and generations, who are often but not always kin. Changes in the population age distribution place stress on some of these links and relax others, affecting the economies of families, social groups, and countries. For autarkic individuals, population age distributions would be irrelevant.

Intergenerational transfers are donations of resources from one age group or generation to another, with no expectation of explicit repayment, and therefore transfers do not include economic exchanges. Examples are resources devoted by parents to childrearing, adult children supporting their elderly parents, and elderly people assisting their adult children or grandchildren.

Intergenerational transfers are central to many important topics in economic demography. For example, expected or planned net transfers to a child by parents are equivalent to the private costs of a child, thus connecting transfers to fertility theory. Transfers to children may also be used for human capital investments (health and education), connecting them to later productivity, well-being, and economic growth. Expectations of future transfers to be received by an individual in excess of transfers expected to be made constitute a form of wealth known as “transfer wealth.” Transfer wealth can substitute for physical wealth or assets in an individual’s portfolio, thereby connecting transfers to saving behavior, capital accumulation, and economic growth (Barro 1974; Feldstein 1974). Transfers can be private, or they can take place through the public sector, as in the case of public pensions or publicly provided health care or education. Where transfers are contingent on traits or behaviors other than age alone, they inevitably create incentives and may require monitoring by families or public agencies. Public pensions often create incentives for early retirement (Gruber and Wise 1998), and publicly provided health care can lead to overuse. Transfers can enable societies to achieve more
balanced allocations of consumption over the life cycle in some circumstances when markets cannot (Samuelson 1958).

Here I will weave together some of these themes and observations across the long sweep of human history. Necessarily much will be speculative, and there are many sweeping generalizations in what follows, but I will draw on data from anthropological studies and the National Transfer Accounts project where relevant. In hunter-gatherer societies the central problem was obtaining sufficient food to rear costly dependent children while sustaining the adults, all of whom contributed to this task. Resources flowed downward from adults of all ages to the young, including from the elderly. In modern industrial societies with low fertility, aging populations, and high consumption by elderly who do not work, the central problem is inverted: how can societies afford to support growing proportions of costly elderly? In some rich countries, resources have begun to flow upward rather than down.

The co-evolution of longevity, menopause, sociality, and transfers

The evolutionary strategy of humans is based on heavy investments by families and other adults, including the elderly, in the human capital of a small number of children through many years of dependency (Hrdy 2009). Figure 1 charts per capita consumption and production (labor income) of food calories by people of different ages in hunting-gathering groups, averaged across sexes. We see that children do not produce as much as they consume until around age 20. Lee, Kaplan, and Kramer (2002) estimated that it cost hunter-gatherers 10 years of adult consumption to raise one child to sexual and economic maturity at age 20, including the cost of children dying before maturity. We also see that older people continue to produce more than they consume throughout their whole lives, on average. Lee, Kaplan, and Kramer (2002) report that 80 percent of individuals above age 50 produce more than they consume.

The long dependency is needed for the development of the brain and its cognitive and emotional aspects, along with acquisition of necessary knowledge (Kaplan and Robson 2002). This evolutionary strategy posed two difficult problems. First, without some form of life insurance, the long dependence of a few children on food and care provided by a parent risked the children’s death following a parental death, with inefficient loss of reproductive fitness (Queller 1994). Cooperative breeding, with inputs from a broader array of relatives and others in the sharing group, provided such insurance because others could fill in following a premature parental death. Such sharing groups also spread the considerable risk of illness and disability. Second, total child dependency became systematically too high in the middle of parents' childbearing ages for a couple to handle on their own (Chayanov 1986; Lee and Kramer 2002). Cooperative breeding funneled resources from
extended family and non-kin to families with high dependency burdens (Gurven 2004). According to Hill and Hurtado (2009), “food provisioning is ubiquitous, generally biased in favour of helping families with large dependency loads and not limited to kin assistance.”

A number of articles by Lee (2003, 2008), Lee and Chu (2012), Chu and Lee (2006, 2012), and Chu, Chien, and Lee (2008, 2010), analyze the way that this human pattern of intergenerational transfers would have shaped the evolution of other characteristics, as summarized in Lee and Chu (2012). The importance of continuing parental investments in offspring appears to explain why, unlike most other species, humans experience quite long survival after reproduction ceases. It also lies behind most theories of the evolution of menopause in humans, which is an exceedingly rare trait among other species. The mounting accumulated value of past transfers to a child as it ages is one reason why mortality drops so sharply from birth until the mid-teen years. Both of these patterns run counter to the analysis in Hamilton’s seminal article (1966). The necessity of male involvement in transfers to offspring explains widespread monogamy and explains why there is relatively little difference in the size, ornamentation, and weaponry of human males and females—that is, humans have low sexual dimorphism.

Intergenerational transfers operate somewhat like a credit market in nature, whereby young offspring are able to borrow to invest heavily in their growth and development, repaying this debt through the transfers they
subsequently make—if they survive—to their own offspring. Because of this possibility of reallocating energy across the life cycle, the marginal impact of resources on reproductive fitness at different ages varies less than otherwise, so the evolved age pattern of time preference (patience versus impatience) over the life cycle also varies less than otherwise, and time preference is lower. Our evolved sociality, including our willingness to share food and to contribute it to others with heavy dependency burdens, even without prospect of repayment, paves the way for our current willingness—ambivalent as it may be—to create and sustain a social welfare state with substantial transfers through the public sector.

Our evolutionary past endowed us with traits that are still evident in modern society. Fertility is relatively low, children are dependent until around age 20, and children are typically raised cooperatively (often by two parents or with help from grandparents, aunts and uncles, or other relatives, or with assistance from the state). Our long lives include long post-reproductive survival, and older people frequently assist their children in various ways including material, financial, child care, and advice. While families and households are more economically independent than in the past, we still show a propensity toward communal assistance when needed, a propensity that in rich countries is mainly mediated by the state. We attempt to plan over long horizons, reflecting our relatively low rate of time preference. We are only slightly sexually dimorphic with similar cognitive capacities but somewhat different physical strength, reflecting a basically monogamous orientation (despite the prevalence of polygamy in some societies) and the substantial contributions by men toward provisioning children.

Agricultural societies

When population was sparse and land was abundant and of little value, swidden agriculture was practiced, and property rights in land were not well-defined (Boserup 1965). From limited evidence (Lee and Kramer 2002) it appears that the elderly continued to be net producers. As density increased, returns to labor declined and land became more valuable, leading to establishment of property rights in land (Boserup 1965). Typically the land was owned by the elderly, and yielded them income, whether it was worked by adult children or by outside labor. Land ownership gave older people leverage over their adult children who hoped to inherit. In the agricultural societies included in the National Transfer Account (NTA) collection (Lee and Mason 2011a), the elderly consume more, on average, than they earn through labor (see poor countries in Figure 1), while their adult children provide more of their material needs. This stage may seem to fit Caldwell’s (1976) demographic transition theory best. But adult children who work on the farm could also be seen as generating labor income for themselves and asset income for the
farm’s owner, their parents. In this case, the elderly are consuming their own labor income plus at least a portion of their asset income. A remaining portion may actually be transferred to their coresident children and grandchildren, rather than the reverse. In NTA data, in some poor countries such as the Philippines, Indonesia, or India, this seems to be the case: on average, the elderly continue to work but consume more than they earn in labor income; at the same time, they make significant net transfers to their adult children and grandchildren, transfers that must be funded by the elder’s asset income, for example from a home or a farm. In other countries, particularly in East Asia, the elderly do receive net transfers from their children. This is so in Taiwan, China, South Korea, and Thailand (Lee and Mason 2011a). These countries have somewhat higher incomes than the first, yet one suspects that this same pattern held when they were poorer. It is also possible that the East Asian practice of transferring ownership of the elders’ assets when they move into their children’s homes clouds the interpretation of the NTA data.

In agricultural societies, the net direction of transfers in NTA is invariably downward from older to younger when we consider the whole life cycle. The direction of flows can be found by calculating and comparing the average ages of consuming and producing based on age profiles like those in Figure 1, weighted by the actual population age distribution (Willis 1988; Lee 1994). In every poor country in NTA, as well as in some other cases (Lee and Mason 2011a, 2011b; Lee 2000), the average age of consuming is younger than the average of producing, indicating that the average unit of output travels down the age distribution before being consumed. This is true despite the fact that children start work early and work long hours as teens, while the elderly consume more than they produce. In these societies, at least before the demographic transition, there are vastly more children than elderly.

Another approach is to calculate the implicit rate of return that a parent earned by rearing a child. Comparing the net consumption of the child to the survival-weighted return earned in old age, Lee (2000) found a large negative value, minus 7 percent. Put differently, on average a parent recovered in old age about one tenth of the net cost of raising a child. And some of that recovered cost could be viewed as asset income earned from a farm or other asset owned by the older parent.

**The fertility transition**

At some point technological progress, capital accumulation, and urbanization raised the rate of return to education. This led altruistic parents to bear fewer children to allow higher investments in the health and education of each, raising their future productivity and improving their health and longevity. A similar calculus may have led some parents who were concerned about their support in old age to invest greater expenditure in fewer children. Support
for both versions can be found in Knodel, Chamratrithirong, and Debaalya (1987). This quantity–quality tradeoff is one view of the driving force behind the demographic transition (Willis 1994; Becker 1981; Galor 2011). But in the most fully developed theory (Willis 1994), the relation of intergenerational transfers and institutions to childbearing motivations is more complex. Parents care about the future welfare of their children, but parents’ altruism toward their children is balanced by concerns for their own consumption and survival. If there are few jobs in which good returns to human capital can be realized, then parents may opt for high fertility and low education. In this case, they may receive little familial support in old age, since some old-age support may be in repayment of earlier parental spending on education beyond the level motivated by altruism. Even if there are high returns to human capital, the limited altruism and limited income of parents may lead them to invest relatively little in the education of each child. Since transfers to children are also the effective “price” of a child, a low level of investment in education per child also makes children cheap to parents, and high fertility prevails, while familial support of elderly parents will be limited. If, however, culture or institutions offer a credible guarantee that children will repay parental “loans” for additional education by supporting their parents in old age, then parents might pay for more education than would be motivated by altruism alone. Then fertility might be lower, but familial old-age support would be greater as earlier “loans” were repaid. This is one interpretation of the East Asian pattern of very low fertility, very high private expenditures on children, and relatively high net familial support of the elderly.5

Using NTA data, we can measure for each country the total societal investment in the human capital of a child by summing the per capita public and the private expenditures on education and health at each age from 0 to 24 for education and 0 to 17 for health (Lee and Mason 2010). For comparison across countries it is convenient to divide this amount by each country’s average labor income at ages 30–49, so that the investment per child is expressed in terms of years of adult labor. This can be compared to the total fertility rate for the country over the five preceding years. Figure 2 shows the result, which is a moderately strong negative association as the quantity–quality theory would predict. A similar association is found for eight Asian countries whether we look at total human capital investment or just private investment (Lee and Mason 2012). We also find a similar pattern over time within a country for Taiwan, Japan, and the US (Lee and Mason 2011a). Taiwan, South Korea, China, and Thailand have the lowest fertility of the developing countries and also have very high investment in human capital per child. These four countries are also the only ones that have net familial support for the elderly. However, one should not make too much of this point, because Thailand and China are close, in Figure 2, to the Latin American countries that do not have familial support of the elderly.
Becker and Murphy (1988) built a different theory on a similar foundation, suggesting that inefficiencies of familial investment in human capital, as a result of the absence of institutions guaranteeing the repayment of parental loans, trapped societies in equilibriums at low levels of education and income. One way to escape, they suggested, was by introducing public education funded by taxing the parental generations. These were repaid in old age through the introduction of public pensions that taxed their higher-earning adult children to pay for parents’ old-age support, and later for their health care and long-term care as well. While the rearing of children (other than human capital investment) remained largely a private familial matter, support for the elderly, who in earlier times would have helped to invest in children, has been taken over by the state in most rich countries and many poor ones. Meanwhile, low fertility and longer life lead to population aging.

Whether or not we credit the Becker–Murphy theory sketched above, there are important regional variations in the basic patterns observed in NTA
data. Figure 3 shows sources of support for the per capita consumption of the elderly (65+) in NTA countries, net of their labor income. The position of the country markers on the triangle indicates the mix of asset income, public transfers, and private transfers that is used to pay for net consumption by the elderly, as explained in the note to the figure. European countries are found along the right edge of the triangle toward the bottom, indicating that they rely two thirds or more on public transfers, with the remainder provided by asset income. The US is farther up on the line, relying two thirds on asset income and one third on public transfers. Latin American countries are also scattered along this line, but all except Mexico rely two thirds or more on public transfers. Mexico is much like the US, except that the elderly in Mexico make larger net transfers to their children. Only Asian countries are well inside the triangle or on the left edge. While Japan looks similar to the European or Latin American countries, China, Taiwan, and South Korea are all fairly close to the center of the triangle, indicating that they rely on all

**FIGURE 3  Support systems for the elderly 65+ according to NTA data**

NOTE: Countries located at a vertex derive 100 percent of support for old-age consumption from the labeled source at that vertex. Countries along an edge derive support from a mixture of the two sources at its ends. Countries within the triangle derive support from a mixture of the three sources, more from the closer vertexes. In countries outside the triangle, the elderly make net transfers to the opposite source. For example, in Mexico the elderly make net transfers to younger people.
three of the sources, and most notably, they rely on the family. Off to the left, Thailand relies on a mix of familial support and asset income. The Philippines relies almost entirely on asset income. The elderly in India fund their own consumption from asset income and have enough left over to make substantial transfers to their children.

We can consider together the patterns of transfers to children and the elderly. In East Asia, public transfer programs for children and the elderly have remained small, while the family plays a larger role than elsewhere. Private expenditures on children’s education complement public, with tutors, cram school, and other educational activities taking place after public education is done for the day. In Latin America a different pattern is observed. Public pensions are generous, sometimes leaving the elderly with resources to transfer to their adult children. Net familial transfers flow downward from the elderly (Brazil, Uruguay, Mexico) or are insignificant. Public education is funded less generously than are pensions, and the rich send their young children to private school, and then to public institutions for higher education. In Europe, the elderly rely heavily on public transfers, which fund virtually 100 percent of their consumption in Austria, Sweden, Slovenia, and Hungary, with Germany and Spain not far behind. But public funding for education is also generous, with very little private expenditure.

The source of consumption funding for the elderly may influence the level of consumption. In those Asian countries where the elderly get substantial familial transfers, the level of consumption of the elderly tends to be similar to or lower than that of younger adults. In the US, the move to more generous Social Security benefits and the inception of Medicare and Medicaid are associated with a great increase in the relative level of consumption by the elderly (NRC Committee, in press: Figure 3-10). The ratio of consumption by an 80-year-old to that by a 20-year-old doubled between 1960 and 2007, and while non-health private consumption began to decline at age 60 in 1960, it did not start to decline until age 80 in 2007. These changes in the age pattern of consumption have made population aging much more costly in the US.

Transfers, incentives, behavior

When the public sector assumes transfer functions that were originally familial, the incentives facing family members may change, as may their choices and behaviors. Publicly funded education reduces the cost of a child and could lead to higher fertility. Alternatively, the effect may be in the opposite direction: by reducing the price of quality, public education may lead to substitution of quality for quantity, causing fertility to fall. Either way, it alters the incentives and opportunities for the family and alters its behavior.

When the public sector makes pension transfers to the elderly, the value of children is reduced to the extent that parents were planning to rely on
them for old-age support. Many researchers have suggested that this could be a cause of low fertility, undercutting the sustainability of the public pension programs. Public pensions have also been structured in such a way that they create incentives, sometimes intentional, for workers to retire at younger ages (Gruber and Wise 1998). This, however, is not an essential feature of public or private pensions, which can be made neutral in this regard, as in Sweden’s Notional Defined Contribution system. For whatever reason, we find in our NTA data that transfer support in old age, either from the family or from public pensions, is associated with lower labor supply than if net consumption in old age is funded from asset income.

Public pensions and other public transfers to the elderly have the essential feature of creating a positive externality to childbearing (the seminal article by Demeny, 1969, is the first discussion of externalities to childbearing, and the first analysis of the economics of population control in a welfare theoretic context). Parents making fertility decisions do not take into account that each child makes the pension system slightly less expensive per worker to maintain. The same is true for publicly provided health care for the elderly and long-term care. The externalities to childbearing generated in this way can be quite large. Wolf et al. (2011) report a fiscal externality for a birth to a parent with above high school education of $247,000 in 1996 US$, which equals $400,000 in 2012 US$.

**Conclusion**

Evolution is partly about optimization of reproductive fitness through mutation and selection. When transfer behaviors evolve, as they did for humans, they create new advantages and disadvantages for patterns of fertility and mortality across ages, and thereby shape the direction in which the age structure of vital rates evolves (Lee 2003, 2008; Chu and Lee 2006; Lee and Chu 2012). In a similar way, patterns of transfers in modern life alter the incentives we face and influence our optimizing behavior, not always in socially beneficial ways.

The sociality of humans, including food sharing, intergenerational transfers, and assistance for families with high dependency ratios, evolved in the setting of small food-sharing groups containing 8 to 25 people, during the lean season when larger groups dispersed to forage (Binford 2001; Lee 2008). These people, who were often kin, would have known each other well and have known one another’s circumstances. Humans have brought this basic psychology into the modern world, though heavily shaped by cultures. The rise of the welfare state could be seen as an expression of these culturally modified evolved propensities, but with the intent of improving on deficiencies in family transfers as a vehicle for investment in human capital, old-age support, and need-based assistance, somewhat along the lines of Becker and
Murphy (1988). However, the anonymous nature of public-sector transfers has led to its own problems. In some cases, the public transfers have helped to create a need, rather than meeting a preexisting one, as in the case of public pensions and the emergence of a prolonged period of retirement and leisure at the end of life.

Humans have an altruistic impulse to assist the young, within limits. This impulse is reflected in the downward direction of net private transfers in all NTA societies, rich and poor (Lee and Mason 2011a: Ch. 4, Ch. 8). Public transfers are a different story, however. In some rich and aging societies, these now flow upward, from working ages to the retired elderly, as gauged by the difference between the average ages of paying taxes and of receiving benefits (Lee and Mason 2011a: Ch. 4). This is true of all European NTA countries, plus Japan, Uruguay, and Brazil, but not the US. In some of these countries the upward public transfers are large enough to overcome the downward private ones, so that the net direction of total transfers has shifted upward, as in Germany, Austria, Slovenia, and Japan. This is a sea change in human history.

We know that the current structures of public transfer programs in many countries are unsustainable in the face of population aging. Certain ages and generations have benefited from windfall gains as these programs expanded, and doubtless some will suffer painful losses as benefits are cut and taxes raised. It remains to be seen, however, whether social support of the elderly will be a dominant function of societies once they have adjusted to the radical population aging of the next few decades.

Notes

The author’s research for this essay was funded by NIA grant R37-AG025247. He is grateful to the country teams for the 23 countries whose data are used in this essay; to Andrew Mason, who co-directs the National Transfer Accounts project; and to Gretchen Donehower, who helped with the data analysis and charts.

1 This is the way the term “intergenerational transfers” is used in this essay. It is sometimes used in other ways, for example in Cox (1987). The definition I give also excludes reciprocal altruism, since that involves expectations of repayment.

2 Each of these examples could also be an exchange, depending on the implicit or explicit terms and expectations of those involved.

3 National Transfer Accounts, or NTA, is a large international project directed by the author and Andrew Mason, with participating research teams in 37 rich and developing countries around the world. Lee and Mason (2011a, freely downloadable through the ntaaccounts.org web site) describes the project and presents results for 23 countries.

4 Demeny (1968: 502) supplied the classic and much-repeated definition of the transition: “In traditional societies, fertility and mortality are high. In modern societies, fertility and mortality are low. In between, there is demographic transition.”

5 This portion of the intergenerational flow that is here viewed as a loan and subject to repayment is not a transfer. Only that part of childrearing costs that is not a loan and is not to be repaid is considered a transfer, according to the definition given earlier.
References


The Intergenerational Welfare State

NANCY FOLBRE
DOUGLAS WOLF

Demography is a realm of private decisions with public effects. In both high- and low-fertility regimes, families make decisions that significantly influence the well-being of others. Excessively rapid population growth can threaten economic and ecological viability. Population decline can threaten cultural continuity and fiscal sustainability. As Paul Demeny has warned us, there is no invisible hand in the family that serves to align private and public interests (Demeny 1986). Yet as he has also pointed out, the visible hands of the state do not necessarily provide a remedy. By socializing the benefits of childrearing more than the costs, the modern welfare state has encouraged fertility decline below replacement levels, lowering the prospects for future intergenerational transfers and perhaps undermining the welfare state itself (Demeny 1987).

When Demeny expressed concern about the effects of unwittingly antinatalist policies in the mid-1980s, most other demographers remained focused on the opposite problem of rapid population growth. Gradually, however, issues of below-replacement fertility began to garner more attention. Demeny’s proposal that we should “privatize” at least some of the economic benefits of childrearing has been taken up by a number of economists and policy experts, including Romesh Ponnuru (2012), Shirley Burggraf (1997), and Phillip Longman (2004). The serious economic slowdown that began in the United States and Europe in 2007 has fueled concerns about ballooning public debt, along with implementation of some strict austerity programs. These current events increase the likelihood that social spending on age-based transfers may be reduced.

In this context it seems appropriate to consider both what we know and what we need to know about intergenerational transfers and the welfare state, with a view toward improving equity, efficiency, and sustainability. In this essay, we focus on the US experience, considering three questions that we deem crucial: 1) Why did a welfare state that includes significant intergenerational transfers emerge? 2) How do public and private intergenerational
transfers interact? 3) What would happen if public intergenerational transfers were “privatized”?

While we cannot pretend to provide complete answers to these questions here, we offer a brief critical assessment of recent research to address them. We begin with a brief discussion of how intergenerational transfers should be defined and measured, emphasizing a number of methodological problems. We conclude with an emphasis on the need for more systematic interdisciplinary research and an outline of several priorities for reform.

Defining and measuring intergenerational transfers

The term “intergenerational transfers,” now widely deployed in both demography and economics, conveys a clear, common-sense meaning: resources transferred from one generation or age group to another. Generally, the literature refers to two countervailing sets of transfers: those made by adults, primarily parents, to children; and those made by grown children, or working-age adults, to the older generation. One factor that contributes to confusion and disagreement in discussions of intergenerational transfers is lack of clarity about the term “generation” and therefore the meaning of “intergenerational.” A “generation” is well-defined in the context of kinship and genealogical studies, where it relates to someone’s location in a lineage. In their textbook of demography, Preston, Heuveline, and Guillot’s very limited use of the term “generation” is confined to this meaning: they develop measures of population growth that “compare the size of successive generations” (2001, p. 113). Ronald Lee’s entry on “intergenerational transfers” in the Encyclopedia of Population acknowledges the difficulty of passing from a focus on the individual (or family) to the population, noting that “generation is a loosely-defined concept, and it can simply mean a different age group…” (2003, p. 542).

Populations contain family members, and those individuals can be associated with well-defined locations in their respective lines of descent, but populations themselves do not have “generations.” Populations can, however, be sorted into well-defined birth cohorts, that is, people born in a specified year or range of years. Laurence Kotlikoff, a leading developer of the idea of generational accounting, defines generational accounts as “the sum of the present value of the future net taxes … that members of a birth cohort can expect to pay over their remaining lifetime” (2003, p. 457).

Thus, addressing the issue of intergenerational transfers and the intergenerational welfare state simultaneously invokes the well-defined flows of private resources between generations within families and the less-well-defined flows between aggregate age groups in a population at a point in time. Moreover, whereas a person’s location in a lineage or in a birth cohort is fixed throughout his or her lifetime (and beyond), that person will, at dif-
different points in the life cycle, occupy different age groups and therefore play different roles—provider at some times, and recipient at others—with respect to societal “intergenerational transfers.”

The theoretical literature on intergenerational transfers often focuses on stylized models with only two age groups, or overlapping generations (Samuelson 1958). John Caldwell (1976) argues persuasively that the changing direction of intergenerational income flows helps account for fertility decline. Ronald Lee, in collaboration with several coauthors, has provided extensive theoretical and empirical analysis of intergenerational transfers (Mason et al. 2006; Lee and Edwards 2002).

Nonetheless, to define and measure transfers is more difficult than it might seem. If there are expectations of “payback,” a transfer to the younger generation may be conceptualized as an implicit loan. Under these conditions, the value of the loan must either be discounted with an appropriate rate of interest or treated as an investment that yields a future rate of return. Either calculation has huge implications for the relative value of transfers over time and helps explain why Gary Becker and Kevin Murphy (1988) challenge the notion that the elderly in the United States fare better than the young. A related issue concerns the value of insurance, which may be greater than the sum of net transfers. That is, individuals and families may be willing to pay a premium to reduce risk resulting from random shocks.

Empirical analysis of differences in consumption between the young and old is often based on the value of physical consumption (e.g., caloric intake), ignoring the value of services provided such as the value of caregiving time (Kaplan 1994). Most estimates of expenditures on children in the United States, including official estimates regularly released by the US Department of Agriculture, omit any consideration of the value of parental time (Folbre 2008). Estimates of government spending on children and the elderly typically focus on transfers that are conditional on age (such as the Earned Income Tax Credit or Social Security), excluding transfers that have the effect of benefiting families with children or the elderly without such conditionality (such as Unemployment Insurance) (ibid.).

Intra-family transfers coexist with, and complicate the impact of, extra-family transfers: an elderly person who receives Social Security may, as a result, provide a more generous bequest to his or her children. Moreover, researchers who focus on intra-family transfers often distinguish among three “currencies” or forms that those transfers take: money, time (in the provision of services), and (residential) space (Soldo and Hill 1993). The latter two “currencies” involve resources that have value, but may not—in the family context—be accompanied by explicit monetary flows. In contrast, societal transfers can only be accomplished through the collection and disbursement of monetary flows. This contrast between the means for effecting intergenerational transfers at the family level and the societal level complicates the
Much of the literature on intergenerational accounting in the US considers only relative tax rates, ignoring educational expenditures and other public benefits (Kotlikoff and Burns 2004). Estimates by Bommier et al. (2010) that take educational expenditures into account arrive at very different conclusions. While these estimates are far more comprehensive than most, they ignore the value of investments in public infrastructure, knowledge, and technology that will benefit future generations (not to mention the value of “ecological capital” that may undergo significant deterioration with climate change). Similarly, a recent effort to estimate the fiscal externalities of parenthood relies on strict assumptions concerning the net present discounted value of future taxpayers, which is difficult to predict accurately (Wolf et al. 2011).

Efforts to build a stronger methodological consensus around measurement of intergenerational transfers could pay off not only in empirical research, but also in theoretical discussions of “efficient” intergenerational transfers and in philosophical discussions of what constitutes “equitable” or fair transfers. They also bear directly on the questions addressed below.

**Why welfare states emerged**

Libertarians critical of the expanded role of government tend to emphasize the self-aggrandizing actions of political elites eager to expand public spending for its own sake, perhaps to curry favor with the electorate. At the other end of the political spectrum, some argue that welfare states are designed to protect human rights and promote egalitarian outcomes. Yet it seems likely that welfare provisions also offer some efficiency gains. Much of the early literature on this topic focused on the need to correct for market failures. Responses to family failures may also have come into play (Folbre 2008). The great significance of spending on education and pensions in modern welfare states suggests that intergenerational transfers played a central role. Like families, welfare states help spread the costs and risks of caring for dependents, including the young, the old, the sick, and the disabled. Unlike families, they offer the advantage of pooling risk for large numbers of people, which improves their insurance function. Unlike private insurance, they do not suffer from the risks of adverse selection.

Yet the differing motives for, and accounting periods relevant to, private intrafamily transfers and collective intergenerational transfers suggest a problem akin to “market failure.” In particular, attempts to calculate a cohort’s “generational accounting” balance lead naturally to a consideration of taxes paid and benefits received throughout a cohort’s lifetime, but not beyond it. Thus taxes paid during a cohort’s adult years (transfers to the younger generation) are somehow to be balanced by a “payback” in the form of benefits re-
ceived late in life (transfers from the younger generation). This approach rests on an assumption that a cohort does not care about those alive after the cohort becomes extinct. However, a family member may care about the welfare of his or her offspring, and of subsequent descendants, well after the family member’s death. Thus it is likely that a dollar spent on one’s own descendants is worth more than a dollar spent on the population age groups that include one’s own descendants. It is also likely that people without descendants wish to transfer fewer resources to younger generations than do those with descendants. For both reasons, aggregate societal transfers from the adult to the younger “generations” are likely to be too low. And a growing prevalence of childlessness in the population (Wolf 2004) could contribute to a weakening of support for key components of the intergenerational welfare state.

A specific rationale for the role of the state in making intergenerational transfers has been articulated by Becker and Murphy (1988). They argue that public investments in education reduce the impact of parental income constraints, improving the overall level of human capital. Such investments are repaid through pay-as-you-go pension systems, which give the older generation a strong claim on the earnings of the working-age population. Although Becker in particular has never been a strong advocate of public spending, he insists that this public contract leads to more efficient outcomes than purely private contracting could achieve.

Another explanation, consistent with Caldwell’s reasoning and research on bargaining power within the family, suggests that the development of capitalist economies with high levels of labor mobility tends to reduce adult children’s willingness to provide for their parents and to weaken paternal commitments to children (Folbre 1994; Caldwell 1976). In other words, private contracts become more difficult to enforce, and the increased risk of declines in private transfers leaves the elderly and mothers vulnerable to what might be termed a default on traditional family commitments. Those individuals and families lacking inheritable wealth might find themselves in a particularly weak position.

The development of markets in labor and capital also reduces the dependence of parents on children as a source of labor and care. Adults can hire young men and women, paying them an hourly wage rather than a share of family income. They can purchase annuities or other financial instruments that guarantee them income over the lifecycle, probably with more efficiency and less risk than relying on children. In most patriarchal systems, sons provide more direct and long-lasting support to parents than do daughters, who come under the authority of their husbands’ families. As a result, the weakening of patriarchal control is associated with a decline in preference for sons.

Yet the weakening of patriarchal control has some negative consequences for women. While wage employment offers them greater independence from parents and more bargaining power within marriage, it also renders
them vulnerable to paternal default—the increased likelihood that fathers will fail to provide for their children, either by abandoning mothers who conceive outside marriage or deserting or divorcing wives. In the US, many states developed Mothers’ Pensions programs alongside Old Age Pensions in the first two decades of the twentieth century (Folbre 1994). Economists Lena Edlund and Rohini Pande (2002) argue that the decline in marriage helps explain why women in the US have become politically more liberal and presumably more supportive of intergenerational transfers.

Another factor in the emergence of the intergenerational welfare state is the impact of military conflict, which typically requires efforts to enhance national solidarity. As Theda Skocpol (1995) shows, public support extended to Union Army veterans of the American Civil War and their families presaged the development of later welfare state programs. Similarly, the G.I. Bill implemented during and after World War II represented a major expansion of public support for US higher education (Mettler 2005). The demographic and economic devastation that World War II wrought in Europe helps explain many of the social programs subsequently instituted there both to support childrearing and to increase income security in old age.

These varying explanations are not inconsistent with one another. Nor do any of them foreclose the possibility that state policies have had unplanned and largely unintended effects that will require major adaptation and reform. Yet, more focused historical research informed by a better theoretical analysis of intergenerational transfers might yield valuable insights into their relative sequence and impact. It would also help answer the next question.

How do private and public transfers interact?

Private and public transfers to children and the elderly are obviously, to some extent, substitutes for each other. Some kind of substitution effect is implicit in the notion, outlined above, that public transfers emerged as a response to increased variability of private transfers, not as some exogenous shock imposed by the state on families. Indeed, there is every reason to believe that the causal linkages between these types of transfers are reciprocal: each can influence the other. Even if public transfers emerged because they were more reliable and efficient than private transfers, it is entirely possible that they had the unanticipated effect of reducing private transfers. But it is by no means clear that this has happened.

Standard neoclassical economic theory suggests that public transfers should crowd out private transfers in the family. That is, if a family has allocated resources according to a joint utility function, and an exogenous change (such as a school lunch program) increases the resources going to one family member who is attending school, and all else remains equal, the family should respond by reducing its own allocation of resources to that member (e.g.,
by not providing breakfast). Yet there is no empirical evidence that public spending on education reduces the amount of income that parents spend on schooling. Studies of parental expenditure usually assume that most spending on children represents a relatively constant share of parental income (Folbre 2008). In an even more extreme (and in our view entirely implausible) version of the family crowding-out argument, Robert Barro (1974) argued that parents respond to an increase in government spending by recognizing that their children will have to pay higher taxes when they grow up, and therefore they set aside more savings to compensate them.

A growing body of interdisciplinary research suggests that family resources are not so fungible. Families tend to set aside a certain amount of money or time for their members in ways influenced by social norms rather than by joint utility functions (Zelizer 1997). Families may be motivated by “warm glow” altruism, in which they derive pleasure from being the source of assistance, rather than mere concern for the well-being of the person receiving assistance (Andreoni 1990). Transfers of money may be less important than commitments to provide emotional support and direct assistance in times of need. Indeed, family members may be better able to meet those commitments if public transfers help them meet their basic living expenses.

Bargaining-power models suggest that increases in the resources available to family members improve their fall-back position, giving them more influence over family decisions (Bernheim, Shleifer, and Summers 1985). From this perspective, increases in the relative income of the elderly in the US provided through Social Security should have increased their bargaining power, which in turn should have increased their ability to obtain care and insurance services from their children. Here, the distinction between the different “currencies” of intergenerational transfers comes into play. While Social Security and Medicare help meet the income and acute health care needs of the elderly, they do not finance long-term care either in the home or in nursing home facilities. Medicaid provides funding for such needs only for those who have effectively spent down their assets. Public intergenerational transfers may reduce private intergenerational transfers of money, while they increase transfers of time, attention, and solicitude.

Social scientists seldom have the opportunity to observe the effect of a simple exogenous change, all else equal. Increases in public transfers to both children and the elderly in the US have taken place over a long historical period characterized, until recently, by increased demand for skilled labor and steady economic growth. We note that the expansion of public higher education in the 1950s and 1960s did not contribute to any decline in private college enrollments, or in parental contributions to college expenses. The opposite effect was, however, visible: states with large, privately funded universities proved less likely than others to expand their public university enrollments (Goldin and Katz 1998).
Critics of Social Security sometimes point to the decline in the propensity of the elderly to live with their adult children as a sign of crowding out, as though the working-age population has become less willing to provide housing for their parents (Ponnuru 2012). In the fifty years before the Social Security Act was passed, most elderly widows lived with their adult children. After 1935, the percentage living alone steadily increased (McGarry and Schoeni 1998). But this change largely reflects their increased income, which gave them the option of living independently.

In short, the older generation was not crowded out. It chose to move out. Karen Holden’s (1988) research suggests that poverty rates among older women living alone fell more slowly from 1950 to 1980 than they would have in the absence of a growing tendency to live alone: in other words, Social Security enabled older women to achieve residential independence at some cost to their apparent material well-being. Also, while working-age adults probably reduced the value of in-kind transfers, such as housing, to their parents, we have no direct evidence that adult children became less willing to help these widows when and if they experienced need (referred to above as the “insurance” function). Indeed, many researchers have refuted the “myth of abandonment” of older people by their family members (Shanas 1979; Brody 1985.) While some types of transfers (housing and other in-kind transfers) may have been reduced, others have taken their place, including informal insurance and care services.

Findings from Australia, Germany, the Netherlands, and the United Kingdom indicate that childless older adults are over-represented in residential care homes or institutions (Dykstra 2009). Evidence from the United States suggests that elderly women without surviving children are more likely than others to become dependent on Medicaid-financed nursing home care (Wolf 1999). Several US studies show that individuals with living children—potential future caregivers—are less likely to buy private long-term-care insurance than those without children (controlling for other economic factors) (Kumar et al. 1995; Cramer and Jensen 2006; Johnson et al. 2007).

Insurance itself—whether provided publicly or privately—represents a substitute for intra-family transfers. Thus, if Social Security, Medicare, and Medicaid have had the effect of crowding out transfers from the working-age to the retired population, it seems logical that private pension programs, private health insurance, and private long-term-care insurance have had similar effects. Such forms of private insurance have expanded enormously in the last three decades. We do not know of any empirical efforts to distinguish the impacts of public intergenerational transfers from these confounding factors.

Too often, the “crowding-out” hypothesis is framed in simplistic terms of public versus private transfers. In reality, individuals face many different options for obtaining health and income security in old age: reliance on public transfers, reliance on private family transfers, reliance on private savings or
pension programs, and reliance on private insurance. All these options are likely to exercise reciprocal effects on each other, and most individuals probably prefer a portfolio of these options to reliance on any single one. Still, it seems likely that the elderly are less dependent on family transfers than they have been in the past. It also seems likely that declining economic incentives to raise children have helped motivate fertility decline. But would modification or elimination of public transfers to the elderly reverse this trend?

**Privatizing intergenerational transfers**

The overall aging of the population increases the fiscal pressure of commitments to provide for the income security and health of the elderly. Possible general responses to the problem include increasing taxes, increasing savings, reducing benefits, increasing retirement age, improving the efficiency of health care provision, and possibly relying more heavily on selective immigration policies. In the long run, however, it seems desirable to bring the intergenerational transfer system into better balance either by providing more public support to parents and children or providing less to non-parents and the elderly. But how best to accomplish this goal?

Many policy suggestions are aimed at greater support for parents during their childrearing years, including higher tax subsidies, paid family leaves, high-quality early-childhood education, and reduced penalties for part-time work (Gornick and Meyers 2003; Gornick et al. 2009). A recent report from the OECD (2011) suggests that policies like these, implemented in several European countries, have had positive effects on both fertility and child outcomes. Related proposals in the US would reduce the economic effect of time parents take out of paid employment, either by using Social Security funds to help finance paid parental leave or by allowing “caregiving credits” that would increase later retirement payments (Center for American Progress 2012; Herd 2006; Jankowski 2011).

Unlike the policies above, efforts to link income security in old age to the specific number of children raised or to the “success” of childrearing efforts represent a form of privatization because—even if they are implemented through the state—they represent a kind of “pay for performance” logic resembling that in the labor market. Indeed, the main impetus behind such proposals lies in the direct incentive to parents to raise more—and more economically successful—children. Proposals run the gamut from minimalist approaches (just eliminate public pensions) to reallocation of existing Social Security or public pension receipts (as in Demeny’s earlier suggestion) and to more detailed proposals designed to offer incentives for marriage or children’s successful graduation from high school.

Consider first the minimalist approach implicit in Romesh Ponnuru’s (2012) suggestion to eliminate all public entitlements. If public pensions
and other forms of support for the elderly, such as Medicare, were simply eliminated, their adult children might assume more responsibility for these expenses. But how much more? As the earlier discussion of child default suggests, parents today have less economic and cultural leverage than they once had. Capitalist development itself has undermined the moral and material basis of parental authority.

It seems likely that some adults would be unable or unwilling to help needy parents. This enforcement problem could be addressed by legislation providing legal redress to needy parents.

Many “filial responsibility” laws remain on the books in the US, and international models include China and Singapore (Ikels 1993; Moskowitz 2002). Most of these laws were adopted with the goal of minimizing public expenditures, with some early advocates in the US promising to reduce total welfare costs by as much as 30 percent (Walters 1999). Actual experiences in the US during the 1950s, however, suggest that the administrative costs were extremely high, along with potential for litigation (Lee 1995). More recently, the US experience with enforcing intra-family transfers at the other end of the spectrum—through enforcement of the child support responsibilities of non-custodial parents—has not been encouraging. Despite new legislation and intensified efforts on the state level, the proportion of eligible children who receive child support payments has changed little since the late 1970s (Case, Lin, and McLanahan 2000; US Government Accountability Office 2011).

This pattern helps explain why most discussions of privatizing intergenerational transfers in the US rely on redirecting some portion of revenues raised through Social Security taxes to strengthen the link between childrearing and income security in old age. In his original proposal, Demeny suggested “allocating a specified fraction of social security contributions made by individual labor force participants directly to their parents, during parents’ lifetimes” (1987, p. 132). He did not elaborate on what share of social security payments would be privatized in this way, or to what extent the parental premium might represent an addition to payments based on their own earnings.

This basic approach was elaborated by Shirley Burggraf, who called for “parental dividends” of around 15 percent of the income of adult children and their spouses (1997, p. 70), about the same as total Social Security taxes (paid by both employer and employee), to be considered a substantial replacement for the existing system (p. 88). The demographic specifics of the proposal are not indicated. For instance, it is not clear whether working adults would pay the same percentage whether they have two surviving parents or not, or whether they are married or not. Further, it is not clear how the parental dividend would be allocated between unmarried or divorced parents who made disparate contributions to childrearing, or how parents would fare if their children predeceased them or were unemployed for long periods of time.
Any reasonable calculation of the expected value of the dividends parents might hope to receive would need to be based on more specific demographic and economic assumptions.

Similar concerns arise from consideration of Phillip Longman’s (2004) rather different approach. Longman proposes to reduce payroll taxes for married parents (one child would reduce them by one-third, two children by two-thirds, and three children would eliminate them altogether). Further, married parents whose children successfully complete high school would receive higher Social Security benefits than others. Unmarried parents would apparently receive no consideration under this scheme, and the fate of divorced parents is unclear. It seems obvious that the effective Social Security tax on non-parents would need to be much higher than it is currently in order to finance such a system. Longman’s emphasis on the desirability of incentives for meeting educational objectives evokes James Coleman’s (1988) earlier proposal (unrelated to Social Security) that parents be socially rewarded for any positive differential between their children’s expected academic performance (based on family and individual-specific factors) and their actual performance.

The mechanism for specifying such a reward might best be specified in a science fiction novel. It is not clear how the polity would decide how to measure parenting performance, and how to net out the many random and uncontrollable factors that enter into the production function for human capabilities. Economic theory suggests that rewards given for measurable performance will reduce effort devoted to immeasurable performance (Kreps 1997). Further, parents might not welcome efforts to remunerate them on the basis of assessments of their children’s quality. Indeed, many might argue that linking parental income to their children’s earnings reintroduces incentives that had many negative as well as positive effects, leading to lower regard for children’s own feelings, proclivities, and expressed priorities.

The limitations of privatized transfers to the elderly highlight some of the underappreciated advantages of a public intergenerational transfer system, which serves not just one but three powerful insurance functions. First, it insures parents against demographic misfortune: the possibility that their children may suffer disability or death. Second, it insures against economic misfortune: the possibility that children may not become reliably employed individuals able to contribute to their parent’s well-being in old age. Third, it insures parents against what might be termed parenting misfortune: the possibility that their children grow up not just unsuccessful but also unhappy and ungrateful to those who raised them, and therefore unwilling to help them. A system of old-age income support that relied exclusively on “individual accounts” would seriously reduce insurance against adverse outcomes resulting from both individual heterogeneity and bad luck. Financial crises such as the recent Great Recession cause even skilled, forward-looking investors...
to experience large losses, which society can potentially overcome thorough collective insurance.

Yet consideration of privatized transfers also highlights inequities in public intergenerational transfers and helps explain growing political resistance to them. As we mentioned at the outset, the aging of the population has created a fiscal imbalance that most countries find politically difficult to redress. Increased economic and racial/ethnic heterogeneity makes individuals less willing to pool risk. Finally, demographic diversity calls into question the existing public transfer system. Men are less involved in the private support of children than they once were; perhaps as a result, they now seem increasingly resistant to providing public support for them. With childlessness on the increase, in particular, among educated women, the costs of parenthood are also unequally distributed among women. How, then, could the intergenerational welfare state be reformed?

**Toward equitable and sustainable intergenerational transfers**

The United States has always had, at best, a partial welfare state compared to its counterparts in northwestern Europe. In general, it has provided more generous and consistent support for the elderly than for children. Educational expenditures per child vary enormously across school districts, often reflecting the local distribution of wealth rather than principles of equal opportunity. Parents in the US generally receive lower levels of public support than their counterparts in northwestern Europe, and this support takes a particularly complex, uneven form (Folbre 2008). As a country with a rich legacy of immigration and the capacity to draw on a global reserve of both skilled and unskilled workers, the US has perhaps been able to postpone considerations of intergenerational equity and sustainability. However, the potential for postponement is limited.

The conceptual problems outlined in this essay are wide-ranging ones. Our political discourse lacks a clear and consistent definition of intergenerational equity, making informed debate difficult. Our social scientific infrastructure offers no clear measures of public and private transfers, and our public accounting system makes it difficult, if not impossible, to track them. The important role of immigration in the US economy creates additional complications, because it weakens the traditional connection between private and public transfers based on residence or citizenship. A concerted interdisciplinary research program to provide a better understanding of intergenerational transfers in the US remains a high priority.

A second, related priority would be to determine how extensive intergenerational transfers should be within welfare state policies, relative to both family transfers and public intragenerational transfers. As we emphasized
above, movement toward increased private provision for old-age health and economic security through either voluntary or forced savings and investment reduces public incentives to invest in the younger generation and would likely also discourage fertility. This is not a reason to dismiss such measures out of hand. Issues of social equity and fairness could, in theory, be addressed by intragenerational transfers, and one might argue that steep declines in fertility throughout the world offer significant environmental benefits, if only social institutions could flexibly adjust.

The question cannot be reduced to either/or. Carving out a smaller but more specific role for intergenerational transfers within the welfare state might allow them to play a continuing role in a larger social portfolio of risk reduction and pooled investment strategies. For instance, the CLASS Act, originally passed as part of the 2010 Affordable Care Act, would have allowed individuals to contribute to an insurance pool for care in old age, and funds provided would have enabled them to hire family members to help provide home and community-based care (Wolf and Folbre 2012).

Presuming that intergenerational transfers will continue to play an important role, a third priority would be to bring the private costs and benefits of parenthood into better alignment. Although most parents undertake childrearing responsibilities for intrinsic reasons, the extrinsic costs have a profound impact on the economic well-being of mothers in particular. Private resources devoted to the production, nurturance, socialization, and education of the next generation yield important social benefits; at the same time, cultural permission to opt out of childrearing responsibilities, either by remaining childless or by becoming a non-custodial, non-contributing parent, is increasing. As aforementioned, increased public support for parenting could take the form of increased tax subsidies, care credits for Social Security, other policies that would reduce the net taxes paid by parents or increase the net benefits they receive, and work/family policies that make it easier to balance the responsibilities of paid and unpaid work. This approach would improve both fairness and efficiency without sacrificing the benefits of social insurance.

Finally, attention to the joint role of the family and the state in intergenerational transfers highlights the importance of designing public policies that reward and reinforce familial commitments, reducing the threat of “crowding out.” For instance, public support for caregiving for adults with disabilities and the frail elderly could be extended to payments for family members, as it has in many countries that have promoted home and community-based care, and as promised by the CLASS act alluded to above. The steep marginal implicit tax rate on low-income couples who marry, pool their income, and quickly lose eligibility for benefits could be reduced. Child support policies for low-income families could ensure that non-custodial parent contributions benefit their children, rather simply reimbursing state expenditures. The
home mortgage interest deduction, a policy that encourages family members to live separately, might be reduced or eliminated.

This list of priorities is by no means exhaustive, and it may not be realistic. Nonetheless, it reminds us of the crucial links between economic demography and public policy that Paul Demeny first brought to our attention several decades ago. In retrospect, intergenerational transfers played a far more important role in the emergence of the US welfare state than has been commonly recognized. Ironically, failure to accurately understand and govern these transfers could play an important role in the decline of the welfare state in the United States.

References


OECD. 2011. Doing Better for Families. «http://www.oecd.org/document/49/0,3746,en_2649_34819_47654961_1_1_1_1,00.html».


LOW FERTILITY,
POPULATION AGING,
AND THE BODY POLITIC
Why Demographic Suicide? The Puzzles of European Fertility

LANT PRITCHETT
MARTINA VIARENGO

A volume of essays on topics in demography honoring Paul Demeny would be incomplete without asking why large parts of the European civilization that produced Paul, a gentleman and a scholar in the finest sense of both words, have decided, through the aggregate decisions of women and men, to commit gradual “demographic suicide.” One would think that potential mothers and fathers, facing the promise of peace and relative prosperity, would choose to have children. But, puzzlingly, many people across Europe are choosing to have fewer children than will produce stable, much less growing, populations.

Demography is destiny. The consequences of shifts in fertility (even as moderated or even partially reversed in the future) will have substantial consequences on many aspects of life: aging, health care costs, marital and family relationships, labor markets, immigration, the fiscal sustainability of social insurance programs, and schooling. If anything like the current fertility trends persists, the outcome will have profound consequences for what it means to be a “nation” and what are acceptable social relationships within a nation-state. Current projections of the rate of natural increase imply either unsustainably large reductions in the working-age population or substantially higher levels of the influx of “non-nationals” (or a little of both).

The fact that women and men are making choices that result in lifetime childlessness must be symptomatic of a revolution in human affairs. Leaving a posterity has, for thousands of years of human history, been an integral part of the very definition of prosperity and happiness for the typical person. Those who remained childless—priests and nuns, monks and eunuchs—were recognized as choosing radically different paths to human meaning and happiness or were regarded as unfortunate.
Below-replacement fertility and disciplines

This “post-modern” demographic transition to below-replacement level fertility (BRF) represents a puzzle for a number of disciplines, including demography, evolutionary psychology, and economics—especially as this involves a considerable fraction of the population who never have any children.

BRF is a bother for demographers

We recently read a paper that began, “The official projections predict that in the year 3023 the population of Japan will be one.” The precision of this projection a thousand years into the future, while far-fetched, reveals a simple fact: by mathematical necessity, long-range population projections tend to infinity (if births consistently exceed deaths), to zero (if deaths consistently exceed births), to some constant level (if they exactly balance), or to cycles (if they vary in certain ways).

While demography is an excellent descriptive discipline, because demographic facts are often the earliest to be established and best documented, it cannot be a predictive science without being a behavioral science. As demographers started making long-run projections, they knew exactly what fertility had to be arithmetically in order for their models to produce stationary populations—the “replacement level”—but they had no positive behavioral theory of why humans would choose to have exactly that many babies. The fundamental problem with demography as a predictive science is that the people whose fertility behavior was being projected had no concern for the knife-edged properties the forecasters faced. Since demography had no behavioral model to explain why people had the number of children they did, they could not model how aggregate conditions of population would feed back into the individual factors that affected fertility. (The one theory which does have a feedback loop from population to fertility, that of Malthus, is irrelevant to explaining below-replacement fertility in conditions of prosperity.)

As long as fertility was above replacement and declining, official projections could pretend fertility rates would do what they needed to do to make projection easy, sensible, and non-controversial: converge to replacement. European fertility has now definitively shown that, at least so far, women feel no pressure to make their fertility decisions add up to reasonable long-run projections of aggregate populations. All of the early demographic projections “predicted” that fertility rates around the world would fall from above 2.1 to exactly 2.1 for no reason other than that fertility levels below that level produced declining populations; and if TFR declined and remained low, then population went to zero at some long time horizon, and that was embarrassing as a prediction (Vishnevsky 1991).
But this is really just a bother for demographers for two reasons. One, nearly everyone accepts that demography is a descriptive science, so its lack of predictive power is not a threat to a disciplinary paradigm. Two, to the extent most demographers became involved in predictions and engaged in purposive activity to change fertility, it was almost exclusively because of concerns that populations were too big and/or growing too fast (as exponential growth of human populations appears environmentally unsustainable). To the extent Europeans are helping to reduce global population stresses—for whatever reasons—good for them.

**BRF is a conundrum for evolutionary psychology**

Over the past 15 years evolutionary psychology, based on Darwin’s theory of evolution by natural selection, has become a major theoretical perspective (Confer et al. 2010). Evolutionary psychology can be regarded as an approach potentially applicable to all areas of psychological research (Gaulin and McBurney 2008). Fundamental to this approach is “understanding the human mind/brain mechanisms in evolutionary perspective” (Buss 1999, p. 3), the rationale being that our minds evolved to solve problems faced by our ancestors within an environment of evolutionary adaptation (Workman and Reader 2008). According to evolutionary psychology, evolution by natural selection has led to morphological as well as psychological adaptations. Central to evolution by natural selection is the theory of parental investment and sexual selection (Buss 1995). Accordingly, evolutionary predispositions affect the nature of the human mind and lead to sex-specific choices that are the result of sex-specific issues faced during mating and reproduction.

Evolutionary psychology provides a micro-level explanation for fertility patterns based on the quantity–quality tradeoff in the context of limited resources. This context seems to be appropriate to explain differentials in high-fertility, high-mortality regimes but not in low- and ultra-low-fertility regimes (Borgenhoff Mulder 1998; Lawson and Mace 2010). Specifically, evolutionary fertility models have failed to explain why fertility declines with wealth and social status (Hill and Reeve 2004).

The surest way to guarantee lack of reproductive success is to choose not to reproduce, and that is what European men and women are doing. Whatever pressures evolution has produced for shaping human psychology, it has produced substantial parts of a population without the characteristic that they choose to reproduce. One would think that a desire to reproduce would be one of the primary traits evolution would select for. But if people can overcome their evolutionary conditioning sufficiently to avoid reproducing, then it would seem they could also overcome their evolutionary conditioning to avoid doing whatever else was merely instrumental to reproducing. BRF in Europe seems a paradigm-shattering phenomenon for evolutionary
psychology. Even if the discipline can save itself intellectually from the failure to adequately explain the spread of BRF in Europe and elsewhere, it seems a decidedly poor model of why people are voluntarily leaving fewer—and in many cases no—genetic heirs.¹

**BRF raises puzzles for economics**

Economics claims to be a behavioral science and has never been burdened by attachment to specific psychological theories, evolutionary or otherwise. It seeks to explain changes in behavior as the result of changes in incomes (or endowments) or changes in prices/costs in their broadest sense. But does economics offer a persuasive explanation of BRF in Europe?

The demographic transition, from high levels of mortality and fertility to low levels of both, produced two puzzles, especially for economists. The first was why fertility fell on average with economic development. The second was why richer households have fewer children than poorer households. Both of these puzzles are easily, if circularly, explained simply by saying that children are an “inferior good”—those for which, unlike “normal goods,” demand declines with income (as it does for low-quality food staples like rice). However, even as heartless economists ourselves we are loath to admit that children might be an “inferior good.”

Becker (1960, 1981) and Willis (1973) proposed two explanations why rises in income over time led to lower levels of fertility and why higher-income families had fewer children. First, they hypothesized that “child services” have a quantity and a quality component so that the same quantity of child services could be produced either by a large number of low-quality children or by a low number of high-quality children. In this way, even if child services are a normal good that rises with income, “quantity of children” could decline with income if “quality” per child rose fast enough. This model was widely invoked as being consistent with fertility falling over time with rising incomes, since formal education per child, taken as a crude index of “child quality,” was demonstrably increasing over time as fertility fell.

Second, Becker noted that the consumption of goods requires time, hence the act of consumption is bundled with the use of time. Therefore as the opportunity cost of time rises, either over time or across households, all time-intensive goods become more expensive relative to less time-intensive goods. So if one could separate increases in income from increases in the opportunity cost of time, then one should discover that demand for children increases with income as a normal good but decreases with price; thus the overall demand from an income increase is moderated (or possibly overwhelmed) by the price increase.

Even if these two concepts—quantity of child services as a combination of quantity and quality, and opportunity cost of time—adequately explained
the fall in fertility from high/moderate to low, these concepts (especially the first) are less applicable when fertility falls to zero.

One of the solutions to the paradox of rising income and falling consumption of a normal good was to deny that consumption of the normal good “child services” was falling, but rather to assert that demand for “quality-adjusted child services” rises with income. However, this exposes the problem that children come in integer quantities and that anything times zero is zero. That is, if one has zero children, then no matter how one “adjusts” quality, this is still zero child services. This means that the transition for an individual woman from “one high-quality child” to “zero children” is a massive discontinuity rather than a gradual transition. In terms of “child services,” one quality-adjusted child is further from zero children than four non-quality-adjusted children.

This discontinuity is conceptually important to how we understand the transition to very low fertility. If we really think “demand for children” has been declining, then the shift from 2 to 1 to 0 is part of a continuous process. On the other hand, if we interpret the transition to low fertility as a rise in demand for child services with incomes, satisfied through a rise in child “quality,” then the transition from demand for exactly 1 high-quality child to the demand for zero children requires something to explain this massive discontinuous drop in demand for child services.

Two different causal mechanisms explaining a fall in observed fertility, “more quality” versus just “less demand (or quantity demanded) for children,” might have the same predictions for average fertility but very different predictions about childlessness. Figure 1 shows the simulation at a normal distribution across individuals of a proxy for their demand for children that takes on continuous values, which then are obviously realized in discrete numbers of children (that match “demand” only loosely because of surprise (twins), mistakes, gender preferences, infertility, and so on). Two different mechanisms are illustrated for a shift from medium to low to below-replacement fertility. One is a generally leftward shift of the index across women of demand/quantity demanded of children. This will cause childlessness (zero children) to grow as a fraction of women as there is only truncation at zero. In contrast, if the falling demand for numbers of children leads to a rise in the demand for child services and an accompanying increase in “quality” of children, then the distribution should “bunch up” on 1 child since this mechanism alone should not cause any increase in the fraction of women wanting zero children (simulated as truncation at one). In fact, the fraction of women with zero should decrease because demand for child services has increased.

This illustration implies that how we explain and interpret the drop to very low fertility hinges on how much childlessness there is—and how much of that childlessness involves ex post regret. At medium fertility the two distri-
FIGURE 1  Different causal mechanisms for decline in the observed number of children have very different implications for childlessness

a. Simulation of falling demand for child services (and number of children) with rising childlessness (truncation at zero children)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Truncation</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRF</td>
<td>1.5</td>
<td>1.3</td>
<td>0</td>
</tr>
<tr>
<td>Low fertility</td>
<td>2.5</td>
<td>1.3</td>
<td>0</td>
</tr>
<tr>
<td>Medium fertility</td>
<td>3.5</td>
<td>1.3</td>
<td>0</td>
</tr>
</tbody>
</table>

b. Simulation of rising demand for child services but falling number of children (truncation at one child)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Truncation</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRF</td>
<td>1.5</td>
<td>1.3</td>
<td>1</td>
</tr>
<tr>
<td>Low fertility</td>
<td>2.5</td>
<td>1.3</td>
<td>1</td>
</tr>
<tr>
<td>Medium fertility</td>
<td>3.5</td>
<td>1.3</td>
<td>1</td>
</tr>
</tbody>
</table>

Distributions of simulated completed parity in Figure 1a and 1b look nearly identical. But at BRF one shows over 20 percent childlessness while the “quality children” narrative shows 50 percent at exactly one child. Figure 2 shows the fraction of women living in a household with no children at various ages for six countries. Even though Mexico, Turkey, and Slovakia have relatively low
TFR, they represent the more traditional pattern of childbearing in women’s mid-to late 20s, becoming nearly universal in the 30s followed by increasing “empty nests” in their 40s.

Finland illustrates the opposite case. Only 30 percent of women in their late 20s live in households with children, and the fraction of women living in households with no children rarely falls much below 30 percent. The Netherlands has just as much late onset of fertility as Finland, but the fraction of households without children falls to 25 percent at its lowest. Other European countries tend to be somewhere in between. While these cohort cross-sections don’t tell the dynamic lifetime story, it is clear that some low-fertility countries experience a significant rise in the number of women who never live in a household with children: a five-fold difference between the fraction of women aged 35–39 in households without children in Turkey versus Finland (6.2 versus 30.3).

At this stage it is difficult to disentangle what part of the very steep declines in fertility in Europe (especially in Southern and Eastern Europe) was the result of a rapid shift in the desired timing of children to later ages for which cohort rates overstated the predicted lifetime declines; and what part resulted from a shift to permanently lower rates of fertility, including a rise in the number of women who intend to have no children over their lifetime (or who are willing to delay childbearing with the accompanying risk of involuntary infertility). There is no question that the percent of women with no children at age 30 has risen sharply over recent cohorts, but the evidence about completed fertility is more mixed.
As shown in Figure 3 in Eastern Europe and across OECD countries, interestingly, there is no correspondence between low fertility and childlessness. In Eastern and South Europe (e.g., Bulgaria, Czech Republic, Hungary, Portugal) low fertility is accompanied by low childlessness (less than 10 percent), while in Finland relatively high total completed fertility (1.9) is associated with very high rates of women with no children (20 percent). This lack of clear association suggests that there are very different causal mechanisms that produce BRF. In particular, it is difficult to explain a rise in childlessness as an increase in demand for child services, regardless of how these services are quality-adjusted. This must instead be indicative of low demand for both children and child services.

What are close substitutes for children?

There are, then, two economic explanations of BRF. One is demand for high-quality children that drives people to choose very few (but positive) numbers of children; the other is a shift in relative prices that makes children less attractive. Only the latter is consistent with BRF in conjunction with high levels of childlessness.

As with all complex capital goods, one can think of children as little hedonic bundles. For instance, a house is a bundle of different attributes that bring well-being in different ways: size (square feet), lot size, location as a determinant of commute times, location relative to neighborhood amenities...
(schools, shopping), location relative to a city with its positives (entertainment, restaurants) and negatives (crime, noise). The overall price of a specific house is determined by the market willingness to pay for its individual features, and the choice of a house reflects a person weighing his or her own valuation of the hedonic characteristics against the market valuation.

One way a fall in the quantity of children demanded can occur is if children become more expensive. Another way is if another good arises that is a close substitute for some important component of the “child services bundle” that reduces its overall value. In those countries with BRF combined with high childlessness, what are the components of the hedonic bundle that is a child, and what are the close substitutes for those hedonic components that have caused the demand for children to fall?

Old-age security: An unstable dynamic of individual and society

One reason to have a child is to have someone to take care of you when you are too old to work. This has a financial element and an emotional element. In some times and places, being left a widow without sons (or offspring) put a woman at risk of destitution, since other relatives had no obligations and the state was too weak to provide income security or care. Especially in Western and Northern Europe, the onset of government-provided income security and health insurance/care has led to a nearly complete disconnect between having children and well-being in old age. The differential generosity and dependability of government-provided care for the elderly has been frequently invoked to explain overall declines in fertility and might explain the differential in childlessness, even at low fertility rates, between Europe and the United States and among Southern/Eastern Europe and Northern/Western Europe.

This explanation for ever-declining fertility in Europe is unstable. In other words, what might be true for individuals—that they can benefit in old age from tax-financed transfers whether they had a child or not—cannot be true of society as a whole. If no one has a child, then eventually there are no taxpayers; and unless social programs are financed entirely out of savings, the programs must fail. On their current trajectories, European countries are headed for a demographic “fiscal cliff” (quite apart from the temporary fiscal problems they currently experience) in which (especially without migration) the ratio of workers to those dependent on transfers becomes far too small. And, with the inversion of the demographic pyramid that low fertility has created, the actual labor requirements of both health care and non-medical elder care become a significant part of the labor force.

This may mean that the greater the number of people who believe and act on the premise that the state will support them in their old age, the
more quickly that promise becomes untenable. This will also complete the
cycle from children being regarded as positive externalities, to children be-
ing regarded as negative externalities, back to children as positive externali-
ties—with the accompanying shifts from pronatal to antinatal (or neutral) to
pronatal policies and social norms.

Externality is a key concept in normative economics. The fundamen-
tal welfare theorems (a formal expression of Adam Smith's key insights)
articulate the conditions under which private individuals making self-
interested choices lead to a social outcome that is Pareto-optimal—defined
as an allocation from which no one person can be made better off without
making another person worse off. One of the key conditions under which a
decentralized market produces a Pareto-optimal allocation is that the costs
and benefits of decisions fall upon the person (or firm) making the decision.
If private decisions are truly private, then public policy can ignore people's
choices across goods. Similarly, public policy needn't meddle with what
people choose to produce and sell either as individuals or as firms as they
attempt to maximize their returns. However, if private decisions don't fully
reflect the costs and consequences of decisions, there is an externality. The
classic example is pollution.

The concept of an externality was recognized long before its formaliza-
tion in economics, and societies have always had two mechanisms to respond
to the misalignment of incentives from externalities: informal mechanisms
of social norms and formal mechanisms of reward and punishment. Simply
put, people who don't recognize the negative externalities of their behavior
are labeled and treated as boors. Conversely, those who engage in unreciproc-
cated kindness are publicly lauded for their positive externality-incorporating
behavior. Formally governments can engage in taxing (or regulation or out-
right prohibitions) of behaviors with negative externalities or offer subsidies
(or regulation or mandates) to attempt to align private and public incentives
when there are externalities.

How does this relate to having children? Nearly all of the costs and ben-
efits of having children are privately borne, and both costs (monetary, time,
and psychic) and benefits are massive. Any reasonable theory of childbearing
involves people weighing these anticipated private costs and benefits among
other considerations. But are there externalities having to children, and, if
so, are children a positive externality or a negative externality? More particu-
larly, under what conditions are children a positive or negative externality?
Answers to these questions have come in three phases.

There is no question that historically births were regarded as a positive
externality. Hence a wide variety of social norms celebrated and encouraged
childbearing, and parents who were unable to have children were regarded as
objects of pity. In most traditional societies high fertility was not only associ-
ated with private wealth and power (Demeny 2003; Sinding 1991), but also
generally celebrated as making the society (from kith and clan up to empire) better off. Governments also tended to regard population as a boon.

It is only very recently that children began to be perceived, even in “normal” times, as a negative externality. The fall in mortality accelerated population growth while fertility was temporarily maintained at high levels. Neo-Malthusian theorists (Coale and Hoover 1958; Myrdal 1968) suggested that this unprecedented population growth in the developing world was detrimental to socioeconomic development. The onset of the environmental movement led to a concern for “sustainability” and popularized jeremiads like *The Population Bomb* (Ehrlich 1968), *The Limits to Growth* (Meadows et al. 1974), and *The Twenty-ninth Day* (Brown 1978).

These views of children as negative externalities justified actions to create new social norms favoring smaller families, subsidize means of reducing fertility by making contraception widely available, offer financial incentives, and, in extreme cases like China, enact prohibitive regulations. Needless to say, programs premised on the notion that children were a negative externality created tension with existing social norms and perceived national interests. If in fact children are a negative externality, then a perfectly legitimate response is for societies to encourage people to “internalize” these effects in their childbearing decisions.

The decline of European fertility to well below replacement levels has again put the shoe on the other foot. Many observers now feel that children are a positive externality justifying pronatalist actions by European governments. Fears of “fertility implosion” and “baby deficits” that started in the 1990s, and increasing awareness of the consequences of these demographic changes, have led the European Commission to set a “return to demographic growth” as one of its “three essential priorities” (European Commission, Green Paper 2005). More recently, some European countries, in particular those with ultra low fertility, have adopted pronatalist policies. Measures to increase fertility include direct financial incentives such as “baby bonuses”—introduced in Spain in 2007 and in Russia for the second child in 2006. Other policies such as tax deductions or welfare provisions like childcare, child allowances, and other in-kind benefits have been introduced in varying degrees. In addition, an increasing number of European countries have adopted policies to support parents and their labor force participation, in the form of parental paid leave and gender equality and other family-friendly policies.²

**Children as a source of meaning and social validation**

Children provide men and women with a source of life meaning and a personal definition of their “role” in society that brings them both internalized
and external reinforcements of their behavior. In some societies the social construction of gender roles produced positive reinforcement for men who were good providers and for women who were good nurturers, and the roles of father and mother were constructed in these terms.

Wide-ranging changes in opportunities, in the structure of economic production toward more traditionally “feminine” work skills, and in social attitudes have decreased the cost to women of seeking meaning and life satisfaction from their economic lives. We deliberately frame these changes as much larger than the shift in “wages” (where women’s wages continue to lag behind men’s) and instead consider them in the broader sense of the intrinsic reward from work and career.

What we are suggesting is that typical economic formulations look at well-being as a choice between bundles of child services (with a quantity/quality dimension) and material goods. In this sense the tradeoff is between more or fewer children and more or fewer consumption goods, and between work hours and leisure hours where work hours are seen as providing not positive utility but a means to purchase goods (including consumption and investment goods for children, like toys and university education). In this calculus, increases in women’s wages have the following income and price effects: 1) they make the woman (and the couple if the woman is in a couple) wealthier, which should raise the demand for (quality-adjusted) children; 2) because children are time-intensive, rising wages reduce the demand for children even for a given level of earned income; 3) rising wages increase the relative price of gaining utility through children versus all other material goods, so would increase work hours. As we pointed out above, there is no unambiguous prediction here about the number of children, but the first two points argue against childlessness in favor of fewer children with more investment of time and resources per child.

We are pointing to a different type of effect of the shift in gender roles and expectations, namely that women are more likely to achieve life satisfaction from market work that is inherently rewarding. High-status occupations are just that: “high status.” Status in and from these occupations tends to be performance-dependent, and time invested is a key component of success in nearly every high-status field.

Children are only available in discrete, integer-numbered, units. So the choice of a “little bit” of child services is difficult. Data suggest that having one school-aged child (compared to none) adds about 10 percent to the time devoted to care activities for women aged 25–44 in Europe (and only 4 percent more care time for men) (OECD Social Policy Database 2011). The difference in this time allocation between gender-equitable countries and other countries is quite small. Adding a second child only adds half as much time. So a big choice is between “childless” and “children,” since a second child adds much less incremental time commitment.
Marriage, family, children

Kathryn Edin (2005) explores the question of why non-marital teenage childbearing in very poor neighborhoods in the United States is so high. Her answer is that the teens want someone to love and to be loved by someone; and in the midst of the negative social and economic circumstances they face, the only way they feel they can achieve that is by having children of their own. Another component of the demand for children is to create another human being with whom to have an intimate and loving relationship. The question is, if people are having fewer (or possibly no) children, what is substituting in their lives for this love and intimacy?

This question is even more complicated as one Western notion of marriage was of a romantic attachment of a man and a woman of which children would be the product. Marriage as a social institution has historically been connected with the social regulation of sex and especially of childbearing, as it created obligations of caring for children born to the couple. Sexual activity, childbearing, and marriage have become disconnected so that increasingly it is socially acceptable to have one without either of the other two. But this has happened in different ways in different parts of Europe.

We can classify European countries into three categories (low, medium, high) by the total first marriage rate and by the proportion of births that are extramarital. Significant changes in both of these indicators have occurred since 1960: the median first marriage rate has nearly halved by 2003, and extramarital births have risen from 6 percent in 1960 to 29 percent in 2003 (Council of Europe 2005; OECD 2010). So both marriage as an institution that is an expected part of the life cycle and the expectation of an association between formal and/or legal marriage and childbearing are no longer social norms. In cross-section, the countries with the highest TFRs tend to be those with the highest extramarital childbearing. All countries with a TFR over 1.7 (Sweden, UK, France, Norway, Iceland, Finland, Denmark) have shares of extramarital births above 40 percent. In these societies, marriage is passé but having children is not. At the opposite pole, a number of countries (most in Eastern Europe) have high rates of marriage and low shares of extramarital births but very low fertility.

Billari and Kohler (2004) argue that the generally positive relationship between marriage rates and fertility rates disappeared in Europe in the 1990s. Alternative models of household formation and parenting have emerged: the Nordic model, where childbearing is disconnected from marriage but persists; and the Southern model, where marriage remains the main characteristic of union formation but fertility has reached very low levels (Kohler, Bilari, and Ortega 2006; RAND 2009). A decline in marriage rates goes hand in hand with an increase in informal unions such as cohabitation or “living apart together” (Kohler, Bilari, and Ortega 2006).
Drawing on family and household surveys, Perelli-Harris et al. (2012) investigate women’s family and fertility histories in 11 European countries. They find that only in Norway and France are a significant share of women who were cohabiting at the start of a union still cohabiting three years after the child’s birth, as opposed to Italy and Eastern European countries, where the share still cohabiting declines sharply, suggesting that the institution of marriage has not been displaced.

It is not at all obvious to us what is going on with the “demand for intimacy.” Unlike the cases of old-age security (where the state was an obvious substitute) and life meaning (where professional or career accomplishment became more widely available for women), it is not clear that there has been a reduction in the cost of any close substitute for the human intimacy of family (marriage and child) relationships. In some parts of Southern and Eastern Europe (e.g., Spain, Greece, Italy, Portugal, Croatia) we find the persistence of marriage as a source of intimacy and near universal childbearing. In other countries (e.g., the Baltics) we find low marriage rates, high extramarital fertility, and ultra-low fertility.

But in at least some countries in Northern Europe there seems to be a declining incidence of marriage (and of stable conjugal unions more generally) and moderate rates of fertility but combined with a high level of childlessness. Finland, for instance, has a TFR well below replacement (1.76), a high rate of extramarital childbearing, and substantial childlessness (20 percent in the 1965 cohort).

What is substituting in the lives of women and men for the love and intimacy that came from parent–child relationships? It is certainly not a significant increase in marital love and intimacy without children substituting for less marital love and intimacy with children—particularly in countries where marriage and long-term cohabitation have declined.

Conclusion

Economics is about making choices given constraints. Gains in productivity ease constraints and hence create new possibilities. The usual consequence of increasing possibilities is that people consume more of every good thing—whether in quantity, quality, or both. Human civilizations have typically regarded children as a good thing. Therefore the fact that people in civilizations at the pinnacle of possibilities are choosing not just fewer children, but fewer children than required to sustain the civilization itself, is a major puzzle for the human sciences.

The standard economic explanation of the fertility transition resembles our explanation of higher food expenditures but falling consumption of staples. As income expands from initially very low levels, people consume more calories; then as incomes increase further, food expenditures continue
to rise but mostly by consuming a higher-quality bundle of food, which has a higher cost per calorie. At even higher levels of income, total calories consumed stop rising, and may even start falling, even as total expenditures on food continue to rise. So “food consumption” as measured in calories might fall even though food is a good thing, but “food expenditures” continue to rise because cost per calorie rises. In a similar manner, TFR might fall but demand for children—properly measured—is really rising with income.

One might not see a fall from low to very low fertility as a puzzle because it is a continuation of that trend. However, this is a very poor explanation of voluntary childlessness, since “quality adjustment” obviously stops working at zero children.

As we have seen, the pattern of low fertility across Europe appears to take several different forms when it comes to childlessness. In Southern and Eastern Europe there are very low levels of cumulative fertility but with very few women ultimately childless. In Northern and Western Europe similar (or even higher) fertility rates are associated with lifetime childless rates approaching 20 percent—so one in five women are making choices that lead to no children at all. This cannot be the result of increased demand for children.

Therefore we cannot avoid the question of why demand for children—even for quality-adjusted children—would fall in countries with high levels of prosperity. Other economic answers are that children became much more expensive or substitutes for children became much cheaper. It is hard for us to believe that the reduction to very low levels of fertility is primarily substitution from “child services” to “other material goods.” That is, while some of the choice is to have fewer children but bigger houses and faster cars, this explanation seems too shallow.

The other alternative is that close substitutes for individual elements of what children provide have become much cheaper. Certainly the steady expansion of the state as caregiver plays a role (though one that the induced BRF choices undermine by making state support unsustainable). The expansion of other domains in which women are given social status and intrinsic satisfaction (over and above the pure wage component) is another likely factor—one that fits especially well with low fertility and high childlessness. Finally, there remains a big open question of how children fit into an overall pattern of “family” in the post modern era.

Notes
Recognizing this, some attempts have been made to address this conundrum. The “relaxed fertility-selection” hypothesis (Kaplan et al. 2002; Keister 2003; Aarssen 2005) has been proposed, according to which in modern societies a more intensive investment is required to ensure the offspring success. On the other hand, the “transmission competition hypothesis” (Aarssen and Altman 2006) relies on a tradeoff parents face between “gene transmission” (offspring production) and “meme transmission” (legacy drive), implying that career development, accumulation of wealth, and focus on social status may explain below-replacement fertility in modern societies. (But this is obviously no longer based on evolutionary biology as “genes” are a biological fact in the way that “memes” are not.)

Analyses of the effect of pronatalist policies on fertility suggest that the impact at best is very limited (Neyer 2003; Gauthier 2005; McDonald 2006; Goldstein and Kluesner 2009). Such policies also were not successful when implemented in the interwar years (Demeny 2003). Many of these policies have been perceived as being too narrow (Coleman 2012).

References


Low Fertility in
Historical Perspective

Massimo Livi Bacci

In the last few years, several Asian populations have joined Europe in the low-fertility league. In 2010 total fertility rates in Japan, South Korea, Taiwan, Hong Kong, and Singapore were below 1.4, around the level found in Germany, Italy, Spain, Ukraine, and Poland. At the slightly higher level of 1.6 are China, Russia, and Thailand, as well as the total populations of East and North East Asia (1.5 billion) and of Europe (0.7 billion). The worried voices of demographers, sociologists, economists, and political scientists have given increased attention to what they perceive as the deleterious consequences of sustained very low fertility. The crucial question is whether low fertility is becoming a long-lasting feature not only of Western societies but also of all those populations that are slowly emerging from secular poverty: poverty of knowledge and poverty of resources. In this essay I offer some reflections on low fertility and on its incidence in the past.

Low fertility and low replacement, for whom?

For Johann Peter Süßmilch, one of the fathers of demography, a low-fertility population was one in which couples bore fewer than four children on average. He wrote in 1741: “We will see, therefore, that each couple has a certain number of children and that, on average, we may attribute four children to a married couple. If this average were lower in such a way that each family had only three children or if the disordered behavior of men lowered marital fertility at this level, multiplication of mankind would be impossible.”¹ But even apart from nuptiality, what the German pastor had in mind was “replacement” of the population from one generation to the next.

Ask a demographer to explain low fertility and, at a minimum, he or she will want to know whether the question relates to a specific cohort or to all cohorts within a population, and whether to a given year or to a specific time period. The easiest response is to say that low fertility occurs when fertility is “below replacement,” or NRR (net reproduction rate) < 1; but in this
case the designation of “low” is made conditional on the level of mortality. This definition is not satisfactory. Let us suppose that a population suffers a new and deadly disease (as in times of plagues in the fourteenth and fifteenth centuries or AIDS in our time) and that NRR falls below replacement in spite of unchanged high fertility: will we say that this population has low fertility? Clearly, not. A more general and acceptable solution would be to define as low the level of fertility that in combination with a “normal” pattern of mortality, as dictated by the prevailing constraints, results in NRR below 1. But even this definition is unsatisfactory: what is a normal mortality pattern? In pre-modern times, a life expectancy of 30 years was normal, but mortality crises also were an expected, recurrent, and therefore normal occurrence, lowering life expectancy over the long term to 25 or 20 years. Can we define the normal frequency and extent of mortality crises in past times?

Our hypothetical demographer answering the question “what is low fertility?” has as a point of reference: twenty-first century populations that have attained a high expectation of life and whose mortality throughout the reproductive years is close to zero. For these populations replacement depends only on fertility, and populations averaging less than two children per woman may qualify as having low fertility. Fertility closer to one child per woman may then be classified as lowest-low or ultra-low.

The concept of low fertility familiar to demographers may not coincide with low fertility as seen from other viewpoints. Let us consider three perspectives. The first is subjective, related to the individual’s, the couple’s, or the family’s perception. The second concerns the “group,” or any sub-population with particular goals or strategies of its own. The third perspective is that of the “state” as the entity representing national interests.

Since the mid-twentieth century, surveys have produced a growing body of evidence related to individual expectations, ideals, and preferences. In many developed countries over the last few decades, the number of children that women, men, or couples consider “ideal” tends to be higher than the actual number born. In other words, they perceive their fertility as low. This is illustrated by Figure 1, which shows the relation between actual and ideal fertility for 15 European countries. The figure also shows that the difference between ideal and actual fertility does not depend on the level of the TFR. For example, German cohorts born in the late 1960s, with a TFR of 1.4, declared an ideal of 0.7 children higher than actual fertility, the same as in Sweden, Finland, and Ireland, where the same cohorts had a TFR around 2. In similar contexts—in this case Western Europe at the beginning of the millennium—the concept of low fertility varies from country to country, and in some of these countries a fertility level around replacement may be perceived as low.

If instead of the individual we consider groups and communities of individuals bound by specific characteristics (social, cultural, religious), low fertility could be defined as a level below the strategic targets set by the group
Itself. A typical example in earlier centuries was found among the aristocracies, landed gentry, or élites, whose titles or landholdings were transmitted through a male heir. In these groups, fertility may be defined as low if it does not ensure that a significant proportion of families have at least one surviving son. In order to satisfy those expectations, women bearing baby girls will be induced to bear additional children until a male heir is born. Therefore the group may define as low a fertility level that, although above replacement, remains below the collective expectations of the group. Another example is migrants settling in virgin territories with abundant land, who need large families and many children in order to be productive, ensure transmission of property, and acquire additional land. On the other hand, migrants in urban or post-industrial settings might prefer very low fertility, in order to improve their chances of social and economic advancement.

The third perspective on low fertility is the nation’s. The expansionist, imperialistic policies of nations such as Italy, Germany, Russia, and Japan in the 1930s would implicitly define as “low” a fertility level above replacement but inadequate to meet the policies’ targets. More recently, if we accept the official responses offered by governments to the United Nations’ survey on population policies, we find, for instance, that in 1985–89 Bolivia and Equatorial Guinea had fertility rates of 5.0 and 5.9 respectively but both viewed their fertility as “too low.” On the other hand, China enacted its one-child policy more than 30 years ago and now describes its fertility level (well below replacement) as “satisfactory.” ²
Individual, group, or national definitions or perceptions of low fertility depend on context, history, and strategic targets. They may diverge considerably—in excess or deficit—from a demographer’s definition based solely on the concept of population replacement.

**Four examples of low-fertility populations**

Let us return to the demographer’s definition of low fertility. Figure 2 sketches the possible occurrence of low fertility in past and present times. We consider four typologies: extinct populations; populations in times of crisis or catastrophe; pre-transition populations from the nineteenth century to the present; and rich populations in the twentieth and twenty-first centuries.

**Extinct populations**

The historical literature has numerous examples of populations that became extinct. Very often, however, the causes and mechanisms of extinction are obscure: short of massive depopulation resulting from human or natural factors, it is difficult to disentangle the complex causes that have led to the extinction of a population. Was the cause emigration, the mixing with other groups and the subsequent loss of identity, catastrophic mortality, depressed fertility, or—more likely—a combination of these factors?

The history of the Americas (and of Oceania) after contact with Europeans offers many examples of groups that became extinct. Seldom have
the mechanisms been examined in detail, however. This was true not only because of the lack of solid information but also because the “epidemiological paradigm” posited that the demographic catastrophe was the (almost) exclusive consequence of new viral diseases, such as smallpox, imported from Europe. Yet, upon closer scrutiny, things often happened differently. In the case of Caribbean populations—particularly the Tainos of Hispaniola island—the population was undoubtedly on a course toward extinction before the first smallpox epidemic affected the island in 1518. The 1514 Repartimento (census) of Hispaniola showed some evidence that the entry of the Spanish conquerors into the indigenous reproductive pool, the continuous dislocation of the local population, the fragmentation of villages and clans, and the forced labor in the search for gold had reduced the reproductive potential of the population. While the colonization of the island probably led to increased mortality among the Tainos, compromised fertility was a powerful concurrent factor in their extinction.

In the Amazonian region, in the two centuries following the initial settlement of Spaniards in the upper valley and of the Portuguese in the estuary of the river, a large number of “nations” (or groups, tribes, or communities), each having distinctive identities and numbering in the hundreds or thousands of individuals, disappeared. Some retreated to the forest away from the major waterways, others lost their identities and merged with other groups, and still others became extinct. One of the reasons for extinction was the fragmentation brought about by the intrusion of the Iberians, and the numerical decline of many groups below their minimum viable size (a few hundred members). Below such a minimum threshold, reproduction could be impaired by random asymmetries in the age structure of the reproductive pool and the reduction of mating opportunities. Thus, it is likely that population decline and extinction might, in many instances, have been the consequence of diminished reproductive potential rather than excess mortality.

**Populations in times of crisis**

Times of mortality crises, so frequent in populations of the past and in the poorest ones today, were also inevitably times of low fertility. A “textbook model” is sketched in Figure 3. A common scenario was a shortage of food (because of conflicts, weather vagaries, or other exceptional occurrences) leading to rising prices, hunger, famine, and the outbreak and spread of infectious diseases. Conceptions usually decline when mortality increases, reach a minimum when mortality peaks, and rebound to a maximum one or two years after the crisis; births follow the same course with a nine-month lag. The reasons for the decline in conceptions and births are mainly involuntary, although some are the result of individual decisions. Among the
former are widowhood, a decline in sexual intercourse caused by stress, the
decline of fecundity because of starvation, and an increase in fetal losses.
The voluntary factors include postponement of marriages and deliberate
fertility control. The rebound in conceptions and births may be due, among
other factors, to the resumption of marriages and an increase in marital
fertility. Even in populations that do not control births, there is evidence
that cohorts born after a crisis had higher natural fertility than pre-crisis
cohorts. Marriages, postponed during a crisis or precluded by the death of
one partner, rebound once the crisis ends. Marriages that were postponed
are celebrated; widowers (and to a lesser extent widows) remarry; and new
marriages are made possible by the transmission of property of deceased
parents. As a consequence, low fertility during a crisis may be compensated
by higher-than-normal fertility once the crisis ends. Over a number of years,
the negative effects of the crisis on population wane and population growth
resumes.

In this textbook model, low fertility during a crisis appears to be a
physiological consequence of the mortality shock, followed by an aftershock
rebound, somewhat in the way that an infection may temporarily affect the
growth of a child but without permanent consequences on its health and

FIGURE 3  Paradigm of mortality crisis

SOURCE: Massimo Livi-Bacci, “Mortality crisis in a historical perspective,” in Giovanni Andrea Cornia and
ultimate body size. There have also been historical instances in which the severity of the crises, and their frequency, were so high that fertility was unable to rebound and reproduction remained well below replacement. Such was the case during the plague pandemic in Europe and the Middle East in the 6th and 7th centuries CE (the Justinian plague) and, seven centuries later, during the Black Death in Europe. Empirical evidence can be drawn from the case of England. J.C. Russell conducted an original analysis of the *inquisitiones post mortem*, or the nature of the succession to deceased males, under three possible cases: no direct heirs, sons or their heirs, and daughters as heirs.5 T.H. Hollingsworth, on the basis of Russell’s data, has calculated male net reproduction rates from the mid-thirteenth century to the end of the fifteenth century.6 These are shown in Figure 4. Reproduction falls below replacement (NRR = 1) in cohorts that were at mid-life in 1335–39 and stays almost continuously below replacement until the cohorts of 1425–29. For almost a century mortality was so high (with any given generation suffering two or three outbreaks of plague during its lifetime) that fertility was unable to offset the effects of mortality.

A few centuries later, another population suffered a period of prolonged crisis leading to a catastrophic population decline. The Guaraní populations of the 30 Jesuit Missions of Paraguay experienced a long period of growth (between 1640 and 1732) followed by several decades (1732–67) of smallpox epidemics, war, and dislocation of the population.7 In the period of expansion expectation of life was 26 years and the TFR was 8; in the time of crisis, these two indicators were 22 and 7.6. Thus, fertility, although unable to compensate for catastrophic mortality, remained very high, notwithstanding the adverse conditions.

**FIGURE 4  Male net reproduction rates, England, 1250–1500**

![Net reproduction rate graph](image-url)

Pre-transition populations in normal times

There are few known instances of sustained low fertility in pre-transition populations. In part this is probably due to the scarcity of long historical series that may be trusted. But even in those areas where Malthusian checks were found, fertility seemed in general sufficiently high to ensure long-term population growth. We again take the case of England, whose population dynamics have been reconstructed in detail by Wrigley and Schofield over a period of more than three centuries (1541–1875). Birth rates, even in difficult periods, were fairly high. The gross reproduction rate fell below the level of 2 (corresponding to a TFR of 4.1) in only six of the 67 five-year periods considered (in 1651–55 and 1661–85, with an average of 1.9); if we consider NRR, there were also six five-year periods below replacement (1566–70, 1661–75, 1681–86, and 1731–35; average 0.94). Even in the period 1661–75, punctuated by a reappearance of plague in 1665, the intrinsic rate of increase was only marginally below zero.

An extensive historical survey would certainly show places and periods of moderate or even low fertility in other regions of Europe, but most of the time this happened because of exceptional external and totally involuntary conditions that crippled the reproductive potential of the population. An exception here is Ireland, which in the aftermath of the catastrophic potato blight of 1845–46 faced an urgent need to reduce its population. This was achieved through massive emigration, very high proportions unmarried along with late age at marriage, and a birth rate (around 25 per thousand) that was low even by nineteenth-century European standards.

Pre-modern Japan underwent a long period of moderate fertility between the beginning of the eighteenth century and the start of the Meiji era in 1868. Between 1720 and 1870 the population remained almost stationary (growing slowly from 30 million to 35 million). The causes of this stagnation are the subject of considerable debate. There is definite evidence of intentional control of the “production” of children, not so much by delaying marriage but by the practices of abortion and infanticide, and of a “destructive” role played by the cities with regard to the rural population surplus (Edo, today Tokyo, was the largest city in the world at the beginning of the nineteenth century). Detailed studies of several Tokugawa-era villages supply ample documentation, as a complement to literary and legal reports, attesting to the widespread practice of abortion and infanticide in all social classes. Beyond those practices another interesting explanation for the slow population growth of the late Tokugawa era and the Meiji era that followed is the well-documented agricultural transformation that led to an ever-greater intensification of farming methods. This transformation improved the general conditions of rural life but was accompanied by notable increase in workloads for men and even more for women. This trend “must have had unfavorable effects on marital fertility,
as well as on infant and maternal mortality, and thereby must have counterbalanced some of the favorable demographic effects of long-term agrarian development. Whatever the explanation of demographic stagnation, Japanese society gradually discovered means to limit demographic growth as the expansion of cultivation encountered natural limits. In any case, reproduction over this long period remained at or slightly above replacement.

Populations in the twentieth century and beyond

Few social phenomena have been investigated more thoroughly than the demographic transition in the West during the twentieth century. That process has led, with some interruptions, to the present situation of widespread low fertility. Since the 1930s the theoretical foundations of the fertility transition have been laid with convincing arguments, confirmed by abundant empirical analysis. Less satisfactory is the interpretation of the long rebound of fertility (the baby boom) after the low levels reached in the 1930s, when fertility in most Western European populations fell substantially below replacement. That rebound, which began during the World War II, progressed at a time in which the diffusion of contraception increased the proportion of couples able to avoid unwanted pregnancies, a factor that in itself would favor the decline of fertility. Figure 5 summarizes the fertility performance of ten Western countries that experienced the baby boom. On average, the minimum fertility level occurred in the cohort born in 1906 (TFR 2.01) and the maximum

![Figure 5: Indicators of the baby boom, ten Western countries](image-url)

among those born in 1930 (TFR of 2.65), with an increase of 26 percent, and a duration of the baby boom of almost a quarter century (24 years). An average increase in TFR of 0.64—the experience of these ten countries during the baby boom—would return those European countries now facing the prospect of depopulation to a reasonable demographic equilibrium.

The experience of the baby boom tells us that fertility may not, therefore, stay low forever. On the other hand, the downward trend of fertility raises another question: is there a minimum threshold below which fertility cannot fall? There is no scientific answer to this question besides the obvious statement that the minimum is zero. But recent history tells us that the minimum period TFR in Europe has not gone below 1 (with the exception of the Länder of the former German Democratic Republic (GDR), where TFR fell to 0.8 in 1993–94). In many countries of Eastern and Southern Europe, minimum levels between 1.1 and 1.2 have occurred in the two decades around the turn of the century; in Asia, minimum levels between 0.9 and 1.1 have been recorded in Hong Kong, Taiwan, and Korea. Not even during the two world wars did fertility fall below 1.5 in the major European countries involved in the conflict. In East Germany in the early 1990s, a time of high unemployment, loss of security, and abrupt transition to the market system, women postponed or avoided childbearing, and period TFR fell below 1 as a result. It is implausible, however, that these unfavorable circumstances will affect the full life cycle of a cohort. So the very nature of period TFR justifies taking its lowest level as an empirical cohort minimum.

Concluding remarks

This succinct review of past events does not provide an answer to the crucial question we posed at the beginning. Is low fertility becoming a permanent feature in a growing number of countries in the twenty-first century? The reversal of low fertility in many Western populations during the postwar baby boom shows that reversals are possible. As for the recent upturn of fertility in Europe, a careful demographic analysis that assessed the role of tempo and parity distortions concluded that “there was little or no increase in the level (quantum) of fertility between the late 1990s and 2008, while most of the observed TFR rise can be attributed to a diminishing pace of the postponement of childbearing.” The current economic crisis is likely to delay further signs of recovery for at least the next several years.

Our historical survey, however, suggests a few reflections. First, the measure of fertility is relevant mainly for demographers. Individuals, couples, families, groups, and countries—in other words, different aggregates of persons—may conceptualize fertility in different and sometimes contrasting ways. What is “low” fertility for the individual may not coincide with the evaluation of the group, and what is low for the group may not be so for the
country. Thus, policies promoted by governments are often ineffective because they run counter to the reproductive strategies of couples or groups.

However, the most relevant consideration is that in the history of mankind the powerful driving force has been mortality. Fertility has always been an extremely robust and resilient characteristic of past populations. Only the destruction of the foundations of fertility (a lack of mating opportunities; separation of couples; loss of libido or decrease of fecundity because of infections, hunger or stress) ever resulted in a seriously diminished and insufficient reproductive capacity.

Notes


7 Massimo Livi-Bacci and Ernesto J. Maeder, “The missions of Paraguay: The demography of an experiment,” *Journal of Interdisciplinary History* XXXV(2), Autumn 2004. The population of the 30 missions increased from about 38,000 inhabitants in 1640 to 141,000 in 1732, declining to 89,000 in 1768, when the Jesuits were expelled from the Spanish colonies.


11 In France, for instance, the lowest TFR during the two world wars was 1.2 in 1916 and 1.8 in 1941; in the United Kingdom, 2.0 in 1918 and 1.7 in 1941.

12 This line of reasoning is symmetrical, and we could take the highest period TFR as an empirical maximum (in populations having completed their fertility transition). The highest TFR was reached, in Western Europe, in the early 1960s, with values close to 3. But because the “contraceptive revolution” had not been completed by that time (in most countries contraception was illegal except for medical purposes; family planning programs were unheard of; abortion was still a criminal offence) a non-negligible proportion of births was unplanned. Discounting these, 2.5 could be taken as a possible “maximum” benchmark.

On Demographic and Democratic Transitions

TIM DYSON

This essay considers the implications of the demographic transition for the emergence of modern democracy. Although the idea of representative democracy was present in Europe and North America in the late eighteenth century, here I am concerned with its realization—and this occurs much later. I argue that consideration of the demographic transition in Western countries sheds light on what is sometimes called the “democratic transition.” Indeed, I believe that few aspects of socioeconomic development were more closely associated with the appearance of democracy than was progress with respect to the demographic transition.

The relationship between the demographic and democratic transitions has been largely overlooked. One recalls Paul Demeny’s still highly apposite observation—made in a paper now almost half a century old—regarding the extraordinary degree of neglect, shown by historians and other social scientists, of the consequences of the demographic transition. Also germane are his remark that study of history can provide valuable insight into forces that are still at work, and his reflection on the non-autochthonous character of many of the development processes that are affecting developing countries (see Demeny 1968: 502–504).

After providing background on previous research, the essay 1) discusses why a population’s progress through the demographic transition might be expected to promote a movement to democratic politics; 2) considers issues of measurement in relation to both the demographic transition and democracy; 3) examines the emergence of democracy with particular reference to Europe; and 4) appraises the potential relevance of my argument to contemporary international conditions.

Background

Accounting for the emergence of democracy is a central concern of political science. Much research on the subject reflects arguments articulated several
decades ago by Seymour Martin Lipset. He argued that economic development, including various factors it subsumes—especially the attainment of higher incomes, but also industrialization, urbanization, and the spread of education—is the most important general force behind the transition to democratic politics (Lipset 1960).

Of course, many non-economic considerations, such as the nature of pre-existing political institutions, cultural orientation, and the degree of social homogeneity, can also affect the position of countries with respect to democracy. And there has been much debate about the ways in which economic development may promote democratization. Thus Lipset saw the growth of an educated middle class as being crucial, whereas Rueschemeyer and colleagues (1992) stressed the role played by an expanding urban working class. Some analysts have questioned whether economic growth assists democratization (e.g., Londregan and Poole 1996; Acemoglu et al. 2008). Indeed, research using similar data and methods sometimes comes to different conclusions on the issue (e.g., Przeworski et al. 2000; Epstein et al. 2006). Nevertheless, most quantitative studies conclude that economic development—as reflected, in particular, by higher incomes, but also by higher levels of non-agricultural employment and education—tends to promote democracy (e.g., Barro 1999; Feng and Zak 1999). In this respect they concur with a large body of more qualitative work that identifies economic growth and processes like urbanization and industrialization as being important (e.g., Moore 1967; Dahl 1989).

In contrast, little research has been done on the relationship between the demographic and democratic transitions. Most work relevant to the subject deals with contemporary conditions and is framed not in terms of the factors that promote democracy, but in terms of factors that threaten socio-political stability. Research in this vein argues that rapid growth in the number of young adults can raise difficulties related to employment opportunities, and therefore social stability (e.g., Weiner and Russell 2001; Weber 2012). While this work focuses on rapid growth in the number of young men, it also highlights the role of young age structures in weakening the basis for political order. Thus, Cincotta argues that the prospects for democracy tend to rise as the proportion of young adults in the working-age population falls. He suggests that this is because support for autocratic rule among a country’s elite—with the aim of avoiding instability—is likely to diminish as the proportion of young adults falls. In addition, he argues that improved economic conditions resulting from changes in age structure may also promote democracy (Cincotta 2008; 2008/09).

**Theoretical considerations**

The theoretical considerations informing the present essay are consistent with the view that the young age structures and rapid population growth rates
experienced by many countries in recent decades have been—and often still are—inimical to democracy. However, here I am mainly concerned with the emergence of democracy, rather than with its retardation. And I am also interested in considerations additional to those arising from young age structures and rapid population growth.

The case for why progress through the demographic transition might be expected to promote democratization is best summarized against the background of the transition’s common stylized depiction, in which mortality decline occurs first—producing population growth—and fertility decline follows, which leads to both a falling rate of population growth and population aging. The start of fertility decline marks the onset of what might be termed the second half of this transition. The following discussion addresses these demographic processes in turn. It does so from the standpoint that, before the demographic transition, all political systems were mainly autocratic. No pre-transitional society was democratic in the sense in which the word is used here.

It is difficult to conceive of the emergence and consolidation of democracy in circumstances where life expectancy is very low. If, as some maintain (e.g., Lipset 1960), release from poverty facilitates intelligent political engagement, then so, surely, does release from disease and sudden and early death. Mortality decline tends to encourage an increase in the level of savings and investment in society, an outcome that is likely to reflect an increasing concern with the future (Bloom and Canning 2001). Improved health may also mean that greater thought is given to matters of justice and social equity. And a healthier citizenry may be more capable of questioning autocratic rule. It is true that the influence of life expectancy on democracy is sometimes hard to demonstrate using contemporary data (Barro 1999). But this may be partly because mortality decline has been so pervasive. Moreover, its implications for democratization may be greater at lower life expectancies than at higher ones. While mortality decline does not guarantee democracy, some degree of decline has invariably preceded the establishment of democracy.

Population growth can test the management capacities of any state—for example, in administration, taxation, and the maintenance of social order (Goldstone 2003). The rise in the rate of natural increase resulting from mortality decline represents a long period of demographic de-stabilization. This raises challenges for existing political systems, all of which at this stage are essentially autocratic. When fertility decline begins and the rate of natural increase falls, however, these challenges are likely to be reduced. Although the closing of the period of de-stabilization does not ensure that democracy is established, the conditions for this to happen are likely to be improved if other processes are also working in this direction. For instance, the basis for establishing a consensus for rule by elected representatives
in proportion to the numbers of people is likely to be facilitated by slower population growth.

Fertility decline causes the fall in the rate of natural increase and leads to population aging. But it also reflects and contributes to a transformation in the lives of women and men. Illustrative calculations suggest that in Europe and the United States in the late eighteenth century, women might spend 70 percent of their adulthood bearing and rearing children. In these societies today the figure is about 15 percent (Lee 2003). Fertility decline has underlain the increasing movement of women out of the domestic domain into formal employment and their growing independence vis-à-vis men (Davis and van den Oever 1982). In this context, it is often suggested that increased female employment stoked the demand for, and realization of, the female franchise: in Western countries, voting rights were commonly extended to women after they had been given to men (Ramirez, Soysal, and Shanahan 1997). With few exceptions, women only began to gain equal voting rights in the 1920s and 1930s. Significantly, this was when conditions of low fertility were first achieved.

Initially in the demographic transition, falling child mortality usually means that populations become slightly younger. Moreover, lower child mortality often introduces a lag between the onset of fertility decline and the onset of population aging. Nevertheless, once underway the process of population aging involves a long-term reduction in the proportion of children and adolescents in society and a corresponding rise in the proportion of people in the main working ages. Perhaps because it is relatively slow, the possible contribution of this development to the widening distribution of political power in society has scarcely been considered. Yet the present idea linking aging with democratization is simple: adults are more likely to want a voice in political affairs, and, with greater proportional numbers, they have greater capacity to achieve this goal. It seems reasonable to suggest that the redistribution of political power away from ruling elites and toward the middle and working classes in Europe and the United States occurred partly because these classes became increasingly adult in their age composition.

Finally, although urbanization is often viewed as resulting from economic growth, there are reasons to regard it as mainly the outcome of demographic transition (Dyson 2011). While urbanization is frequently seen as supporting democratization (e.g., Moore 1967; Acemoglu and Robinson 2001), some scholars note that towns with growing numbers of poor people can be very receptive to autocratic rule (e.g., Huntington 1968). Research using contemporary data does not uniformly conclude that urbanization promotes democratization (e.g., Epstein et al. 2006). Here one should recall that the urban growth rates experienced by modern developing countries are higher than those experienced by developed countries in the past. Moreover, the fact that political activity is usually centered in urban areas does not necessarily mean that urbanization itself promotes democracy.
In summary, there are reasons to think that, in the long run and in underlying ways, most of the aforementioned demographic processes promote democratization. In particular, population growth may raise challenges for political regimes that initially are largely autocratic. Then, in the second half of the demographic transition, fertility decline—which helps to empower women—and consequent population aging both seem likely to fuel demand for a more equal distribution of political power. Moreover, the slowing of population growth should increase the chances for socio-political stability, which may also benefit democratization.

**Issues of measurement**

In what follows, I use a population’s median age (MA) to gauge its progress in the demographic transition. I considered alternative measures for this purpose (e.g., life expectancy), but none performed as well. The estimates of median age come from United Nations (1956, 2011) sources and are based on census data. While a population’s age structure may have an independent effect on its chance of being a democracy, I make no attempt to gauge this here. Instead, median age is regarded as a composite measure of overall demographic progress. Thus in international cross-section (and over time), median age is strongly and positively associated with life expectancy and urbanization. And, because of the direct causal connection from fertility decline and slowing natural increase to population aging, median age is negatively and even more strongly associated with total fertility and the rate of natural increase.³

While my chief concern here is with historical experience in Europe, the analysis also draws on data for the 151 countries (35 in Europe) that, according to the United Nations, contained more than one million people in 2005. The average median age of the ten youngest countries in that year was 16.3 years. The corresponding life expectancy and total fertility figures were 50.4 years and 6.7 births. The average annual rate of natural increase was 3.1 percent, and an average of 26 percent of people lived in urban areas. In contrast, the average median age of the ten oldest countries was 40.9 years. The life expectancy and total fertility figures were 78.2 years and 1.4 births. The rate of natural increase was –0.05 percent, and 72 percent lived in urban areas. These statistics summarize contemporary variation with respect to the second half of the demographic transition. A country where the median age is 16 is only on the brink of fertility decline, whereas one where it is 40 has reached the end of the transition.

It is relatively straightforward to gauge a country’s status with respect to the demographic transition. But assessing its position with respect to democracy is more problematic—not least, because democracy itself is a difficult concept (Inkeles 1991). Furthermore, I need a measure of democracy
that is available for the past as well as for more recent times. In this context, two datasets provide national measures of democracy from the nineteenth century onward—those of Vanhanen (2000, 2003, 2009) and the Polity IV project (e.g., Marshall and Jaggers 2009). Both datasets were used here, and they both support my general argument. However, the results presented here derive from Vanhanen’s index of democratization (ID). This dataset is preferred as being more objective.

Vanhanen’s index of democratization is the product of two variables. The first, C, measures the extent to which political competition occurs in society. This is reflected by the proportion of electoral votes that goes to parties other than the largest one. The second, P, measures the degree to which there is political participation. This is indicated mainly by voter turnout, expressed as a proportion of the population, although allowance is also made for the taking of referendums (Vanhanen 2003: 59–63). The upper limit of both proportions is restricted at 0.7. Therefore the value of ID (= C*P*100) ranges between zero and a theoretical—and unattained—maximum of 49. To illustrate using the 2010 United Kingdom election results: 29.69 million votes were cast (10.73 million for the largest party) in a country of 62.04 million people (BBC News 2010). Therefore, ID = 30.6 (i.e., 0.639*0.479*100).

ID is a minimalist measure. Among other reasons, it has been criticized because 1) it neglects the degree to which government offices are actually filled by elections and 2) it may be biased where party systems are highly fragmented (Munck and Verkuilen 2002; Gates et al. 2006). However, ID also has a number of strengths. Thus, in contrast to most other democracy measures, it is relatively free of subjective evaluation. It is also fairly straightforward and is transparent in terms of the data on which it is based. It is a continuous rather than an ordinal scale, and it reflects the extent to which people actually vote in elections. In particular, it values the votes of women equally to those of men (Paxton 2000). Thus, importantly, if the franchise is extended to women, the value of ID can potentially double (other things being equal). However, perhaps its greatest strength is that ID attempts to capture what many see as democracy’s two key dimensions: the competitiveness of elections and the extent to which people vote (Dahl 1989).

Nevertheless, a problem with ID is that participation, P, is calculated with reference to the total rather than the adult population. Therefore, for any given level of voter turnout, ID will be biased downward in a society with a young age structure. This problem is acknowledged to exaggerate differences between countries (Vanhanen 2000: 255), but it does not seem to have been addressed in previous research. Consider, for example, that in 2005 in the ten youngest of the 151 countries, the average proportion of the population aged 20 years and over—taken here to represent those of voting age—was just 42.0 percent, whereas in the ten oldest countries it was 79.5 percent. Fortunately, however, the problem can be remedied. For the 151 countries,
United Nations (2011) data reveal a stable and tight relationship between the proportion of a population that is aged 20 years and over, \( y \), and MA.\(^6\) It is: 
\[
  y = 0.4097 \ln(MA) - 0.7255 \quad (R^2 = 0.991).
\]
Therefore, given MA, the value of ID can be divided by an estimate of the proportion of the population aged 20 years and over to give an *age-adjusted* index of democratization (AID). For the UK in 2010, the United Nations puts MA at 39.8 years—which, using the equation, implies that 78.4 percent of the population was aged 20 and over. Therefore, AID = 39.0 (i.e., 30.6/0.784).

In what follows, the main focus is on the relationship between AID and MA. Therefore, it is important to stress that the adjustment of ID to produce AID makes the task of establishing a positive relationship with MA *more* difficult. This is because, using ID, in a young population (low MA) there will be relatively few potential voters, whereas in an old population (high MA) there will be many more.

Finally, some remarks are required regarding the other measures used in the historical analysis below. Clearly, the extent to which socioeconomic measures could be included depended on their availability, and it proved difficult to obtain consistent estimates of urbanization. All estimates of real per capita GDP were taken from Maddison (2009). For European countries, the estimates of the percentage of the economically active population employed in agriculture (forestry and fishing) were calculated from data in Mitchell (2007); and estimates of the average number of years of schooling attained by people aged 15 years and over were taken from Morrisson and Murtin (2009). For the United States, estimates of the percentage of the population living in urban areas come mainly from the US Census Bureau (2008); estimates of the percentage of the labor force working in agriculture were calculated from Mitchell (1983), updated using statistics in Carter et al. (2006). Lastly, the education measure used for the US is the high school graduation rate; a series for 1870–1995 is also provided by Carter et al. (2006).

**Historical experience**

According to Vanhanen’s dataset, there was virtually no democracy anywhere in the world during the first half of the nineteenth century. In Europe, this was a period of demographic de-stabilization—one overshadowed by the French Revolution. Autocratic elites clung to power, and what political progress occurred was extremely limited (Ziblatt 2006). Thus in 1850 only two countries had ID values much above zero: France (7.4)—briefly, following the uprising of 1848—and the United States (6.8). Accordingly, Table 1 presents estimates of AID and MA for Europe and the United States starting in 1850.

The AID values in Table 1 are presented along with the ID values from which they are derived. For Europe, the table also indicates the number of countries used to obtain the ID values on a population-weighted basis.
The 14 countries with ID values for 1850 contained 84 percent of Europe’s population. By 1995 all 35 European countries are represented. Producing the MA values for Europe involved rather more bricolage. The estimates for 1950–2005 are those of the United Nations. For the period 1850–1950 it was possible to construct a relatively consistent series of population-weighted MA values using estimates obtained for 15 countries that are mostly in northwestern and southern Europe. Comparison for 1950 of the value from this series with the UN estimate for the whole of Europe allowed the MA values for 1850–1940 to be adjusted to make them consistent with the UN estimates.

### TABLE 1

Trends in the age-adjusted index of democratization (AID) and the median age (MA) in Europe and the United States, 1850–2005

<table>
<thead>
<tr>
<th>Year</th>
<th>Europe</th>
<th></th>
<th>United States</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AID (ID, n)</td>
<td>MA</td>
<td>AID (ID)</td>
<td>MA</td>
</tr>
<tr>
<td>1850</td>
<td>2.3 (1.3, 14)</td>
<td>24.0</td>
<td>14.2 (6.8)</td>
<td>18.9</td>
</tr>
<tr>
<td>1860</td>
<td>1.0 (0.6, 14)</td>
<td>23.7</td>
<td>18.2 (8.9)</td>
<td>19.4</td>
</tr>
<tr>
<td>1870</td>
<td>3.7 (2.1, 16)</td>
<td>23.2</td>
<td>13.8 (7.0)</td>
<td>20.2</td>
</tr>
<tr>
<td>1880</td>
<td>5.2 (2.9, 18)</td>
<td>22.9</td>
<td>18.3 (9.5)</td>
<td>20.9</td>
</tr>
<tr>
<td>1890</td>
<td>5.8 (3.2, 18)</td>
<td>22.7</td>
<td>18.1 (9.8)</td>
<td>22.0</td>
</tr>
<tr>
<td>1900</td>
<td>7.2 (4.0, 18)</td>
<td>22.8</td>
<td>18.5 (10.3)</td>
<td>22.9</td>
</tr>
<tr>
<td>1910</td>
<td>10.1 (5.7, 19)</td>
<td>23.1</td>
<td>36.0 (20.8)</td>
<td>24.1</td>
</tr>
<tr>
<td>1920</td>
<td>14.9 (8.8, 26)</td>
<td>24.9</td>
<td>29.1 (17.4)</td>
<td>25.3</td>
</tr>
<tr>
<td>1930</td>
<td>17.7 (10.9, 27)</td>
<td>26.5</td>
<td>37.9 (23.4)</td>
<td>26.5</td>
</tr>
<tr>
<td>1940</td>
<td>7.6 (4.9, 16)</td>
<td>28.3</td>
<td>36.4 (23.8)</td>
<td>29.0</td>
</tr>
<tr>
<td>1950</td>
<td>25.3 (16.8, 25)</td>
<td>29.7</td>
<td>34.2 (22.9)</td>
<td>30.1</td>
</tr>
<tr>
<td>1955</td>
<td>25.0 (16.8, 25)</td>
<td>30.1</td>
<td>34.3 (23.0)</td>
<td>30.2</td>
</tr>
<tr>
<td>1960</td>
<td>21.7 (14.7, 25)</td>
<td>30.7</td>
<td>35.8 (23.7)</td>
<td>29.6</td>
</tr>
<tr>
<td>1965</td>
<td>23.1 (15.8, 25)</td>
<td>31.2</td>
<td>27.6 (17.8)</td>
<td>28.4</td>
</tr>
<tr>
<td>1970</td>
<td>22.9 (15.9, 25)</td>
<td>31.8</td>
<td>42.8 (27.5)</td>
<td>28.2</td>
</tr>
<tr>
<td>1975</td>
<td>26.0 (18.1, 25)</td>
<td>32.1</td>
<td>29.8 (19.4)</td>
<td>28.8</td>
</tr>
<tr>
<td>1980</td>
<td>29.3 (20.6, 25)</td>
<td>32.7</td>
<td>41.8 (28.0)</td>
<td>30.1</td>
</tr>
<tr>
<td>1985</td>
<td>29.2 (20.9, 25)</td>
<td>33.7</td>
<td>34.7 (23.9)</td>
<td>31.5</td>
</tr>
<tr>
<td>1990</td>
<td>34.7 (25.3, 24)</td>
<td>34.8</td>
<td>43.4 (30.6)</td>
<td>32.8</td>
</tr>
<tr>
<td>1995</td>
<td>42.9 (31.9, 35)</td>
<td>36.1</td>
<td>55.6 (40.0)</td>
<td>34.0</td>
</tr>
<tr>
<td>2000</td>
<td>41.1 (31.2, 35)</td>
<td>37.6</td>
<td>47.0 (34.4)</td>
<td>35.1</td>
</tr>
<tr>
<td>2005</td>
<td>37.2 (28.8, 35)</td>
<td>38.9</td>
<td>46.5 (34.5)</td>
<td>36.0</td>
</tr>
</tbody>
</table>

**SOURCES:** ID values are from Vanhanen (2009). For Europe, the number of countries (n) with available ID values is also given. See Maddison (2009) for the populations used to weight these values. The MA values for Europe for 1950–2005 are from United Nations (2011). The MA values for 1850–1940 derive from population-weighted averages of estimates made for Austria, Belgium, Bulgaria, Denmark, France, Germany, Greece, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom using age data in United Nations (1956). Data are available for six countries in 1850; by 1900 the number is 12. Comparison for 1950 of the MA estimate based on these 15 countries with that for Europe in United Nations (2011) led to the 1850–1940 values being pro-rated downward by 0.929—reflecting their scant representation of Eastern Europe. For the United States, the MA values for 1850–1940 are from Haines and Steckel (2000) and those for 1950–2005 are from United Nations (2011). See text for the procedure used to derive AID.
As one would expect, in the early decades after 1850—when the populations of Europe and the United States were young—the calculation of AID involves a large upward adjustment of ID in proportional terms. The scale of the adjustment then declines as the populations get older. Note too that the median age of Europe’s population may have been falling in the late nineteenth century, before starting its long-run rise after 1890. For both Europe and the United States, the strength of the association between MA and AID declines as AID is increasingly lagged on MA—that is, the correlation is strongest with no lag.7

Considering the emergence of democracy in Europe, the AID values in Table 1 indicate an initial rise between about 1890 and 1930. Of course, the abrupt fall in the AID value for 1940 reflects the appearance of fascist regimes. After World War II, there is a (more than) full recovery by 1950, relative stability in the period to 1985, and a further sharp rise between 1985 and 1995—reflecting the collapse of the Soviet Union. Lastly, the AID value falls slightly after 1995, mainly the result of adverse events in Russia. It is worth noting that although Europe experienced considerable economic growth and socioeconomic development between 1950 and 1985, there was little change in the level of democracy: essentially, political progress was “frozen” by the Cold War.

The AID values for the United States suggest that the level of democracy has always been higher than in Europe. However, one should recall that the emergence of democracy in the US was closely related to the experience of countries in northwestern Europe, rather than to that of Europe as a whole. The AID values for the United States suggest that, again, the period between about 1890 and 1930 was particularly significant. Indeed, the value of AID almost doubles between 1900 and 1910.8 It is noteworthy that, with some fluctuation, the level of democracy then remains fairly unchanged for much of the twentieth century. The modest upward shift in democracy seen after 1990 partly reflects the influence of the increased holding of state referendums in the calculation of ID.

Table 2 summarizes the results from multiple regression analysis of the trends in AID shown in Table 1. For Europe, it was only possible to compile additional time series for per capita income. However, as noted, series for the United States can also be assembled for urbanization, employment in agriculture, and education. The difficulties of drawing firm conclusions in circumstances where all variables are moving in one general direction should require no emphasis here. Nevertheless, in both cases MA performs relatively well.

The rise of democracy in Europe during the “democratic age” of the late nineteenth and early twentieth centuries is of particular interest. Exactly when the rise began is hard to say, but it was definitely underway from 1890 (see Table 1). Comparison of the ID values for 1890 and 1930 for the 18 European countries for which estimates for both years are available re-
TABLE 2  Regression coefficients (ordinary least squares) of the effect on AID of median age (MA) and other socioeconomic measures: Europe 1870–2005 and United States 1870–1995

<table>
<thead>
<tr>
<th></th>
<th>Europe</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age (MA)</td>
<td>1.912**</td>
<td>2.38*</td>
</tr>
<tr>
<td></td>
<td>(0.717)</td>
<td>(1.081)</td>
</tr>
<tr>
<td>Log (GDP)</td>
<td>2.128</td>
<td>1.480</td>
</tr>
<tr>
<td></td>
<td>(5.100)</td>
<td>(7.964)</td>
</tr>
<tr>
<td>Percent urban</td>
<td>—</td>
<td>1.895</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.289)</td>
</tr>
<tr>
<td>Percent in agriculture</td>
<td>—</td>
<td>1.652</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.417)</td>
</tr>
<tr>
<td>High school graduation rate</td>
<td>—</td>
<td>-0.109</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.215)</td>
</tr>
<tr>
<td>Constant</td>
<td>-53.141**</td>
<td>-182.481</td>
</tr>
<tr>
<td></td>
<td>(23.140)</td>
<td>(117.638)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.897</td>
<td>0.807</td>
</tr>
<tr>
<td>Observations</td>
<td>21</td>
<td>18</td>
</tr>
</tbody>
</table>

*Significant at p < 0.10; **p < 0.05. Standard errors in parentheses.

SOURCES: Full series of per capita GDP (in 1990 international Geary-Khamis dollars) in Maddison (2009) start from 1870. For the United States, statistics on percent urban were derived mainly from US Census Bureau (2008). Percent in agriculture relates to the economically active population engaged in agriculture, forestry and fishing; figures for 1870–1945 are from Mitchell (1983); estimates from 1950 are from Carter et al. (2006). The high school graduation rate is secondary school graduates as a percentage of 17-year-olds, and is also from Carter et al. (2006).

reveals that the rise was almost entirely attributable to substantial advances in eight countries, mostly in northwestern Europe. These countries are Austria, Belgium, Denmark, Germany, Netherlands, Norway, Sweden, and the UK. More limited progress occurred in France, Greece, and Switzerland. On the other hand, ID remained close to zero in the remaining countries, which are mostly in southern and eastern Europe.9

In this context, it is notable that the onset of fertility decline in northwestern Europe often occurred in the 1870s and 1880s—that is, somewhat earlier than elsewhere (Demeny 1968; Chesnais 2001). Table 3 contains the median ages for 1890 and 1930 of the eight countries that experienced major rises in ID. Between 1890 and 1930 their average (unweighted) AID value rose from 3.2 to reach the high level of 42.1, while their average median age increased from 24.0 to 28.8 years. Even in 1950 the median age of Eastern Europe’s population was only 26.4 years (United Nations 2011).

Table 3 presents the MA and AID values for all countries for which estimates of both measures can be made for 1890 and 1930. Figure 1 plots the relationship between the changes in MA and AID. The relationship is posi-
### TABLE 3 Values of AID and median age (MA) for 13 countries in Europe, 1890 and 1930

<table>
<thead>
<tr>
<th></th>
<th>AID (ID) 1890</th>
<th>AID (ID) 1930</th>
<th>MA 1890</th>
<th>MA 1930</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>0.4 (0.2)</td>
<td>48.0 (32.3)</td>
<td>26.4</td>
<td>30.4</td>
</tr>
<tr>
<td>Belgium</td>
<td>1.0 (0.6)</td>
<td>26.1 (17.8)</td>
<td>24.3</td>
<td>31.1</td>
</tr>
<tr>
<td>Denmark</td>
<td>2.1 (1.2)</td>
<td>37.0 (23.5)</td>
<td>23.9</td>
<td>27.7</td>
</tr>
<tr>
<td>France</td>
<td>14.3 (9.4)</td>
<td>23.5 (16.3)</td>
<td>29.2</td>
<td>31.9</td>
</tr>
<tr>
<td>Germany</td>
<td>5.2 (2.9)</td>
<td>48.9 (32.3)</td>
<td>23.0</td>
<td>29.5</td>
</tr>
<tr>
<td>Greece</td>
<td>1.4 (0.7)</td>
<td>15.1 (8.7)</td>
<td>20.6</td>
<td>24.0</td>
</tr>
<tr>
<td>Italy</td>
<td>2.1 (1.2)</td>
<td>0.5 (0.3)</td>
<td>24.4</td>
<td>26.1</td>
</tr>
<tr>
<td>Netherlands</td>
<td>5.5 (3.1)</td>
<td>50.4 (30.4)</td>
<td>23.2</td>
<td>25.6</td>
</tr>
<tr>
<td>Norway</td>
<td>4.6 (2.6)</td>
<td>46.8 (29.2)</td>
<td>23.3</td>
<td>26.9</td>
</tr>
<tr>
<td>Portugal</td>
<td>0.3 (0.2)</td>
<td>0.0 (0.0)</td>
<td>24.7</td>
<td>24.4</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.3 (0.2)</td>
<td>37.1 (24.4)</td>
<td>25.3</td>
<td>29.3</td>
</tr>
<tr>
<td>Switzerland</td>
<td>11.8 (7.0)</td>
<td>26.7 (17.5)</td>
<td>25.2</td>
<td>29.2</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>6.2 (3.4)</td>
<td>42.8 (28.6)</td>
<td>22.6</td>
<td>30.1</td>
</tr>
</tbody>
</table>

**SOURCES:** AID values, shown in parentheses, are from Vanhanen (2009). See text for the procedure used to derive AID. MA values were estimated using census age data in United Nations (1956). The age data provided for Austria for 1910 and earlier years clearly relate to Austria-Hungary; therefore age data for Austria available at «http://sdb.statistik.at/» were employed to estimate MA for these years. Interpolation was sometimes used to obtain estimates of MA for the exact years shown.

### FIGURE 1 Relationship between change in AID and change in median age (MA), 13 countries in Europe, 1890–1930

**NOTE:** The changes in AID and MA are calculated from the values given in Table 3.
tive and significant ($p = 0.049$). In general, countries that experienced larger increases in MA became appreciably more democratic. The geographical dimension of Figure 1 is noteworthy. Thus Denmark, Norway, and Sweden form a Scandinavian cluster that underwent similar changes. Also, Greece, Italy, and Portugal—together with France—can be seen as comprising a "southern" cluster that experienced relatively minor changes in MA and AID. Perhaps the most notable outlier is the Netherlands, which had become the most democratic country by 1930. The Netherlands experienced a major rise in AID, even though, with somewhat higher fertility than most neighboring countries, its median age increased by just 2.4 years. It is possible that political developments elsewhere in northwestern Europe influenced democratization in the Netherlands despite the country's slower demographic progress.

For 11 of the countries in Table 3 it is also possible to assemble estimates of per capita income, employment in agriculture, and the average number of years of schooling attained by the population aged 15 years and over, for the period 1890–1930. Table 4 summarizes the results obtained when these variables are included in the analysis, both for the period as a whole and on a pooled basis when decadal data are considered. Again, changes in MA are most closely associated with changes in AID (in the pooled analysis, $p = 0.043$).

<table>
<thead>
<tr>
<th>TABLE 4 Regression coefficients (ordinary least squares) of the effect on changes in AID of changes in median age (MA) and other socioeconomic measures: Europe, 1890–1930</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>1890–1930</td>
</tr>
<tr>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td>Median age (MA)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Log (GDP)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Percent in agriculture</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Years of schooling of population aged 15+</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>R-squared</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

*Significant at $p < 0.10$; **$p < 0.05$. Standard errors in parentheses. 
SOURCES: For the sources of the AID and MA values used, see the notes to Table 3. Per capita GDP estimates were taken from Maddison (2009); percent working in agriculture was calculated from Mitchell (2007); the education estimates are from Morrisson and Murtin (2009). See also endnote 11.
Comparison of the time paths of AID and MA for individual countries reveals interesting variation. While, as one would expect, the median ages of populations increase relatively smoothly, the progress of AID is more variable—for example, reflecting legislative change on voting rights. France is a notable case. The country’s population was perhaps the oldest in Europe during 1890–1930, but its level of democracy, as indicated by AID, while increasing, remained relatively low (see Table 3). The main reason was that French women could not vote until 1945. Similarly, female suffrage in Belgium was not attained until 1948, and the modest rise in AID for Switzerland again reflects the late extension of voting rights to women. These cases illustrate that while demographic progress, as summarized by MA, is fairly regular, changes in the rules that govern political participation (and competition) can be both delayed and abrupt. Among other factors, legislative change is the outcome of struggles between different interest groups, and it is also influenced by the nature of pre-existing institutions. Therefore, one might expect to find evidence of “stickiness” between demographic change on the one hand and the emergence of democracy on the other.

For a few countries, however, it is clear that increases in AID occurred ahead of increases in MA. The case of Netherlands has been noted, although the explanation at least partly reflects the country’s comparatively late fertility decline. But the experience of Denmark and Norway was different. Their time paths of MA are not unusual, but their rises in AID clearly began somewhat earlier. This is especially clear for Norway. One might suggest that the country’s struggle for independence from Sweden may have facilitated its early realization of democracy. Of course, a key aspect of this phenomenon was the gaining of votes for women, which occurred in 1913. As in Australia and New Zealand—the first countries to introduce female suffrage—it appears that the creation of political institutions to some extent de novo in Norway meant that problems of institutional stickiness were easier to overcome.

Lastly, using estimates from Chesnais (2001), I can also examine trends in fertility for a few of the countries in Table 3. For illustration, Figure 2 compares changes in total fertility, MA, and AID in Germany between 1871 and 1940. To facilitate comparison, the fertility, MA, and AID series have been indexed between zero (the lowest value) and 100 (the highest value), with the index for fertility reversed (i.e., an increasing index signifies falling fertility). As one would expect, fertility decline in Germany preceded the rise in MA, the delay reflecting the offsetting effect on age structure of declines in child mortality. Notice that MA and AID rose together (i.e., with no lag), before democracy collapsed in the 1930s. The experience of other countries exhibits some variation in the timing of the rise in AID relative to changes in both fertility and MA. Thus, and exceptionally, in Norway fertility decline and the rise in AID occurred more or less simultaneously—and well before
the rise in MA. On the other hand, the experience of Sweden and the UK was more akin to that of Germany, with MA and AID rising at about the same time. Finally, for reasons already noted, the attainment of high AID values for France and Switzerland was delayed even longer compared to these demographic changes.

In summary, it appears that fertility decline almost always preceded the rise in AID. Moreover, in several cases MA and AID rose broadly together, although the concurrence shown in Figure 2 for Germany is exceptional. Returning to theoretical considerations, one could clearly develop an additional argument that fertility decline reflects an increase in human agency and that, other things equal, it is this same increase in agency that, with varying lags, eventually makes itself evident in the political domain.

**International variation**

The main concern of this essay is with historical trends. But analysis of contemporary cross-national data has supported the conclusion of a close as-

**FIGURE 2  Trends in AID, MA, and total fertility, Germany 1871–1940**

NOTE: For purposes of comparison, the total fertility index has been reversed.

SOURCES: Total fertility rates from Chesnais (2001). For the derivation and sources of the AID and MA values see the notes to Table 3.
association between per capita GDP and democracy (Ziblatt 2006). Therefore, it is worth briefly considering the recent international relationship between MA and AID.

Many factors affect the status of countries with respect to democracy in the modern world. And, partly as a result of pressure from Western countries, the process of democratization often reflects foreign influences. In this context, the AID values for 2005 for some countries with very young populations—such as Malawi (39.7) and Mauritania (45.8)—compare favorably with those of long-established democracies with older populations—such as Canada (34.5) and the UK (38.2). Such considerations illustrate the limitations, perhaps especially in modern circumstances, of AID as an indicator of democracy having become consolidated in a given country. A high AID value does not necessarily mean that democracy is secure.

Nevertheless, in what follows the focus remains on AID and on the relative importance of key socioeconomic variables that are thought to promote democracy. The period considered begins in 1980, partly because of the paucity of employment and education estimates for earlier decades. Moreover, for most of the 1950s, 1960s, and 1970s Vanhanen’s dataset contains ID values for relatively few countries, chiefly because it excludes populations then under colonial rule.

From 1980 onward, however, it is possible to assemble a relatively comprehensive set of estimates relating to AID, MA, per capita GDP, urbanization, employment, and education. Given what is sometimes claimed regarding the influence of urbanization on democratization, the inclusion of United Nations (2010) urbanization estimates should be noted. Estimates of the percentage of the workforce employed in agriculture are taken from the International Labor Organization (2008). Estimates of the percentage of the population aged 20 years and over having at least secondary school education are from Lutz and colleagues (Lutz, Goujon, and Sanderson 2007).

With this as background, Table 5 summarizes the results from cross-sectional regressions relating to 1980–2005. For each point in time, the analysis was conducted for the 151 countries for which estimates of all of the measures could be assembled. The regressions were run both with and without median age. Notice that without MA, per capita income is statistically significant at each point in time. But when MA is introduced, the strength of the association between AID and income weakens considerably, and disappears for 1990 and 2005. There is no consistent indication that a country’s level of urbanization, employment structure, and level of education are associated with its position with respect to democracy. In contrast, the relationship between MA and AID is statistically significant throughout. These results tentatively suggest that a country’s status with respect to the demographic transition may still exert an influence on whether it is a democracy.
Discussion

This essay has proposed that progress in the demographic transition is a significant factor in the process of democratization. Drawing on experience in Europe and the United States, I have suggested that demographic and democratic developments were associated with each other in the past and that they probably remain so today. Historical research can shed light on factors that help determine whether new democracies will succeed.

Some degree of mortality decline is almost certainly required for stable democratic conditions. However, through the population growth that it generates, mortality decline raises challenges for pre-existing—autocratic—regimes. Fertility decline is crucial here. It causes population aging, brings about major alterations in the lives of women, and produces a fall in the rate of natural increase—gradually enhancing the conditions required for socio-political stability. Median age is a good measure of cumulative demographic progress in most of these respects. It is difficult to regard any country with a low median age as being firmly democratic.

Demographic and democratic gains can both be viewed as reflecting and representing advances in human agency. Much of the present argument has been that as the structure of a society becomes increasingly composed of adult men and women, autocratic political structures are likely to be increasingly

### TABLE 5 Regression coefficients (ordinary least squares) of the effect on AID of median age (MA) and other socioeconomic measures: All countries with available estimates of all measures, 1980–2005

<table>
<thead>
<tr>
<th></th>
<th>1980</th>
<th>1990</th>
<th>2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age (MA)</td>
<td>—</td>
<td>1.22***</td>
<td>—</td>
</tr>
<tr>
<td>Log (GDP)</td>
<td>10.59***</td>
<td>7.46**</td>
<td>6.16**</td>
</tr>
<tr>
<td>Percent urban</td>
<td>–0.41**</td>
<td>–0.42***</td>
<td>0.06</td>
</tr>
<tr>
<td>Percent employed in agriculture</td>
<td>–0.24</td>
<td>–0.26</td>
<td>–0.12</td>
</tr>
<tr>
<td>Percent with secondary education</td>
<td>0.19**</td>
<td>0.02</td>
<td>0.22**</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.50</td>
<td>0.55</td>
<td>0.67</td>
</tr>
<tr>
<td>Observations</td>
<td>92</td>
<td>92</td>
<td>94</td>
</tr>
</tbody>
</table>

*Significant at p < 0.10; **p < 0.05; ***p < 0.01. Standard errors in parentheses.
SOURCES: GDP estimates from Maddison (2009); percent urban estimates from United Nations (2010); employment estimates from International Labor Organization (2008); education estimates from Lutz, Goujon, and Sanderson (2007). See also endnote 14.
challenged and replaced by more democratic ones. Here demographic processes work in remote (i.e., underlying) ways. Debating whether the middle or the working class is the vehicle of change is a secondary matter.

I should underscore that this essay has been concerned with factors that contribute to the establishment of stable democracies. In this context, research on contemporary cross-national data suggests that countries that have recently become democratic are much more likely to regress if their age structures are young (Weber 2012)—like Malawi and Mauritania, for example. I would also emphasize that the present results do not rule out the beneficial influence of other socioeconomic variables on democratization. Thus, using age-specific estimates of educational attainment, work by Lutz and colleagues on recent data also finds that declines in youth dependency are favorable for democracy, but suggests that improvements in education may be even more important (Lutz, Cuaresma, and Abbasi-Shavazi 2010). These authors too doubt whether per capita income has a significant influence on democracy. Indeed, there is growing evidence that demographic transition is at least as beneficial for democratization as is economic growth.

Of course, demographic and socioeconomic factors are not the only influences on democratization. Historical and geographical considerations can be very important. Moreover, political developments frequently have a powerful—and sometimes adverse—dynamic of their own. While I have not addressed such factors here, they often explain the variable progress of countries in their demographic and democratic transitions. India and Cuba provide examples. Democracy in India advanced somewhat ahead of demography. India’s first general elections were held in 1951–52, when the country’s median age was only about 20 years. However, external influences were significant in the establishment of democracy in India. One should also recall that parliamentary rule was suspended during 1975–77: democracy was fragile. The fact that India’s median age currently exceeds 25 and is rising surely signifies a strengthening basis for democratic politics. By contrast, Cuba is an example where, for political reasons, there has been no real prospect of democracy for decades, despite substantial demographic advance (median age is about 38). Nevertheless, the demographic basis for democracy clearly exists in Cuba, and it may take only a seemingly small event, such as the departure of a few national leaders, to trigger significant democratic change.

It is important to acknowledge, therefore, that progress in the demographic transition does not guarantee the transition to democracy—even on time scales that can stretch for many decades. In particular, political processes can intervene. One has only to think of the rise of fascism in Europe in the 1930s, the influence of the Cold War, or the continuing legacy of Communism in countries like China, Cuba, and Russia. Nevertheless, if the argument has force, it should bode well for democracy. In recent decades many countries, poor as well as middle income, have experienced fertility decline and begun
to age. Such changes are likely to help foster the emergence of stable democratic systems.

Notes

This work was undertaken through a Fellowship from the UK Economic and Social Research Council (ESRC Award RES-063-27-0159). I thank Steve Broadberry, Sean Fox, Elliott Green, Lloyd Gruber, Mike Murphy, Ken Shadlen, Máire Ní Bhrolcháin, Bas van Leeuwen, and especially Ben Wilson for help and advice.

1 Of 125 countries, only New Zealand (1893), Australia (1902), Finland (1906), and Norway (1913) had the female franchise before 1915 (Paxton 2000). As Rueschemeyer, Stephens, and Stephens (1992: 140) remark, New Zealand benefited from earlier franchise struggles in Britain and was “virtually ‘born free.’”

2 Indeed, because scholars often attribute democratization to growth of the middle or working classes, it might be said that population aging is—albeit unconsciously—implicit in their argument insofar as aging contributes to growth of the adult membership of both classes.

3 For support for these statements, see Dyson (2010).

4 For the results of earlier work using the Polity measure “Democ,” which derives from evaluation of institutional structures, see Dyson (2010).

5 Democracy is often operationalized with respect to male suffrage. The Polity measures have many weaknesses (Munck and Verkuilen 2002)—for example, in this context, Switzerland receives the maximum “Democ” score of 10 from 1848 onward, although women only gained voting rights in federal elections in 1971.

6 Data for 1950 and 2005 revealed little change in the relationship. Therefore, data for both years were pooled to produce the equation.

7 This statement is based on the decadal data for 1850–2000 in Table 1. Lags of up to 50 years were considered.

8 Earlier versions of Vanhanen’s dataset suggest a slightly less abrupt—and arguably more plausible—increase in ID for 1890–1930.

9 These countries are Hungary, Italy, Portugal, Romania, Russia, Spain, and Yugoslavia/Serbia.

10 Greece and Norway were excluded owing to lack of employment data.

11 The coefficient for per capita GDP in Table 4 becomes negative (p=0.093) in the pooled analysis. This is partly because times of hardship, especially 1910–20, seem to have accompanied, and indeed prompted, democratization (there may be parallels here with Latin America in the 1980s).

12 See endnote 5.

13 See also the comment on New Zealand by Rueschemeyer, Stephens, and Stephens (1992) in endnote 1.

14 In Table 5, the 92 countries for which estimates of all measures could be assembled for 1980 were: Argentina, Australia, Austria, Bangladesh, Belgium, Benin, Bolivia, Brazil, Bulgaria, Burkina Faso, Cambodia, Cameroon, Canada, Central African Republic, Chad, Chile, China, Colombia, Costa Rica, Côte d’Ivoire, Cuba, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Finland, France, Gabon, Ghana, Greece, Guatemala, Guinea, Haiti, Honduras, Hungary, India, Indonesia, Iran, Ireland, Italy, Japan, Jordan, Kenya, Laos, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mauritius, Mexico, Mongolia, Morocco, Mozambique, Myanmar, Netherlands, Nepal, New Zealand, Nicaragua, Niger, Nigeria, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Rwanda, Saudi Arabia, Singapore, South Africa, South Korea, Spain, Sri Lanka, Sweden, Switzerland, Syria, Tanzania, Thailand, Togo, Turkey, Uganda, United Kingdom, United States, Uruguay, West Germany, Zambia, and Zimbabwe. For 1990, Germany replaced West Germany, and measures were also available for Namibia and Vietnam (bringing the total to 94). For 2005, Morocco and Romania were omitted because Vanhanen (2009) does not provide ID values; however, estimates of
all of the measures could be assembled for Armenia, Czech Republic, Croatia, Estonia, Latvia, Lithuania, Kazakhstan, Kyrgyzstan, Macedonia, Russia, Slovakia, Slovenia, Tajikistan, Turkmenistan, Ukraine, and Uzbekistan (bringing the total to 108).

15 Similar conclusions emerge from regressions using the Polity measure “Democ” as the dependent variable.

16 Educational data of the kind employed by Lutz and colleagues (2007) are unavailable for the historical contexts considered here.

References


A Demographic Perspective on Japan’s “Lost Decades”

REIKO AOKI

The Japanese economy has not grown since the asset bubble burst in 1990. Stock markets have not yet recovered and the unemployment rate has remained high, by Japan’s standards, despite a zero real interest rate (held down by deflation). Initially, the years of economic underperformance were called the “Lost Decade.” A widely held view was that Japan had not bounced back because the country lacked an effective economic policy. But after another decade passed without significant improvement, those years are now called the “Lost Decades.” The emerging consensus is that the causes of the Lost Decades are to be found in fundamental aspects of Japanese society and economy.

My argument in this essay is that the economic stagnation of the Lost Decades is a symptom of problems brought on by demographic changes in Japan since the end of World War II, notably the population aging caused by low fertility and increased longevity. Institutions that had functioned well under a growing supply of labor in expanding markets, such as social security schemes, labor market practices, and the political system, failed to adapt. In addition, the shifting age structure worked against badly needed economic adjustments, as resources continued to be directed toward the aging members of the society. The burden of old-age support has become the responsibility of the relatively smaller young age cohorts, creating a situation that is very difficult for the society to correct. Japan’s traditional respect for the elderly means that taking resources away from the old is socially unacceptable; it is also politically difficult, because the older generation forms a very large voting bloc. In the conclusion I describe the country’s demographic prospects and take note of possible interventions that might yield better outcomes: some of them specific to Japan, others that would apply to any aging society. There may be lessons for other societies that eventually must follow Japan’s demographic path.
Background

The foundations of Japan’s current economic and legal system were established immediately after World War II. In 1947, the year in which the revised constitution became effective, the total fertility rate (TFR) was 4.5, life expectancy was 50.1 (men) and 54.0 (women), and the median age of the population was 22 years. An economic boom associated with postwar reconstruction and the Korean War followed, along with a brief surge in fertility paralleling the Western baby boom. A ten-year plan to double the national income, implemented in 1960, achieved its target in six years. During this time, the TFR steadily declined while life expectancy and the median age of the population steadily increased.

The economic downturn began in 1989, when the Bank of Japan, fearing inflation and overheated assets markets, raised its interest rate in a series of steps from 2.5 percent to 6.0 percent by the summer of 1990. The Nikkei average stock price index fell from over 30,000 to below 20,000 by 1995. The unemployment rate increased from 2.1 percent in 1990 to 3.2 percent by 1995. The bubble had burst. The Bank of Japan then began to cut the interest rate until it reached 0.3 percent in 2001, but the Nikkei average remained below 20,000 and unemployment reached 5 percent. The nominal Japanese GDP has remained flat for over two decades. The Nikkei average dropped below 10,000 after the Lehman Brothers crisis in 2008 and has not recovered.

The year 1990 was also a demographic landmark. By then, life expectancy had risen to 75.9 (men) and 81.9 (women), and the median age of the population was 37. It became evident during 1990 that the TFR for the previous year had dropped to 1.57—a fact now referred to as the “1.57 shock.” The number was significant because it was lower than the 1966 TFR, 1.58. According to the Chinese zodiac, 1966 was the “Year of the Fire Horse” (Hinoe-Uma), in which parents prefer not to have a child. That fertility could be so low in a non-Hinoe-Uma year became a source of public concern, calling wide attention to the fertility situation.

In response, the government introduced various pronatal policies. However, unlike the economy, which had begun to show some improvement, the TFR continued to decline. In 2010, the TFR was 1.39. Life expectancy had continued to rise through the Lost Decades, reaching 79.7 (men) and 86.4 (women). The median age had risen to 47.

This time is different

Japan’s economy has recovered from major economic shocks in the past. The last one was the OPEC petroleum price hikes in the 1970s. At that time, Japan’s recovery was assisted by demand for fuel-efficient cars, a market that
Japanese manufacturers dominated. Japan also had an ample supply of engineers, scientists, and skilled workers to support innovation and production to meet the global demand for Japanese manufactures.

Why has Japan been unable to recover economically this time? Several factors have changed since the 1970s. Internally, population growth has halted and the population is aging, with implications not only for labor markets and social services but also for the knowledge base for innovation and production. Externally, Japan’s comparative advantage has changed. China has replaced Japan as the major manufacturer of electronics and electric appliances. China also can provide labor comparable to Japan’s in quality.

**Macroeconomy**

A direct consequence of an aging population is that the pension system becomes unsustainable. This topic has been discussed in detail elsewhere, and I will not dwell on it here other than to point out the unwillingness of society as a whole to correct the problem. A few adjustments have been made to accommodate the age shift. The age of eligibility for *kosei nenkin*, one of the two public pension schemes, has been raised from 55 to 65. A law was passed to index pension payments to the price level, but an addendum was made, with support of all political parties, not to apply it for deflation. As result, pensions have been overpaid in terms of purchasing power since 2005. (Some difference is allowed between price change and pension levels). A law was finally passed in 2012 to correct the overpayment. Politicians have been very reluctant to vote to reduce nominal pension payments—an unpopular move, especially among older voters.

Stopping deflation to prevent further stagnation is one of the key aims of a successful economic policy. Deflationary expectations make people wait for their money to increase in value, which reduces current demand for goods. Excess supply leads to price cuts, fulfilling deflationary expectations. Deflation often involves a cut in wage income for those who work, from either lower wages or fewer opportunities (less work and less overtime). However, deflation benefits those on fixed incomes, such as pensioners, since the same amount in Yen buys more. Deflation adjustment would undo this appreciation, holding purchasing power constant.

Similarly, the recent appreciation of the Yen against other major currencies hurts the younger generation much more than the older generation of pensioners. Many Japanese firms depend on exports, and thus profits decline with a higher exchange rate. Firms adjust by reducing labor costs by either hiring fewer staff or reducing wages. There is downward rigidity in base wages but Japanese firms are able to adjust using “bonuses,” which are paid twice a year and may amount to several months’ wages. Bonuses are contingent on firm performance, not individual performance. But pensions are not adjusted
to reflect firm performance. Premiums and tax revenue may fall, but pension payments are not indexed to any sort of revenue. Hence the current Japanese economy is not necessarily bad for those on pensions. In contrast, workers are likely to be harmed by deflation. Inflation (in moderation) would be a very good way of reducing the country’s large national debt, which will otherwise be a burden on future generations.

Human capital and the labor market

With few natural resources, Japan’s people are its greatest resource—a fact drilled into Japanese from an early age. There has always been a strong emphasis on investment in human capital. Education is given a high priority within the family, and labor practices have promoted firm-specific on-the-job training. The expectation of “lifetime employment,” at least as an ideal, meant that firms could recoup their investment over their employees’ working lives.

This combination of labor practices made Japanese firms highly competitive in the world market of manufactured goods. However, this system relied on population growth and became unfeasible when that growth ceased. In its place, firms have turned increasingly to “non-regular workers,” those on short-term contracts. Such workers have little opportunity to accumulate human capital and limited chances of promotion. Non-regular workers are disproportionately young and their numbers have risen significantly in the past two decades. Since older workers often have lifetime employment and job security, the young bear most of the adjustment costs.

Initial wages under lifetime employment were very low because workers implicitly paid for being trained. Nevertheless, wages were in effect guaranteed to rise in the future to allow the young to make long-term commitments. The recent change in employment status among the young has several long-term implications. First, firms are no longer willing to invest in the human capital of their employees. Workers accumulate fewer skills and have poorer wage prospects, whether in their current job or in the outside labor market. They are forced to delay marriage and starting a family, putting further downward pressure on the fertility rate and hastening population aging. Long-term planning has become very difficult for the young generation.

Less on-the-job-training means that firms rely more on the skills and knowledge of their older workers. Hence firms continue to do what they did well two decades ago: traditional manufacturing, such as cars and electric appliances. The aging workforce and its dated knowledge of science and technology make switching to new industries difficult and slow.

These traditional manufactured products are no longer profitable for Japan. They have become commoditized, relying for profit on low-cost production based in low-income countries. Higher value-added products, such as
sophisticated electronics, require cutting-edge knowledge and expertise—embodied in younger workers and researchers who have started their careers with the latest technological training. Few firms are able to engineer this shift.

Science, technology, and innovation

There is a keen awareness in Japan that the country faces many challenges from emerging economies. Japan needs to find new sources of comparative advantage. More immediately, there is an urgent need to find alternative sources of energy to replace the idled nuclear power plants. These issues have been debated repeatedly, but change has been slow. Shifting workers from one sector to another is impeded by lifetime employment. And researchers and academics with stakes in the older technologies are poorly suited to supporting the development of new industries. Recruitment patterns in Japanese universities do not support the risky, but high-return, staffing bets on young researchers in the way that American universities use tenure-track appointments. The difficulty in updating skills and knowledge, partly a consequence of population aging, is a serious factor in preventing recovery from the Lost Decades.

Female work force

Women in Japan are well educated but not fully integrated into the general labor force. About 50 percent of women obtain four-year college degrees, compared to 70 percent of men. (Another 15 percent of women obtain junior college degrees—an uncommon option for men.) The ratio of women to men in tertiary education is lower than in most other OECD countries, and this ratio worsens as careers progress. At each stage of higher education and career, proportionately more women than men drop out and fewer women are promoted. Some women, of course, opt to become full-time mothers.

One offset to the declining labor force is for more women to enter. Recently, terms such as “gender equality” and “promotion of women in the workplace” have been replaced by the term “work–life balance.” The shift reflects the fact that female labor participation has become a macroeconomic necessity, needing to be encouraged independently of political concepts such as equality. A work–life balance is also relevant for men. International comparisons of time-use surveys have shown that household chores and childcare (measured in time) are concentrated among women in Japan while the burden is spread more evenly between men and women in other countries. The evidence implies that changes are needed at home as well as at work to induce women to work outside the home.

This time-use evidence also suggests that, compared with women, men devote too much of their time to work and not enough to home life. Greater
participation by men in the pleasures and challenges of parenthood would also mean a better work–life balance for their partners. Fortunately, that balance has become easier to achieve in the present economic circumstances in which there is less opportunity for overtime work. (A further impetus to reduce overtime in Tokyo resulted from the electricity shortage that has followed the 2011 Fukushima disaster. Tokyo firms kept shorter hours and fathers found themselves spending time on chores such as picking up their children from daycare.)

Foreign workers

While many countries depend on immigrant labor, Japan has no tradition of accepting immigrants. Citizenship is by birth (*jus sanguinis*), and naturalization is comparatively rare. It is not impossible, as demonstrated by the existence of un-Japanese-sounding names among Japanese Olympians, but it is reserved for exceptional cases. The numbers that would hypothetically be required to make up the labor shortage from changing demographic circumstances are several orders of magnitude more than the level of immigration that Japan has experienced so far. Public opinion is overwhelmingly of the view that the values of Japanese society are not consistent with immigration. This view prevents any move toward increasing immigration as part of “policies toward offsetting low fertility,” although that possibility has often been debated. A more realistic option is the admission of foreign temporary workers who will eventually return to their home country. These guest workers contribute to Japan’s economy while young, but long-term obligations to them are limited: they may not collect pensions and do not draw on Japan’s health care system in their old age.

Officially, Japanese acceptance of foreign workers has been unenthusiastic. For instance, a program admits trainee nurses and care workers for the elderly from the Philippines and Indonesia to study and eventually work in Japan. The country confronts an acute shortage of health care workers; however, the foreign candidates must take the same qualifying examination as local candidates, requiring mastery of Japanese—a serious obstacle. While some language skill is needed, there would seem to be room for compromise if there was sufficient interest in the foreign recruitment. The problems implementing such a program send a signal to other potential immigrants that Japan does not welcome foreign workers.

Nevertheless, Japan is bound to increase its intake of foreign workers, and some evidence suggests that foreign workers find ready acceptance despite the cautious official stance. Reliable statistics on foreign workers have been recorded only for the last two decades: over this period, despite rising unemployment in the general population, the number of foreign workers has increased sevenfold. Among university faculty—a critical group for the
country’s technological future—the numbers of foreigners is very small by US and European standards, however, particularly in the top universities.

Fertility policy

Following the 1989 Hinoe-Uma shock, various “policies to address low fertility” were adopted. Typically, the word “problem” is omitted from their title, so as not to imply that low fertility is a problem. Similarly, they do not refer to “pronatal policy”: the idea that women could be strongly urged to have children is seen as almost reactionary. As Paul Demeny pointed out in 1986, family planning—aimed at lowering unwanted fertility—is part of a progressive ideology that is often associated with the liberation of women. It would follow that pronatal policies, in that respect the reverse of family planning, are antiprogressive: seeking to induce women to have children. (Indeed, some people, including politicians, claim that Japan’s low fertility is a result of women’s selfish behavior.)

Needless to say, this dichotomy is false. If women were being coerced into not having children, then having children would be progressive. The proper question is rather: should women be expected to behave a certain way, be it having children or not having children, or should they be free to choose to have children or not? Having the latter choice is progressive, as is family planning if it gives women freedom not to have children even when that is counter to the social norm. A further question is whether fertility decisions should be made solely by women. If fathers are expected to play an active role in childrearing, their interests should presumably also have an influence on the fertility decision. This kind of thinking is the underlying rationale for a work–life balance.

Another reason that the word “problem” is excluded in reference to low fertility is that many of the policies potentially bearing on fertility are not strictly pronatal. Increasing the number of daycare centers is one example. This will certainly reduce the cost of having children: a shortage of publicly subsidized daycare makes both the pecuniary and the opportunity costs of having children very high. However, building additional daycare centers is also a construction project and so doubles as part of an economic stimulation package. Given the current economic environment, this emphasis on stimulus is understandable, but as a result any pronatal effect of such projects is at most a minor consideration.

Overall, the amount of money allocated for purely pronatal or profamily policy is too small to be effective. An example is the childcare payment for each child. Initially this was ¥5000 per child per month. To put that figure into perspective, music lessons at Yamaha Music School are ¥4700 per month (three lessons). Although the maximum per-child amount has been increased to ¥15,000, there are new eligibility restrictions based on family income.
Cost of rearing children

Childrearing in Japan is very costly to families. Both the direct pecuniary costs and the costs in terms of time required of parents are substantial. Children usually have various after-school lessons, such as piano and ballet or additional academic tutoring. Preparation for entrance exams to selective junior high schools starts in the fifth (or even fourth) grade and often extends late into the evening. A parent must escort the child home, even if the school is within walking distance. At weekends there are mock practice exams, often located further away from home. These private expenses have become more necessary as the quality of public schools has deteriorated—perhaps another example of the lack of public investment in the younger generation.

The market for household and childcare services in Japan is very small and underdeveloped. If a woman wants to work outside the home, she must rely on institutional childcare, of which there is a chronic shortage. Lack of household help also makes it more difficult for women to work outside the home. In other developed countries domestic service workers are often immigrants or foreign guest workers; there is no supply of such people in Japan.

In the case of childcare, there is no practice of older schoolgirls babysitting to earn pocket money. Indeed, there is also some societal disapproval of a mother paying someone else to take care of her child. From the supply side, people are uncomfortable with the idea of demanding money, particularly from an acquaintance, to take care of a child: that should be done out of kindness, not as a business transaction. A century ago it was common for girls to work in childcare; now, as in all developed countries, there are legal restrictions on child labor. What is different about Japan is that informal babysitting hardly exists. Ironically, this protective stance toward children may help to reduce the size of the next generation of children.

Social values that restrict informal babysitting are part of Japanese social norms regarding family. Parents are expected to be very devoted to their children and willing to make sacrifices. Many people do not want to become parents until they can provide well for their children, especially a high-quality education. That requires money and time. In the absence of a market for household services, children are seen as nearly prohibitively expensive for those planning a career outside the home.

Policies and prospects

Japan’s demographic predicament is in some respects a product of its unique cultural circumstances and in other respects the familiar outcome of post-industrial affluence and a completed demographic transition. The range of plausibly beneficial policy responses reflects both these aspects. Success-
ful policies must also take into account the technological environment, in particular the fact that communication and transportation technology has changed the nature and effect of globalization.

The country’s demographic and economic circumstances need to be addressed simultaneously. Policies promoting greater work–life balance could lead to changes in both labor practices and the childrearing environment. Changes in labor practices and lifetime wage profiles will also make it possible for people to remain in the workforce longer.

Policies promoting work–life balance will also allow married couples to share work both in and outside the home. Sharing work outside the home means a household will have various job options, allowing couples to pool the risk of unemployment. A household with a single earner having the prospect of lifetime employment will no longer be the norm. With access to the global labor market, Japanese firms no longer need to invest in human capital themselves as they did in the days of lifetime employment. In fact, such a strategy is more risky for them as workers who have greater access to global markets can leave the firm whenever they choose. By relying more extensively on women, firms can rapidly draw on a large pool of highly educated workers.

Dependence on workers born abroad is likely to continue and expand, but Japan will never have an immigration policy akin to countries like the United States or Australia. Even in those countries, however, immigrants now maintain much closer ties to their country of birth than they did for most of the last century. The lower cost of communication and transportation makes it possible to work in one country for a significant length of time without becoming a citizen or putting down permanent roots. The diminution of the institution of “regular workers” with lifetime jobs is conducive to an increase in foreign temporary or guest worker employment. This form of employment may become the new face of immigration worldwide in the twenty-first century.

Successful promotion of work–life balance and an increase in the numbers of foreign workers may help to generate economic growth and prosperity, which in turn may encourage more couples to have at least one child. Greater economic prospects will also increase the expected lifetime earnings of children. But the effect of any change of this kind will not occur soon enough to address the country’s pension problem. (On the other hand, one might speculate that the potential insolvency of the pension system could itself serve to increase fertility. Having children may become a more attractive means of saving for old-age security than depending on an uncertain pension.)

If the aforementioned policies fail, or are not adopted in the first place, it is likely that Japan’s population will decline at a faster rate than at present. Skilled workers will look for work abroad. Single women in particular may move to other countries where opportunities are greater. Men and women who have a strong preference for children may also have an incentive to relocate elsewhere. Japan will then reach a point of pension insolvency at
which the government is unable to fund pension payments at the promised level. One can hope that before such an endgame results, the country will muster the political will to undertake the institutional reforms needed for a fundamental intergenerational reallocation of resources. The design of such reforms, a topic on which Paul Demeny’s writings have made notable contributions (e.g., Demeny 1986; see also Aoki and Vaithianathan 2009), remains a major task for public policy.

References


POLICY AND PROGRAM:
CHINA
One of the main puzzles of modern population and social history is why, among all countries confronting rapid population growth in the second half of the twentieth century, China chose to adopt an extreme measure of birth control known as the one-child policy. A related question is why such a policy, acknowledged to have many undesirable consequences, has been retained for so long, even beyond the period of time anticipated by its creators.

With the world’s population growth rate now at half its historical peak level and with nearly half of the world’s population living in countries with fertility below replacement level, we can look back at the role politics played in formulating, implementing, and reformulating policies aimed at slowing population growth (Demeny and McNicoll 2006; Robinson and Ross 2007; Demeny 2011). In this context, an examination of China’s unprecedented government intervention in reproduction offers valuable lessons in appreciating the role of politics in the global effort of birth control in the twentieth century.

Aside from the rise and fall of Communism, family planning programs along with the Green Revolution could be considered two of the most consequential social experiments of the twentieth century. These two experiments differ, however, in both content and approach. The Green Revolution was aimed at feeding the population, while family planning programs were designed to curtail its growth. The Green Revolution was technological, economic, and global, while family planning programs were social, political, and often country specific.

Nowhere in the world did politics and policies figure more prominently in the effort to control population growth than in China. The policy of al-
lowing all couples to have only one child finds no equal in the world and it may be one of the most draconian examples of government social engineering ever seen. In this essay, we cast China’s one-child policy in the changing global context of population policymaking, we revisit the supposed necessity of such a policy by examining the claim that the policy was responsible for preventing 400 million births, and we discuss the reasons such a policy, with all its known negative consequences, has been allowed to stay in place for more than thirty years since its inception.

Entry: Population and politics

The world in the second half of the twentieth century saw the fastest rate of population growth ever in human history. While the main driver of this growth, the unprecedented lengthening of human life expectancy, had its seeds in the previous centuries, the momentum of population growth by the early twentieth century was spreading from industrialized countries in Europe and North America to the rest of the world, especially Asia and Latin America, where the majority of the world’s population resided (Reher 2011).

Rapid population growth created challenges in many societies and led to a global movement of birth control, soon with the active involvement of national governments. This movement can be dated to the early 1950s, shortly after the end of World War II, with events such as adoption of the first national population policy in India, the establishment of the International Planned Parenthood Federation (IPPF), and the founding of the Population Council with the support of the Rockefeller Foundation (Sinding 2007).

Two views of birth control programs emerged: the earlier one focusing on providing contraceptives for voluntary use without joining forces with governments in order to avoid interfering with individual reproduction choice; the later one legitimating and promoting the role of government in restricting individual reproductive freedom for the sake of controlling population growth (ibid.).

By the mid-1960s, the population control movement had gained new allies, with more and more international organizations identifying population growth as a hindrance to economic growth and a potential risk to political stability in the world’s less developed areas. Funding increased from both governments of the developed world and private foundations. The United Nations established its Fund for Population Activities (now the UN Population Fund) in 1969. The first UN-organized World Population Conference was held in 1974 in Bucharest. Although the conference ended in some controversy, it represented a “great step toward better national solutions and ultimately toward a better world” (Demeny 1984, p. 359).

China was an example of the contradiction between rhetoric and action evident in Bucharest. At the conference, the Chinese delegation denounced
the call for population control from the West as part of an imperialist agenda and claimed that population is not a determinant of national economic growth and well-being (Demeny 1985). At home, however, a nationwide birth control program was already well underway. Ten years before the Bucharest conference, China had established the Birth Planning Commission within the State Council to lead its birth control efforts, with a long-term goal of reducing the population growth rate to 1 percent by 2000 (Yang, Liang, and Zhang 2001, p. 34). (The growth rate in 1962 was 2.7 percent.) In talking with the American journalist Edgar Snow in the mid-1960s and later, Mao Zedong complained that few rural women were practicing contraception. In 1971, a Chinese State Council document made reducing the population growth rate to 1 percent in cities and 1.5 percent in rural areas by 1975 part of China’s fourth Five-Year Plan. By the time of the Bucharest conference, fertility in urban China, at 1.98, was already below replacement. A year after the conference, another goal was set to reduce the rural population growth rate to 1 percent and the urban rate to 0.6 percent (Liang 2007, p. 116). The national growth rate in 1975 was 1.57 percent.

China was not the only country where rhetoric did not match reality. Eight years before the Bucharest conference, India’s Ministry of Public Health had been renamed the Ministry of Public Health and Family Planning, and a special department on family planning was formed within the ministry. Two years after the conference, at which an Indian delegate famously argued that “development is the best contraceptive,” a National Population Policy statement proclaimed that sharply reducing fertility was a “top national priority and commitment.” India’s notorious forced sterilization campaign was carried out between 1975 and 1976, during which more than 8 million sterilizations were performed (Harkavy and Roy 2007; Visaria 2002). The campaign led to the collapse of Indira Gandhi’s government in January 1977 and resulted in new approaches to birth control. In 1977, family planning in India was renamed family welfare, and sterilization targets were sharply reduced. By the mid-1990s, all targets related to birth control had been removed (Harkavy and Roy 2007).

In a number of other countries with large populations and rapid growth rates, such as Bangladesh and Indonesia, coercive methods of a lesser degree were also used (Sinding 2007, p. 8). All around the world, governments in countries experiencing rapid population growth recognized the challenges they faced and considered the level of fertility in their country too high. Between 1976 and 1996, the number of governments viewing their population growth rate as too high increased from 55 to 87 (UN 2010).

By launching a nationwide policy of limiting each couple to only one child, China, of course, established an unprecedented level of government control of births. What forces and political contexts led China to adopt the one-child policy? Moreover, why did opposition to intrusive measures of birth
control lead to a quick reversal in policy in India, while in China the one-child policy continues to the present? Three features of the Chinese political system and associated policymaking help explain the Chinese exception: a statist tradition and bureaucratic institutions that treat individual citizens as subjects of the state; a post-revolutionary regime that places birth control at the heart of its political legitimacy; and a policymaking process that is carried out among political elites shrouded with secrecy and lacking public scrutiny. Combined, these three features of the Chinese political system help explain not only the design but also the survival of the one-child policy.

The one-child policy was launched at a time when China’s political leadership was reestablishing its political legitimacy following the death of Mao Zedong in 1976. In contrast to Mao and his generation of leaders, who based their political legitimacy on armed revolutions that led to the founding of the People’s Republic, the post-Mao leadership could no longer count on such a legacy. The devastating aftermath of the Cultural Revolution at the time of Mao’s death presented the greatest crisis for the Chinese Communist Party (CCP) in its nearly three-decade rule. The post-Mao leadership consequently made economic development the new mandate and the fundamental basis of its political legitimacy. The measure used to gauge its success in increasing living standards of the population was per capita GDP growth. With this new mandate and focus, controlling population growth became a critical component of this legitimacy. In 1982, birth control was announced as a “basic state policy” (jiben guoce).

Historical studies of the design of the one-child policy reveal a fluid, complex, and at times confused process involving many actors. The core calculation leading to the policy began soon after Mao’s death (Sun 1987; White 2006; Greenhalgh and Winckler 2005; Scharping 2003; Liang 2007, 2010; Greenhalgh 2008). As part of the modernization drive, a goal was proposed in September 1977, of reducing China’s annual population growth rate to below 1 percent within three years. The goal, formally stated in the government’s Working Report in February 1978 (Yang, Liang, and Zhang 2001, p. 62), was not achieved until 1998. The planning mentality, treating population simply as a number that can be planned and regulated, was clearly evident.

The idea of one-child families, as a means to achieving the goal of reducing the birth rate quickly, came from political leaders as early as 1978, two years before the policy was announced nationally. In June 1978, the policy measure of “encouraging couples to have one child, at most two” was formally stated in a CCP Central Committee document, to be followed by concrete measures adopted in over ten provinces in the same year. These measures included rewarding parents having only one child and punishing those with two or more (Liang 2007, 119). By January 1979, official Chinese news outlets, the People’s Daily and Xinhua News Agency, reported and printed editorials that clearly called for one child per couple, with “at most two.”
They also reconfirmed the measures already proposed in some provinces, namely to reward couples pledging to have only one child and to punish those with two or more. In June 1979, the headline in the *People’s Daily* report of a speech by Chen Muhua, then head of the National Family Planning Leading Group, read: “To move the focus of birth control work to having one child is the best approach.” The leaders who lent the strongest support to this goal, not surprisingly, were those in charge of economic planning.4

A coherent set of demographic goals emerged by February 1980 in support of the economic developmental goals advanced by the Chinese leadership. These goals included a total population of 1.2 billion by 2000 and a population growth rate of zero, supported by concrete measures of restricting 95 percent of urban couples and 90 percent of rural couples to having only one child. The 1.2 billion population size figure, while still unclear how calculated, clearly conformed to the overall goal set by Deng Xiaoping of quadrupling China’s per capita income between 1980 and 2000, to $1,000. Given that a per capita income goal was already set and the growth of the total economy (the numerator for calculating the per capita GDP) could be projected at a certain rate, determining the denominator, namely the total population, should not be difficult.

A strategy of GDP growth has remained the chief goal at all levels of the Chinese political system, and population has been made a lifeless number, not an aggregation of individual lives, in this pursuit. Such a pursuit is further reinforced by the CCP’s cadre performance evaluation system that measures government officials’ performance with numeric indexes, such as GDP per capita or government revenue per capita.5 This per capita metric, using population size simply as part of the political calculation for achieving political goals, has been a constant driving force to justify the continuation of the one-child policy.

With an overall political goal set and with population size entered into the calculation of achieving that goal, the policymaking process led directly to the promulgation and continued existence of the one-child policy. Here, an extremely important policy justified by the goal of maintaining Communist Party legitimacy was left in the hands of elite political leaders, whose decisions few could challenge. Dissent and reservations notwithstanding, Party elders were instrumental in implementing the policy. Seizing on these signals for an even stricter birth control policy than those implemented by other governments in the 1970s, Chinese scholars, especially those with updated knowledge and skills in the natural sciences, offered more definitive calculations and population projections under different fertility scenarios, and pushed for the formalization of the one-child policy (Greenhalgh 2008). Both the temporal sequence of policy formulation and the prevailing style of decisionmaking of the CCP, however, suggest clearly that the idea of the one-child policy came from leaders within the Party, not from scientists who offered evidence to
support it (Liang 2007). But the work of the scientists nevertheless provided what the leaders wanted: population numbers that suggested a doomed future for China if extreme measures were not taken. Moreover, these scientists provided fertility numbers to achieve the goal the planners wished, under a cloak of scientism that evinced authority, confidence, and elegance (Greenhalgh 2008). The fears of neo-Malthusian crisis increased, fanned not only by the scientists but also by academia and the state-controlled media (Lee and Wang 1999).

The one-child policy strangely was never issued as a government order nor written explicitly in Chinese law, but rather was announced in the form of an Open Letter to members of the Chinese Communist Party and the Communist Youth League, in the language of “advocating” such a policy. Worries about the consequence of the one-child policy, such as population aging and lack of old-age support, sex imbalance, and future labor shortages, were acknowledged in the Open Letter. Demographers argued that there were alternative ways of reaching the population target of 1.2 billion by 2000 (Liang 1983; Bongaarts and Greenhalgh 1985). Facing resistance both within and outside the CCP, China’s legislative body, the National People’s Congress, resisted suggestions for enacting a national law requiring one child per couple. The Open Letter contained both recognition of the sacrifices that Chinese families would have to endure and a promise to change the policy in 25–30 years. As for the difficult task of enforcing the policy, it was left to each province to translate the policy into concrete requirements and to specify the exemptions allowed (Gu et al. 2007).

Legacies: What contributions has the one-child policy made?

At the time the one-child policy was announced to the Chinese public, the designers of the policy anticipated a faster pace of population aging as a result and acknowledged the policy’s deleterious effect on the kinship structure of Chinese families. Over the course of the policy, unanticipated consequences also emerged, such as physical, often violent abuses, sex-selective abortion, and, more recently, a risk of a depressed fertility level that could lead to irreversible population aging, labor shortage, and economic slowdown and stagnation (Wang 2011). For its critics, the one-child policy is a blunt violation of basic human rights and an intrusion into individuals’ and families’ reproductive freedom.

Supporters and sympathizers, even some of the critics, however, have accepted the policy in the belief that it has played a positive role in containing global population growth and in saving the world from unbearable population pressure. A common claim is that the policy has prevented some 400 million Chinese births. This claim originated with the Chinese government,
which credits the policy for undergirding China’s economic miracle. In the Party’s view, the fertility reduction helped create a population age structure favorable for economic growth. The government has even suggested that its population control effort was a major contribution to the global effort to address the negative consequences of climate change. No one, however, has seriously examined the validity of this claim.

Where did the claim of 400 million births prevented come from, and how credible is it? The original calculation of the number of the births prevented came from a project sponsored by China’s National Population and Family Planning Commission, completed in the late 1990s, with a goal to evaluate the inputs and the benefits of China’s birth control program (Yang, Chen, and Wei 2000). The authors used a simplistic extrapolation method, projecting what the crude birth rate would have been by 1998 had China’s birth rate followed the trajectory of decline between 1950 and 1970. Their projected crude birth rate for China was 29.7 per thousand by 1990 and 28.4 per thousand by 1998. Given that the observed birth rate after 1970 fell far below these rates, the authors calculated the number of births attributed to the difference between the predicted and the observed birth rates as the number of births “averted.” For the period 1970 to 1998, the number derived was 338 million. In the decade after the publication of this number, the number of births averted was raised to 400 million.

In stating that the one-child policy averted 400 million births, the promoters of the policy first misinterpreted the original results from the study mentioned above. The number of births averted was for the period since 1970, not from 1980, when the one-child policy was formally implemented nationwide. This mistake is crucial because most of China’s fertility transition was completed during the decade of the 1970s—that is, before China’s one-child policy was enacted. Within that decade, China’s total fertility rate dropped by more than half, from 5.8 in 1970 to 2.8 in 1979. Most of the births averted, if any, were due to the rapid fertility decline of that decade, not to the one-child policy that came afterward.

Even the number of 338 million births averted appears to be a result of over-generous assumptions, not reality. To examine what China’s birth rate or fertility would have been in the absence of the one-child policy, we use data from countries that had a birth rate similar to China’s in 1970 to compare the trajectory of change in these countries with that of China. There were 16 countries in 1970 with a population of at least one million that had a birth rate between 30 and 38 per thousand, 12 of which had a level that was higher than China’s at the time, 33.4 per thousand (Table 1). The average of these 16 countries was 35.6 per thousand. The average birth rate in this group of countries declined to 26.6 per thousand by 1990 and 22.0 by 1998, both significantly below the levels predicted by the authors who arrived at the “births averted” numbers. This comparison shows that in other countries
without a one-child policy the birth rate also declined, and it declined below the level predicted for China.

To further examine the claim that China’s one-child policy was necessary to lower fertility, we also explore a counterfactual: what would China’s fertility be if there were no one-child policy? We use the Bayesian model developed by Alkema et al. (2011), the same used for the latest United Nations population projections (United Nations 2011). The Bayesian model projects China’s future fertility scenarios based on: 1) the fertility trend in China before launching the one-child policy and 2) fertility trends in all other countries.

Fertility in China would have continued to decline if the country’s rapid fertility decline in the 1970s offers any hint about the country’s future fertility trajectory. According to the Bayesian model, the decline would have continued after 1980. By 2010, fertility would have fallen to its currently observed level of around 1.5 children per woman.

Moreover, the counterfactual scenario shows that the fertility level observed for the early 1980s was somewhat higher than what would be expected from the empirical Bayesian model. This seems to be counterintui-

---

**TABLE 1** Assumptions and comparison of birth rates, China and 16 other countries, 1970, 1990, and 1998

<table>
<thead>
<tr>
<th>Country</th>
<th>Births per 1,000 population</th>
<th>1970</th>
<th>1990</th>
<th>1998</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Korea</td>
<td></td>
<td>31.2</td>
<td>15.4</td>
<td>13.8</td>
</tr>
<tr>
<td>Costa Rica</td>
<td></td>
<td>32.8</td>
<td>27.0</td>
<td>21.2</td>
</tr>
<tr>
<td>Albania</td>
<td></td>
<td>33.0</td>
<td>24.3</td>
<td>18.6</td>
</tr>
<tr>
<td>Lebanon</td>
<td></td>
<td>33.0</td>
<td>25.7</td>
<td>21.9</td>
</tr>
<tr>
<td>Jamaica</td>
<td></td>
<td>34.4</td>
<td>25.2</td>
<td>22.2</td>
</tr>
<tr>
<td>North Korea</td>
<td></td>
<td>34.6</td>
<td>20.6</td>
<td>17.9</td>
</tr>
<tr>
<td>Brazil</td>
<td></td>
<td>35.0</td>
<td>24.2</td>
<td>21.4</td>
</tr>
<tr>
<td>Uzbekistan</td>
<td></td>
<td>36.5</td>
<td>33.7</td>
<td>23.4</td>
</tr>
<tr>
<td>Malaysia</td>
<td></td>
<td>36.6</td>
<td>30.4</td>
<td>24.3</td>
</tr>
<tr>
<td>Thailand</td>
<td></td>
<td>37.0</td>
<td>19.7</td>
<td>16.4</td>
</tr>
<tr>
<td>Venezuela</td>
<td></td>
<td>37.1</td>
<td>28.6</td>
<td>23.9</td>
</tr>
<tr>
<td>Turkmenistan</td>
<td></td>
<td>37.2</td>
<td>34.7</td>
<td>24.6</td>
</tr>
<tr>
<td>Paraguay</td>
<td></td>
<td>37.4</td>
<td>33.5</td>
<td>29.0</td>
</tr>
<tr>
<td>Panama</td>
<td></td>
<td>37.5</td>
<td>26.1</td>
<td>24.0</td>
</tr>
<tr>
<td>Colombia</td>
<td></td>
<td>37.6</td>
<td>27.2</td>
<td>23.8</td>
</tr>
<tr>
<td>South Africa</td>
<td></td>
<td>37.8</td>
<td>29.3</td>
<td>25.1</td>
</tr>
<tr>
<td>Average of 16 countries</td>
<td></td>
<td>35.6</td>
<td>26.6</td>
<td>22.0</td>
</tr>
<tr>
<td>Government projection for China</td>
<td></td>
<td>32.9</td>
<td>29.7</td>
<td>28.4</td>
</tr>
<tr>
<td>China observed</td>
<td></td>
<td>33.4</td>
<td>21.1</td>
<td>15.6</td>
</tr>
</tbody>
</table>

itive, but it might be explained by the anxiety caused by policy uncertainty, which prompted many couples to have children earlier than they would have otherwise. This tempo effect is shown in the declines in ages at first marriage and first childbearing in the 1980s. In addition, the policy overemphasized meeting short-term (period) targets (e.g., reducing the population growth rate to 5 per thousand by 1985) and overlooked the tempo and cohort effects in reducing population momentum. Calculations of a universal two-child policy with spacing, as advocated by Liang Zhongtang, have demonstrated that such an approach would not only have achieved China’s population control goal, but would also have produced more favorable social and demographic conditions (Gu and Wang 2009). If we apply the same empirical Bayesian model to contemporary China, lifting the one-child policy is unlikely to have a major impact on population size. Experience from other countries shows that raising fertility is probably an even more challenging task than reducing it (Frejka, Jones, and Sardon 2010; Goldstein, Sobotka, and Jasilioniene 2009).

If the one-child policy did not achieve its goal of preventing hundreds of millions of births in China, the question then becomes why is it widely criticized as a cause for many of China’s short- and long-term social problems, and why have so many both in China and abroad urged that it be phased out. The one-child policy is criticized not just because of the demographic uncertainty it has caused. More fundamentally, it is seen as an ill-conceived policy that has prevented Chinese individuals and families from having the number of children they desire. It is a policy that has forcefully altered family and kinship for many Chinese, has contributed to an unbalanced sex ratio at birth, and has produced effects that will be felt for generations, with its burden falling disproportionately on those many couples who were forced to have only one child. In other words, at the same time as the demographic effect of the one-child policy in controlling population growth has been exaggerated, its long-term impact on Chinese society has been underappreciated.

A particularly long-lasting consequence is the hundreds of millions of families with only one child. Although, with declining fertility, some families would opt for one child voluntarily even if the government had not imposed the draconian restriction, there is little reason to believe that the proportion of families with only one child in China would ever have reached the level we currently observe. To quantify the social effect of the one-child policy, we estimate the proportion of Chinese families with only one child. We restrict our calculation to women aged 35 to 44 in 2005. We focus on this group for two reasons: first, these women spent their prime reproductive years under the one-child policy; second, by age 35 and over, they have mostly completed their childbearing.8

Among this group of women, numbering just over 110 million, who were born between 1961 and 1970 and by now are all aged over 40, a large proportion will complete their childbearing with no more than one child. In
2005, nationally, over 40 percent of them had only one child. The proportion of one-child women varies widely by province, from a high of nearly 80 percent in the urban centers of Shanghai and Beijing, to a low of less than 20 percent in the western provinces of Guizhou, Yunnan, and Tibet. In 11 of China’s 31 mainland provincial-level administrative regions, over half of women aged 35–44 had only one child. If we treat each married woman as representing a single-couple family, then roughly 50 million families in this age group will have only one child.

The share of families with only one child is even more pronounced in China’s cities, where the one-child policy is most strictly enforced (Gu et al. 2007). Couples with non-agricultural household registration status (hukou) not only face a stiff fine for violating the one-child rule, but also risk losing employment and being unable to register their children for health care and education. Among women in our sample, roughly 18 percent have non-agricultural household status. Among these urban families, the proportion of one-child families is 86 percent. In ten Chinese provincial regions, this share exceeds 90 percent, and in the three provincial-level urban centers, Beijing, Shanghai, and Tianjin, and in Liaoning province, the share is over 95 percent. Among urban Chinese families of the 1960s birth cohort, who will begin to reach age 60 in less than a decade from now, a virtually universal one-child family world is already a reality.

In contrast to the demographic effects of the one-child policy in reducing population growth, which can at best be very small, the policy’s social effect in depriving a large proportion of Chinese families of a second child is enormous. A simple calculation illustrates the magnitude of this impact: allowing 15 million families in this age group to have a second child results in 1.5 million more births per year over a ten-year period and could virtually eliminate all one-child families in this age group of urban families. Applying the figure of 15 million to the whole country, the share of one-child families nationwide in this age group would drop from 42 percent to 28 percent. The figure of 15 million is trivial in comparison to the total number of births over a ten-year period in China. Assuming most of the women aged 35–44 in 2005 had their children during the years 1985 to 2000, and given that the annual average number of births for this period was about 22 million, the additional 1.5 million births amounts to only 6.8 percent more births, a number that is inconsequential for total births for each year or for population growth, but extremely consequential for a large number of Chinese families.

The policy’s long-lasting social consequences will be judged by the future lives of China’s one-child generation and their parents. China today has about 150 million families with only one child, accounting for more than a third of all families. Among the post-one-child generation of parents, the share is even higher. This is the population structure with which China will have to confront the challenge of rapid population aging at the societal level.
Exit: People vs. population

In less than half a century, the alarm that unprecedented population growth caused in many counties in the mid and late twentieth century has turned to concerns over population decline and aging. Moreover, the twentieth century saw not only population growth but also extraordinary growth in material wealth (Lam 2011). While, in parts of the world, especially sub-Saharan Africa, rapid population growth still poses serious challenges to poverty reduction and economic prosperity, more and more countries are now concerned with sustained low fertility and its consequences. The number of countries where the government considered its fertility level too low had increased from 16 in 1976 to 47 in 2006. Among 48 countries with a total fertility rate below 1.8 children per woman in 2009, over 70 percent considered their fertility too low (United Nations 2010). Countries in Europe and Asia have begun incentive programs to encourage couples to have more children (Demeny 2003; Frejka, Jones, and Sardon 2010; Demeny 2011). The latest entry is Iran, where total fertility dropped from 4.8 in 1990 to 1.9 in 2006 (Roudi 2012).

China today is also economically, demographically, and socially vastly different from 35 years ago, when the one-child policy was officially initiated. The goal of quadrupling per capita income from $250 to $1,000 by the year 2000, seemingly wildly ambitious when first set by Deng Xiaoping in 1980, was achieved in roughly half the time. Today, with per capita GDP at over $8,000, China has joined the ranks of the high-middle-income countries. China’s economy is second only to the United States in total size, and the question for the future is not whether but when China will become the world’s largest economy. While China’s economic growth exceeded the goal set at the start of the post-Mao era by a wide margin, the goal of not permitting the population to exceed 1.2 billion was not achieved, even with the draconian one-child policy. China’s 2000 census reported a total population of 1.26 billion. However, in light of the much faster economic growth, the extra 60 million should be of little concern.

Even with the better-than-expected economic growth, however, Chinese policymakers have not revised their original population goals to allow couples to have more children. Unlike other governments around the world, under which declines in fertility level have led to revisions and reversals of their population policy, in China the process of policy review and revision has been surprisingly long and slow. Almost 35 years after it was ushered in as an emergency response to population growth, the one-child policy, with its many known deleterious consequences, is still in place.

Why has China been so slow in phasing out a policy that has become increasingly unpopular among the public—a policy that was initially defended by deeply flawed science? This policy was only able to survive because of those three features of the Chinese political system and policymaking process
that we outlined earlier. Moreover, the huge birth planning apparatus that was created in the process of implementing the one-child policy has become a main force of resistance to policy change. Not only does it produce and disseminate misinformation such as that the one-child policy has averted 400 million births, but it continues the practices of coercion and other heavy-handed tactics.

The question, however, is not whether, but when, the one-child policy will be phased out. Politically, the Chinese leadership has come to realize that the country can no longer continue the old development model that sacrifices long-term benefits for short-term gains, that ignores individual rights and welfare, and that elevates economic development above everything else. Socially, China’s younger generations, especially those born under the one-child policy in the 1980s and 1990s, have become the most critical segment of the population not only economically but also socially and politically. These generations are the best educated in Chinese history, are highly mobile geographically, and exert their power and voice in the sphere of social media. China’s single-child generations are no longer willing acceptors of the country’s statist orientation. While many of them are voluntarily postponing marriage and childbearing, and hence form the driving force of today’s below-replacement fertility, they also increasingly regard it as their right, not the government’s, to decide whether and when to have children. To them, population is no longer a number of aggregated subjects, but individual lives that should be respected and cherished.11

The opening up of Chinese society and the revival of population studies as an academic discipline have also produced an independent force that can confront government pressure. A decade and more of research by scholars has shown consistently that fertility remains well below the replacement level (Retherford et al. 2005; Guo and Chen 2007; Cai 2008; Morgan, Guo, and Hayford 2009), and has revealed that the one-child policy is not a main driver of China’s low fertility (Cai 2010; Gu and Wang 2009; Zheng et al. 2009). Leading demographers have formed a collective voice calling for phasing out the policy (Gu and Li 2010; Zeng 2007). Insistence on maintaining the policy has become increasingly a political liability.

History will remember China’s one-child policy as the most extreme example of state intervention in human reproduction in the modern era. History will also likely view this policy as a very costly blunder, born of the legacy of a political system that planned population numbers in the same way that it planned the production of goods. It showcases the impact of a policymaking process that, in the absence of public deliberations, transparency, debate, and accountability, can do permanent harm to the members of a society.

The one-child policy will be added to the other deadly errors in recent Chinese history, including the famine in 1959–61 caused largely by the industrialization and collectivization campaigns of the late 1950s, and the Cultural
Revolution of the late 1960s and early 1970s. While those grave mistakes both cost tens of millions of lives, the harms done were relatively short-lived and were corrected quickly afterward. The one-child policy, in contrast, will surpass them in impact by its role in creating a society with a seriously undermined family and kin structure, and a whole generation of future elderly and their children whose well-being will be seriously jeopardized.

Notes

1 There were two earlier World Population Conferences, in 1954 and 1965. Unlike its predecessors, the delegates to the 1974 meeting had the authority to commit their governments to decisions taken at the conference.


4 Two CCP leaders in charge of the economy, Chen Yun and Li Xiannian, were the most vocal in their support. Chen Yun, then Vice Chairman of the CCP, made it clear in conversations with other party officials on 1 June 1979 that clear laws and regulations needed to be formulated to require only one child per couple (Liang 2010, p. 5).

5 In November 1979, the CCP released the “Notification on the Implementation of the Cadre Evaluation System,” marking the formalization of government officials’ performance evaluation.

6 The work by Song Jian and his colleagues, who were the main scientists providing support to the one-child policy, was not brought to the attention of Chen Muhua until February 1980 (Liang 2007, p. 193; Greenhalgh 2008, pp. 269–270).

7 It took two decades, until 2001, for China’s National People’s Congress to pass a family planning law, but still “advocating,” not “requiring,” each couple to have no more than one child.

8 We use data from China’s 2005 mini-census to calculate the proportion of women with only one child. The calculation was done using a 20 percent micro sample from the 2005 mini-census, equivalent to 2 per thousand of China’s total population. The total number of women included in the calculation is 220,747. Inflating this number with the sampling ratio gives a number for the women in this age group of 110,373,500, or about 110 million.

9 This number is derived by multiplying the share of women with non-agricultural household registration status and the total number of women in this age group, which gives a number of women in cities of 19,867,230. Our figure of 15 million accounts for 75.5 percent of these women.

10 The total number of Chinese households counted in the 2010 census was 402 million.

11 A recent example is the case of Feng Jianmei, a woman in Shaanxi province, who was forced, in July 2012, to abort a seven-month fetus because the child would have been her second. Her case caused national outrage and produced a quick government apology (see Wikipedia page on Feng Jianmei: «http://en.wikipedia.org/wiki/Forced-abortion_of_Feng_Jianmei»).
References


Financial magazine Press.
Patriarchal Demographics?  
China’s Sex Ratio  
Reconsidered  

Susan Greenhalgh

Since Amartya Sen published his pathbreaking research on the “missing women” of Asia two decades ago (Sen 1989, 1990), the distorted sex ratio at birth and consequent masculinization of Asian societies have become a growing focus of scholarly and media attention. Across East, South, and even West Asia, the masculinization of sex ratios has been proceeding at a pace unprecedented in recorded history. Christophe Guilmoto has documented male-heavy infant sex ratios that range from 112 boys per 100 girls in India, Pakistan, and Vietnam, to 120–121 in China and tiny Azerbaijan and Armenia (the international average is 105; Attané and Guilmoto 2007; Guilmoto 2009). Seeking to draw attention to this relatively neglected “revolution” of “rampant demographic masculinization,” Guilmoto foresees a future in which Asian sex ratios will continue to rise over the next few decades, spreading in epidemic fashion, before eventually falling below 110 before 2050 (Hvistendahl 2011, p. 5; Guilmoto 2009). If high sex ratios are part of the world’s post-transition demographic future—they have already boosted the global sex ratio at birth from 105 to 107—how we understand them is critically important.

The prospect of a “world without women” has attracted keen interest in the news media. Journalists, of course, add human interest by enlivening the scholarly narratives with provocative words and images. For example, in a recent commentary in Newsweek, historian Niall Ferguson (2011) suggests that Ernest Hemingway’s 1927 collection of short stories, Men Without Women, provides a preview of tomorrow’s Asia. The stories feature gangsters, bullfighters, wounded soldiers, and killers—clearly what Ferguson envisions for Asian men.

Fundamental to the growing body of work on the sex ratio at birth (below, simply sex ratio) is the explicit or implicit notion of patriarchy. Although rarely defined, the term tends to be deployed as an overarching concept to signify a fundamental power differential between men and
women in which women are invariably the victims and men the unnamed perpetrators of gender wrongs.\footnote{In this literature, the term often implies not just gender inequality in social resources, but also an essential, biologically based difference between men and women. Perhaps precisely because it is not defined, its meaning assumed to be already established or just obvious, the term patriarchy (or a surrogate) has proven remarkably supple, fitting comfortably into the otherwise distinct discourses of demography, public health, anthropology, and security studies (among other fields), where it quietly structures—and limits—data collection, analysis, and interpretation. In demography, patriarchy, “based on the submission and exploitation [of women],” is a central concept in Guilmoto’s (2009) theory of sex ratio transition: the sex ratio rises in response to the patriarchal value of son preference (in combination with falling fertility and the spread of sex-selective technology) and then falls when “the patriarchal system” is undermined. Demographers Poston and Zhang (2009) see the high sex ratio as in part a product of a Confucian patriarchal tradition marked by strong son preference and female subordination. In anthropology, Barbara Miller offers the term “patriarchal demographics” to describe the social and cultural roots of uneven sex ratios among infants (Miller 2001). In security studies, Hudson and den Boer (2004) do not use the term patriarchy, but their depiction of a world of violent, testosterone-driven unmarried males who threaten social stability embeds core assumptions associated with the term. The examples are countless.\footnote{Population specialists are familiar with the troubling political effects of such loaded constructs as “fertility crisis” and “population bomb.” Yet few have reflected on the political work done by more mundane, seemingly neutral, and even politically progressive terms such as “patriarchy.” Few have considered the possibility that that term (and its surrogates) has not only reflected the world they are trying to understand, but also actively shaped the field of demographic reality, constraining their views of it, narrowing public policy responses, and perhaps even reproducing the wrongs they are trying to undo. In this essay I consider such possibilities.} Population specialists are familiar with the troubling political effects of such loaded constructs as “fertility crisis” and “population bomb.” Yet few have reflected on the political work done by more mundane, seemingly neutral, and even politically progressive terms such as “patriarchy.” Few have considered the possibility that that term (and its surrogates) has not only reflected the world they are trying to understand, but also actively shaped the field of demographic reality, constraining their views of it, narrowing public policy responses, and perhaps even reproducing the wrongs they are trying to undo. In this essay I consider such possibilities.}

Such possibilities emerge from an analytic perspective that emphasizes how scientific and political discourses on social problems are humanly shaped and, in turn, shape social and demographic reality. Described elsewhere, this “governmental” problematic sees science as situated within politics, and it views power as not only negative and repressive, but also positive and productive, giving rise to new fields, new subjects, new bodies, and more.\footnote{Stressing the power of discourses—historically specific bodies of knowledge that structure how things can be said and that produce effects—I take a close look at scientific and official state discourse to see how particular social problems are probematized—that is, framed or formulated, and intervened in in particular ways (cf. Ong and Collier 2005). As elements of discourse, scientific and of-}
ficial framings of the problem are highly consequential, helping to constitute the very political and sociodemographic field that is emerging and to shape the policies and other interventions that are adopted. By specifying particular “subject positions” for their targets (“rural girls,” “two-daughter couples,” and so on), policies also shape people’s subjectivities, or identities, and the life options available to them. My interest here then lies not in the demography or anthropology of the “surplus men,” but in how the problem of excess men is understood and acted upon by actors with the power to shape dominant societal discourses and policies toward them.

I examine how the problem of excess masculinity in China is being framed and intervened in and with what effect. Studying the problem in China, where the unmarried men are called “bare sticks” (guanggun), will allow us see how that huge nation is coping with one of the most adverse long-term social effects of its extremely rapid, essentially forced transition to low fertility. I ask four sets of questions. First, how is the problem being defined in scientific and political discourse? What assumptions about women and men do these framings make? Second, what larger historical, cultural, and political forces have shaped the formulations that have emerged? What notions of gender or patriarchy have emerged from each of these sites and been imported into the construction of the men? Third, how is the problem of the involuntary bachelors being managed (by what measures, policies, laws, and so on)? And fourth, what political work are the discursive framings and real-world measures doing? What is at stake in the making of this new field of biopolitics? I begin with a brief portrait of the “patriarchs” who are the focus of so much (or so little) attention.

The rural marriage crisis is still a very sensitive topic in China, one that remains little studied. The reality of the men’s lives, as well as official, scholarly, and popular perceptions of them, can be only dimly perceived at best. I draw here on media items, official sources, leader speeches, scientific research, and my own interviews in China to sketch the outlines of the emergent field of thought and practice surrounding the “bare sticks.” What I offer, of course, is only a partial picture and preliminary interpretation based on the limited information available at this time. My hope is that my arguments will spur rethinking of the categories used in this work as well as new research on how those categories are reshaping life on the ground.

**Rural patriarchs? A portrait of China’s bare sticks**

China is often called a patriarchal society, but the rural bachelors who dot the countryside are unlikely to see themselves in that term. Let us start with their numbers.
Some numbers

Since the introduction of the one-child policy in 1979–80, the sex ratio at birth has been steadily rising, climbing from 108.5 boys per 100 girls in 1982 to 118 in 2011. The gender gap is especially pronounced in poor, rural areas. An analysis of the ratio among children aged 0 to 4 reveals a national average of 120.2, but spatial clusters of counties in which the ratio ranges from 150 to 197.2 (Cai and Lavely 2007).

The disappearance of girls from China’s male-centered society, coupled with the rapid rise in marriage costs, has led to a growing crisis for men unable to find brides. Demographic research indicates that, of the cohorts born between 1980 and 2000—those expected to marry between 2005 and 2025—there is an excess of 22 million men, meaning that 10.4 percent of all men will fail to marry in the conventional way. Unless the sex ratio at birth declines, by mid-century over 15 percent of men will face that fate (Ebenstein and Jennings 2009). The situation is particularly severe in rural and remote areas, where men face massive out-migration of women to the cities as well as razor-thin incomes that make it impossible to meet the often heavy demands of brides’ families. China’s “surplus men” are thus overwhelmingly poor, illiterate, and rural. For these men, these numbers portend a real-world social crisis of monumental proportion.

Not “real men”

Although small numbers of urban Chinese are now opting to remain childless, marriage and fatherhood remain essential to being a “real Chinese man.” Without doubt this is especially true in the countryside (Brownell and Wasserstrom 2002). Despite the rise in divorce, marrying and perpetuating the family line remain a social imperative for men. By definition, men who do not marry and rear children cannot be “good men.” Despite rapid social change, in the countryside having a wife and at least one child continues to be essential for social and even physical survival. Wives and children are critical parts of the family labor force, children provide vital support in old age, and sons are essential to carrying on the ancestral line.

Although some men have migrated to the cities in search of work and wives, those left behind, the limited research suggests, lead lives of social and economic marginalization. In one small-scale study in Hebei, bachelors were allotted poor-quality land and housing at family division. As the last in their families to find wives, they were responsible for the support of parents. Unable to take jobs outside the village, they were often forced to work for others, leading to a loss of face. Lacking support from their families and their communities, the bachelors faced lives of severe social discrimination and economic destitution (Huang 2007).
An important survey conducted by Xi’an Jiaotong University in 2009 (below, Xi’an survey) provides the first nationwide picture of the fate of involuntary bachelors throughout rural China. The survey covered 264 villages in 28 of China’s 31 provincial-level units. The survey reveals gender imbalance to be a pervasive feature of life in rural China today. In the villages surveyed, the sex ratio was 121.5 and there were 9 involuntary bachelors per village (or 2.7 per 100 households). Those men most likely to remain unmarried were poor or disabled. Unmarried older men (the average age was 41) suffered discrimination in nearly half (48 percent) the villages. In a society in which parents are responsible for finding brides for their sons by a reasonable age, the parents suffered as well, facing not only the lack of a son and daughter-in-law to support them, but also social disapproval and feelings of shame from not fulfilling one of their most fundamental duties.

Obtaining a bride, legally or otherwise

Given the personal and familial stakes, older rural men seem to be trying every means conceivable—legal and illegal—to secure a bride or, failing that, simply a child. Informal discussions with Chinese researchers suggest that men in different areas are dealing with the bride shortage in different ways. In the border areas of the northeast and southwest, the dearth of local brides has been met by importing women from North Korea, Vietnam, and Myanmar and, more recently, also from Laos, Thailand, Cambodia, and Mongolia (Xinhua 2011c). Men in border provinces sometimes travel abroad to personally select their bride; in other cases, they work through middlemen to acquire “mail-order brides.” Both ethnographic research and some press reports suggest that many women from these countries are eager to marry men they perceive as wealthier than men in their home countries (Belanger, Lee, and Wang 2010). In poor interior provinces, interviews suggest the scarcity of marriageable women has given rise to culturally despised forms of union, including polyandrous arrangements (yiqi duofu) in which the wife of one man informally services several others.

Since the late 1980s, the urgent need of poor peasant men for brides has led to the proliferation of underground smuggling networks engaged in the long-distance purchase and sale of young girls and women (Zhuang 1993; Han and Eades 1995; Fan and Huang 1998; Chao 2005). Such trafficking appears to be on the rise. To pay the high prices for brides, poor rural men often save for years and borrow from family members. These arrangements have sometimes proven disastrous, as wily “brides” have turned to crime themselves, absconding with the bride payments (Fong 2009; He 2010). In the Xi’an survey, men in 40 percent of the villages had experienced abandonment by brides who were either stealing the brideprice or escaping forced marriages. In rural China today, even smuggling often fails to produce the needed wife.
In some poor and remote areas, men cannot find wives by any means. Some work through informal networks to adopt daughters to support them in old age (Kay Johnson, personal communication, August 2012; also Johnson 2004). Others have no marital or parental prospects at all. Scattered reports suggest that these men may live together in bachelor communities where they join forces to manage life’s problems, or form a spatially dispersed bachelor underclass (He 2010).

The term patriarchy would seem to have little relevance in this context. When masculinity is rural and poor, many men do not even have the power to secure a bride, the essential condition for social adulthood. Although few individuals in rural China are winners in the game of Chinese modernization, rural women seem to benefit from the new environment more than men, for they can migrate to cities to find jobs and look for higher-status husbands. The older rural men, far from perpetrators, would seem to be victims of China’s gender culture and population politics, forced to turn to illegal means to survive. These brutal realities of rural bachelorhood, however, are rarely acknowledged by the political center.

**Official framings**

Official framings of social problems bear close attention, for they shape public policy and, in turn, individual identities and life prospects. The official construction of the sex ratio problem in China has changed over time, as the issue has moved from party to scientific discourse.

**1979–93: An undiscussable issue**

In the 1970s, when Chinese reproduction was guided by the later-longer-fewer policy, the sex ratio was not an issue because fertility limits were high enough that couples could achieve their preferred gender composition. That would change with the advent of the one-child policy in 1979–80. Unlike the age structure of the population, which was made a main target for normalization, the makers of China’s new policy on population did not establish a normative sex ratio and make it a goal of official policy. This meant that, initially at least, the sex ratio would not be officially measured; in both literal and political senses, it did not count.

At the same time, the strong possibility that a strict one-child policy would lead to female infanticide rendered the sex ratio a politically sensitive and undiscussable topic. Throughout the long 1980s (1979–93), when female infanticide came to light and the sex ratio was beginning to rise, scholars were forbidden to do research or publish on it. Throughout the decade it would be framed in party, not scientific, discourse.
1993–2000: Rural girls and women as victims of feudal patriarchal culture

Despite the prohibition, during the 1980s some demographers began to use large-scale surveys done by the state to measure infant sex ratios. Far from some random killings of baby girls, the demographers discovered a worryingly large and growing gender gap at the national level. This demanded central-level attention. In the early 1990s, just as the party-state was winding up its last nationwide mass birth control campaign, some Beijing-based demographers took advantage of the political opening provided by the decline in fertility to replacement level, and energetically pressed then–Population Minister Peng Peiyun to bring the issue to the attention of top leaders. Their efforts bore fruit. (The results of their research were published in *PDR*; see Zeng et al. 1993.) In response, the state quietly reversed course, authorizing scholars to quietly study the sex ratio problem and how to fix it. In the early 1990s the party-state began addressing the growing gender gap. But given the political sensitivity of the problem, how would it be framed?

Initially, attending to gender issues was part of China’s response to the new emphasis within international population policy on women’s reproductive health and rights, which emerged from the UN Cairo Conference in 1994. In a world in which major transnational development agencies (UNICEF, UNFPA, and the World Bank, for example) were stressing gender equity and the “missing girls,” attending to the gender question—which in China included women’s health, gender equality, and the sex structure—was a crucial aspect of China’s emergence as a responsible member of the world community.

In the late 1990s, the gender imbalance issue was given further emphasis by a new concern about “population quality.” Since the advent of the one-child policy, China had sought to foster a new generation of healthy, well-educated “quality” single children. In the 1990s it became clear that the rapid decline in fertility was producing rapid aging and rising sex ratios at birth. A distorted age-sex structure would create havoc in state development planning and in people’s lives. An abnormal sex structure in the reproductive age group would mean some would be unable to marry—a disaster for the regime as well as for the individuals involved. Worries about these issues gave rise to a new concern about creating a “quality population,” one with a modern age and sex structure.

Given China’s adherence to Marxian ideology and its longstanding commitment to women’s liberation and male–female equality, it is not surprising that these gender issues, including the sex ratio at birth, were lumped together and fitted into pre-existing Marxian framings of the woman question. Following well-established constructions of “women’s subordination” in the official and scientific framing, the missing girls problem was attributed most basically to feudal culture (zhongnan qingnu, valuing males, devaluing
females). Men were simply the unmarked perpetrators of gender wrongs. They did not appear in official discourse.

**2000 to present: Men as threats to social stability**

As their numbers began to rise and their presence to trouble the rural landscape, around 2005 official concern started to focus for the first time on unmarried older men. If previous constructions had focused on women as victims, these newer formulations emphasized men as perpetrators of various wrongs and dangers, biologically driven to violence. Unhampered by state restrictions, Western scholars in security studies, demography, and public health began delving into the issue, creating a narrative in which growing numbers of bachelors will form a mobile army of violent males who will threaten China’s sociopolitical stability and perhaps make it more bellicose abroad. Lacking access to concrete data on how the “surplus men” were coping on the ground, the scholarly literature drew on limited sources—certain theoretical insights, historical precedents, scattered journalistic reports, and survey data on other groups—to assess the implications. In their influential book, *Bare Branches: The Security Implications of Asia’s Surplus Male Population*, security scholars Hudson and den Boer (2004) foresaw the spread of violent crime—from smuggling and prostitution to robbery, rape, and murder—and the export of violence to neighboring countries (also Edlund et al. 2007; Ebenstein and Jennings 2009). Elevated testosterone levels helped explain the tendency of unmarried males to be more violent, indeed to possess the potential to create some of the worst violence in Chinese history (Hudson and den Boer, pp. 195–196). Public health researchers and demographers warned of the potential for an HIV/AIDS epidemic of previously unimaginable scale, as the surplus men migrate to cities to have sex with commercial sex workers, risking contracting HIV and becoming a bridge population from high- to low-risk individuals (Poston and Zhang 2009; Tucker et al. 2005; Ebenstein and Jennings 2009).

The Western-scientific figure of the sex-starved, violence-prone rural bachelor may have had an interested audience in Beijing, for it accorded with the party’s own rural imaginary, at least the one that became public for a short time a few years ago. In January 2007, the Central Committee and State Council issued a report saying that the gender ratio imbalance amounted to a “hidden danger” for society that “will affect social stability,” an obsession of the ruling party. Reflecting the top leadership’s official security framing, in 2007–08 the population establishment began articulating a narrative of impending demographic crisis in which a large mass of potentially violent unmarried men constitutes a “social time-bomb.” Commissioned by the government, some university-based population centers began studying the matter, labeling the topic “surplus men and social stability.” As a report by People’s University put it, “When their basic biological demand is unable to be satis-
fied, the sex-starved male adult will become more violent” (Xinhua 2007). Although the language of the Population Commission7 is less virulent now, the involuntary bachelors continue to be seen as threats to national stability. In a mid-2012 statement, Population Minister Wang Xia expressed concern about the sex imbalance because it causes a “series of social problems,” including sex crimes and trafficking in women (Xinhua 2012).

Sources of anti-bachelor bias

If the violent tendencies of the rural male have not been scientifically or empirically established, what, besides some Western sciences, drives the animosity toward rural men? Although associations of violence and masculinity are found in many parts of the world, each society has its own variant of the image of “the violent male.” China’s is shaped by particular features of its culture, history, and politics. Here I offer a partial account of some of the elements that matter most.

Cultural histories: The “bare sticks”

The “bare sticks” were some of the most poignant figures on the social landscape of pre-Communist China. Poor and ill educated, village men who had no wife, no children, and no way to fulfill their filial duties had no place in the social order. In the eyes of most villagers, an unmarried and sonless man could never be seen as a true adult or real man (R. Watson 1986). In the early years of the PRC, such people-out-of-place largely disappeared, only to reappear in the post-Mao years.

In Chinese culture, guanggun has meant not only unattached, but also outcast and subtly or not so subtly threatening to public order. Throughout late imperial and Republican-era Chinese history, the “bare sticks” were widely disparaged and even feared. Historical research on bandits and rebels indicates that unattached men on the margins of lineage and village life, and unable to fulfill gender expectations, often engaged in petty violence and were known as “village bullies.” Sometimes they formed heterodox groupings such as rebel bands and secret societies (Ownby 1996, 2002). At other times they formed fraternal associations that involved people banding together for mutual aid and protection (J. Watson 1989). In popular lore, however, the guanggun were known as bullies, bandits, and rebels. These associations appear to linger today, perhaps encouraging framings of the men as threats to social stability.

Class prejudice: “Backward peasants”

The rural bachelors suffer not only from their maleness, but also, and perhaps more importantly, from their peasantness. If, under Mao, peasants were the
makers of the Chinese revolution, in the reform decades rural people have been positioned as “backward” in the Chinese scheme of things, obstacles to the country’s modernization and global rise. As the divide between rural and urban has widened, rural people, and especially poor, ill-educated villagers, are seen not as resources who might contribute to the nation’s goals, but as problems to be dispensed with as quickly as possible (Cohen 1993; Kelliher 1994; Gaetano and Jacka 2004; Whyte et al. 2010). Far from deserving official support, they are viewed as deeply unworthy of official consideration. For rural bachelors, gender differences have interacted with rural–urban inequalities to place them apparently beyond official care and support.

**Political constraints: An uncriticizable policy**

A third reason for the hostility toward the bachelors and the silence regarding their plight can be found in the complex politics surrounding the one-child policy, which remains in place today (with significant exceptions for second children). Despite the many social costs the policy has incurred, in official discourse it remains a sacrosanct gift of the party to the Chinese nation, responsible for accelerating modernization by averting 400 million births. Clearly, talk of social costs such as the surplus men is never welcome, for it serves as an uncomfortable reminder that the policy’s cost–benefit scoresheet is not as favorable as the party’s glowing assessment suggests. Tellingly, only after the policy had done its most important work and fertility had dropped to replacement level was the regime willing to listen to its demographers’ warnings about the sex ratio. It was only after an official determination that the rise in the sex ratio was due not only to the one-child policy, but also to a range of other features of China’s development, that relatively open discussion of the problem was allowed. Given the continued sensitivity of the one-child policy, the bachelor issue must be framed in such a way that that policy escapes culpability. Alternative framings of the men—such as “unfortunate victims of the one-child policy”—are unlikely to be politically viable.

**Party priorities: Stability above all**

Finally, the regime’s obsession with social stability makes it ever alert to threats to social harmony, especially in remote areas, and keen to smash them. During the 1980s and 1990s the party pursued economic growth at any cost, ignoring the social and cultural problems that inevitably emerged in its wake. Deeply worried about political unrest and social instability, the Hu Jintao–Wen Jiabao leadership (2003–12) redirected the party’s priorities to solving those new social problems and to containing threats to political security. Given this larger concern, any evidence that one element of society poses threats to social order is likely to be seized on and acted upon. The
apparent receptivity of the leadership to the social science research portraying the men as social threats and disease vectors makes sense in light of this regime focus.

**Interventions**

How, then, has the party-state tackled this thorny problem? A review of a wide array of measures shows a close connection between the gendered framing of the problem and its official solution.

**Helping rural women and girls**

After years of public denial, around 2000 China’s state finally began to openly acknowledge the gender imbalance problem and place it on the policy agenda. The socially oriented administration of Hu and Wen made halting the rise in the sex ratio at birth a priority and, drawing on the work of expert advisors, introduced numerous laws, policies, and programs to enhance the well-being of young girls and women (Greenhalgh and Winckler 2005; Shen 2008). Following the Marxian framing of the gender question as women’s subordination, the missing girls problem was attributed most basically to feudal culture. The solution was for the party-state to promote advanced gender-equitable culture, support women’s continued liberation, and protect women, guaranteeing their constitutional equality with men. In support of this agenda, in 1995 “male–female equality” was made one of a handful of top-priority “basic state policies” (Zhao and Qiu 2008).

Formalizing this broad approach, an important 2007 Decision on population set out a large set of educational, social, economic, and legal responses to “comprehensively address the abnormal sex ratio at birth” (Decision 2007). The population establishment has initiated a wide range of activities designed to eliminate discrimination against girls and women and improve their status in the family and society. These include a massive propaganda effort aimed at reducing son preference and promoting gender equality (1998–2000); wide-ranging programs to improve job and other opportunities for women; and a large-scale Action to Foster (or Care for) Girls designed to boost their well-being through preferential treatment for rural girl-only families that have accepted birth planning (2000 to present, with official launch in 2003). The state has worked continuously to popularize legal knowledge about the protection of the legitimate rights and interests of women and children.

In the 2000s, girl care has become a major industry. With state support, many foreign organizations (such as UNFPA, UNICEF, UNIFEM, Ford and Asia Foundations), as well as Chinese nongovernmental organizations, have flooded into the field of girl care, offering programs that treat rural girls as
victims of patriarchal culture who require care and support. To take but one example, the website of the organization All Girls Allowed calls for ending gendercide, educating abandoned orphans, rescuing trafficked children, and defending mothers («allgirlsallowed.org», accessed 29 August 2012). These activities, of course, are worthy, but they embed and perpetuate certain myths about the gender problem in the Chinese countryside, ones that seem to exclude men as figures deserving of assistance.8

“Perfecting” policies and programs

Beyond these cultural and socioeconomic approaches, the state has introduced two other measures designed in part to normalize the sex ratio among infants. First, it has greatly improved the rural social security system (including old-age pensions), hoping to discourage parents’ preference for boys to support them in old age. Second, since around 2000 it has quietly supported a two-child policy for certain couples (a change seen as conforming to the one-child-with-exceptions policy). As of late 2011, all 31 provincial-level units had adopted policies allowing couples composed of two single children to have two children of their own (Xinhua 2011b).

Criminalizing prenatal sex selection, sex-selective abortion, and human trafficking

The state has also relied on law-and-order measures to crack down on medical professionals who engage in illegal prenatal sex determination and sex-selective abortion for non-medical reasons. Since the early 1990s, all organizations and individuals have been “strictly forbidden” to perform these procedures, but these rules have been notoriously hard to enforce. In August 2011, the Population Commission, working with several other ministries, launched an eight-month nationwide campaign to reduce the incidence of these “two illegals” by revoking medical licences and meting out jail sentences. Officials were clearly not satisfied with the results, for the campaign was extended into 2012 (China Daily 2012; Xinhua 2012).

Although no official policy toward rural bachelors has been announced, the construction of these men as violent threats to public security implies harsh, authoritarian measures. Reflecting precisely that approach, the party has criminalized and actively prosecuted the trafficking in women and children. Although police efforts to eliminate smuggling networks and maimai hunyin (marriage by purchase) started as early as the 1980s, these efforts have become increasingly public and strident as the number of men unable to find brides has grown in recent years. Frequent media reports announce how many kidnapped women and children have been rescued and returned to their homes by the police (e.g., Xinhua 2011d). In July 2011, the party
announced a “people’s war” against infant traffickers, who are now targeting rural transients in the cities, who are too busy to watch their young children. Of course, there is no way to verify the numbers or to know how many of the women did not want to be “saved,” but instead willingly left their home counties or countries in an effort to escape poverty (Chao 2005). What does seem clear is that, in official discourse and practice, unmarried rural men are being treated like anti-state, quasi-criminal elements, while the party appears as the heroic rescuer of vulnerable women and children.

**A growing concern**

The distorted gender structure is a growing concern to the central government, as the number of men unable to find brides rises year by year and the social problems they face—and appear to cause—become more visible. Today the gender gap is one of “five major population problems” that the population establishment is addressing (Xinhua 2011a). Reflecting concern that the state will not achieve its goal of lowering the sex ratio to 115 by 2015, 2012 was designated the Year of Focused Management of the Sex Ratio at Birth (Wang 2012).

Despite the many resources devoted to reducing gender inequality and lowering the infant sex ratio, the results have been discouraging. Although the officially measured ratio has fallen recently—from a high of 120.6 in 2008 to 117.8 in 2011, it is not clear whether this decline is real or an artifact of measurement procedures (Xinhua 2012). In a 2007 report, the Chinese demographer Li Shuzhuo indicated that the efforts had had “a preliminary impact on the attempt to control SRB,” yet were marred by continued weaknesses in many areas (Li 2007, p. 12). Li pointed to a host of obstacles, including the persistence of regulations supporting gender inequality, weak implementation of related laws and policies, the difficulty of carrying out policies in rural China (where sex-selective abortion is conducted in secret), the absence of an efficient evaluation system, and the difficulty of raising funds for the sex ratio issue in a political environment with many other issues competing for state attention and resources. As Li’s report suggests, given the now-entrenched nature of the gender-imbalance problem, finding effective measures will be difficult indeed.

Effectiveness aside, what is striking about this large collection of measures devoted to lowering the sex ratio at birth is that, adhering to dominant party framings, all are geared to helping women and girls; none is aimed at alleviating the problems of rural men, in particular the men who cannot find brides. These efforts—which are part of a much larger package of policies, programs, and multi-year development plans for women—are critical to the legitimacy of a party-state that has made “male–female equality” part of its foundational charter and that sees its woman-work as important to its
identity as a responsible member of the world community. Men are the un-
marked, presumably advantaged, comparison group. There are no ten-year
development plans for them. This difference in official attention is rooted in
part in Chinese-Marxism, in which gender equates with women and gender
policies with helping women (and girls). Men (carriers of “ feudal patriarchal
culture” and beneficiaries of “son preference”) are positioned as the problem,
the object of party and state ire, and subject to criminal restraint. In practice
as well as in policy, China’s rural men are treated as much less worthy of hu-
manitarian care or support and concern than are women. They are, in short,
targets to be punished, not helped.

Political productivities

How, then, is the problem of rural bachelors being framed? And what is that
framing producing?

A Chinese articulation of the male marriage problem

Echoing China’s distinctive culture, politics, and history, the guanggun prob-
lem is being largely subsumed under the master problem of the “ sex ratio
imbalance,” which in turn means “women’s subordination.” Surrounded
mostly in a shroud of silence, the rural bachelors appear in public discourse
not as sympathetic figures to be helped, but primarily as threats to the socio-
political order. As for interventions, the dominant one is a cluster of cultural,
socioeconomic, and legal measures aimed at helping women become the
equal of men. Clearly, the leadership is concerned about the unmarriage-
ability of China’s older rural men, taking quiet measures (relaxing the one-
child policy, strengthening the social security system) to ease the situation
for future generations. What is striking is how few measures are directed
at helping older men resolve their marriage and fatherhood problems. Al-
though rural local officials in some areas have occasionally broken the law
to help men secure families and thereby become stable forces in the village,
at the level of the political center little is being done. The growing number
of bachelor villages suggests that the de facto policy seems to be to leave
the men to their own devices. The most visible central-level measure is to
criminalize men’s efforts to purchase brides from intermediaries, making un-
available one of the only ways open to many to secure a family and ensure
themselves a life that accords with the conventions of Chinese culture.

Not quite human

Far from “ mere discourse,” the Chinese notion of patriarchy—women as
needing state protection and support, men as inherently threatening and
requiring state restraint—has done a lot of political work. It has helped to
narrow the range of framings that are thinkable, leaving only one way to
understand and intervene in the rural bachelor problem. In principle, there
are at least four ways the problems of rural men could be framed in Chinese
official thought. Given the critical role of wives in the farm family economy,
the bachelors could be labeled “poverty households,” deserving of limited
state support. Alternatively, they could be deemed innocent victims of the
one-child policy; like couples having only one child or two girls, they could
be labeled “sacrificers for the nation” who deserve state support. Third, they
could be understood as “victims of their families’ bad planning,” undeserving
of assistance from anyone. Finally, they could be considered potential threats
to village stability to be dealt with by harsh, authoritarian measures. Each of
these would fit into contemporary party discourse. As far as I can tell, only
one—the last, and least humanitarian, framing—has been used to understand
and respond to their situation.

The current approach to the issue works to dehumanize the bachelors,
leaving them socially marginalized; economically at risk, with no one to sup-
port them in old age; politically excluded from the circle of citizens deserv-
ing care; and, for the many who will never have children, reproductively
extinguished, with no offspring and no one to carry on the family line. The
current strategy effectively treats them as not quite worthy of the category of
human being. As poor, uneducated peasant men, they are already by defini-
tion “low quality” persons who not only do not help, but may even hinder,
the nation’s modernization. The state may even secretly welcome the men’s
non-reproductivity, fearing that any offspring would further lower the over-
all quality of the population. But there is a risk involved: by using legal and
other means to prevent the men from obtaining brides, and closing off the
only route for many to marriage, the current approach may end up producing
the very group of “sex-starved violent bachelors” the state seeks to check. Put
another way, by offering the men only one subject position, one way of being
in the world—that of “violent offender”—and treating them in that way, the
policy encourages them to adopt that subjectivity and encourages others to
marginalize and discriminate against them.

Rural realities: Complicating the story

If the current framing sees the world through only one gender lens, what
aspects of rural reality remain outside the frame and thus invisible to poli-
cymakers? Although some women are indeed “vulnerable victims,” this
frame neglects the calculating women who cheat men by manipulating the
marital inequalities in the countryside to their advantage, as well as those
who are willing and eager to be transported long distances to marry. As for
the men, do the bachelors threaten village security and social stability? Until
very recently, the claims that they do have been largely speculative, based on other types of data and historical analogies that are problematic at best. The 2009 Xi’an survey provides some empirical answers to these questions. Compared to married men, involuntary bachelors were seen by village cadres as more likely to be irritable (31.8 percent). Perhaps because, with no families, they had time on their hands, they were seen as loafers (45.6 percent) and gamblers (35.6 percent). In a very small number of villages they reportedly engaged in commercial sex (6.7 percent), harassed women (5.0 percent), and damaged others’ marriages (4.7 percent). If such individual threats to public security and welfare were rare, group security incidents were even rarer. The proportions of villages experiencing group trouble-making, stealing, and fighting were 10.6, 8.4, and 7.8 percent, respectively. As the authors state, “Most village cadres report that problems caused by bachelors do not exist or are not serious.” Only 6.3 percent indicated that the bachelor problems are “very serious,” while another 15.5 percent indicated that they are “somewhat serious” in their villages. The authors conclude: “Thus, the problems caused by gender imbalance and the marriage squeeze at present appear not to be too serious...so there is still an opportunity for the government and society to take positive measures to prevent further deterioration” (Jin et al. 2012, p. 26). In other words, at least so far, the threat narrative is overwrought. Although it is impossible to know what the future will bring, I am suggesting that the state can influence the future by how it chooses to frame this growing problem. By reframing the bachelors as poverty households deprived of a family labor force or by labeling their plight an “unfortunate social cost of the one-child policy,” the state would have the political grounds to provide the kinds of support to the bachelors that would discourage them from turning to violent or other anti-social means to meet their needs.

The need for new constructs

This discussion, though limited by a lack of data, suggests the need to rethink the core concepts used in the study of the sex ratio at birth and subject them, including their underlying assumptions about the genders, to empirical test. Such concepts must reflect social life in countryside. When scholarly research and public policy rely on the notion that rural men are violent and threatening, most of the data gathered on them and made available to policymakers seem to relate to their violent behaviors. But the Chinese bachelors have other attributes, including admirable traits that might contribute to village welfare. In the Xi’an survey, village cadres viewed the involuntary bachelors as being helpful (40.4 percent), thrifty (52.5 percent), and, above all, filial (62.2 percent). While many scholars and government officials have railed against the perceived negative behaviors of the “bare sticks,” few have (at
least openly) considered the possible benefits involuntary bachelors might offer their villages—and society as a whole. But they should do so—and the Xi'an project may have been designed precisely to begin changing the discourse in Beijing.

This discussion also suggests the need to rethink the larger field of research and policy on the sex ratio and male marriage. Certainly, the notion of patriarchy captures the reality of sharp male–female disparities throughout Asia and, indeed, much of the rest of the world. In the literature on the sex ratio, the term patriarchy has served as a useful consciousness-raising device, pointing to a problem that deserves much more attention. But, left undefined and unattached to a robust theory of gender that recognized gender’s already classed, raced, and sexed nature, the term patriarchy has done some not so helpful work, work of analytic, political, and even ideological sorts. Too-easy use of the term has limited understanding and data collection, constricted vision, and narrowed policy choices. Failure to carefully define and delimit patriarchy has allowed ugly biases about certain men, certain ethnic groups, and certain classes to enter the field of discourse unchallenged. In the policy arena, constructions of men as inherently violent and outside the field of humanity both imply and, in China, have encouraged, the application of harsh, authoritarian measures that may be counterproductive, fostering the very violence they set out to contain. The constructs population specialists use matter because, among other things, they inform state perception and public policy. My aim here has been not to present a definitive interpretation, but to point to conceptual lapses and unexpectedly worrying social and political productivities in hopes that others will take up the intellectual and political challenges mapped out above.

Notes

1 In some other demographic research on gender, the term patriarchy is carefully defined and theoretically contextualized.

2 They extend into the media as well. Media treatments of the looming excess of men rarely mention the word patriarchy, but its implicit assumptions of female victimization and biologically rooted male aggression infuse popular discourse on the topic. In his Newsweek essay, for example, Ferguson writes:

   It may be that the coming generation of Asian men without women will find harmless outlets for their inevitable frustrations.... But I doubt it. Either this bachelor generation will be a source of domestic instability, whether Brazilian-style crime or Arab-style revolution—or, as happened in Europe, they and their testosterone will be exported. There’s already enough shrill nationalism in Asia as it is. Don’t be surprised if, in the next generation, it takes the form of macho militarism and even imperialism. Lock up your daughters. (Ferguson 2011)

   Although such language is extreme, this passage illustrates with particular clarity the ease with which patriarchal assumptions can travel between the academy and the wider culture.

3 See Greenhalgh 2008; Greenhalgh and Winckler 2005.

4 In 2000, roughly 4 percent of men aged 40 had never married, while 27 percent of
those with the lowest level of schooling were not married (Wang and Mason 2004; Tucker et al. 2005).

5 Although the use of village cadres as respondents (rather than the bachelors themselves) limits the value of the findings—cadres’ own political interests may well have colored their responses—the data are valuable nonetheless.

6 The Western science work may have had an influence on the 2007–08 framing; this possibility requires further research.

7 Formally, the National Population and Family Planning Commission of China.

8 Urban bachelors are better off in the mating game, having access to internet dating sites and even billboards to advertise for brides.

9 Such a logic about the appeal to the state of bachelor non-reproductivity is consistent with calls over the years for well-educated urban professionals to be allowed more than one child.

References


Poston, Dudley L. and Li Zhang. 2009. “China’s unbalanced sex ratio at birth: How many surplus boys have been born in China since the 1980s?,” in Tucker and Poston (eds.), pp. 57–70.


POLICY AND PROGRAM:  
AFRICA
Fertility Transition: Is sub-Saharan Africa Different?

John Bongaarts
John Casterline

Fertility in sub-Saharan Africa (“Africa”) stood at 5.1 births per woman in 2005–10 (United Nations 2011), more than double the replacement level. This high fertility combined with declining mortality has resulted in rapid population growth—2.5 percent per year—and the UN projects the sub-Saharan population to grow from 0.86 billion in 2010 to 1.96 billion in 2050 and 3.36 billion in 2100. Such unprecedented expansion of human numbers creates a range of social, economic, and environmental challenges and makes it more difficult for the continent to raise living standards. Hence the growing interest in demographic trends in Africa among policymakers.

According to conventional demographic theory, high fertility in the early stages of the demographic transition is the consequence of high desired family size. Couples want many children to assist with family enterprises such as farming and for security in old age. In addition, high child mortality leads parents to have additional children to protect against loss or to replace losses. Fertility decline occurs once rising levels of urbanization and education, changes in the economy, and declining mortality lead parents to desire a smaller number of births. To implement these desires, parents rely on contraception or abortion, and family planning programs in many countries accelerate their adoption (Notestein 1945; Easterlin 1975, 1978; Lee and Bulatao 1983).

This theory is widely accepted as a broad outline of the forces that shape fertility transitions and is consistent with much empirical evidence (Bryant 2007). As countries develop, fertility generally falls and there is a strong inverse correlation between development indicators and fertility in contemporary societies (with Africa characterized by relatively low levels of social and economic development and, accordingly, high fertility). The theory is not without its critics, however, who claim that fertility change can be brought about by ideational change and the diffusion of ideas (Bongaarts and Watkins...
Instead of contributing directly to this longstanding debate, our aim is to add to our understanding of the transition from high to low fertility by examining whether countries or regions differ in their patterns of reproductive behavior at a given stage of the transition. In particular we compare recent patterns in Africa—with most countries at early-transition (or even pre-transition) stage—with patterns in other regions. A motivation for this analysis is to assess the validity of the arguments for what might be termed “African exceptionalism” that have been made by John Caldwell and others (Caldwell and Caldwell 1987, 1988; Caldwell, Orubuloye, and Caldwell 1992; Bledsoe, Banja, and Hill 1998).

We begin by reviewing levels and trends in overall fertility and in the age pattern of fertility by region. Next we examine evidence on family size preferences and their role in shaping reproductive behavior. The last section discusses trends in the demand for and use of contraception. The conclusion comments on policy options and the potential role of family planning programs.

Throughout this analysis we rely on data from Demographic and Health Surveys (DHS) and World Fertility Surveys (WFS). Surveys from all major regions are examined, with surveys in Latin America and in Asia and North Africa providing comparison with the African experience to date. The selection of surveys (and therefore countries) varies according to the purposes of the analysis, in particular whether we are examining onset of transition or mid-transition and whether we are examining levels or trends.

**Fertility levels and trends**

Figure 1 plots trends in the total fertility rates (TFR) of Africa, Asia, and Latin America from 1950 to 2010. In the early 1950s fertility levels were high and nearly stable in all three regions, suggesting all these societies were pre-transitional, that is, there was little deliberate effort to reduce fertility through the use of contraception or abortion. It is notable that the pre-transitional TFR was significantly higher in Africa (6.5) than in the other regions (5.8). This indicates that the pre-transition reproductive regime in Africa was more pronatalist (consistent with pre-transition regional differences in fertility desires, as we document below).

Fertility decline began in the mid-1960s in Latin America and slightly later in Asia. In both regions decline proceeded steadily and rapidly. The trends in these two regions followed roughly similar paths, reaching 2.3 births per woman in 2005–10. In contrast the fertility transition in Africa did not begin until the late 1980s, and the TFR in 2005–10 (5.1) was still more than double the level observed in the other two regions.
The trends in Figure 1 also suggest that the recent pace of fertility decline in Africa is slower than the pace observed in Asia and Latin America in the 1970s. In the 1970s Asia and Latin America were at approximately the same early to middle stages of the transition, as Africa is today. To examine this issue more closely, Figure 2 plots the pace (i.e., the annual change) in the TFR by country. For Asia and Latin America the pace estimates for the 1970s are calculated from UN estimates of the TFR for 1970–75 and 1975–80 (United Nations 2011). For sub-Saharan Africa the pace estimates are the rate of change between the two most recent DHS surveys. Note that in many countries the most recent survey did not figure into the most recent round of UN estimates and projections released in early 2011. Figure 2 excludes countries that are pre-transitional (i.e., TFR at the beginning of the historical interval has not yet declined 10 percent below the maximum value in the UN estimates for 1950–2000).

The results show substantial variation among countries within regions, which is not unexpected: the pace of decline is known to vary considerably across countries (Casterline 2001). The most important finding in Figure 2 is that the median pace of change in sub-Saharan Africa (0.03 per year) is less than one-third the pace in the other regions (0.12 and 0.13, respectively). In all but one African country, the pace of decline is lower than the median pace in Asia and Latin America. To be sure, some of the faster African declines resemble the slower declines in other regions (especially Latin America). More striking is the fact that the estimated TFR was unchanged or rose between the two most recent surveys in 9 of the 20 African countries where fertility had already fallen by at least 10 percent. Some of these apparent stalls may prove to be artifacts of measurement error, as has been common during the
Is sub-Saharan Africa Different?

past decade (Machiyama 2010). But without doubt these African countries are experiencing a pace of decline that is slow in comparative perspective. A clear exception in Africa is Rwanda (the outlier in the right-hand panel of Figure 2), which experienced extremely rapid fertility decline between 2005 and 2010 (we discuss Rwanda in more detail below). Clearly, Africa’s transitions are on average much slower than past transitions elsewhere in the developing world.

A new type of fertility transition?

A widely cited study by Caldwell, Orubuloye, and Caldwell (1992) argues that the fertility decline in sub-Saharan Africa represents “a new type of transition” (p. 211). This African pattern “will be one of similar fertility declines across all age groups” (p. 237), which would be in contrast to “the greater declines among older women that characterized the non-African transitions” (p. 237). In other words, the African pattern would be different because it would “not be dominated by changes among older women” (p. 220).

We begin an examination of this provocative argument by comparing average age patterns of sub-Saharan Africa’s fertility with those of other regions in the developing world (Figure 3). This figure is limited to surveys in which the TFR ranged between 4.5 and 5.5 at the time of the survey; this controls for variation in the age pattern that occurs over the course of the transition.

Although regional differences are not large, the age pattern of Africa is clearly flatter than the corresponding age patterns in Asia and Latin America. That is, compared with other regions, African fertility is slightly lower at intermediate ages and slightly higher at younger and older ages.
Differences in age patterns are related to differences in the duration of inter-birth intervals. According to Caldwell, Orubuloye, and Caldwell (1992) long birth intervals are a key feature of African childbearing: “Traditionally, long periods of postpartum sexual abstinence were probably standard over most of sub-Saharan Africa, and the practice is still adhered to in much of West, Middle, and Southern Africa” (p. 217).

NOTE: Sample is one survey per country, TFR closest to 5.0. Age-specific rates adjusted to sum to TFR = 5.0. n = 38 countries.
To examine this issue further, Figure 4 plots the median duration of birth intervals at birth order two for Africa and the other two major regions, as estimated from surveys yielding moderately high levels of fertility (TFR near 5.0). Most of the median intervals in the African countries range between 30 and 40 months, centered around 35 months (almost three years), whereas

**FIGURE 5 Age pattern of fertility decline, by TFR**

Sub-Saharan Africa

Latin America

Asia/North Africa

*NOTE: n = 49 countries and 92 inter-survey periods; minimum TFR decline = 0.4 births; minimum inter-survey period = 4 years.*
the median intervals in the other two regions range from 20 to 30 months and are centered around 25–26 months (a bit over two years). Thus, birth intervals in Africa tend to be roughly one year longer than in other regions at this level of fertility (which can be regarded as early transition). These regional differences in birth-interval duration are consistent with the age patterns of fertility observed in Figure 3: longer intervals are associated with flatter patterns. An important inference from Figure 4 is that the potential for fertility decline due to lengthening of inter-birth intervals is less in Africa at present than it was in the other two regions at the equivalent stage of decline.

These findings confirm a key difference in childbearing patterns between Africa and non-African countries. But Figures 3 and 4 shed no direct light on Caldwell’s hypothesis that fertility declines at older ages in Africa are less dominant contributors to the overall decline than elsewhere in the developing world. An assessment of this issue requires the measurement of fertility declines at different ages. In particular we compare the decline in TFR at ages 30+ with the overall decline in the TFR between successive surveys. The results from this exercise are plotted in Figure 5, which shows the ratio of the percentage decline among women ages 30+ to the percentage decline in the TFR. (The percentages are plotted against TFR, to control for stage of fertility decline.) According to Caldwell’s hypothesis this ratio should be near 1 in African countries (indicating similar declines at all ages) and above 1 in countries in Asia and Latin America. In the event, the data in Figure 5 do not confirm this; the majority of median ratios are above 1 in Africa and Latin America and near 1 in Asia at TFRs above 4.5. We therefore do not find support for Caldwell’s hypothesis that African declines are distinctive in this respect.

Fertility preferences

Fertility preferences represent a key link in the chain of causation between fertility and its socioeconomic determinants. In low-income agricultural societies, parents tend to want relatively large numbers of children while in higher-income societies with more developed secondary and tertiary sectors, women typically want only about two children. The decline in preferences that accompanies development in turn leads to a decline in actual fertility with adoption of birth control (induced abortion and/or contraception).

A standard indicator of lifetime fertility goals is the ideal family size (IFS). In the DHS women are asked, “If you could go back to the time you did not have any children and could choose exactly the number of children to have in your whole life, how many would that be?” Estimates based on this type of question are widely used because they are straightforward to interpret. There are, however, two potential sources of bias: nonresponse and rationalization (Bongaarts 1990). The fraction of women who report a non-numeric response (e.g., up to God) was substantial in some countries in
the past but in general has declined over time; in a majority of DHS surveys it is now less than 5 percent. The stated ideal number of children can also be subject to “rationalization bias” when a woman gives an ideal family size that is inflated because she is reluctant to provide a number smaller than her current number of living children. Rationalization and nonresponse are typically higher among older women who have been married many years and have large numbers of living children. To minimize these sources of bias, we present IFS estimates for women who are in their first ten years of marriage.

To describe regional trends in IFS over time, we first examine preference levels at the beginning of the transition. Figure 6 plots IFS as reported in surveys with TFRs above 5.5. While differences between countries are substantial, there are clear and marked inter-regional differences, with the IFS higher in Africa (above 5 in almost all countries, and above 6 in a majority of countries) than in Asia or Latin America, where the IFS exceeds 5 in just a few countries and is below 4 in one-half the countries (combining both regions). In addition, while not shown explicitly in Figure 6, looking separately at countries in East and Southern Africa and countries in Central and West Africa reveals that the IFS is especially high in the latter sub-region.

Trends in IFS are plotted in Figure 7. Each country is represented by one line connecting IFS at the first and last available survey. The two estimates of IFS are cross-classified by TFR (x-axis). Several conclusions can be drawn. First, over time the IFS and the TFR decline in almost all countries and, as expected, these two indicators are strongly correlated. Second, at any given level of the TFR, the IFS is higher in Africa than in the other two regions. As countries proceed through the fertility transition, however, the African trends in IFS tend to converge on those of Asia and Latin America. Third,

---

**FIGURE 6**  Ideal number of children at onset of transition among women married <10 years

![Graph showing ideal number of children at onset of transition among women married <10 years](image)

Note: Earliest survey, TFR >5.5. n = 37 countries.
despite substantial declines in fertility desires in Africa, the most recent levels of IFS (i.e., the endpoints of the lines in Figure 7) are still relatively high, with a median of 4.6 children per woman. This level is slightly higher than the median IFS of 4.5 for Asia at the beginning of the transition (see Figure 6). These findings suggest that to this point in the fertility transition Africa remains distinctive in its pronatalism, a finding consistent with past literature (e.g., Caldwell and Caldwell 1988).

**Preference implementation**

In order for declining fertility preferences to have their intended effect of reducing actual fertility, sexually active women must use contraception and/or abortion to prevent unplanned pregnancies. Throughout the developing world a substantial proportion of women who do not want to get pregnant are not using contraception. The reason for this “unmet need” for contraception is that women encounter multiple obstacles to the use of contraception (Bongaarts and Bruce 1995; Cleland et al. 2006; Casterline and Sinding 2000; Casterline, Sathar, and Haque 2001; Westoff and Bankole 1995). These include lack of knowledge of contraceptive methods and sources of supply; low quality and limited availability of family planning services; cost of method, services, travel, and time; health concerns and side effects; objections from husbands or other family members; and concerns about moral and social acceptability.

These obstacles lead some women to forgo contraceptive use despite their wish not to get pregnant. As a result, each year about 80 million unintended pregnancies occur in the developing world, which amount to 40 percent of all pregnancies. These unintended pregnancies end in abortions (40
million), unintended births (30 million), or miscarriages (10 million), with detrimental health and economic effects for many women and their families (Singh and Darroch 2012). Unintended births can be further divided into mistimed and unwanted births; the former occur before women reach their desired family size and the latter occur after.

The DHS surveys have developed a standard procedure for estimating the level of unmet need in a population and, indirectly, the total demand for contraception. The latter has two components: demand = current use + unmet need. All elements in this equation are proportions of women currently in union. The proportion currently using contraception is obtained directly from responses to straightforward survey items. Unmet need is the proportion of in-union women who are fecund and do not wish to become pregnant soon but are not using contraception. It is estimated through a complicated algorithm that has been refined over time (Westoff and Ochoa 1991; Bradley et al. 2012). Estimates of demand, current use, and unmet need include both women who state that they want no children in the near term (“spacers”) and women who have reached their desired family size (“limiters”).

We begin our analysis of levels and trends in demand, use, and unmet need by examining patterns in a set of African countries (Figure 8). The time series for demand and unmet need are available only for dates after 1990 because earlier surveys did not collect comparable data. Three key patterns are evident in Figure 8:

1) Demand rises over time in most countries, as is expected from declines in ideal family size, but the pace of increase is typically not rapid and in some countries demand has stalled since the mid-1990s (e.g., Ghana, Kenya, Niger, Senegal, Zimbabwe).

2) Contraceptive use generally follows a trajectory that is parallel to demand, but at a lower level.

3) The trend in unmet need (the difference between demand and use) differs among countries, increasing in some and declining in others. Where unmet need has hardly changed, this is a consequence of parallel trends in demand and use. Even when use is rising rapidly (e.g., in Namibia), unmet need can remain unchanged because demand is rising at the same pace (which increases the fraction of women at risk of unmet need).

There are also notable differences between sub-regions of Africa. Countries in East and Southern Africa are presented in the upper portion of Figure 8, and countries in Central and West Africa in the lower portion. Demand and use are generally higher in East and Southern than in Central and West Africa, which is consistent with the generally lower levels of fertility and IFS in the former. In addition, in most countries in East and Southern Africa use is increasing and unmet need is falling, and often the two lines cross in the period under observation. In contrast, in the countries of West and Central Africa, the slope of the lines is rather similar and unmet need generally exceeds use.
FIGURE 8  Historical trends in contraceptive demand and use

Ethiopia

Kenya

Madagascar

Malawi

Namibia

Rwanda

Tanzania

Uganda

Zambia

Zimbabwe

Burkina Faso

Ghana

Mali

Niger

Nigeria

Senegal
Rwanda shows the most unusual pattern of change in Figure 8. Before 2005 trends were essentially flat but between 2005 and 2010 demand rose, contraceptive use more than doubled, and unmet need declined by nearly half. A plausible explanation for this rapid increase in use is the impact of a family planning program that has been reinvigorated since the mid-2000s.

FIGURE 9  Trends in unmet need for contraception, by region

NOTE: Sample is onset of fertility decline as of earliest survey. n = 27 countries.
The accompanying rise in demand is probably attributable to a combination of two factors:

1) Rapid socioeconomic change (a booming economy, rising school enrollment, and declining infant and child mortality) since the early 2000s, following a period of political instability and civil war in the 1990s.

2) The information and education messages from Rwanda’s family planning program. In addition to making contraceptive services more widely available, the program has a well-funded and active country-wide media campaign that provides information about the benefits of contraception and smaller families. Such campaigns can have a substantial impact on reproductive behavior and preferences (Bongaarts 2011; Cleland et al. 2006).

Stepping back and considering regional patterns, Figure 9 plots trends in unmet need for contraception by TFR for each of the three major regions. This figure is limited to experience after the onset of fertility decline. Each country is represented by one line connecting unmet need estimates. As was evident in Figure 8, the African results show a mixed pattern, with a few countries experiencing periods of little or no change and others showing decline. The dominant pattern is a decline in unmet need in countries in which fertility is declining. In this respect Africa resembles the other two major regions, which are characterized by downward trends in most countries and periods.

A further important feature of Figure 9 is that unmet need on average is higher in Africa than in Asia and Latin America. These two non-African regions are more developed and further advanced in their fertility transitions. As a result, women have more knowledge about and access to contraception, and they are more motivated to avoid unplanned pregnancies.

Figure 9 does not provide a sufficient basis for assessing regional differences in the level of unmet need controlling for level of fertility. The overlap between the regions is greatest around a TFR of 4.5 (see vertical line), but even at this level of fertility there are too few observations in each of the three regions from which to draw firm conclusions about either levels or trends. It is too early, therefore, to determine whether unmet need behaves differently in Africa over the course of the fertility transition.

Conclusion

Our analysis reveals several important differences in reproductive behavior between African and non-African countries. Most obviously, current fertility is substantially higher in Africa than elsewhere. Africa is still in the early stages of the fertility transition, while Asia and Latin America are approaching replacement fertility. This difference is not surprising in view of Africa’s lower level of social and economic development and weaker family planning programs.

But an examination of trends and patterns in fertility, family size preferences, and preference implementation reveals other key differences. First,
the recent pace of fertility decline in Africa is substantially slower than the pace of decline in Asia and Latin America during the 1970s. In fact, in several African countries the fertility transition appears to have stalled with a TFR near 5. This is a highly unusual event because non-African countries typically experienced an acceleration of the pace of decline in the early stages of their transitions during the 1970s. Second, birth intervals are longer in Africa. The main reason is the well-documented postpartum behaviors (abstinence and breastfeeding) that make African periods of postpartum infecundability longer than elsewhere. Third, ideal family size is higher in Africa. This is expected in part because Africa is still early in its transition, but in fact African societies show a higher ideal family size controlling for stage of the transition. These patterns of ideal family size suggest distinctive pronatalist features of African societies, as argued by Caldwell and his collaborators (1987, 1988). Interestingly, despite its high ideal family size Africa also has a higher unmet need for contraception than other regions. This combination of high ideal family size and high unmet need is the consequence of a low level of preference implementation (i.e., only a relatively small proportion of women who do not want to get pregnant are using contraception), which in turn is the result of multiple obstacles women face in attempting to control their fertility.

We are unable, however, to confirm Caldwell’s hypothesis that sub-Saharan Africa is experiencing a new type of transition in which declines in fertility in the early stages of transition are not due disproportionately to declining age-specific rates at older ages (as is thought to have been the case in Asia and Latin America). Our examination of changes in age-specific fertility rates between successive surveys suggests that Africa does not differ from the other regions—and Asia in particular—in the manner posited by Caldwell.

Several policy implications follow from these findings. The high unmet need for contraception in Africa is indicative of frustrated demand for contraception. The solution to this problem is the implementation of family planning programs. Commitments to family planning in several countries (Ethiopia, Madagascar, Malawi, Namibia, Rwanda) have had the expected result of rising use and declining unmet need. Resource constraints faced by African governments are being addressed with new funding (amounting to $4.6 billion) pledged at the London Family Planning summit in London in July 2012.

Even in the unlikely event that all unmet need could be eliminated, however, Africa’s fertility would remain substantially above contemporary Asian or Latin American levels. The reason is Africa’s high ideal family size, which is clearly an obstacle to rapid fertility decline. In fact, it is one of the main reasons why the current pace of fertility decline is so slow. The conventional view on how to reduce preferences is to invest in social and economic development. There is no doubt that such investments would have a fertility-reducing effect, but this process is likely to take many decades, during which rapid population growth would continue. Fortunately, family planning programs can also bring about changes in preferences through information
campaigns that present evidence on the health and socioeconomic benefits of contraception and smaller families. Such messages are particularly effective when they have the support of political leaders.

The most important step required to make progress in addressing high and unwanted childbearing and rapid population growth is for policymakers in Africa to realize that the current demographic trajectory is a major obstacle to their countries’ development. National and international attention to demographic and family planning issues has increased in recent years, following a period of neglect in the 1990s and early 2000s. These concerns have moved higher on the global development agenda because fertility decline provides not only important health benefits but also economic benefits that improve the lives of women and children and their communities.

Notes

This research was supported in part by grant R24-HD058484 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development awarded to the Ohio State University Institute for Population Research.

1 Age-specific fertility rates within the group of countries with TFRs between 4.5 and 5.5 are adjusted so that the average TFR for each region equals 5.

2 Includes pregnancies following contraceptive failure.

3 Surveys are included if the TFR is 90 percent or less of the country’s maximum TFR as estimated by the United Nations (UN 2011).

4 Note that Figure 8 includes pre-transition countries, which explains the lesser dominance of decline in unmet need in that figure.

References


The Recent Fertility Transition in Rwanda

CHARLES F. WESTOFF

Rwanda is a small, poor country in Central Africa surrounded by the Democratic Republic of Congo, Uganda, Tanzania, and Burundi. With an estimated population of about 11 million, the country has the highest population density in the African continent (415 persons per square kilometer). In addition, at the beginning of the twenty-first century it had one of the world’s highest population growth rates. However, during the first decade of this century, this country, commonly remembered for the genocidal massacres in 1994, has experienced a very rapid demographic transition.

The government first became concerned about rapid population growth in 1982 when a National Office of Population was created to confront the problem and to promote family planning. In 1990, a formal policy with similar objectives was adopted, but progress came to a halt with the 1994 genocide in which an estimated 800,000 to one million persons were killed. Following this crisis, the promotion of birth control was unpopular for several years (BBC News 2007) because of the enormous loss of life during the mid-1990s. At the time it “seemed difficult to believe that overpopulation would ever be a problem” (Kinzer 2007). The genocide and the associated collapse of the health system reduced the total population from 7.1 million in 1990 to 5.6 million in 1995.

This trend reversed quickly, however, and by 2000 the population had reached 8 million with a growth rate of 3 percent per year. As a result, concern about rapid population expansion returned and a new national population policy was issued in 2003. This policy was strongly endorsed by the country’s female legislators, “who have spent years pushing for a serious population control effort” (Kinzer 2007). In the late 1990s, the reconstruction of a devastated health system resumed and by 2001 the Ministry of Health had “implemented a national network of mutuelles, innovative community-based health insurance schemes supported by member premiums and government funding.” Although family planning was not a primary focus, it grew in importance over time (Madsen 2011).
Recent Fertility Transition in Rwanda

This essay first examines the very rapid changes in reproductive behavior and preferences that have occurred over the past decade. Next, trends in socioeconomic characteristics and child mortality are discussed to elucidate these reproductive trends. The concluding section reviews government policies and program developments that appear to underlie these population changes.

The demographic changes described below are based on national Demographic and Health Surveys (DHS) conducted in Rwanda in 1992, 2000, 2005, and 2010. Occasional reference will be made to an Interim survey in 2007–08, which first documented the striking developments that were subsequently confirmed in the 2010 DHS (Ayer and Hong 2009).

Reproductive behavior

Fertility

Figure 1 plots the total fertility rate (TFR) in the three years before surveys conducted from 1992 to 2010. Between 1992 and 2000 the TFR declined from 6.2 to 5.8 before rebounding to 6.1 in 2005. The temporary decline in the late 1990s can plausibly be attributed to the disturbances caused by the civil war in the mid-1990s (Schindler and Brück 2011). The fertility effect of the crisis was moderated by the continued high fertility in refugee camps (Verwimp and Van Bavel 2005).

Between 2005 and 2010, the total fertility rate dropped from 6.1 to 4.6, a 25 percent decline. A more current estimate of fertility trends uses data for the 12 months (instead of 36 months for the standard TFR) preceding the

FIGURE 1 Trends in fertility and contraceptive use, 1992–2010

SOURCE: DHS surveys.
surveys; this measure dropped 27 percent for all women and 33 percent for married women between 2005 and 2010.

Family planning

The recent decline in fertility is primarily the result of an expansion of the use of birth control; changes in proportions of women married and in the age at first birth, on the other hand, have been small.

The use of contraception among currently married women rose sharply from 17 percent in 2005 to 52 percent in 2010 (see Figure 1). This new level is among the highest in sub-Saharan Africa, placing Rwanda in a group of countries that includes Botswana, Lesotho, Malawi, Namibia, South Africa, and Swaziland. The primary method of contraception used in Rwanda is the injectable, which accounts for about half of all use.

A multivariate analysis of the covariates of contraceptive use among married women shows strong associations with education and wealth and with having heard radio messages about the benefits of family planning. There seems to be little impact of religious affiliation, despite the opposition of the Catholic Church to controlling births (Wadhams 2010). The proportion of Catholics using modern contraceptive methods is actually higher than among Protestants (49 vs. 41 percent).

Abortion is legal only on very restricted grounds. As a result, the abortion rate is low (25 abortions per 1,000 women of reproductive age) and has only a small impact on fertility (Basinga et al. 2012).

Reproductive preferences and their implementation

The number of children that Rwandan women consider ideal has declined from 4.9 to 3.3 over the past decade, with the most rapid change since 2005. Declines have occurred at all ages (Figure 2), but for young women under 30 the drop is much greater than among those over 30. The ideal number of children among cohorts of women also shows a sharp decline (not shown here). For example, the cohort born in 1981–85 desired an average of 5.1 children in 2000 (at ages 15–19), 4.1 in 2005 (at ages 20–24), and 3.2 in 2010 (at ages 25–29).

These results are for all women. Confining the comparisons to currently married women does not change the picture. However, the pattern for never-married women shows an even more radical drop. Among never-married women ages 15–19, the average number of children desired declined from 5.1 in 2000 to 4.3 in 2005 and 2.7 by 2010. Large reductions are also evident for never-married women aged 20–34.
An alternative measure of preferences is the desire to stop childbearing. Figure 3 presents trends in proportions of women and men who want no more children (or who are sterilized) by their current number of living children. The shapes of the two sets of curves are similar and indicate a strong increase over time in the intention to terminate childbearing. These trends are fully consistent with the observed declines in desired family size.

As expected, the downward trend in ideal family size is accompanied by a similar trend in the wanted total fertility rate (see Table 1). Between 2000 and 2010 the wanted TFR declined from 4.7 to 3.1 births per woman. Unwanted
fertility remains substantial, which suggests a potential for further declines in actual fertility when women control their fertility more fully.

Preference implementation

The decline in desired family size has led to a rise in the demand for contraception to avoid unplanned pregnancies. Figure 4 plots trends in the demand for contraception as measured by the proportion of women at risk of pregnancy who do not want to get pregnant. Between 2000 and 2010 this demand

<table>
<thead>
<tr>
<th>Year</th>
<th>Wanted total fertility rate</th>
<th>Total fertility rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>4.7</td>
<td>5.8</td>
</tr>
<tr>
<td>2005</td>
<td>4.6</td>
<td>6.1</td>
</tr>
<tr>
<td>2007–08</td>
<td>3.7</td>
<td>5.5</td>
</tr>
<tr>
<td>2010</td>
<td>3.1</td>
<td>4.6</td>
</tr>
</tbody>
</table>

The total wanted fertility rate is calculated in the same way as the standard total fertility rate, except that the numerator excludes unwanted births.

FIGURE 4  Trends in the demand for family planning, unmet need, and the percentage of demand satisfied, currently married women in Rwanda, 2000, 2005, and 2010

NOTE: Demand is the sum of unmet need and contraceptive use.
rose from 49 to 76 percent of exposed women. Among these women who do not want to get pregnant, some are not using contraception, resulting in a so-called unmet need for contraception (the second set of bars). Unmet need was high in 2000 and 2005 (36 and 38 percent) but dropped to 24 percent in 2010, as a result of a sharp rise in contraceptive use. The last set of columns in Figure 4 shows the percent of the demand for contraception that is satisfied (i.e., the percent of women at risk of pregnancy who are using contraception). It rose from 26 to 69 percent between 2000 and 2010. This indicates an increase in the implementation of preferences as actual use of contraception catches up with demand.

**Socioeconomic change**

The trends in reproductive behavior and preferences described above are in part a response to social and economic changes underway in Rwanda. One of the most important of these changes has occurred in education. Between 2005 and 2010 the proportion of women with no formal schooling dropped from 23 to 15 percent, while school attendance above the primary level increased from 10 to 16 percent. A similar picture appears for men, with the percentage with no schooling declining from 17 to 10 percent and the percentage with more than primary school increasing from 11 to 20 percent. Improvements in education levels are even larger among younger men and women.

Social change is also driven by increased exposure to mass media, particularly the radio. Between 2000 and 2010 proportions reporting at least weekly exposure to radio rose from 39 to 68 percent among women and from 63 to 87 percent among men. Exposure to television is still low but rising, reaching 9 percent for women and 24 percent for men in 2010 (when only 10 percent of households had electricity). The spread of the mobile phone—now owned by 40 percent of households—has also become an important source of social interaction and change.

Education and mass media exposure have been shown in many countries to influence contraceptive behavior and the number of children desired (Westoff and Koffman 2011). In Rwanda in 2010, the odds of using a contraceptive method increased strongly with increasing education, while the main influence of radio seems to lie in having heard messages about the advantages of family planning.

Rapid economic growth in recent years has undoubtedly also contributed to reproductive change. During the civil war of the mid-1990s gross domestic product declined, but after 1996 Rwanda experienced a steady economic recovery. Since 2002, the economy has grown at around 4 percent annually, reaching 8 percent by 2010. Nonetheless, about half of the population still lives in extreme poverty (Kinzer 2007).
A final factor with a likely impact on reproductive behavior is the recent decline in child mortality. The mortality of children under age five dropped from 196 per 1,000 live births in 2000 to 152 in 2005 (back to pre-war levels) and then to 76 in 2010. These are substantial declines, especially the 50 percent drop between 2005 and 2010. The 2010 DHS report attributes this decline to “the implementation of integrated management of childhood illness in health facilities and communities and also the introduction of new vaccines” (National Institute of Statistics 2012). As the chances of child survival rise, parents feel less need to have many children to ensure the survival of a few.

The family planning program

The new national population policy developed in 2003 embedded a reinvigorated family planning program within general social and economic development goals. The effort to reduce the birth rate was seen as essential to improving the quality of life for the Rwandan population. Initially there were sensitivities about promoting birth control after nearly a million deaths in the 1994 genocide, but the new policy was a broad-based effort to respond to a rapidly expanding population. Political opposition diminished greatly with the strong leadership of President Kagame and with the increased recognition of the positive effects of family planning in various sectors including health, environment, education, and women’s status. A 2008 review of the policy emphasized the environmental connections: “As population pressure is one of the key drivers of environmental degradation and poverty, the implementation of the population policy, especially aspects that address high fertility rates, gender and reproductive health, migration and human settlements, constitute important triggers for sustainable natural resources management is important” (sic) (Rwanda State of Environment and Outlook 2009, p. 13).

These efforts followed the publication of “Rwanda Vision 2020,” an influential report published in 2000 that set development goals for Rwanda. The report’s section on population refers to the high annual population growth rate, then close to 3 percent. It notes that family planning is crucial for reducing birth rates as well as the prevalence of HIV/AIDS. Targets included a TFR of 5.5 for 2010 and 4.5 for 2020. The 2010 target has already been surpassed with a rate of 4.6 in 2008–10. In the more recently proposed new set of targets, the TFR was set at 3.4 for 2020 (Cabinet Paper for Revised Vision 2020, p. 6).

The National Reproductive Health Priority policy of 2003 is described by the Ministry of Health as a key part of Rwanda’s commitment to the United Nations Millennium Development Goals (Ministry of Health 2006). The document asserts that “unplanned fertility fuels a rate of population growth that is outpacing economic production . . . constituting a real hindrance to the achievement of the Vision 2020 to reduce poverty.” Family planning is regarded as a key intervention; “fertility concerns will be mainstreamed into
health information and adult literacy courses in 2020”; and “the Government will ensure the availability of reproductive health services at the district level.” The policy set the objective of 15 percent use of modern contraceptive methods by 2010, an objective that was greatly surpassed by the 2010 DHS estimate of 45 percent. Another objective, only partly realized by 2010, is the provision of a full range of methods, including sterilization.

The National Family Planning Policy also calls for coordination among different government agencies and for the mobilization of Rwanda’s parliamentarians to assume leadership roles in the advocacy of family planning. The latter task is made easier by the strong female presence in the parliament (56 percent female is the highest proportion in the world). The policy planned to “launch a national targeted media campaign to increase public information about family planning” (p. 16) and to ensure that private health providers include family planning within their services. Media activities maintain the visibility of the government’s commitment to family planning: “The Rwandan leaders have the duty of mobilizing the population about family planning activities at all levels of government and civil society.” Furthermore, the Ministry of Health intends to integrate family planning into youth centers and clubs, to increase the involvement of men, and to train teachers to provide family planning education in nurseries, schools, and universities.

According to the 2010 DHS, 91 percent of current contraceptive users obtain their supplies at government facilities. This statistic confirms the key role of the government’s family planning program in providing women with access to contraceptive services. In addition, the program’s media campaigns have undoubtedly contributed to the observed decline in desired family size. The widespread promotion of family planning has sensitized men and women to the advantages of having fewer children. A study based on the Rwanda 2005 DHS (Ndaruhye, Broekhuis, and Hooimeijer 2009) demonstrated this effect: “As our findings show, the dissemination of information about family planning through personal contact at health centers contributes to a high level of desire for family limitation.”

The costs of all of these activities, including the provision of contraceptive supplies, are being met partly through an increase in health insurance and partly through support from several international donors.

How did all of this happen?

This is a question posed by Julie Solo, who interviewed around 40 key informants and visited several health centers in an effort to identify factors that contributed to the program’s success (Solo 2008). First is the fact that top governmental officials recognized the importance of limiting population growth to reduce poverty and to develop the country. President Kagame declared family planning to be a national priority, and, in the words of the Minister of
Health, “Family planning is priority number one—not just talking about it, but implementing it.” Government officials promote the small-family ideal, which is communicated on the radio and through other means of community outreach. One of the key events that led policymakers to this commitment was a 2005 presentation of the RAPID model to parliamentarians. This model, developed by the Futures Group, communicated the advantages of smaller families for the country’s economic, health, and educational goals.

Another critical factor was the emphasis on coordinating family planning efforts among numerous governmental agencies and private organizations. Family planning has been integrated into all health services including HIV/AIDS programs and immunization programs. The Ministry of Health is committed to this coordination, and a consensus has developed around these concerns among the different agencies in recent years (Thaxton 2009).

Finally, family planning services, including contraceptives, are free in Rwanda. This is in part due to the support of international donors. For example, USAID has funded contraceptive supplies and UNFPA provides support for training family planning workers.

Conclusion

Rwanda has made rapid strides toward the goal of providing universal access to reproductive health and family planning. The driving force has been the strong political determination of the country’s leaders since the early 2000s to reduce the rate of population growth to promote economic development and reduce poverty.

On the occasion of the 2011 London summit on family planning, the prime ministers of Rwanda and Ethiopia noted that “family planning makes a major contribution to improving the health of mothers and children, while also empowering women to participate more in economic productivity … and also helps to slow the high levels of population growth” (Habumuremyi and Zenowi 2012). Strong governmental leadership coupled with concerns about the economic consequences of rapid population growth has also led to substantial expansions of family planning activities in a few other sub-Saharan African countries, in particular in Ethiopia and Malawi. In Nigeria a new National Population Commission has been established following the president’s expression of concern about population growth. Many more African countries could benefit from Rwanda’s experience.
References


Fertility and Population Policy in Algeria: Discrepancies between Planning and Outcomes

ZAHIA OUADAH-BEDIDI
JACQUES VALLIN

Over the last fifty years Algeria has undergone major political, socioeconomic, and demographic upheavals. In population terms, the current situation and the concerns it raises are the very reverse of the position in the 1960s. Despite the loss of lives during the war of liberation from France in the late 1950s until independence in 1962, the first census taken in independent Algeria confirmed the extent of the country’s population growth. Between the 1954 and 1966 censuses the country’s predominantly Muslim population had risen from less than 9 million to 12 million, an average annual increase of over 2.6 percent. At that rate, a population doubles in 30 years, and Algerian researchers at the time were alarmed by a study by the Algerian Association for Demographic, Economic, and Social Research (AARDES) which indicated that this growth would accelerate and might approach 4 percent a year by 1985. This would mean a doubling of the country’s population in less than 20 years. Despite these warnings, far from introducing a policy to reduce population growth, the Algerian government adopted a discourse hostile to the notion of family planning, which was seen as imperialist in nature.

Although the census of 1977 provided some evidence for the alarmist thesis, with an intercensal average annual growth rate of 3.2 percent for 1966–77, civil registration data showed that the rate of natural increase began falling in the early 1970s. In fact, the predicted 4 percent annual rate was never reached. Although officials and decisionmakers were fairly quick to understand the economic and social challenges posed by a 3 percent population growth rate, the fact that the growth rate was significantly lower than had been feared delayed discussion of the need to introduce a population policy in Algeria. As a result, it was not until 1983 that the first national program to control population growth was enacted.
By the first decade of the twenty-first century, some two decades after the start of this program, Algeria’s rate of natural increase had fallen to 1.5 percent, despite the sharp decline in mortality since independence and the high potential for growth resulting from large cohorts reaching childbearing age. Algeria’s crude death rate had fallen from 21 per thousand in the years before the war of liberation (Negadi, Tabutin, and Vallin 1974) to 4.6 in 2001 (ONS 2004a), and the crude birth rate had fallen from 47.4 per thousand to 20.4. Further, the average number of children per woman, more than 8 in the early 1970s (Negadi and Vallin 1974), had shrunk to 2.2 in 2000–01 (ONS 2004b). Ironically, the government still feared the consequences of unsustainable population growth and remained determined to reduce the total fertility rate. For the second time, policymakers were at odds with actual demographic developments, since by this time fertility had fallen to near replacement level. The 1998 census had even shown that fertility in some Algerian cities had fallen below the replacement level as early as the mid-1990s, and the 2002 Papfam survey produced a figure of 1.4 children per woman in Algiers in 2001 (MSP et al. 2004b).

Since then Algerian fertility has risen, to a total fertility rate of 2.9 children per woman in 2010 (ONS 2011a). This unexpected turnaround only emphasizes the gap between policy and population change. In Algeria policy seems to have been applied in the wrong way at the wrong time, but this appears to be irrelevant, since policy has had so little influence on the course of events.

The decline in fertility preceding population policy

1962–1975: Government opposition to reducing rapid population growth

As revealed by civil registration statistics,1 Algerian fertility peaked after the end of the war of liberation, exceeding 8 children per woman in 1966. This figure was confirmed by the 1969–71 multi-round survey, which recorded 8.1 children per woman for 1970 (CNRES 1974b). The rise in fertility could easily be explained by family reunion after the war as well as by the euphoria of liberation and the return to peace. But since mortality was falling rapidly at the same time, the population’s rate of natural increase rose to around 3.5 percent by the mid-1960s, one of the highest rates in the world at that time. In the opinion of AARDES, this explosive growth was likely to accelerate further: Algerian demography was likened to “a sports car with no brakes” (AARDES 1968, p. 54). The need for a population policy began to be discussed as part of preparations for the first national development plan. Given the cost of educating young people and the high rates of unemployment and under-employment, AARDES concluded that “Algeria’s economy
is unable to meet the needs of all of its population and cannot form a sufficiently sound basis for proper economic development. We cannot envisage without serious concern any further rapid increase in the population growth rate” (1968, p. 52).

Despite this call to action, the government appeared to ignore the problems associated with rapid population growth. The 1966 call of the national union of Algerian women (UNFA) for the right to family planning received some support from President Boumediène, who recognized that family planning was one way of liberating women and protecting the health of mothers and children. The President approved the opening in 1967 of the first experimental center for birth spacing at Mustapha Bacha hospital in Algiers. “Birth spacing” was the object rather than “birth control,” because its justification was a matter of health not of population.

An AARDES survey in 1967–68 revealed that public attitudes toward family planning were generally favorable, especially among the young, city-dwellers, and the literate. The survey also identified some reluctance about birth control, largely the result of religious belief, suggesting that a favorable public position by religious authorities could have a significant influence on opinion and practice (AARDES 1968). Such a position was taken in 1968, when Algeria’s High Islamic Council issued a fatwa encouraging birth spacing, a radical departure from the ideology of the 1950s and 1960s that had been hardened by seven years of war against a colonial power.

In 1969, two more birth spacing centers were opened in other major cities, Oran in the west and Constantine in the east (Remili 1972). But even as awareness of the population issue appeared to be growing, President Boumediène declared later in the year, “Our aim is to ensure for our masses... a standard of living equal to that of the most developed nations in the world of the future... we do not accept false solutions such as birth control, which amounts to removing difficulties rather than finding appropriate solutions. We prefer positive, effective solutions, namely, creating jobs for adults, schools for children and better social amenities for all” (Kouaouci 1996). Meanwhile Algeria asserted on the international stage its categorical rejection of any population policy, on the grounds that development is the best contraceptive. At the 1974 World Population Conference in Bucharest, Algeria vied with China for leadership of the movement firmly opposed to the idea that reducing population growth was a precondition for development.

Nevertheless, Algerians who supported birth planning continued their efforts. Training courses for public health professionals were conducted in the United States. At the three major centers in Algerian cities, birth spacing was being integrated into mother-and-child protection, and in 1971 the Ministry of Public Health presented a proposal for a national birth spacing program to the World Health Organization. The program was officially launched in 1974, under Dr. Malika Ladjali, head of the central office for mother-and-child
protection, involving however only seven centers for the whole of Algeria (Ladjali 1985). Under the pressure of demand, this number slowly increased, although not until 1984 was even sparse national coverage achieved, with a total of 340 centers. This low-profile program, which was exclusively devoted to birth spacing for health reasons (Paulet 1994), had only limited effect on the overall increase in use of contraception.

1975–1983: Fertility begins to fall without any birth limitation program

And yet, fertility declined sharply. The TFR fell from 8.1 children per woman in 1970 (Negadi 1975) to 5.3 in 1981–85 (see Figure 1). Thus, the official position opposed to a national population policy did not prevent the start of a long downward trend. Not only did contraceptive use increase significantly more than had been expected in the unambitious national birth spacing program, but the rise in the age at marriage, unrelated to that program, was the prime factor in the fall in fertility.

In a society where contraception is hardly ever used and women are strictly kept apart from any sexual relationship outside marriage, a rise in the marriage age is automatically a factor for lower fertility. The mere fact of delaying a union by a few years reduces the duration of exposure to the risk of procreation and lowers TFR by several points. In Algeria, the 1969–71 ENSP (CNRES, 1974) and the 1986 ENAF (Kouaouci, 1992) surveys can be used to measure marital fertility by age and compare it with the general fertility rate (Table 1). From 1970 to 1981–85, marital fertility fell 19 percent as a result of contraception, but general fertility fell by more than 33 percent, which can only be explained by the rise in marriage age. The phenomenon is even more marked, naturally, for the fertility of young women: while the rise in marriage age lowered general fertility by 45 percent at age 20–24 and even by 78 percent at age 15–19, other factors (including contraception) lowered marital fertility at these ages by only 21 percent and 12 percent respectively.

After World War II, women’s mean age at first marriage fell to 18 years in the early 1950s7 from levels around two years higher in the preceding decades (see Figure 2A). However, the 1969–71 demographic survey showed that the decline in age at marriage reversed in the mid-1950s, probably as a result of the absence of young men because of emigration or the war of liberation. Thereafter, age at marriage continued to rise rapidly, to 21 in the early 1960s and nearly 24 in the early 1970s. The trend is a broad one from the 1966 to the 1987 censuses, with a higher proportion of women unmarried at all ages: increasing from 53 percent to 90 percent at ages 15–19, from 11 percent to 52 percent at ages 20–24, and from 4 percent to 22 percent at ages 25–29 (see Figure 2B). Most of this change occurred even before a 1984 law established a minimum legal age for marriage (18 years
for women and 21 for men). The law only confirmed a trend attributable mainly to the expansion of education, which lengthened time at school and facilitated women’s access to the labor market, two developments that discourage early marriage.

SOURCES:

Civil registration
1965–69: dotted line, rate by age published by the Sub-Directorate for Statistics (SDS 1968; Negadi 1975); solid line, same data corrected by the under-recording rate estimated from the 1969–71 multi-round survey.
1976–1991: raw (dotted line) and corrected (solid line) TFR published by ONS (ONS 1999).

Censuses
1966: children under one year of age and survival rate (authors’ calculation); estimate based on fertility and proportion of unmarried (Navelet and Nizard 1972).

Surveys

ONS estimates
The 1983 ideological shift is not the main reason for the accelerating drop in fertility

After much delay and debate, in 1983 the Algerian government launched an ambitious population program. As we will see, however, the program did not greatly contribute to lowering fertility.

An ambitious program to manage population growth

As we mentioned earlier, the 1974 national birth spacing program was intended not to limit births but only to enable couples to space them. Although it had no major effect on fertility, it helped promote a more positive image of modern contraception. Furthermore, by the early 1980s policymakers no longer feared negative reaction from religious authorities, since a new fatwa from the High Islamic Council in 1982 had reasserted the lawfulness of contraception, on condition that it was voluntary and reversible. In political

---

### TABLE 1  
Age-specific general and marital fertility rates during four periods: Algeria 1970–2008

<table>
<thead>
<tr>
<th>Rate and age group</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>Change (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15–19</td>
<td>109</td>
<td>24</td>
<td>5</td>
<td>9</td>
<td>−78.0</td>
</tr>
<tr>
<td>20–24</td>
<td>330</td>
<td>181</td>
<td>51</td>
<td>75</td>
<td>−45.2</td>
</tr>
<tr>
<td>25–29</td>
<td>378</td>
<td>252</td>
<td>111</td>
<td>139</td>
<td>−33.3</td>
</tr>
<tr>
<td>30–34</td>
<td>345</td>
<td>259</td>
<td>126</td>
<td>149</td>
<td>−24.9</td>
</tr>
<tr>
<td>35–39</td>
<td>272</td>
<td>218</td>
<td>102</td>
<td>118</td>
<td>−19.9</td>
</tr>
<tr>
<td>40–44</td>
<td>144</td>
<td>110</td>
<td>40</td>
<td>51</td>
<td>−23.6</td>
</tr>
<tr>
<td>45–49</td>
<td>37</td>
<td>31</td>
<td>8</td>
<td>8</td>
<td>−16.2</td>
</tr>
<tr>
<td>TFR</td>
<td>8.1</td>
<td>5.4</td>
<td>2.2</td>
<td>2.7</td>
<td>−33.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15–19</td>
<td>412.8</td>
<td>308</td>
<td>290</td>
<td>354</td>
<td>−25.4</td>
</tr>
<tr>
<td>20–24</td>
<td>455.0</td>
<td>413</td>
<td>296</td>
<td>347</td>
<td>−9.2</td>
</tr>
<tr>
<td>25–29</td>
<td>430.4</td>
<td>353</td>
<td>237</td>
<td>297</td>
<td>−18.0</td>
</tr>
<tr>
<td>30–34</td>
<td>385.0</td>
<td>312</td>
<td>183</td>
<td>239</td>
<td>−19.0</td>
</tr>
<tr>
<td>35–39</td>
<td>316.4</td>
<td>251</td>
<td>123</td>
<td>163</td>
<td>−20.7</td>
</tr>
<tr>
<td>40–44</td>
<td>176.8</td>
<td>129</td>
<td>47</td>
<td>64</td>
<td>−27.0</td>
</tr>
<tr>
<td>45–49</td>
<td>48.4</td>
<td>39</td>
<td>10</td>
<td>10</td>
<td>−19.4</td>
</tr>
<tr>
<td>TFR</td>
<td>11.1</td>
<td>9</td>
<td>5.9</td>
<td>7.4</td>
<td>−18.9</td>
</tr>
</tbody>
</table>

FIGURE 2  (A) Mean age at first marriage and (B) proportion of unmarried women by age according to various sources available since World War II

A

RGP 1998 (last 12 months)
MICS3 2006
PAPFAM 2002
RGP 2008
PAPFAM 2002 (last 12 months)
RGP 1998
RGP 1977
RGP 1948
RGP 1954
RGP 1966
RGP 1977
RGP 1987
RGP 1998
PAPCHILD 1992

B

Proportions (per 1000)

NOTE: RGP = Recensement général de la population (general population census).
SOURCES: See Table 3, column “Reference.”
terms, public opinion was ready for a change in the official position, especially following the death of President Boumediène in 1978.10

What occurred was more than a policy change, it was an ideological shift. In Algeria’s 1980–84 Five-year Plan, the population issue had been identified as crucial, since the country’s potential resources were considered inadequate for the needs generated by population growth and increased by social demands. Economic and social development was deemed to be impossible without some control of population growth (Paulet 1994). As the State Secretariat for Social Affairs wrote at the time, “Much more than in the past, demographic constraints will continue to largely dominate the conduct of economic policy. Action to actively reduce the birth rate has become an essential way of improving efficiency in constructing our economy and hoping to meet the population’s social needs in a satisfactory and sustainable fashion” (SeaS 1983). In February 1983, the government adopted the national program to control population growth (Programme national de maîtrise de la croissance démographique). In the same year, Information et education sociales, a review published by the State Secretariat for Social Affairs, reported, “The population explosion in Algeria has only just begun. Most of it has yet to come. Examination of the balance between births and deaths (natural growth) reveals an average annual increase in the resident population of 404,000 people per year in 1965–70, approximately 675,000 people per year in 1980–85 and a projection of as much as 1,114,000 per year in 2000–2005” (SeaS 1983).

 Shortly afterward, at the International Conference on Population, held in Mexico City in 1984, the Algerian delegation reversed its earlier position, even asking the international community for support for the country’s new population policy, since it considered population growth as “the main obstacle to development.” Now the line was: “The repercussions of excessive population growth, both individual and collective, on the welfare and self-fulfilment of the family, the quality of life and social environment, and the ability of the country’s economic fabric to satisfactorily meet basic social needs, urgently require widespread, concerted action to ensure the right conditions for controlling population growth” (SeaS 1983). This new position, radically opposed to the one advanced in Bucharest in 1974, was issued against an economic and social background different from that of a decade earlier. In addition to the death of President Boumediène, now criticized for his “sudden reversal” at Bucharest, the objective of full employment by 1980 had still not been met. The country was under increasing pressure from all sides: housing shortages, unemployment, problems in the health care system, and the high cost of education at schools and universities. Population growth was blamed for the inability of economic policy to meet the country’s basic needs.

The national program to control population growth (hereafter, the population program) was based on experience with the 1974 birth spacing program. Its main agent was the midwife because of her strategic position as
intermediary between the health care system (supply of contraception) and individual women (demand for contraception). Emphasis was placed on training. In 1988, a contraception module was included in the initial training for midwives, and training courses for health care staff were held into the 1990s. Priority was also given to developing primary health care infrastructure. The number of birth spacing and mother-and-child protection centers increased from 260 in 1980 to 2,500 in 1994. Contraceptives were made available free of charge. In 1985, the program was publicized with a national awareness campaign, presenting clear messages publicly endorsing family planning (Oussedik et al. 1988). However, this publicity campaign was not successful and was quickly dropped.

In the late 1980s, the population program was subjected to an evaluation, which determined that its performance was unsatisfactory. In addition to insufficient contraceptive supplies, difficulties were reported in management, information, and awareness. The effectiveness of the program was challenged (Oussedik et al.; Paulet 1994; Mokkadem and Bouisri 1998; Benkhelil 1999; Boumghar 1998). Steps to strengthen it were taken in the early 1990s: in 1991, for example, midwives were authorized to prescribe contraceptive products and methods. Following the International Conference on Population and Development, held in Cairo in 1994, the population program was extended to include reproductive health, a new concept based on international recommendations (Mokkadem and Bouisri 1998), and in 1995 a national committee for reproductive health and family planning was established (MSP 2000). These steps were taken to no avail. Within a few years the assessment was even more severe: the program should be totally revised. “Services were better structured under the birth spacing policy. Once a population policy was adopted, grassroots work became less effective” (Belkhodja, quoted by Kouaouci 1996).

In 1995, an attempt was made to strengthen and broaden the population program. Its national character was reasserted: the central government provided its organization, regulation, and funding. While the principle of voluntary family planning was preserved, emphasis was placed on improving access to family planning services integrated with reproductive health and reducing cultural and social obstacles to adoption of family planning. The creation in 1995 of the central hospital pharmacy (Pharmacie centrale des hôpitaux) was designed to secure the public sector’s supply of medical products including contraceptives. Funding for contraceptives, previously part of the social security budget, was transferred in 1996 to the health care sector, with specific budgets allocated for purchases from the central pharmacy. A budget of 161 million dinars (equivalent to US$8 million at that time) was allocated in the 1996 finance law, and oral contraceptives were exempt from tax and duty to lower their cost (MSP 1998). The government reoriented the program two years later, with population policy becoming part of a multi-
sector strategy for comprehensive sustainable development. The program now had three basic functions: development of family planning, intensification of communication, and improved capabilities for the collection and analysis of demographic data.

The accelerating fall of births had little to do with the family planning program

Despite the severe criticism of the population program, its introduction was followed almost immediately by a sharp acceleration in fertility decline. The TFR fell from 6.2 children per woman in 1985 to 3.5 in 1995 and 2.2 as early as 2000–01, a fall of 60 percent in 15 years (Figure 1). In fact, however, neither government policymakers nor researchers attributed this fall primarily to the population program. The program did help to increase the use of contraception, from 25 percent among married women of childbearing age in 1984 to 57 percent in 1995. But this figure leveled out at roughly 60 percent (Table 2). The accelerating fertility decline was only partly attributable to marital fertility control; age at marriage continued to rise and probably contributed at least as much if not more to the decline.

The rise in age at marriage accelerated in the 1980s and 1990s, from 23.7 years in the mid-1970s to 30.5 in the late 1990s (Table 3 and Figure 2A), an increase of nearly seven years in a little over two decades. Figure 2B shows that the proportion of unmarried women by age group rose most sharply between 1987 and 1998.

Contraceptive use largely contributed to lowering the TFR to near replacement level by the turn of the century. Without contraception, the decline would have required a much greater rise in age at marriage. From 1981–85 to 2000–01, marital fertility fell by 35 percent, mainly as a result of contraceptive use. But general fertility fell by nearly 60 percent. The difference clearly results from the rising age at marriage. This difference is even more striking when one examines the fertility of young women: for example, for the 20–24

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Use some method</td>
<td>8.0</td>
<td>25.0</td>
<td>36.0</td>
<td>50.7</td>
<td>56.9</td>
<td>64.3</td>
<td>57.0</td>
<td>61.4</td>
</tr>
<tr>
<td>Modern, of which</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pill</td>
<td>31.5</td>
<td>42.9</td>
<td>49.0</td>
<td>50.1</td>
<td>51.8</td>
<td>52.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IUD</td>
<td>2.1</td>
<td>2.4</td>
<td>4.1</td>
<td>4.3</td>
<td>3.1</td>
<td>2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional</td>
<td>8.0</td>
<td>4.5</td>
<td>9.7</td>
<td>7.5</td>
<td>13.9</td>
<td>5.0</td>
<td>9.4</td>
<td></td>
</tr>
<tr>
<td>Use no method</td>
<td>92.0</td>
<td>75.0</td>
<td>64.0</td>
<td>49.3</td>
<td>43.1</td>
<td>35.7</td>
<td>43.0</td>
<td>38.6</td>
</tr>
</tbody>
</table>

*The health ministry’s 2000 survey probably overestimated total contraceptive use by overestimating the use of traditional methods, where the prevalence rate doubled from the 1995 figure and fell even more sharply in the 2002 survey.

We use Bongaarts’s framework model (1978, 1993) to decompose the contribution of various factors to the decline in fertility (Table 4). In 1970, fertility in Algeria was lower than theoretical maximum fertility by 25 percent as a result of breastfeeding alone. This contribution has barely altered since then. However, the fertility-reducing effect of age at marriage rose from 27 percent in 1970 to 63 percent in 2001. The effect of contraception, only 3 percent in 1970, rose to 51 percent in 2001. In other words, before the initia-

<table>
<thead>
<tr>
<th>Date of calculation</th>
<th>Mid-point of reference period</th>
<th>Mean age</th>
<th>Source</th>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1948</td>
<td>1934.5</td>
<td>20.0</td>
<td>1948 census</td>
<td>Hajnal</td>
<td>SDS 1968</td>
</tr>
<tr>
<td>1954</td>
<td>1940.5</td>
<td>19.4</td>
<td>1954 census</td>
<td>Hajnal</td>
<td>SDS 1968</td>
</tr>
<tr>
<td>1966</td>
<td>1952.5</td>
<td>18.4</td>
<td>1966 census</td>
<td>Hajnal</td>
<td>Vallin 1973</td>
</tr>
<tr>
<td>1987</td>
<td>1973.5</td>
<td>23.7</td>
<td>1987 census</td>
<td>Hajnal</td>
<td>ONS 1989</td>
</tr>
<tr>
<td>2002</td>
<td>1988.5</td>
<td>29.4</td>
<td>Papfam 2002</td>
<td>Hajnal</td>
<td>MSP et al. 2004a</td>
</tr>
<tr>
<td>1998</td>
<td>1997.5</td>
<td>30.5</td>
<td>1998 census</td>
<td>Period first marriage probabilities</td>
<td>Authors’ calculations</td>
</tr>
<tr>
<td>2002</td>
<td>2001</td>
<td>29.4</td>
<td>Papfam 2002</td>
<td>Period first marriage probabilities</td>
<td>Authors’ calculations</td>
</tr>
</tbody>
</table>

TABLE 4  Fertility-reducing effect of contraceptive use and age at marriage as estimated by the Bongaarts framework: Algeria 1970–2001

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed fertility</td>
<td>TFR</td>
<td>8.1</td>
<td>6.1</td>
<td>4.4</td>
<td>2.2</td>
</tr>
<tr>
<td>(children per woman)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect of breastfeeding</td>
<td>C_i</td>
<td>0.75</td>
<td>0.73</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td>Effect of contraception</td>
<td>C_c</td>
<td>0.97</td>
<td>0.68</td>
<td>0.56</td>
<td>0.49</td>
</tr>
<tr>
<td>Effect of marriage</td>
<td>C_m</td>
<td>0.73</td>
<td>0.68</td>
<td>0.53</td>
<td>0.37</td>
</tr>
<tr>
<td>Estimated fertility</td>
<td>TFR^a</td>
<td>8.2</td>
<td>5.2</td>
<td>3.5</td>
<td>2.1</td>
</tr>
<tr>
<td>(children per women)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference between estimated and observed fertility</td>
<td>TFR–TFR^a</td>
<td>-0.1</td>
<td>0.9</td>
<td>0.9</td>
<td>0.1</td>
</tr>
</tbody>
</table>

^aEstimated by the model.  
tion of the population program, the contraception effect increased faster than the marriage effect, whereas after the program began, the marriage effect increased more rapidly. Of course, it was largely because its initial rate was so low that the contraception effect rose so fast in the earlier period, but this comparison still puts the specific role of the population program into perspective. Although the effect of contraception was significantly less than the effect of age at marriage, birth planning did play a major part in the steep reduction of TFR from 8.1 to 2.2 children per woman over some three decades.

Furthermore, it would be misleading to attribute the increasing use of contraception to the population program alone. Although the program encouraged contraceptive use by providing unrestricted free access to modern methods (mainly the pill) for the entire population, women and couples nevertheless had to choose to practice contraception. That choice was in large measure the result of the transformation of Algerian society brought about by three major and closely interconnected phenomena: the expansion of education, particularly for women; women’s access to the labor market; and urbanization. These are standard features of demographic transition (Joshi and David 2002) and fertility theories (de Bruijn 2002), but are particularly applicable for developing countries, including Algeria (Ouadah-Bedidi 2004; Ouadah-Bedidi and Vallin 2006). These powerful factors in modernizing demographic behavior would most likely have caused Algerian women to limit their childbearing even in the absence of the population program, just as they had begun to do before the program was introduced. This outcome can be seen, for example, in the even sharper fall in fertility in Libya over the same period in the complete absence of government action in support of family planning (Ouadah-Bedidi, Vallin, and Bouchoucha 2012). Last but not least, Algeria, like many developing countries, experienced the severe effects of the economic and social crisis of the 1980s, exacerbated by the civil war of the 1990s. The “crisis Malthusianism” (Lesthaeghe 1989) combined its effects with those of the preceding development (Ouadah-Bedidi 2004).

Thus, there is not necessarily a contradiction between official criticism of the program in the mid-1990s and the concomitant acceleration in the fall in fertility. What is more surprising is that in the early 2000s Algerian policymakers continued to believe that fertility was too high and to call for a further decline in TFR when that figure had reached the replacement level by 2001.

The rise in fertility over the past decade despite the reinforcement of the family planning program

Of the two quantitative goals set when government fertility policy changed in 1995—a TFR of 2.5 by 2005, and 2.1 by 2010—the first was nearly achieved
by 1997, a year for which the 1998 census reported a TFR of 2.7; and the second was achieved nearly ten years ahead of time, with the 2002 Papfam survey reporting 2.2 children per woman in 2001. Despite this success, in 2001 the Algerian government considered that fertility was still too high and posed an obstacle to development, nor have they changed their view since then (Table 5). In the first decade of the century, therefore, further steps were taken for reinforcing family planning. In 2001, the organization chart of the Ministry of Health and Population was revised and a report on the demographic situation and trends for 2010 was submitted to the Council of Ministers on the basis of the 1998 census results. To allow for the regional variations observed, a readjustment of the population policy’s strategies and objectives had become necessary. A process of devolution of national population programs to the local level was started. Population Committees were formed in each wilaya (province) with the goal of implementing policy according to national objectives and local conditions (MSP 2003).

In recent years the Algerian government’s perception of fertility trends may have been influenced more by such indicators as the birth rate or the rate of natural increase than by the TFR. The rate of natural increase is clearly the figure to consider at a given point in time to assess the obstacles that population growth may pose for economic development. With 1.5 percent annual growth, a population doubles in 47 years. But for that to happen, the rate must persist for at least 47 years, a circumstance that no demographer would reasonably have predicted in 2001 (Ouadah-Bedidi and Vallin 2000). On the contrary, a continued fall in fertility below the replacement rate and consequently a slowdown in population growth were much more likely occurrences. This was clearly the case in all developing countries that had already reached the replacement rate after a rapid fall in fertility. As early as the mid-1990s, for example, the TFR was 1.8 in Thailand, 1.6 in Cuba and Singapore, and 1.5 in South Korea. And in all of these countries, fertility has

| TABLE 5 Official government positions and population policies, 1976–2006 |
|-----------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| **Growth and fertility indicators** |        |       |       |       |       |       |
| Natural growth rate         | 3.1    | 2.7   | 1.7   | 1.5   | 1.7   | 2.0   |
| Total fertility rate        | 7.3    | 5.4   | 3.1   | 2.2   | 2.3   | 2.8   |
| **Official government position** |        |       |       |       |       |       |
| Population growth           | Satisfactory | Too high | Too high | Too high | Too high | Too high |
| Fertility                   | Satisfactory | Too high | Too high | Too high | Too high | Too high |
| **Policy applied**          |        |       |       |       |       |       |
| Population growth           | No action | Reduce | Reduce | Reduce | Reduce | Reduce |
| Fertility                   | No action | Reduce | Reduce | Reduce | Reduce | Reduce |
| Support for contraception   | Direct support | Direct support | Direct support | Direct support | Direct support | Direct support |

continued to fall, while other developing countries have also seen fertility fall below 2.1: Iran, United Arab Emirates, Lebanon, Chile, Brazil, Vietnam, and China. If the government continued to encourage birth planning, then, Algeria could be expected to follow the same course.

Against all expectations, the reverse happened. Not only did Algerian fertility stop falling once it reached the replacement rate, but it has continuously risen since (Figure 1). It rebounded to 2.3 children per woman in 2005 2.8 in 2007–08, and 2.9 in 2010. At the very least, this is yet another reason to reconsider the ability of family planning programs to exert any measurable effect on fertility trends by itself only. But what can explain this unexpected reversal?

As we stated above, since 1995 the contraceptive prevalence rate appears to have plateaued at 60 percent, and modern contraceptive practice barely exceeds 50 percent (Table 2). In addition, calculations based on the 2002 Pap-fam survey indicate that the mean age at marriage fell by about one year, from 30.5 to 29.4, between 1998 and 2002 (Figure 2A). Although we must await more detailed findings from later surveys, this sudden fall in marriage age may be the prime or sole cause of the rise in fertility observed since 2001.

We cannot exclude the possibility that increased marital fertility is also the result of the declining use of contraception observed in the early 2000s. The MICS-3 survey, on the other hand, shows an increase in contraceptive use between 2002 and 2006. If in fact the lower age at marriage is the only factor contributing to rising fertility, then this phenomenon might be temporary and the fall will resume when the marriage age stabilizes.

It is instructive to note that this recent rise in fertility occurred first in cities and was later and less marked in rural areas. Similarly, the rise occurred first among better-educated women and only later among the illiterate. The same social classes that once pioneered lower fertility are now responsible for the upward trend.

**Conclusion**

After long rejecting any idea of population policy at a time when fertility was extremely high and rapid population growth posed serious obstacles to economic and social development, in 1983 the Algerian government established a national program to control population growth, at a time when fertility decline was underway. Fertility reached the replacement level early in the twenty-first century. The government continued to assert the necessity of pursuing its population policy, even as fertility appeared to be falling still lower. Ironically, not only did fertility stop falling but it rose again sharply.

This account of mismatched correlation between fertility transition and the measures taken by the Algerian government casts doubt on the ability of
family planning programs to have a decisive impact on reproductive behavior. Clearly fertility may fall in spite of government opposition to family planning; the adoption of a population policy does not necessarily produce a fall in fertility, and its continuation does not necessarily prevent a subsequent rise.

In fact our account strongly supports the view that family planning programs are in no way the key for lowering fertility in those countries where it is considered to be too high (Vallin and Locoh 2001; Vallin 2011). Such programs, we would argue, have never done more than accompany basic changes in attitudes, at best anticipating them but mainly providing some support for changes in behavior. The basic causes of changes in fertility, whether upward or downward, are to be found elsewhere—in economic, social, and cultural transformations in society. Consequently, the effectiveness of fertility policy depends closely on other policies directed at encouraging social changes favorable to reducing fertility: improvements in health, development of education, higher status for women, and women’s access to remunerative employment and economic and social responsibility.

We do not contend that family planning programs are pointless nor that the Algerian government should have refrained from establishing its program to control population growth. The program did at least liberalize and facilitate access to modern contraception, providing moral and material support to couples who wanted to limit the number of their children. We only emphasize that couples mainly reduced their fertility as a response to the economic, health, welfare, and cultural policies undertaken by the government.

Note too that in the particular case of Algeria, the fall in fertility was largely due to a rise in the age at marriage that was not encouraged by any policy. At present young adults, especially young women, pay a psychological price for this trend toward later marriage. Young people in Algeria must cope with the consequences of the demographic upheavals Algeria has undergone in the past forty years. Not only do they confront high levels of unemployment and housing shortages, but, unable to marry, they are constrained by the social norms and taboos that forbid sexual relations outside marriage. Of course, this taboo is often evaded, but often at high personal and social cost.

Notes

1 Comparison of data from the 1969–71 multi-round survey (CNRES 1974b) with those from civil registration showed that the latter recorded an estimated 89 percent of births. We have consequently revised recorded births upward to estimate fertility from civil registration from 1964 to 1969. The figures obtained closely resemble two estimates based on the 1966 census data, one combining marital fertility and proportion of unmarried women (Navelet and Nizard 1972), the other dividing the number of children under the age of one by their estimated probability of survival (authors’ estimate).

2 Interview in 2006 with Jeanine Belkhodja, professor of gynecology and obstetrics at the University of Algiers.

3 The center was also called CPMI (center for the protection of mother and child).
References

AARDES (Association algérienne pour la recherche démographique, économique et sociale).
1968. Impératifs de la planification familiale en Algérie. Algiers: AARDES.

4 Connaissances, attitudes et pratiques en matière de contraception (contraception knowledge, attitudes, and practices).
5 Interview Belkhodja, 2006.
6 Programme national d’espacement des naissances.
7 This figure for mean age at first marriage, based on the 1966 census (Vallin 1973), is calculated by applying Hajnal’s method (1953) to the proportions of unmarried women. It is the only method usable for much of the period studied in this article. Unfortunately, mean age at first marriage calculated in this way depicts a reality much earlier than the observation date, since it covers marriages occurring in the previous 30 years. By making special use of the individual data from the 1998 census, we have observed that the weighted mean of actual year of marriage occurred 13.5 years before the census. Consequently, the result taken from the 1966 census is attributed to 1952–53, and so on.
8 Under sharia law the legal age had been puberty. A 1963 law first set it at age 16 for women and age 18 for men (Khemisti Law). The 1984 law was then amended in 2005 by a decree setting the minimum legal age at 19 for both sexes.
9 Although access to the labor market does not necessarily mean finding gainful employment, women routinely become job-seekers on completing or interrupting their education.
10 Interview, Belkhodja, and with Mustapha Benzine, Director of the Social Studies Division of the Algerian National Economic and Social Council (CNEES) 2006.
11 Unfortunately, the comparison of figures taken from two different sources (a census and a demographic survey) may produce a large margin of error. However, the comprehensive data on weddings provided by civil registration support the assumed fall in age at marriage, since the annual crude marriage rate (all ages) nearly doubled from 2000 to 2010, which could hardly occur without a lowering of the marriage age.


Fertility and Population Policy in Algeria


Why do development projects, and AIDS projects in particular, take the forms they do? In this essay we argue that it is because the conflicting interests and world views of the key actors involved—donors, brokers, and villagers—leave only a narrow range of themes and practices that can “work” on the ground. By “work” we do not mean, in the AIDS case, that they help prevent HIV transmission; indeed, the non-medical approaches to HIV prevention are largely ineffectual. What needs to be explained is why such approaches are nonetheless repeated so consistently.1 We show that these themes and practices work in the sense that they satisfy the varied agendas of the major actors sufficiently to sustain their day-to-day cooperation. Understanding these actors and their varying interests and world views is key to understanding why development projects rely on such a narrow repertoire of approaches, particularly the arcane and ubiquitous practice of training. HIV prevention projects are our case in point.

HIV prevention projects generate complex misunderstandings and conflicting motives among the critical actors in the AIDS enterprise: the deep-pocketed altruists who fund the programs, the local brokers who implement them, and the villagers who are the programs’ ultimate targets. Much has been written about such misunderstandings and motives, both during the colonial period in Africa (White 1987; Hodge 2007; Cooper 1998) and in recent times (Luke and Watkins 2002; Gibson et al. 2005; Burchardt 2012; Maes 2012). The surprise in our story is the degree to which these disparate actors fumble toward accommodations that allow them to get along, however awkwardly. Using an unusual range of data that we, our colleagues, and graduate students collected in Malawi, we describe some of the cultural themes the AIDS prevention and mitigation enterprise produces, and the remarkably
narrow set of concrete practices through which those themes are enacted. The case of “altruism from afar” thus illustrates the interaction of culture and practice in development aid, where interdependent actors cannot draw on a framework of common meanings and shared motives.

Although our focus is AIDS, and particularly HIV prevention and mitigation (rather than treatment, which has extended millions of lives), we should note that in many respects the global response to the AIDS epidemic (which we term the AIDS enterprise) is similar to its predecessor, the global population movement (Cleland and Watkins 2006). Each addressed what was defined as an unprecedented emergency with global consequences. Solutions to both were understood to require fundamental changes in intimate behavior, changes that decoupled sex and birth in one, sex and death in the other. In both, the solutions were designed primarily for women, perceived to be vulnerable and powerless, rather than for men (Hodgson and Watkins 1997; Dodoo and Frost 2008; Higgens, Hoffman, and Dworkin 2010; Poulin and Watkins 2012). The organizational machinery constructed in response to both emergencies was similar: a standardized choreography that mobilized huge resources, the creation of dedicated UN agencies and functionally identical national coordinating commissions in developing countries, pressure on recalcitrant governments to take action, and blueprints for action (Putzel 2004; Robinson 2011).

In both movements, the first efforts were to disseminate information and provide services (modern family planning, condoms), followed by a period of disappointment that the demand for services was inadequate. The slowness of behavioral response led to re-conceptualizing the problem in terms of steep barriers to change. Although international actors in both movements identified poverty, gender inequity, and “traditional culture” as major barriers, in the case of HIV prevention attention was also given to individual characteristics such as fatalism, denial, and stigma. The agents of the global actors also differed. Governments were to do much of the managing of policies and programs to stem population growth, making use of what eventually became extensive networks of clinics to distribute modern contraception in the course of providing routine maternal and child health services. In contrast, nongovernmental organizations were assigned much of the work of HIV prevention. The distinction is important: while government institutions persist in situ, NGOs come and go, and any structures they create, such as youth clubs to dramatize the dangers of AIDS or microfinance projects to ameliorate poverty and empower women, usually evaporate when the project funding ends and the NGO departs (Watkins, Swidler, and Hannan 2012).

The AIDS enterprise

AIDS in Africa variously inspires empathy or at least sympathy (Kristof and WuDun 2010; Rothmyer 2011), or disillusionment (Moyo 2009; Easterly
2009). Our reaction is different. Rather than criticizing the massively funded organizations of the AIDS enterprise for their failures or praising the efforts of the small-scale altruist who alights for a few weeks to assist a grandmother caring for AIDS orphans, we have sought to describe, analyze, and interpret. In doing so, we highlight features of the enterprise that others have not commented on.

Strikingly for a humanitarian endeavor, the organization of the AIDS enterprise has much in common with the organization of an international capitalist firm: a head office in one country that outsources production and distribution to subcontractors in other countries (Watkins, Swidler, and Hannan 2012; Wallace et al. 2007; Cooley and Ron 2002). In Malawi, as in many other countries receiving foreign aid, the organizational field of institutionalized altruism is complex, chaotic, and frenetic. Money often goes first to an international nongovernmental organization (INGO), which then subcontracts to either a branch office or a local NGO in the nation’s capital. Since these offices are still too distant from the intended beneficiaries, a chain of smaller subcontractors is constructed, until finally the donors reach an organization that interacts—or claims to interact—directly with the intended beneficiaries in the villages. At each level the NGO may engage “partners,” subcontracting specific programs to other NGOs working on the same issue. Thus, money flows chaotically both downward and sideways. And because projects are typically of short duration—and even then funding may be abruptly withdrawn—there are frenetic efforts to spend the money before it disappears, to increase what is called “the burn rate.”

In addition to institutional altruists, there are also swarms of freelance altruists. We met some of these on the plane flying to Lilongwe, in the guest houses and small motels where we stayed, or walking around a market. Some were church groups; others were Mom and Pop altruists, such as a Scottish couple who collected school materials during the year and then brought them to Malawi in the summer; others were individuals, like a retired school teacher who had invented a better wheelbarrow, and a charlatan who was promoting his cure for AIDS.

The donors, large and small, have fantasies about what poor villagers need; the villagers too have their own fantasies and, like the brokers, their own aspirations. In Malawi, where most villagers depend on subsistence agriculture, occasionally supplemented by small cash crops, any source of assistance is eagerly sought. Testimony to the basically materialist orientation of the villagers is seen in the very terminology of the aid enterprise: the words “development” and “empowerment” have come to be understood as code words for material help, albeit help that usually remains tantalizingly out of reach. The newspapers regularly announce donations to the cause of HIV prevention and AIDS mitigation: “Bush pledges $500m... to help fight HIV/AIDS” (Reuters 2002); “Canada grants K700 million for HIV/AIDS” (Nyoni 2002);
UNFPA, BLM launch K437m Youth Project” (Times Reporter 2006); “Agrofund gives Malawi US $93m” (Sabola 2012); “Concern Universal embarks on K640m project” (The Nation 2011). Many rural villagers read these announcements in the newspapers, but they can also see signs of donor wealth in the Toyota 4x4s with NGO logos that speed along the roads, or foreign research projects that deposit interviewers and graduate students in villages year after year to gather what the villagers think must be valuable information.

The primary way that donor money reaches the pockets of poor Malawians, however, is when an NGO trains volunteers to disseminate HIV prevention messages in the villages, to provide psychosocial support for orphans, or to promote better nutrition for people living with AIDS. Those who are selected for a “training” are considered lucky: stories circulate about a friend of a friend who received a per diem of 600 kwacha/day, plus accommodation and transportation. At the current exchange rate, a person who attended a three-day training would return home with nearly US$6.00, in a country where annual per capita income is below US $200 (even less in the rural areas where 85 percent of the population lives).

In our research we came to appreciate the roles of intermediaries: the cadres of relatively educated Malawians who serve as brokers, translators, and fixers (Lewis and Mosse 2006). Like Janus, the brokers look in one direction at the altruists who pay their salaries and in the other direction at the poor villagers the altruists seek to help. Just as there are layers of subcontractors, there are layers of brokers. The most educated and fortunate brokers staff the offices of the NGOs in the capital, writing strategies, frameworks, log frames, work plans, and proposals for more funding and producing financial reports for donors; others—less well-networked—may staff a district office, running specific projects or conducting monitoring and evaluation for a project in their district. When the aim is both ambitious and vague, as it is when the donor aims to prevent HIV by transforming villagers’ behavior through community mobilization, participation, and empowerment, there is a final step. The last mile to the villages is to be walked by volunteers who are trained by an NGO to deliver messages to their relatives, friends, and neighbors about the goodness of family planning or the badness of multiple sex partners. Without the volunteers, donors cannot envision their projects as “sustainable” once the NGO inevitably decamps (Swidler and Watkins 2009). Without brokers, then, nothing that the altruists imagine will happen does happen.

Despite all the attention to the altruists and their successes and failures, remarkably little attention has been given to the brokers, the actual individuals and organizations that provide the crucial link between the good deeds imagined by the altruists and their imagined recipients in the villages. These intermediaries, however, have complex aspirations and motivations of their own—some of them created by the AIDS industry itself.

As the AIDS industry grew, participating NGOs could offer opportunities for formal employment outside of government. This contributed to expanding
the middle class and enhancing the perceived value of education. In Malawi there is now a clearly understood hierarchy of career possibilities based on educational credentials, from a volunteer in a small village community-based organization (CBO), appealing to youths with secondary education and no other hope for escaping village life, to a Ph.D. in a UN agency in Lilongwe with a salary large enough to provide first-world health care for his or her family and first-world education for his or her children. Brokers may also have other aspirations: to maintain their status in the local community, to manage the support of myriad relatives and other dependents, to cultivate local networks, to attain the next-higher educational credential.

Our account of the roles of local actors can be viewed as a narrative of seduction—and sometimes, of heartache and betrayal—characterizing the fraught embrace between altruists who come from afar and their significant others, the locals. Brokers all up and down the hierarchy play the role of “authentic African” for foreign visitors to embrace. For altruists, these direct experiences often shape beliefs about the real nature of Africans and their problems. The educated cosmopolitans like to tell foreigners about the exotic customs of the villagers, such as particular sexual behaviors or widow inheritance. Recounting these “harmful cultural practices” (the typical term used by the elites) and telling stories about “foolish villagers” seem meant to accentuate the differences between the brokers and what they refer to as the “backward masses” and to solidify their credentials as valuable cultural insiders (Englund 2006).

The fraught relationship of brokers and donors, and of both to the villagers, is ironic, since the partners in these dalliances really want to get along: they want—and sometimes need—to make it work. Nonetheless, their different views of the world and of rural Malawians, their differing experiences and constraints, and their sometimes conflicting interests and aspirations mean that they must fumble toward the themes and practices that can be made to work for everyone.

**Themes that make everyone happy**

**Fighting stigma**

“Fighting stigma” is the mom-and-apple-pie of AIDS interventions. Everyone loves the idea. To Western donors, the term signals a commitment to human rights, to fighting discrimination, and to an aggressive self-assertion by otherwise despised groups, all responses to encounters with AIDS in the West (Shilts 1987; Epstein 1996; Allen 2004). To Malawians, its meaning is very different. Here, the message to fight stigma reinforces the normal obligation of reciprocity and interdependence among kin and neighbors. For villagers the message echoes their own understandings of the responsibility to care for those who are hungry or ill.
Many donors sponsor programs to fight stigma, but the emphasis on compassion and care fits especially well with the mission of one set of donors, the major religious denominations. Local churches embrace the appeal to fight stigma because it gives them a way to be active in combating AIDS without having to raise awkward issues about sex and condoms or having to discuss openly what everyone knows: that the wealthier men in the congregation (and indeed the pastor as well) probably have outside sexual partnerships (Hunter 2010; Hirsch et al. 2009; Poulin and Watkins 2012). It allows churches to maintain comity and fellowship among their members by advocating that people should “love each other,” should not blame or “isolate” each other, and should treat each other with kindness.

The adjuration to be kind to those with AIDS, to care for others, to take care of those who are ill, resonates with the importance villagers themselves place on taking care of the needy and maintaining harmonious relationships among neighbors. When we studied proposals written by villagers in 2005–06, requesting money from Malawi’s National AIDS Commission to fund Community Based Organizations, many of the proposals—often those written in naïve language rather than the jargon villagers later mastered—claimed that the organization would “love” and care for those with AIDS, as well as for the elderly or orphans, as indeed villagers had been doing since the epidemic began (Chimwaza and Watkins 2004). For example, Trinitapoli (2011), in a study of rural churches and mosques, found that pastors and sheikhs strongly advised against divorcing a spouse with AIDS. The advice was not a matter of political correctness or human rights. Rather the pastors and sheikhs felt that someone had to care for the person who was ill.

Elizabeth Colson, an anthropologist who has worked for decades in Zambia, argues eloquently that Africans do not need to be taught to fight stigma: “Somehow they have maintained a respect for their own moral integrity. Many have shown abundant compassion, not just for the moment but over the long months of illness of those for whom they have accepted responsibility…. I would argue that they conceive of themselves as responsible human beings, meeting their obligations as best they can, and so maintain their own self respect” (Colson 2006: 124).

From a public health perspective, the persistence of calls to fight stigma is odd. We have found no rigorously designed studies that demonstrate empirically that reducing stigma reduces HIV transmission. Moreover, if the goal of HIV prevention is behavior change, then persuading people not to condemn behaviors that lead to HIV transmission could be counterproductive. Stigma nevertheless looms large in the imagined arsenal of AIDS prevention interventions because it sounds good to all and offends no one: everyone, from donors and their sponsors, to brokers, to churches and villagers, can get on board. But even though donors, brokers, and villagers have all signed on to the same program, they “misunderstand” each other in the sense that the fight against stigma means something quite different to each group.
Orphans and vulnerable children

The designation Orphans and Vulnerable Children (OVC) “works” for everyone in a different sense than stigma does. Rather than legitimating an intervention, it defines a category of people to whom help may be legitimately directed.

For international organizations, the image of “orphans”—those bereft of both parents, alone in the world, facing desperate poverty and the emotional devastation of terrible loss—are the perfect poster children for attracting funding. For those living in circumstances where many children are poor, privileging orphans over other vulnerable children makes little sense. For the donors’ purpose of stimulating compassion and charity in international audiences, real orphans—children who have lost both parents—are too small a category. Orphanhood as a category was therefore expanded to cover loss of one parent as well as both. The resulting increase was substantial. Survey data from Malawi records that 13.2 percent of children aged 0–17 in 2004–06 were orphans by this wider definition. Only 3.0 percent had lost both parents, a category now termed “double orphans” (Pullum 2008: 5).

Reports on AIDS orphans often acknowledge the enormous influence of the 1997 USAID report *Children on the Brink*. Its “breakthrough” was a new methodology for estimating both paternal and maternal, as well as double, orphans, creating high figures for the total number. The more capacious orphanhood definition allows international organizations to win the sympathy that the image of a parentless child evokes, while claiming very large numbers of children as in need of help and support. (A second *Children on the Brink* report was issued in 2000.)


Expanding the orphanhood category did not, however, solve the problems of those with the task of distributing resources in the villages. World Vision found that its child-sponsorship program created problems of inequality and envy in the villages: why should one poor child be favored over another just because that child was an orphan? (Bornstein 2003: 67, 95). The category was thus further expanded to take account of village understandings that it was not only orphans who were vulnerable children. As Chirwa (2002: 96) explains:

In most Malawian languages the terms used to define an orphan and orphanhood include loss of parents; the rupture of social bonds; lack of family support; the process and situation of deprivation and want; and the lack of money
or means of livelihood. Some of these are, indeed, the effects of orphanhood. However, the Malawian equivalents of orphanhood treat these as integral parts of the totality of the process of orphanhood.

As a purely practical matter, local understandings determine the distribution of resources to village families. Since there are no records of the economic status of children in a village, and since interventions rarely begin with a survey that distinguishes between poor orphans and other poor children, the village chief or a village committee is given the task of determining who deserves aid. For them, however, singling out those who happen to be orphans makes no sense. Orphans are usually taken in by relatives—indeed often by the wealthier relatives in an extended kinship system—and thus may be better off economically than children who are still living with their biological parents (Ainsworth and Filmer 2002; Republic of Malawi and World Bank 2007). The flexibility offered by the term Orphans and Vulnerable Children means that brokers working on the ground can fold child sponsorships into broader programs meant to help whole communities, as World Vision does. They thus avoid creating envy, envy that may lead to accusations of witchcraft against those who are seen to benefit unjustly.

Thus the capacious term Orphans and Vulnerable Children or OVC is not simply a random piece of organizational jargon. Rather it represents a strategically calibrated, if unacknowledged, compromise among a number of agendas, meanings, and interests. Like many other awkward neologisms and acronyms that litter the AIDS landscape, the term also illustrates enormous cultural creativity as program beneficiaries and donors try to find an umbrella large enough to cover their multifarious needs, while retaining a symbol powerful enough to legitimate their activities.

Vulnerable women

Of all the themes that stimulate the imaginations of donors and brokers, one of the slipperiest is that of “vulnerable women.” Here, Malawians and their international partners appear to be singing the same song: because of poverty and culture, women are particularly vulnerable to HIV infection, and as a result “something must be done.” But although the melody is the same, the lyrics are different. Both partners seem to identify the same problem, but each has a very different conception of what must be done. Western donors imagine women as poor and weak, victims who urgently need empowerment so that they can “just say no” to unsafe sex and thereby turn the tide of the epidemic in sub-Saharan Africa (USAID 2002; European Union 2005; Watkins 2010). Malawian brokers agree that women are “poor,” but they see mercenary women in miniskirts tempting good men: women who need to be reined in, rather than empowered.
The disconnect is, we believe, a matter of perspective (Tawfik and Watkins 2007). For international donors, women's vulnerability has moved to the center of their understanding of the AIDS epidemic, especially in Africa (United Nations 2001; Higgins, Hoffman, and Dworkin 2010). From afar, Western altruists see desperately poor women, without legal or social rights, engaging in survival sex to feed themselves and their starving children. Malawians, close to the epidemic, recognize that many women are poor and have few alternatives to depending on a sexual partner. But because in their view men cannot resist a woman who offers them sexual favors, what preoccupies Malawian villagers—and elite brokers—is the personal threat posed by rapacious women who take men's money and give them, and thus their wives, a fatal disease. Both parties, international and national, believe that women are more likely to be infected than men: thus, the solutions they offer, while often appearing gender neutral, are not.

Rural Malawians have come up with solutions to the problems posed by women. Some of these are on the international agenda, others are not. A council of town fathers, including “representatives of the District Executive Committee, a representative of religious organizations, Non Governmental Organizations and female youth representative,” recommends abstinence and fidelity to counter the spread of AIDS, but also restrictions to combat moral disorder: “3.3.5 To control the proliferation of pornographic materials” (Balaka District Assembly 2000). Other solutions are based on the assumption that when unrelated men and women are together at night, sex is inevitable; thus, there are recommendations to ban night meetings of religious groups and night dances after weddings. These solutions appear to be gender neutral. But villagers have also proposed ways that men can protect themselves that are not gender neutral: closing bars—where sex is sold—in the evenings, locking up sex workers, and disclosing the names of those found to be HIV-positive (the majority of whom they assume are women) so that others can avoid having sex with them.

So how do elite NGO brokers, who see themselves as threatened by disorderly women but who also partake in the globalized discourse about vulnerable women, manage? One solution is to speak in different voices in different contexts. Those who sit in a meeting making AIDS policies and implementing programs do not speak of women in miniskirts seducing men. Rather, they echo the language of the international community in attributing women’s and girls’ vulnerability to HIV to poverty. But in our experience, when they leave the room where formal policy is being deliberated, they spontaneously talk of predatory sex workers and schoolgirl temptresses. From one room to another, the images of women shift shape. Despite the rhetoric of women’s vulnerability, NGO brokers seek to restore moral order by suppressing vice, reducing temptation, and—especially—restraining women’s behavior.
Currently, proposed legislation in several African countries would subject anyone who knowingly infects another person with HIV to criminal penalties (such legislation has been introduced in several high-HIV-prevalence countries in the region). Further, Malawi’s parliament is considering legislation for the mandatory testing of sex workers, domestic workers (typically female), and pregnant women. While human rights advocates oppose this legislation, the concept behind it—that those who know they are HIV-positive but nonetheless have unprotected sex are willfully endangering others—makes sense not only to villagers but also to elites.

In response to the human rights advocates, elite brokers have adopted a clever solution to envelop the issue of controlling infected women in the rhetoric of human rights promoted by international organizations. In October 2009, police in Malawi arrested and forcibly tested sex workers, stimulating an outcry from human rights organizations. As reported in the newspapers, the official who had final responsibility for AIDS policy, Dr. Mary Shawa, began by criticizing the police, but then, even as she invoked “human rights,” her view was transformed into a characterization of mercenary women who endanger their fellow citizens:

Dr Mary Shawa, Secretary for HIV/AIDS and Nutrition in the Office of the President and Cabinet, said while Malawi had to employ every available and legitimate way to contain HIV/AIDS, testing the sex workers alone, and not their clients as well, was not justice enough.

“We all know the importance of human rights. But when we are demanding our rights, we also have to keep in mind our responsibilities. It is the sex workers’ right to make money but if we are all careful, we will not make that money at the expense of another person’s right to life,” Shawa said. (Mpaka 2009)

A high official of the Malawian Human Rights Commission, speaking of the need for mandatory testing of sex workers, put it more bluntly. He said, “It’s not easy to take a human rights lens to some of these issues. It’s like having a person with a machine gun with bullets. The bullets shoot out, bang! bang! bang! And many people are dead” [Interview with Watkins, 2009, authors’ files].

In the case of vulnerable women, then, the misunderstandings that separate the perceptions of Malawian brokers (and villagers) from those of international donors are harder to bridge than in the cases of stigma or orphans. Here the “working misunderstandings” often don’t work, as local actors—both the district-level town fathers who want to ban pornography and forbid night gatherings where men and women might meet, and the national elites who wish to lock up sex workers—use the rhetorics of “empowerment” and human rights while continually returning to policies that would constrain rather than enhance women’s autonomy.

In all three of these examples (and we could provide many more), gaps between expectations and understandings of different participants in the AIDS
enterprise are finessed, side-stepped, or obscured. Categories and concepts are unstable in their meanings, and groups with very different agendas “get along” because they do not need to confront the different meanings they attach to the same words.

The fact that the actors who must get along to make HIV prevention projects happen have such different interests and ideas nonetheless has real consequences. Because projects must meet the needs of donors, brokers, and villagers, there is a surprisingly narrow range of techniques used to conduct AIDS prevention interventions in the field. The confusions created by the different meanings that villagers, brokers, and donors attach to the key themes of the enterprise are resolved only in the shared understanding that, whatever the theme, the dominant practice will be “training” (Smith 2003).

Practices that make everyone happy

In contrast to the malleability of categories and concepts are practices that seem set in stone. Every AIDS organization that seeks financial support from outside donors, even groups of church women caring for orphans or those living with AIDS, has to adopt standardized practices. They have to learn how to write proposals, log frames, and work plans; to provide monitoring reports of their spending; and to evaluate their programs (DiMaggio and Powell 1983). Every CBO in Malawi has a standard array of committees corresponding to donor themes (OVC, Home Based Care, PLWA [Persons Living with AIDS], Youth), each with an executive committee, and all must submit regular reports to their funder.8 “There is an untheorized consensus on what an HIV/AIDS programme should look like: it should be founded on voluntary counseling and testing, education (preferably by peers), provision of condoms, efforts to overcome denial, stigma and discrimination, and care and treatment for people living with HIV and AIDS…. Even when undertaken by a government ministry, army, or private company, it is essentially an NGO model of public action” (de Waal 2003: 254).

Training

The most pervasive, most standardized, and to an outside observer the most arcane in its detailed choreography is the practice of “training.” Trainings have become the ubiquitous social practice through which development aspirations are enacted jointly by donors, brokers, and villagers fortunate enough to be included in a training. Whatever donors might imagine is being done with the money they allocate to prevent HIV, to improve agriculture or nutrition, to initiate early childhood education projects, or, now, to combat climate change, the recipients—and villagers at large—know that one specific requirement is training. We have seen it in the hundreds of budgets submitted by CBOs seeking funding from Malawi’s National AIDS Commission (NAC), but also
in the myriad trainings we have observed at rural motels and city conference centers, sponsored by the World Bank, USAID, the Norwegian Agency for Development, and many other donors. Some trainings impart substantive information, as clinical officers are trained in a new procedure for resuscitating infants. Others, with goals like creating “AIDS competent communities” or making communities “resilient to the impacts of climate change,” seem unlikely to have any impact at all. A UNDP project to “mainstream gender” was implemented by training “gender focal persons” (often the lowest women on the professional ladder) in government ministries. Agricultural extension agents are trained in how to train farmers how to cook nsima (the staple food, which even men know how to cook) and other local foods. Villagers are trained to be volunteers, spreading the word about how to care for the sick and elderly, or how to prevent HIV transmission, or how to empower women by changing the norms that make women vulnerable.

All of the trainings are constituted by a ritualized set of practices. Just as a Catholic Mass would not be a Mass without wine and wafers, a training is not a “training” without allowances for attending the training, flip charts and magic markers, a “bun” and a “Fanta” at mid-morning and mid-afternoon breaks, and an ample lunch. Not only are the equipment and the food ritualized, the activities are ritualized as well. The better-educated facilitator speaks of familiar concepts and practices in new terminology (“decision making,” “multiple and concurrent partners”). Forms of participation are stereotyped, as when participants break into small groups, with each group taking a large sheet from the flip chart to write down ideas supporting the trainings’ theme, and then reassemble to share the products of their efforts. Other activities seem more peculiar to an outside observer: hymn singing, jumping around playing children’s games, skits and songs. But donors and the many layers of brokers have become convinced that whatever the problem, training is the solution.

In 2005, we studied a NAC program aimed at mobilizing community groups to address the causes and impacts of the AIDS epidemic. Existing community groups, such as members of a church group that visited the sick or cared for orphans, were invited to turn themselves into CBOs with a committee and a bank account, and then write proposals to the NAC to support their activities. Others, hearing of the funding opportunities, created a CBO de novo. We reviewed the proposals submitted through the district office of one of the international NGOs that had been hired by the NAC to manage the proposal process. The proposals from village groups were poignant in what they asked for and in their attempts to put what they really wanted—material resources—in terms that NAC might agree to fund (Swidler and Watkins 2009). Early in the process, proposals would ask for resources to feed the sick and the elderly, but these were likely to be turned down as “not sustainable” or “not compatible with NSF [the National Strategic Framework of the NAC],” a document not available in the villages.
By the time we reexamined the process in 2008, again looking at both accepted and rejected proposals, much in the process had changed as communities learned what would be funded: a lot of trainings and a small amount of resources. The Nsamanyada CBO proposed nine activities, each with a training and a budget. The list of activities begins with training for HBC [Home Based Care], which also includes a small amount for panadol and band-aids for the HBC kit. Next was the HIV/AIDS Support Group training for “Positive Living” and for “Nutritious Training,” along with an “Awareness Campaign.” In addition to the standard list of necessary items without which a training would not be legitimate (the flip chart, pens, buns and Fantas), the budget included a modest expenditure for “procurement of other nutritious food, cooking oil and sugar.” Third is “Elderly and Disabled Training,” followed by “Child Protection Training,” and then “Youth Peer Education Training,” and training in care for OVC. Then there is “Business Management Training,” again with the usual apparatus and a modest amount to support an income-generating project: the group proposed to earn income by raising pigs, and requested funding to build a pen, purchase three pigs, and buy feed and drugs for the pigs. But this, of course, requires “Piggery Training,” which turns out to be considerably more expensive than buying the three pigs or even feeding them. The grand total in the Nsamanyada CBO’s budget proposal came to 1,500,000 Kwacha (about $11,000 at the 2008 exchange rate). Of this budget, 1,478,040 Kwacha was for the multiple, carefully elaborated varieties of “training.” [document in authors’ files]

In the proposals, the budgets are detailed but the content of the trainings is not. Nor do they need to be: what matters to the proposers as well as to NAC’s budget is the number of people who will benefit by receiving allowances, the number of days of training, and the amount of the allowances. In 2008 the Interfaith Helpers Ministry submitted a proposal to train 144 people in nine villages: “44 female church leaders; 60 male church leaders; 37 youth leaders; and 3 Traditional leaders [chiefs].” The actual activities budgeted were: “Sensitization meetings on legal, ethical, and social rights of the PLWA and the affected people” and “training of key social groups in HIV/AIDS management.” The project was to culminate in a “social mobilization campaign on gender/human rights and cultural values” for which five members of a drama group and five facilitators of the trainings were to be paid allowances for four days. [Document in authors’ files.]

What goes on in a training? We have observed trainings of youth peer educators, volunteers who are to fan out to the villages to teach others what they have learned. The trainings are led by a facilitator (sometimes two) who follows a manual. Like teachers, they stand in front of the seated trainees, lecturing, pointing to the blackboard, and asking questions with known answers (Kendall 2004). When participants are asked to demonstrate what they have learned, they replicate this performance: lecturing, pointing to words on a flip chart, asking questions.9
The structure of a training, which donors might imagine as an equalizing activity, in fact mirrors local understandings of the social hierarchy. The more educated enlighten the less educated, who defer to their expertise. Village chiefs and other participants of higher status, such as a schoolteacher or a pastor, are budgeted for a larger per diem and a more expensive lunch than ordinary villagers, and are served first. Such distinctions of influence and status are respected all the way down the line.

Why does the training as a social form have such an amazing hold? We think that trainings satisfy the practical needs of the participants, of the brokers, and of the donors, but also provide meanings that satisfy the fantasies of each group.

Who benefits from the trainings? For the poorest participants—the villagers who volunteer their services to mobilize their community for one NGO goal or another—per diems are a financial windfall (currently a day’s allowance for a training of rural volunteers in Malawi is about US$1.90). Although this is not the only way that participants benefit from trainings, it seems to be the most important. Villagers with some education but no job might attend a training with little direct remuneration—perhaps only a lunch or a snack—because volunteering is seen as a route to the possibility, however remote, of eventual employment (Swidler and Watkins 2009).

Trainings are also important for brokers who have a government or an NGO salary. The amount of broker per diems is such a contentious issue that it is likely that without any financial benefit no one with a regular salary would attend a training, even one related to the broker’s duties. An evaluation of a USAID project to improve maternal and child health by training staff reported that “There has [sic] been persistent problems with the allowances awarded to the DHO [District Health Office] staff, which has prevented or slowed down many project activities. Many MOHP [Ministry of Health and Population] staff refuse to conduct activities with the level of allowances approved by USAID…” (USAID 2003; see also Vian et al. 2012). Trainings can also serve in lieu of salaries: a skilled technician working at a government hospital as a volunteer because there were no openings on the staff was assigned by his boss to attend successive trainings on anemia and on prevention of mother-to-child transmission—both outside his area of expertise—indeed on anything that might offer him even a modest amount of support. [interview 2010] This fits well with a patron–client economy: the boss becomes a patron who will expect something, sometime, in return, and the counselor becomes a client, in debt to his patron (Smith 2003).

For donors, the trainings of village volunteers to teach others are the only practical way that they can imagine transforming the billions of poor people into participating, self-actualizing, empowered individuals. Equally important, the training satisfies the donor fantasy that the interventions are sustainable—knowledge is transmitted, participants sensitized—without requiring recurrent expenditures. Lastly, for both donors and brokers, trainings
have another practical advantage: the number of people trained provides an easy measure of success to report to funders, at least as long as donors and their funders have faith in the power of training to transform the identities and behaviors of poor villagers.

Training also benefits donors and high-level brokers in another, less obvious way. Because training is so flexible (anyone can be trained about almost anything) and so easy to arrange (partly because the conventions are so well understood), trainings can be used as “walking around money” through which a program’s sponsors can create broad political support for their programs among various “stakeholders” and others whose good will may be useful. To build support for AIDS programs, for example, Malawi’s National AIDS Commission allocated funding to virtually every government ministry for “mainstreaming” AIDS, in effect spreading the AIDS largesse around in the form of training (with allowances and per diems) in HIV/AIDS.

Trainings, however peculiar they may seem, are not some arcane donor invention foisted on resistant locals. At all levels brokers as well as village volunteers embrace, even demand, them.10 The appeal of training lies not just in the material rewards, but in the mystique of its games and songs and its hours of scripted minutiae. Adults play children’s games or sit on the ground drawing root-and-branch diagrams to discover the that “root” of a particular problem is, in the well-recognized conventions of the training, precisely the thing the training is meant to combat: whether AIDS, malnutrition, illiteracy, or infant mortality. Participants listen attentively as facilitators rehearse jargon and pose rhetorical questions, adding their own suggestions to the inevitable lists of “next steps” or recite problems for the “problem tree.” All this activity helps donors believe they are doing something about problems they define as urgent, even if the problems themselves are hardly likely to be solved through training. Elite brokers go to trainings, although these are called “workshops,” which signals the status of the attendees. These help elite brokers keep abreast of the latest donor concerns, distinguishing those who know the current importance of climate change, or the role of “multiple concurrent partnerships” in the spread of HIV, from those who are not as up-to-date. But even the aspiring local brokers—those whom we have called “interstitial elites” because they link the world of the village to the larger cosmopolitan world outside (Swidler and Watkins 2009)—eagerly seek the sense of identity and of connection to authoritative global knowledge (Frank and Meyer 2007) that trainings provide.

Trainings are thus the quintessential example of a practice that “works” because it satisfies the multiple agendas of different actors, even when the concepts and meanings they bring to it may be very different. There are other such practices: the pervasive evaluations that produce mainly “success stories” and lists of “challenges” calling for more resources; the “monitoring” that carefully enumerates the number of members in youth drama clubs that circulate from one village or secondary school to another, performing AIDS
dramas for students who in their turn perform AIDS dramas for others; or the precise number of people “reached” at an outdoor AIDS Day celebration featuring a band, comedians, and skits (1,410 men, 1,634 women, 1,940 boys, 2,010 girls) (Balaka Acting District AIDS Coordinator 2010). These activities share a similar property of providing satisfaction for actors up and down the aid chain without forcing them to confront the contradictions in their goals, aspirations, or understandings.

Conclusions

By focusing attention on the cultural themes and practices that dominate the AIDS prevention enterprise, we have emphasized the generative power of the conflicting motives, aspirations, and understandings of the central actors in such efforts in poor countries. Like family planning before it, AIDS prevention has generated layers of local brokers who mediate between international donors and the poor villagers whose lives are seen as requiring transformation. Rather than simply being critical of the self-delusion of some donors, or the mercenary motives of some brokers and villagers, we have emphasized that in trying to get along, to find ways of making their shared enterprise work, these actors have, without taking conscious account of it, generated new themes and practices that flourish because at one level or another they “work” for everyone involved. Combating stigma may not be an effective way to reduce HIV transmission; focusing on Orphans and Vulnerable Children may not target the parentless children the altruists imagine; efforts to empower vulnerable women may be foiled by Africans’ views that women are more perpetrators than victims of the AIDS threat. But these focal points of the AIDS enterprise have nonetheless created new realities on the ground in Africa. Training, as we have sought to show, is even more thoroughly institutionalized. As Marian Burchardt (2012) has noted of institutions of AIDS governance in faith-based organizations in South Africa, “the resulting assemblage of people, resources, and opportunities is not some form of halfway modernization or uneasy connection of two pure forms but a form of practice in its own right.”

Our argument suggests that even mismatched fantasies can create new realities as people attempt to enact their ideals and aspirations. Like a passionate love affair that eventually settles into a reasonably satisfying marriage, imagination can animate the creation of real partnerships. Each party compromises; neither ends up in the situation he or she originally imagined; but their mutual accommodations and compromises create new possibilities.

How then should we evaluate the influence of the AIDS industry in Africa? If we ask whether these organizations have succeeded in doing what they set out to do, the answer is no; the record of HIV-prevention NGOs is largely one of failure (Watkins, Swidler, and Hannan 2012; Potts et al. 2008). But if we take a broader perspective, we find that AIDS altruism has
changed the landscape in fundamental ways. There are new understandings of possible careers; new cultural themes to be mastered; and the adaptation and multiplication of imported social structures, such that no one is far removed from a professionalized Community Based Organization or from an indigenous group of church members that has transformed itself into a Faith Based Organization.

It is possible, as Cleland and Watkins (2006) argued for the family planning movement, that even without transforming poor rural villagers’ identities and turning them into more self-actualized, empowered actors, the programs and projects of international donors may, over a longer time, have at least some of their intended effects. The terminology of “gender” is already becoming domesticated (Merry 2006), sometimes incorporated into the flirtatious repartee that characterizes interactions between women and men in rural Malawi. The AIDS epidemic itself has also stimulated the formulation of new strategies in local social networks, such as limiting the number of sexual partners, choosing partners more carefully to avoid infection, or divorcing a high-risk spouse, strategies that in principle could reduce the force of the epidemic (Watkins 2004).

In the study of institutional forms, scholars recognize that institutional capacities created for one purpose may be animated by new purposes and may come to serve new interests (Skocpol and Feingold 1982; Hall and Taylor 1996; Thelen 2004; Mahoney and Thelen 2010). Such a process appears to be happening, for example, with the CBOs and FBOs, structures meant to replicate traditional forms of village self-help, but largely invented as formal organizations in response to the flood of funding for AIDS prevention and mitigation. But precisely because these organizations could be integrated with existing patterns of patron–client ties, because they served both the status aspirations and to some degree the material interests of at least some local populations, they have become increasingly an institutionalized fixture of Malawi’s rural landscape. Currently these organizations exist largely to apply for outside donor funding, rationalizing at least one small corner of the enterprise of altruism. But when structures are created, no one can fully anticipate what these new capacities may be used for. Possibilities range from new avenues for political mobilization to increased penetration of global governance directly into village life, bypassing national states.

There is one other unanticipated consequence of the AIDS enterprise. The new opportunities that the donor-driven economies have created depend crucially on the acquisition of an array of minutely stratified educational credentials. Donors have thus given substance, however fragile and elusive, to Malawians’ passionate belief that education is the route to a “bright future” (Frye 2012). This reshaping of identities and aspirations may ultimately be the most lasting legacy of the struggle against AIDS. In one of the ironies of complex human enterprises, this reshaping in turn becomes a major impe-
tus to the fulfillment of another donor fantasy: persuading the poor to want fewer children.

As the AIDS enterprise, and the NGO presence more generally, have opened up career paths that appear to depend on educational credentials, even poor parents may start to see focusing their resources on fewer children as crucial for their own futures and those of their children. It may be precisely NGO practices like training, which so resemble school, which require moderate literacy but no high-level technical skills, and which reward achievement even in a woefully out-of-date school curriculum such as Malawi’s, that could motivate parents—who have long resisted the blandishments of the family planning NGOs—finally to limit fertility.

Notes

Susan Cotts Watkins is grateful for the support of a John Simon Guggenheim Foundation fellowship. Ann Swidler is grateful for support from the Canadian Institute for Advanced Research and the Russell Sage Foundation.

1 Few HIV prevention interventions are designed to permit rigorous evaluation; if there is an evaluation, it typically consists of collecting anecdotes—success stories and testimonials—from the people who implemented the intervention and from those who were meant to benefit from it (Watkins, Swidler, and Hannan 2012). A recent comprehensive review of HIV prevention activities found that when the study design is rigorous, that is, a randomized controlled trial with biomarker outcomes rather than self-reports on a survey, the only effective interventions were biomedical. The seven behavior change randomized controlled trials with biomarker outcomes had no significant effects (Padian et al. 2010; see also Potts et al. 2008).

2 The purchasing power parity estimate is higher, $834 per person per year. Cash income remains hard to come by for most villagers, so even the tiny allowances for meals and transport are extraordinarily valuable.

3 Exemplary exceptions are Anders 2006; Merry 2006; Mosse and Lewis 2005; Olivier de Sardan 2006; Temudo 2005). For historical accounts of brokers in colonial Africa, see Cooper 1998 and Hodge 2007.

4 “Preaching Wednesday to more than 3,000 people in a stadium in the mining town of Kitwe, Archbishop of Canterbury Rowan Williams called on Christians not to abandon people living with HIV/AIDS. ‘The church should deal with HIV and AIDS. The church in Africa deals with HIV and AIDS, and we should not forget or forsake those that have the disease,’ Williams said in his address, which was broadcast live on community radio stations. ‘We are sent to share the suffering and joys with our neighbors; we are sent to be alongside them and listening to their needs. We should give our lives to others, as Jesus did’” (Agence France Presse 2011).

5 Trinitapoli (2011) writes: “the primary reason given [by pastors and sheikhs] for opposing AIDS-related divorce is the mandate to care for the sick. If an adulterous spouse is caught ‘red-handed,’ one can divorce, as a matter of self-protection; but if a spouse is known to be HIV+ (through an HIV test or obvious symptoms), one must not divorce. One leader who strongly and emphatically opposed divorce under suspicion of AIDS emphasized the wife’s role to care for her husband: ‘No! It is not appropriate! Who are you going to leave him with? Who would take care of him?’”

6 The report analyzed data from 34 “developing countries,” mostly in Africa, but also including several in Latin America and Asia. It constantly merges the idea of orphans in general and AIDS orphans. The 1997 report covered 23 countries; the 2000 report added 11 more, mostly in sub-Saharan Africa, for a total of 34.
7 A special meeting of the United Nations in 2001 (United Nations 2001) labeled women and girls as particularly vulnerable to AIDS. “Stressing that gender equality and the empowerment of women are fundamental elements in the reduction of the vulnerability of women and girls to HIV/AIDS” (article 14). Higgins et al. (2010) label this the “vulnerability paradigm,” noting that it has become the dominant discourse in the global AIDS response.

8 The distribution of CBos in Malawi was recently reorganized by the National AIDS Commission, such that there would be only one in each Group Village, an administrative unit that includes multiple villages.

9 A more elaborate version of training is set out in a facilitator’s manual for a large project funded by USAID. The facilitator’s manual instructs the facilitator to distribute chewing gum to all the participants “so that students should learn what problems they could meet from sleeping/sex with many people. In short, what happens? A teacher or facilitator should give chewing gum to the audience, when they have chewed for a little the facilitator should take all the chewing gum and put it together in one place. The audience will refuse to take back the chewing gum. This will trigger discussions or debates concerning body fluids.”

10 A newly installed chief, for example, might demand that he and his sub-chiefs be trained again, in exactly the same program they were trained in only months earlier, as an acknowledgment of his new status. Stakeholders in a new project—those whose goodwill or cooperation is needed even though they are not direct subcontractors of the project—may each insist on being trained separately, at a separate training with allowances and per diems, rather than simply attending a stakeholders meeting.

References


ENVIRONMENT, TECHNOLOGY, AND WEALTH
Expecteding that more and richer people will demand more from the land, cultivating wider fields, logging more forests, and pressing Nature, comes naturally. The past half-century of disciplined and dematerializing demand and more intense and efficient land use encourage a rational hope that humanity’s pressure will not overwhelm Nature. Beginning with the examples of crops in the large and fast-developing countries of India and China as well as the United States, we examine the recent half-century. We also look back over the past 150 years when regions like Europe and the United States became the maiden beneficiaries of chemical, biological, and mechanical innovations in agriculture from the Industrial Revolution. Organizing our analysis with the IMPACT identity, we examine the elements contributing to the use of land for crop production, including population, affluence, diet, and the performance of agricultural producers.

India and China

In 1960 the population of India was about 450 million. In 1961, Indian affluence, as measured by GDP, equaled about 65 billion recent US dollars (World Bank 2012). The average Indian consumed 2,030 food calories (kilocalories) per day, a level that approaches minimum calorie thresholds for hunger.1 Indian farmers tilled 161 million hectares (MHa) of land to grow crops, while the country imported a net 4 million to 10 million tons2 a year of cereal grains, over 6 percent of its demand on average during the decade of the 1960s (Food and Agriculture Organization [FAO] 2012). In the United States in 1960, youngsters were admonished to finish their peas and be grateful that theirs was not the lot of the hungry children in India.
The years between 1960 and 2010 saw more babies, more affluence, and better nutrition. India’s population rose over two and a half times, while national income rose 15 times. By 2010, the average Indian ate a sixth more calories than in 1960. While a majority of Indians still lived in the countryside, many moved to the cities, where they depended less on forest resources (Chandramouli 2011; Ministry of Environment and Forests, Government of India 2009). The 15 MHa added to Indian forests from the 1960s to 2000 exceeds the size of the state of Iowa in the US. The reversal of deforestation hints at an associated peak in farmed land.

India, which had net imports of wheat throughout the 1960s, had net exports in four years between 1970 and 1989 and 11 years between 1990 and 2009 (FAO 2012). Because of agricultural technologies introduced in the 1960s and 1970s and persistent efforts to raise yields since, cropland occupied only 5 percent more land, 170 MHa, in 2009 (FAO 2012).

As depicted in Figure 1, the actual land harvested for wheat in India rises as a plateau against a Himalayan-like background, the amount of land that would have been harvested absent the productivity gains since 1960. The 65 MHa of land spared is an area the size of France or more than four Iowas. After the Green Revolution received its name during the 1960s, the land sparing continued into the twenty-first century. Unlike some other revolutions of that era, this one has proven enduring and provides the continuing benefit of reducing cropland expansion to feed ever more mouths.

**FIGURE 1** Actual and potential land used for wheat production, India 1961–2010
Upper segment shows the hectares farmers would have tilled to produce the actual harvest had yields stayed at the 1960 level.

A comparison of FAO reports of meat consumption with GDP per capita shows that Indians, like Pakistanis and Japanese, eat little meat considering their level of affluence (FAO 2012). Do the many vegetarians in India enable the nation to cover fewer hectares with crops, notably, crops for feed? The sparing of land from maize (corn) production in China, where two-thirds of the maize is fed to animals, tests the necessity of vegetarianism for land sparing. As in India, China’s population and income grew, and many moved to the cities since 1961. During the last half-century, China’s population doubled, while GDP multiplied over 45 times (World Bank 2012). As they multiplied and prospered, the average Chinese consumed twice as many calories, including calories from 8 times more eggs and 14 times more milk and meat. Although rising meat consumption in China did contrast with that in India, the multiplication of meat consumption was far slower than the multiplication of affluence: 45 times more affluence gave rise to only 14 times more meat consumption.

While the area of harvested Chinese corn doubled during the half-century, each harvested hectare became more than four and a half times more productive (Figure 2). The 120 MHa of land spared is the equivalent of 2 Frances or 8 Iowas. While disciplined consumers and effective farmers restrained the expansion of cropland area, the extent of Chinese forests reportedly expanded 30 percent from 1990 to 2010 (FAO 2010a, 2010b).

Despite more and wealthier mouths to feed, Indian and Chinese farmers fed their populations while restraining the expansion of cultivated area.

**FIGURE 2** Actual and potential land harvested for maize production, China 1961–2010
Upper segment shows the hectares farmers would have tilled to produce the actual harvest had yields stayed at the 1960 level.

![Graph showing actual and potential land harvested for maize production, China 1961–2010](source: FAO (2012)).
Rising incomes brought better nutrition, but average human food consumption grew much more slowly than rising incomes and began to plateau. Even the appetite for meat in China grew more slowly than affluence. In both countries, agricultural techniques improved the yields for all crops using new high-yielding seed varieties, new crop rotation schedules, synthetic fertilizers, irrigation, weather forecasts, and better management. As these nations became more technologically competent, better information allowed for more precise agriculture and greater resource efficiency.

A look at development of cropland from 1700 to very late in the twentieth century shows that traditionally fertile areas of India and China saw their established cropland areas undergo early intensification (Ramankutty and Foley 1999). Cropland expanded little outward from areas of agricultural development. Unlike southern and eastern Asia, most of North America supported a sparse human population before 1700, and arable land was plentiful to acquire, respectful or not of prior claims. North (and South) American farmers initially favored technologies such as horses and later tractors to extend their reach, producing larger crops from larger areas with little lifting of yields. Greater productivity on North American farms would have to wait.

**United States**

The United States provides a longer, century-and-half trajectory with reliable data. During the longer span, expanding settlement and transportation corridors along with soil exhaustion and new crops affected the course of agricultural production. Increasing the immediate meaning of the example, the United States and other wealthy countries produce much of the world’s food supply today. For example, in 2007, Canada, the European Union, and the United States accounted for about half of global maize and a third of global wheat production (FAO 2012).

Between the 1860s and 2010, the population of the United States grew nine times. Income, as measured by GDP, grew 130 times. Corn production in the United States rose 17-fold from 1866 to 2010 (US Bureau of the Census 1975 and 2012). Yet, more land was planted with corn in 1925 than in 2010 (Figure 3). Note the arc in the corn harvested area separating from the corn production ascent around 1940. For several decades before the 1970s, the area of land used for corn cultivation in the United States declined absolutely, despite growing production. The rise over the last several decades in the extent of the US corn harvest accompanied rising demand for biofuels made from corn. The percentage of US corn used for ethanol grew from about 2 percent of total usage in 1990 to 20 percent by 2006/07 (Trostle 2008) and nearly 40 percent in 2011 (Agricultural Marketing Resource Center 2012). Globally in 2006/07 12 MHa of corn was harvested for ethanol, which approaches the land area of Iowa or one-quarter that of France.
From about 1850 until 1910, expanding numbers of farmers, helped by energetic loggers filling demand for wood as fuel and railroad ties, cleared US forests at an unprecedented rate. Later in the twentieth century, the extent of US forests stabilized while the volume of timber standing in US forests increased appreciably from 1952 to 2007 (Smith et al. 2009).

Over the course of three centuries after European settlement, agriculture in the United States spread rapidly in extent to feed expanding populations, domestically and globally. Subsequently, as farmers learned, agricultural activity shifted to the more productive geographic regions, yields increased, and the expansion of cropland slowed and reversed. Production and cropland area also went separate ways in France, China, and Egypt in the twentieth century—in France nearly simultaneously with the separation in the United States, and, in China and Egypt, within a few decades (Figure 4).

The ImPACT identity

Are the trajectories of the countries examined so far typical? Can we generalize that stabilizing, and perhaps falling, cropland area accompanies rising population and affluence for all regions? To broaden the analysis beyond three disparate countries, we examined world data and reports produced by the World Bank and FAO, generally for the years from 1961 to 2010. We summarize performances with the parameters of an identity that separates factors levered by parents, workers, consumers, and farmers (Waggoner and Ausubel 2002). The identity we use equates the amount of cropland used
with the product of population, affluence, food calories consumed per GDP, crop production per calorie, and land required per unit of production. These factors can be expressed as follows:

\[ Im = Impact = P \cdot A \cdot C_1 \cdot C_2 \cdot T \]

where:

\( Im \) = Cropland (in hectares) taken as the amount of arable land and permanent crops\(^3\) (representing land currently used for crop cultivation, not land that is potentially cultivable), as defined and reported by the FAO.
\[ P = \text{Population (persons)} \]
\[ A = \text{Affluence (in GDP per capita)} \]
\[ C_1 = \text{Consumption 1 (in kcal/GDP), where kcal refers to the annual national or global food supply in kilocalories from both vegetal and animal sources.} C_1 \text{ tracks the dietary response to affluence in calories, an indicator of both hunger and excess.} \]
\[ C_2 = \text{Consumption 2 (in Crop Production Index [PIN]/kcal) using the FAO Crop Production Index, which measures the relative level of aggregate volume of agricultural crop production indexed to a base year.} C_2 \text{ tracks the ratio of crop production for food, feed, fuel, fiber, and tobacco to the supply of food calories. It monitors farmers’ planting choices relative to caloric supply.} \]
\[ T = \text{Technology (in hectares/Crop PIN) tracks how much land farmers use relative to total crop value.} \]

Our identity in this case states that
\[ Im [\text{hectares}] = P [\text{persons}] \cdot A [\text{GDP per capita}] \cdot C_1 [\text{kcal/GDP}] \cdot C_2 [\text{PIN/kcal}] \cdot T [\text{hectares/PIN}] \]

Changes in these factors encompass food demand and production. Straightforward mathematical considerations allow for estimating the annual changes in this product by adding together the logarithmic changes, and hence approximate percentage changes, of the factors. Representing the annual changes by lowercase letters, one can write
\[ im = p + a + c_1 + c_2 + t \]

We recognize the importance, especially over the long term, of the interactions of the variables. For example, one needs prosperity \( A \) for technical change \( T \); with little research and development or capital, yield and \( T \) will suffer. Exporting grain or growing crops not eaten will lower \( C_1 \) and raise \( C_2 \). While the PACT variables are not independent of one another, analytically we can use them to attribute their leverage or contributions to a particular impact, such as land used for crops. The dimensions of the percentages in change per unit of time are commensurable and can be compared, equated, added, or subtracted.

**Global patterns**

The values of the annual changes for the world during the period from 1961 to 2010 of the five ImPACT factors reveal diverse, durable patterns and explain collective impact. Figure 5 displays the patterns created by parents, workers, consumers, and farmers with ten-year moving averages.

Around 1970, the annual increase of global population began to slow. Population continued to grow after that time, but more slowly as seen in the positive, but falling, \( p \) in Figure 5. Averaged over the surrounding decade,
annual population growth dropped from about 2 percent centered on 1970 to about 1.3 percent centered on 2004. Although parents have chosen slower growth, farmers will need to accommodate annual population growth near 0.7 percent for the next half-century according to standard UN forecasts.

The factor $a$, affluence, fluctuated but always grew 1–2 percent per year during the half-century alongside a steady decline in population growth. The economic downturn beginning in 2008 disrupted the general rise of affluence, but that rise persisted.

The change in available food calories per GDP, $c$, fell about 1 percent a year on average during the period. In other words, consumers did not increase caloric consumption in step with increases in GDP. Industrial ecologists call declining $c$ “dematerialization.” Economists use the ratio of dematerialization to changing affluence, $c/a$, plus 1, to generate what they call “income elasticity.” In this case, the global elasticities were 0.25 from 1961 to 2010 and only 0.12 from 1995 to 2010, both low values characteristic of staples.

A comparison of the composite world of Figure 5 with three large countries argues for the general validity of the fact that the poor will use more new income for food calories, while the wealthy will spend it elsewhere. Manifesting Engel’s Law, in country after country after calories exceed minimum levels, caloric intake rises, slows, and may eventually level off as affluence grows. The range charted in Figure 6 from China in 1961 to the United States in 2007 covers a rise in GDP per capita of 498-fold and a
drop in kcal/GDP of 187-fold, which together correspond to a global income elasticity of 0.16. Of course, in the long run, excursions or pauses may occur. Amid the general rise of affluence, an example of declining affluence raising income elasticity for food comes from Romania, a formerly Communist economy on the European margins of the former Soviet Union. During the decade 1989 to 1999, when affluence fell as much as 24 percent in Romania, the number of food calories per person fell no more than 3 percent, and \( C_1 \), the food calories per GDP, actually rose 33 percent. The similar, though not as stark, experiences of Hungary, Poland, and Bulgaria demonstrate that the dematerialization of calories reverses when affluence declines. Returning prosperity restores the downward trend of dematerialization.

Producing a grain such as maize to feed animals represents an alternative to growing a crop such as wheat that directly adds calories to the human food supply. Growing grain for feed increases the ratio \( C_2 \) of crop production per calorie in the food supply because the feed consumed exceeds the amount of animal product. Meat for many is a luxury rather than a staple, and affluence increases meat consumption more than it increases calorie consumption. As the Chinese grew more affluent after about 1970, their meat consumption grew rapidly with little dematerialization. By the 1990s, however, the FAO reported Chinese meat consumption rising less than half as fast as affluence and dematerializing 6 percent per year from 1995 to 2007. As Indian consumers grew more affluent, they behaved differently. They scarcely increased their meat consumption during the half-century to 2010, causing rapid dematerialization and even exhibiting income elasticities below zero. Globally,

FIGURE 6  Dematerialization of food, 1961–2007
The plot of kcal/GDP as a function of GDP per capita for China, India, the United States, and the world shows globally consistent behavior over a range of incomes and cultures.

SOURCES: FAO (2012); World Bank (2012).
average meat consumption per capita dematerialized little from 1980 to 1995, but then as in China, it rose only half as fast as affluence from 1995 to 2007.

The change in crop production per calorie in the food supply \( c_2 \) has risen recently. Starting around 1990, farmers began raising crop production more rapidly than the calories of the food supply rose, lifting the factor \( c_2 \), indicating the diversion of production into higher-value, lower-calorie food crops, feed crops, and especially fuel crops, a phenomenon also noted elsewhere (Jensen and Miller 2011).

From cornfields in Iowa to sugar cane plantations in Brazil and monoculture palm forests in Indonesia, farmers around the world have dedicated cropland to “energy” crops, particularly in the last decade. Encouraged by expensive oil and by government policies, global fuel crops expanded from 4 to 25 MHa in a recent dozen years, as we detail below. Looking ahead, however, environmental and economic objections seem set to discourage expansion of energy crops (Michel 2012; National Research Council 2011).

Finally, over the last half-century, farmers around the globe have consistently squeezed more crop from the same area, annually lowering the hectares of cropland per unit of production, \( t \), by around 2 percent. A combination of agricultural technologies raised yields, keeping downward pressure on the extent of cropland, sparing land for Nature.

Countering the global rise of population and affluence by parents and workers, consumers and farmers restrained the expansion of arable land by changing tastes and lifting yields. The noticeable shrinkage in the extent of cropland as a function of the Crop Production Index since 1990 (Figure 7) provides encouragement that farmers will continue sparing land.

**FIGURE 7  Arable land/Crop PIN, \( T \), for the world, 1961–2009**

To produce an equivalent aggregate of crop production (PIN) in 2009 required only about 35 percent of the land needed in 1961.

While the ratio of arable land per unit of crop production shows improved efficiency of land use, the number of hectares of cropland has scarcely changed since 1990. Absent the 3.4 percent of arable land devoted to energy crops (Trostle 2008), absolute declines would have begun during the last decade.

Looking back and looking forward

If humanity has not already passed its peak use of farmland, are we near? Table 1 presents annual changes of ImPACT factors corresponding to past data and future projections. A comparison of the first two rows in the table shows notable differences between the full half-century and the most recent 15 years. Population growth slowed consistently during the recent period. Always with fluctuations, affluence continued to grow, but more slowly in the recent period. The ratio of food supply to income, and the ratio of arable land to crop production both continued to fall at slightly steeper rates. Unexpectedly, the ratio of crop production to food supply sped up. The net result of all the changes was a sixfold reduction in the growth rate for land used for crop production, which slowed from 0.24 percent per year in 1961–2010 to 0.04 percent per year in 1995–2010. For the coming half-century, 2010 to 2060, and following the organizing principle of the ImPACT identity, we examine changes that can be rationally projected.

Population, changing gradually and steadily, is the most foreseeable of the factors affecting cropland. The 2010 Revision of the UN’s World Population

<table>
<thead>
<tr>
<th>ImPACT factor</th>
<th>Population (P)</th>
<th>Affluence: GDP per capita (A)</th>
<th>Food supply/GDP (C1)</th>
<th>Crop PIN/Food supply (C2)</th>
<th>Arable land/Crop PIN (T)</th>
<th>Arable land (Im)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data for 1961–2010</td>
<td>1.68</td>
<td>1.67</td>
<td>–1.20</td>
<td>0.24</td>
<td>–2.15</td>
<td>0.24</td>
</tr>
<tr>
<td>Data for 1995–2010</td>
<td>1.24</td>
<td>1.53</td>
<td>–1.35</td>
<td>1.04</td>
<td>–2.42</td>
<td>0.04</td>
</tr>
<tr>
<td>2001 projection for 1997–2050a</td>
<td>+0.91</td>
<td>+1.80</td>
<td>–1.26</td>
<td>0.0</td>
<td>–1.70</td>
<td>–0.25</td>
</tr>
<tr>
<td>2012 projections (and alternatives) for 2010–2060b</td>
<td>+0.9 (+0.7)</td>
<td>+1.8 (+1.5)</td>
<td>–1.6 (-1.4)</td>
<td>+0.4 (0.0)</td>
<td>–1.7 (-2.1)</td>
<td>–0.2</td>
</tr>
</tbody>
</table>

NOTES: Rows 1 and 2 show actual average annual changes during the half-century from 1961 to 2010 and during its concluding 15 years from 1995 to 2010. Row 3 shows projections for the half-century from 1997 to 2050 made by two of the authors writing in Population and Development Review (Waggoner and Ausubel 2001). Finally, after a decade of experience since 2001, Row 4 shows our updated projection for 2010–2060, with alternative projections in parentheses.

*aIn the 2001 projection (Waggoner and Ausubel 2001) C1 and C2 differ slightly from those used here. In that publication, C1 denotes Food PIN/GDP, not Food supply/GDP used here; and C2 denotes Crop PIN/Food PIN, not the Crop PIN/Food supply used here.

*bSee text for bases of alternative projections.
Prospects (United Nations Department of Economic and Social Affairs 2012) projects a slowing from the 1995–2010 rate of 1.2 percent per year to a 0.7 percent annual increase during the next 40 years, so our projection of 0.9 percent for the next half-century seems conservative. The UN projection of 0.7 percent for the next 40 years provides a realistic alternative.

Affluence, fluctuating from booms to busts, challenges projection from anything but a long view. From 1961 to 2010 and from 1995 to 2010, affluence rose a bit more than 1.5 percent a year on average. The general upward trend evident in Figure 5 for the past half-century suggests that projecting 1.8 percent in the future seems reasonable and provides a margin of safety for the impact of additional wealth in demanding land for crops. The factual 1.5 percent during the 1995–2010 boom and recession provides a realistic alternative.

Together, population and affluence present a challenge to the environment. Our projections of 0.9 percent and 1.8 percent combine annually to increase this challenge by the GDP growth rate of 2.7 percent. The forces for restraining and possibly reducing land use must be the remaining factors, C1, C2, and T, driven by the restraint of consumers and the effort of farmers.

From 1961 to 2007, per capita food supply rose 27 percent, with meat slightly increasing its share of the whole (Table 2). Changes in the factor C1 measure how many more calories people eat as their income rises. Globally from 1961 to 2010 while affluence rose 1.67 percent annually on average, the number of kilocalories per capita rose at an annual rate of only 0.5 percent (Figure 5). Because the average global citizen today is richer, we expect that demand for new calories will fall even more with new income, resulting in an elasticity of 0.1, just below the 0.12 of 1995–2010. An income elasticity of 0.1 combined with affluence rising at 1.8 percent per year corresponds to a c1 of –1.6 percent. The alternative of affluence rising only 1.5 percent per year and income elasticity of 0.1 corresponds to a c1 of –1.4 percent.

Changes in the factor C2, crop production per kilocalorie of food supply, measure how much more farmers produce relative to the food supply. Simply looking at the ratio of food calories produced to those available for consumption from cereal and oilseed crops shows the large difference between what

<table>
<thead>
<tr>
<th>Year</th>
<th>Total (kcal)</th>
<th>Percent distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>1961</td>
<td>2,201</td>
<td>100</td>
</tr>
<tr>
<td>2007</td>
<td>2,798</td>
<td>100</td>
</tr>
</tbody>
</table>

Source: FAO (2012).
farmers produce and what consumers eat. In 2007, the FAO reported that farmers produced 2.35 billion tons of cereals and 0.79 billion tons of oilseeds. At 4,000 kcal per kg, this translates to the production of 12.57 million billion food calories or nearly 5,000 kcal/person/day for 7 billion people from cereals and oilseeds alone. Because the FAO reported 2,798 kcal/person/day in the world food supply in 2007, we estimate that only half the calories in these crops—and less than half of all calories from crops—entered the food supply. The trend in $C_2$ reflects the changes in the fraction reaching the food supply. Table 1 shows that for the period 1961–2010 $C_2$ rose modestly, and for the period 1995–2010 it rose sharply.

Knowing the causes for the last 15 years of rising $C_2$ would improve rational projection of its future course. Globally, trade cancels out as countries exporting nutrition are balanced by those importing it. An alternative cause, production of more nonfood crops like cotton, hemp, or tobacco, is an unlikely cause of rising crop production per kilocalorie of food supply because FAO’s Nonfood PIN rose more slowly than its Crop PIN, both between 1961 and 2010 and between 1995 and 2010. The FAO’s Livestock PIN also rose more slowly than its Crop PIN, leaving increased animal production as an unlikely source for the rapid rise.

Trying to identify the cause for the sharp rise in $C_2$, we arrive at alternative uses for crops, exemplified historically by George Washington Carver’s discovery of 202 products in the peanut (Merritt 1929) and chemurgy, the industrial use of crops. The new products offered new markets to absorb the bounty produced by farmers. More recently, expensive petroleum, the desire for energy independence, and climate fears combined with carbon accounting have encouraged biofuel mandates and support policies. These policies encourage production with the consequential use of cropland, some formerly used to grow other crops (Wallander, Claassen, and Nickerson 2011). Starting from a baseline of less than 4 MHa in 1995, by 2007 according to the USDA and FAO, nearly 25 MHa worldwide were devoted to crops used for fuels (Trostle 2008). This number exceeds the additions to arable land globally from 1995 to 2010, suggesting that much of the addition to cropland over this period was used to grow fuels.

In Figure 5 we saw that until about 1990, $C_2$ oscillated around zero, more often than not in negative territory. The entry of biofuels as major crops in the mid-1990s helps explain the fourfold increase in $C_2$ from 0.24 percent in the 1961–2010 period to 1.04 in the last 15 years of that period. Cheaper oil, penetration of natural gas into the market for mobility either directly or through electricity, and removal of subsidies and environmental considerations could discourage biofuels, thus sparing land for Nature, and realize the low or zero growth of $C_2$ predicted in the final row of Table 1. As the shortcomings of biofuels become evident to governments and champions of the environment alike, we conservatively project $C_2$ as slowing to 0.4 percent annually, slightly
less than half the 1995–2010 level. This value would allow the growth in nonfood crop production to still exceed growth in the food supply by more than 20 percent. A biofuels bust would lead to a negative value. Alternatively, the ratio of crop production to food supply could very well resume its oscillation around zero, absent new nonfood markets for crops profitable at a global scale. The steady ratio of crop production to diet and its oscillation near zero before 1990, seen in Figure 5, support an alternate $c_j$ projection of 0.0 percent.

Although the global average of the caloric supply—which FAO uses to indicate hunger—has risen more than a quarter and stands well above the requirement of around 2,000 kcal/person/day, the FAO still finds several African countries undernourished. In six representative African countries, each having a food supply of less than 2,000 kcal/person/day in the early 1990s, five increased calories per person and five increased crop production faster than population since Sen reported their low intakes in 1992 (Sen 1994). Despite improvements, however, the FAO reports showed that poor nutrition remained a problem in these countries, where neither increasing affluence nor crop production raised the number of kilocalories per person per day to a level of 2,100. Large differences in rates of population increase and in levels of nutrition lurk within global averages and projections (Alexandratos and Bruinsma 2012). Because demographers now expect that population growth rates in Africa will be responsible for much of the future increase in global population, the extent that such African countries as Nigeria, Congo, and Ethiopia raise yields, extend their fields, import food, and export citizens warrants attention by those concerned about land use. The forces behind the distribution of the food supply lie outside the forces we encompass. Although we cannot say whether distribution will be more equal in the future, our global projections of factors $a$ plus $c_j$ raise average supply to about 3,100 kcal/person/day after 50 years, considerably above the requirement of about 2,000.

Factor $T$, the cropland per unit of crop production or PIN, measures farmers’ intensity versus sprawl, and it is the inverse of tons per hectare, or yield. In the United States, farmers merely maintained corn yields from 1.53 tons/ha in 1866 to a similar value of 1.47 tons/ha on average in the 1930s. Then they lifted the yields over seven decades (Figure 3) to 10.07 tons/ha in the bumper year of 2004. Accordingly, their intensification lowered the hectares per ton, factor $T$, from 0.65 in 1866 and 0.68 in the 1930s to 0.10 in 2004 (US Bureau of the Census 1975 and 2012). The past half-century of $t$ charted in Figure 5 shows the continuation of intensification (i.e., land sparing), globally and for all crops. The joint Organisation for Economic Co-operation and Development (OECD)/FAO outlook for 2011–2020 anticipates agricultural production growing 1.7 percent annually, slower than the 2.6 percent in the previous decade (OECD/FAO 2011).
Although farmers persistently exceed predictions, just as their 1995–2010 performance exceeded our own expectations of 1.7 percent in 2001, we shall adopt the 1.7 percent as a standard, below the actual rise in world average corn yields of 1.8 percent per year achieved between 1983 and 2011 (Figure 8). Annual improvement of 1.7 percent sustained to 2060 would multiply world production per area by 2.3 times. For corn, the average global yield in 2060 would resemble the average US yield in 2010 (Figure 8). Farmers maintaining their 1961–2010 rate argues for the alternate projected rate of −2.1 percent per year.

Looking ahead, one must ask whether a biological limit on photosynthesis will soon constrain the rise in yields and accordingly slow the decline of T and farmers sparing land. The curves of production in Figure 4 suggest saturation or S-curves. However, these show production, not yield, and probably express lack of profitable markets or other incentives to produce bigger crops. Surpluses have long vexed farmers. (They may largely explain the diffusion of the hamburger since the 1950s, as farmers sought markets for their corn.) A clue about biological limits lies in the rise of winning Iowa yields in the National Corn Growers Association contest, which continue to rise and maintain their margin far above both the US and world average yields of this important crop (Figure 8). For, say, the next half-century, this clue suggests

**FIGURE 8  Corn yields, 1983–2011**
The highest maize yields in Iowa entered in the National Corn Growers Association contest compared with US and world averages. The percentages show average annual increases.

no approaching biological limit and supports our projection of improvements of $T$ shown in row 4 of Table 1.

The changes of all global ImPACT factors shown in row 4 of Table 1 project a trajectory in global cropland change over the next 50 years. Figure 9 shows the combined consequence of the predicted values for each of the ImPACT factors on the extent of global arable and permanent crops. Of course, wild cards may confound projections, but we contend that our assumptions are conservative, transparent, and based on historical trajectories. After detailed regional analysis, Alexandratos and Bruinsma (2012) sound the theme of inevitably slower growth of demand for food production similar to our theme of peak farmland. According to our projection, by 2060 some 146 MHa could be restored to Nature, an area equal to one and a half times the size of Egypt, two and a half times France, or ten times Iowa.

The alternate values in the last row of Table 1, one by one, raise the predicted 146 MHa restoration of cropland. Slowing population growth to 0.7 percent instead of the 0.9 percent or more rapid dematerialization from −1.4 percent to −1.6 percent would each spare an additional 132 MHa. Uninterrupted yield growth, −2.1 percent instead of −1.7 percent, or the almost complete demise of nonfood agricultural production, could spare an additional 252 MHa.

Slower population growth, restraint in taste perhaps reflecting concern about obesity and the increasing popularity of a vegetarian diet, abandonment of biofuels, and continued improvement of technology sum to reversal in demand for land. Sustained for 50 years, such a performance would take humanity from its current peak use of farmland into an era of land sparing.

**FIGURE 9** Peaking farmland: Extent of global arable land and permanent crops 1961–2009 and our projection for 2010–2060

![Graph showing peaking farmland from 1960 to 2060](image)

Global arable land and permanent crops spanned 1,371 MHa in 1961 and 1,533 MHa in 2009, and we project a return to 1,385 MHa in 2060.

Undoubtedly, the use of irrigation and fertilizer has fueled the growth in yields. Their use has grown more efficient over the last half-century and especially over the last 20 years. The magnitude of irrigation can be measured as the area equipped for irrigation in millions of hectares, or as the volume of water consumed in cubic kilometers. Globally, irrigated area expanded at 1.4 percent per year from 1900 to 1950, accelerated to 2.1 percent per year from 1950 to 1995, and then slowed to 0.5 percent per year from 1995 to 2003 (Freydank and Siebert 2008). In 2010, the total area actually irrigated was 253 MHa globally, and total consumption of irrigation water was 1,277 km³/yr (Siebert et al. 2010). Historical data for irrigation water use are difficult to find, but according to the FAO the annual rate of growth for all water withdrawals was about 1.4 percent (FAO 2012; United Nations Environment Program 2002). In the United States, the withdrawal of irrigation water peaked in 1980, and has since declined relative to crop production by an average of 2.0 percent per year (Kenny et al. 2009). Water-conserving practices and shifting irrigation to more humid and cooler areas achieved this improvement, which may foreshadow global adaptation to a stable irrigated area while still allowing the continued decline of cropland per crop value, $T$.

Nitrogen fertilizer, like water, raises yields and shrinks the area of land needed to produce a particular quantity of crop. After World War II, economical synthetic nitrogen became available for farming, and its increasing consumption outraced crop production in the 1960s by as much as 10 percent, but the increase slowed by the 1980s to an annual rate between 0.5 percent and 2.5 percent. Global data show that fertilizer consumption per unit of crop production rose only 0.72 percent annually from 1970 to 2009 (International Fertilizer Association 2012; FAO 2012). Because nitrogen comprises 16 percent of protein, neither humans and other animals nor plants grow and survive unless roots extract it from the soil, a process that soon exhausts soil stocks. The 50 g per day of protein that each person requires equals 3 kg per year of nitrogen. Room for greater precision and less waste in nitrogen application can be deduced from the fact that, globally, the average fertilizer application per capita of 14.6 kg per year still far exceeds the human requirement for nitrogen in protein.

Conclusions

Our analysis encompasses the leverage on cropland exerted by parents, workers, consumers, and farmers. Since 1960, their combined behaviors have spared areas of land that are immense when compared with what continuation of birth rates, appetites, yields, and other factors might have led us to expect. India and China alone have spared an area more than three times the
size of France or a dozen times Iowa. Absent the slowing population growth, evolving tastes, and improving agricultural practices, unimaginable destruction of Nature would have occurred.

The past 50 years have already witnessed important peaks for environment and resources. The rate of increase of world population peaked around 1970 and has slowed considerably since then. Peaks of forest destruction also have passed with a transition from less to more forests in many countries and regions. By the 1980s wooded areas in all major temperate and boreal forests were expanding. After 1990, growing stock expanded in many forested countries (Kauppi et al. 2006), and during 1990–2010 the density of forests grew in all world regions, albeit unevenly (Rautiainen et al. 2011). Like farms and their crops, the productivity of forests providing wood products has risen. Meanwhile consumption has fallen as e-readers replace paper and as demand for other wood products, such as railroad ties and telephone poles, has declined. As we hinted above, peaks of farmers’ use of nitrogen and water may also have passed.

The peak of cropland anticipated in Figure 9 does not derive from depletion of the resource. The envisioned cropland peak rises in part from another peak, that in the rate of population growth. Whether affluence will peak depends on the continuing competition between seemingly boundless desire for more and acceptance of the essential and possible.

In any case, the calories in the food supply per GDP, the use of affluence for nutrition, begins the inventory of tools to counter the environmental challenge of population and affluence. And unlike humanity’s striving for affluence, its striving for food has limits that help meet the challenge. The survival level near 2,000 kcal/person/day sets a lower limit. The upper limit at, say, 4,000 set by obesity is the one that moderates the ratio of food to GDP. While the dematerialization common to staples such as food and calories helps counter the challenge of population and affluence, the limit of obesity adds another effect. Producing grain to feed animals represents an alternative to crops that directly adds calories to the food supply and so increases the ratio of crop production per calorie in the food supply. Fortunately for the sparing of cropland, meat consumption is rising only half as fast as affluence.

We were surprised by the recently rising ratio of crop production to calories in the food supply. The growth of biofuels illustrates the wild cards that can disrupt projections. If government policy or opportunity encourages farmers to grow alternative crops that do not reach the food supply, less cropland will be spared than anticipated by thinking only food comes from cropland. Cotton and flax illustrate that alternative crops are not new. In the past, proponents have encouraged alternative crops to relieve farm surpluses and depressed prices.

This broad sweep should not obscure the crucial, final role of yields and the shrinking of hectares per unit of crop production. The new varieties of
the Green Revolution in the 1960s, bred to exploit better fertilization, water supply, and crop protection, accelerated the shrinking of cropland. Precise interventions in DNA, fertilizer, irrigation, pest control, and weather forecasts offer improving tools to help continue lifting yields.

Again, however, wild cards remain part of the game, both for and against land sparing. As discussed, the wild card of biofuels confounded expectations for the past 15 years. Most wild cards probably will continue to come from consumers. Will people choose to eat much more meat? If so, will it be beef, which requires more land than poultry and fish, which require less? Will people become vegetarian or even vegan? But if they become vegan, will they also choose clothing made from linen, hemp, and cotton, which require hectares? Will the average human continue to grow taller and thus require more calories? Will norms of beauty accept obesity and thus high average calories per capita? Will a global population with a median age of 40 eat less than one with a median age of 28? Will radical innovations in food production move humanity closer to landless agriculture (Ausubel 2010)? Will hunger or international investment encourage cropland expansion in Africa and South America? (Cropland may, of course, shrink in some countries while expanding in others as the global sum declines.) And will time moderate the disparities cloaked within global averages, in particular disparities of hunger and excess among regions and individuals?

Allowing for wild cards, we believe that projecting conservative values for population, affluence, consumers, and technology shows humanity peaking in the use of farmland. Over the next 50 years, the prospect is that humanity is likely to release at least 146 MHa, one and a half times the size of Egypt, two and a half times that of France, or ten Iowas, and possibly multiples of this amount.

Notwithstanding the biofuels case, the trends of the past 15 years largely resemble those for the past 50 and 150. We see no evidence of exhaustion of the factors that allow the peaking of cropland and the subsequent restoration of Nature.

Our analysis of cropland, concentrating on sparing land for Nature, overlooks everything about farmers except their efficient choice of crops and yield. Wilderness wanderers enjoying Nature at a relic cellar hole, perhaps with a lilac or rusting tool nearby, are witnessing farmers’ hopes dashed by surplus crops.

Another 50 years from now, the Green Revolution may be recalled not only for the global diffusion of high-yield cultivation practices for many crops, but as the herald of peak farmland and the restoration of vast acreages of Nature. Almost 20 years ago we made a wild surmise about land sparing (Waggoner 1994). Now we are confident that we stand on the peak of cropland use, gazing at a wide expanse of land that will be spared for Nature.
Notes

1 The FAO measures chronic hunger by calories: “The average minimum energy requirement per person is about 1800 kcal per day. The exact requirement is determined by a person’s age, body size, activity level and physiological conditions such as illness, infection, pregnancy and lactation.” [http://www.fao.org/hunger/en/].

2 Throughout this essay, “tons” refers to metric tons.

3 When corn is processed into ethanol, a considerable fraction remains as dry distillers grains and solubles (DDGS), a medium protein feed ingredient. Correcting the 40 percent of the corn production used for ethanol production for this feed ingredient lowers the percentage to about 31 percent of corn production in 2011, raises the corn fed to animals to about 48 percent, and moderates an alarm that ethanol production was equaling the corn fed to animals. See the Agricultural Marketing Resource Center Ethanol Usage Projections & Corn Balance Sheet at [http://www.extension.iastate.edu/agdm/crops/outlook/cornbalancesheet.pdf].

4 To reach 2010, it was necessary to estimate cropland area for 2010 as equal to 2009 and kilocalories per person per day for 2008–2010 as the average of 2002–2007.

5 For the rest of this essay, the terms “cropland” and “arable land” refer to the FAO category of “Arable land and Permanent crops.”

6 To estimate annual changes denoted by lowercase letters, we use the convention \( x = \ln(X_f/X_i)/(f-i) \) where \( X \) denote the value of \( X \) for final and initial years of the period being analyzed and \((f-i)\) is the number of years in the interval. This operation is justified for small changes (i.e., \( x \ll X \)) such that higher orders can be neglected. In our case, we get the equation \( im = p + a + c_1 + c_2 + \ell \).

7 Parents’, consumers’, and producers’ leverage of impact by \( p, a, c \), and \( \ell \) is identical to changing impact. Nevertheless, we might have chosen other dimensions than GDP, calories, and crop production. For example, we might have chosen a dimension of median family income rather than GDP for \( A \), food expenditures rather than calories for \( C_1 \), and tons of a crop rather than the crop production index for \( C_2 \). Replacing GDP with a slower rising median income would slow \( a \). That choice, combined with replacement of calories and its upper limitation with almost unlimited expenditure on food, would speed the rise of \( C_1 \) and lift income elasticity. Then combining that choice of faster rising food expenditures with replacement of the crop production index with a slowly rising production of a single sort of crop would lessen producers’ \( C_2 \). While alternative dimensions might shift the blame and credit among parents, consumers, and producers, the consequent impact of changing hectares would remain the same. We chose GDP because it encompasses all economic activity, is generally used to measure national economies, and is almost universally used in industrial ecology. We chose calories because, as FAO has decided, they measure hunger and excess. And we chose the crop production index because it encompasses all crops and reflects farmers’ combined ability to lift yields and respond to demand.

8 If the dimension of GDP is removed from the vertical axis of Figure 6 and it becomes instead kcal/person, the correlation is less and the slope becomes income elasticity \( b \) rather than changing intensity of use \( C_1 \). The new coordinates and alternate estimation still support the conclusion that the elasticities are low and similar among the three countries and the world. For 1961 to 2007, the elasticities thus estimated are: China 0.2, India 0.1, US 0.3, and world 0.3. For the three countries and world combined, \( b \) is 0.1.


10 Angola, Central African Republic, Chad, Ethiopia, Mozambique, and Zambia.

11 The Contest Winners in Figure 8 are the highest yields for all classes, irrigated or non-irrigated, in the National Corn Yield Contest, except the remarkable yields of 20 to 25 tons/ha grown by Francis Childs during 1997–2005.
References


United Nations Department of Economic and Social Affairs. 2012. World Population Prospects, the 2010 Revision. Reference to 0.7% projection for medium variant 1010-2060 «http://esa.un.org/unpd/wpp/unpp/panel_population.htm».


National Wealth

Partha Dasgupta

National accounts are descriptors. They describe the state of an economy and form the raw material for both assessing its performance and prescribing policy. National accounts embody information essential for economic evaluation.

The framework for national accounts currently in use throughout the world, however, suffers from extreme narrowness. Vast quantities of information relevant for economic evaluation do not appear in them. Some don’t because the appropriate data are difficult, even impossible, to collect; but others don’t because the theory and practice of economic evaluation haven’t asked for them. The widespread demand for “green” national accounts has arisen in recent years because of a growing recognition that contemporary national accounts are an unsatisfactory basis for economic evaluation. The qualifier green signals that we should be especially concerned about the absence of information on society’s use of the natural environment.

In recent years collaboration among a few economists has given rise to a comprehensive theory of national accounts based on primary normative principles. Attempts have also been made by these authors to put the theory to work on international data.1 In this essay I review that theory in a non-technical manner and illustrate it with data from India.

The theory I am reviewing here takes as its starting point the view that the ultimate purpose of economic evaluation is the protection and promotion of human well-being across the generations. It instructs governments and international agencies to go beyond even green national accounts by reclassifying certain classes of goods and services and adding others that are currently missing. The theory goes on to show that the objective of economic evaluation should be a comprehensive notion of wealth (adjusted for population and the distribution of wealth in the economy), not gross domestic product (GDP) nor the many other ad hoc indicators of human well-being that have been advanced in recent years, such as the United Nations’ Human Development Index.

1 See Hamilton and Clemens (1999); Dasgupta and Mäler (2000); Arrow, Dasgupta, and Mäler (2003); Dasgupta (2004); and Arrow et al. (2012a, b).
The theory is even more specific. It says that under a wide range of conditions, changes in the circumstances of an economy should be judged on the basis of their effect on the economy’s wealth per capita, adjusted for the distribution of wealth. A corollary is that net domestic product, NDP, should be the object of interest in flow accounts, not GDP.

Wealth is the social value of an economy’s stock of capital assets, comprising (i) produced capital (commonly known as “manufactured capital”), (ii) human capital (the size and composition of population, education, and health), and (iii) natural capital (e.g., ecosystems, sub-soil resources). To be sure, the notion of “capital assets” extends today to institutions, knowledge, culture, religion, and even norms and practices. The theory I am reviewing acknowledges that those intangible objects are indeed assets, but that the appropriate procedure is to regard them as the “social infrastructure” within which the more grounded capital assets (categories (i)–(iii)) get allocated and are put to use.

The objects that link well-being across the generations and wealth are shadow prices, which are the social values of capital assets. The social infrastructure just alluded to is a factor determining those shadow prices. Since wealth is a stock, the theory I am reviewing here tells us that particular attention should be paid to an economy’s capital accounts. That suggests there should be a shift in interest from gross investment to net investment, the latter being the value of the change in the quantity (or quality) of capital assets. It bears emphasis that in widening the range of capital assets to health, education, and natural capital, the theory extends the notion of investment from its commonplace use, namely, deferred consumption. The way to view net investment is to regard it as any decision that enhances future consumption.

The next several sections develop the theory. These are followed by a discussion of ways to estimate shadow prices and tentative estimates from Indian data that serve as illustrations.

Economic evaluation

Economic evaluation is an essential activity. We evaluate our own household’s economic circumstance, we worry that green is a vanishing color in our cities, we are puzzled by the growing inequality in living standards in much of the world, we are concerned about the character of economic development in our own country, and we have views on the role mutual trust, personal honesty, economic incentives, and governance play in our lives.

Economic evaluation is also a compulsive activity. We commonly discuss the economic performance of nations and of states within our nation with our friends and neighbors, read reports and commentaries in newspapers and magazines, listen to the views of experts on radio and television, and occasionally take our concerns to our political representatives. We evaluate economies so as to assess and prescribe, and we do so because we are curious to know answers to five questions:
(1) How is the economy doing?
(2) How has it been doing in recent years?
(3) What should our forecast of the economy’s performance be under “business as usual”?
(4) How is the economy likely to perform under alternative policies?
(5) Which policies should we support?

Questions 1 and 2 direct us to assess the performance of economies. Questions 4 and 5 are prompted by the need to prescribe policy. Because any change in policy is a perturbation to “business as usual,” obtaining an answer to 3 is essential if we are to respond to 4 and 5.

Assessment should precede prescription. It could be foolish to look for improvements if an economy is doing well, for there would be a risk of making things worse. If on the other hand the economy has been doing badly, we should search for policies and institutional changes that are likely to improve matters. An immediate problem would arise if we don’t know what to look for when evaluating the economy’s performance. National accounts are crucial for this evaluation. They are needed if we are to address any of the five questions.

The Brundtland Commission’s view

In their classic Report on humanity’s dependence on Nature, the Brundtland Commission (World Commission 1987) took the future explicitly into account. The Commission recast questions 2 and 3 by interpreting the object of interest as current and future human needs. In an oft-quoted passage, the Report defined sustainable development as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs.”

Say an economy has enjoyed economic growth by investing in buildings, factories, mines, and transport facilities. If those investments have been accompanied by a depletion of the economy’s forests, wetlands, and aquifers, we should wish to ask whether economic growth there is sustainable or whether future needs will be compromised. The widespread acknowledgment today that national accounts should include the use and depletion of the natural environment—we call the latter natural capital—can be traced to the Brundtland Commission’s work.

Notice that the quoted passage from the Report mentions human needs, not human well-being. The requirement is that future generations will have no less of the means to meet their needs than we do ourselves; it doesn’t ask for more. As needs are the austere component of well-being, a country’s economic development could be sustainable in the Commission’s sense without it being judged to be optimal, even efficient.

Notice also that the Commission’s definition is directed at sustaining the determinants of human needs (food, clothing, shelter, and so on). Observe,
though, that the determinants would be available only if the society in question were to have access to adequate quantities of assets for producing them. Sustainable development in the Commission’s sense would then seem to require that, relative to their populations, each generation should bequeath to its successor an access to at least as large a quantity of what may be called a society’s “productive base” as it had itself inherited from its predecessor. The stress is on “access,” not “possession,” because trade with others enables an economy to depend on assets it does not itself possess. In the world economy, however, access means possession.

The term “productive base” suggests a full array of capital assets. The Commission could be interpreted as saying that sustainability analysis involves tracking movements in an economy’s portfolio of assets over time. The problem is that it doesn’t say how the portfolio should be aggregated to produce a numerical index. We need a numerical index because it is important to know the rates at which assets can be traded off for one another while preserving the economy’s productive base. The viewpoint I adopt here is that the criterion for sustainable development should be intergenerational well-being and that the required index of an economy’s productive base should be derived by analysis. That derivation would then yield the rates at which the assets can be substituted for one another in economic evaluation.

The source of current and future consumptions in an economy is its productive base. The economy inherits its productive base from its past. However, “productive base” isn’t another name for “capital assets.” An economy’s productive base certainly includes the assets it has inherited from the past, but it includes a lot more. There are components of a productive base—the quality of an economy’s institutions is an example—that impart value to capital assets but are not assets themselves (at least not in a sense that proves fruitful in economic evaluation).

**The idea of wealth**

In contrast to the Brundtland Report, I adopt the view that by sustainable development we should mean that well-being across the generations does not decline over time. Imagine now that we are successful in constructing a numerical index of an economy’s capital assets, with the property that its

\[ V(t) = \sum_{s=t}^{\infty} \frac{U(s)}{(1+\delta)^{s-t}}, \quad \delta \geq 0. \]  

(F1)

I have deliberately left the time horizon unspecified. In discussions of climate change, it is now common to consider a horizon of 100 or more years.

The theory I am sketching here does not depend on equation (F1). It is nevertheless useful to keep the “utilitarian” formulation in mind because it is very commonly in use.

---

2 A common formulation of intergenerational well-being is a weighted sum of each generation’s well-being. To formalize, assume time is discrete. Let \( s \) and \( t \) denote time \( (s \geq t) \). If \( U(t) \) is an aggregate measure of the well-being of all who are alive at \( t \), intergenerational well-being at \( t \) is taken to be

\[ V(t) = \frac{U(t)}{(1+\delta)^{t-t}} = U(t), \quad \delta \geq 0. \] (F1)
movement over time faithfully tracks movements in intergenerational well-being. The index could then be used in sustainability analysis.

It transpires that a suitably weighted sum of an economy’s capital assets is the required index. The weights are “shadow prices” and the index is a comprehensive measure of wealth. An asset’s shadow price is the contribution an additional unit would make to intergenerational well-being, other things being equal.

Dasgupta and Mäler (2000) and Dasgupta (2004) identified circumstances under which intergenerational well-being averaged across the generations and wealth per capita track one another exactly. Put another way, the authors identified conditions under which wealth per capita increases if and only if intergenerational well-being averaged across the generations increases. I invoke that proposition in what follows.3

In order to estimate wealth, we need first to list the items in an economy’s productive base that are to be included (I have named them the economy’s “capital assets”; categories (i)–(iii) above). We are then required to weight those items (the shadow prices), and then estimate the weighted sum of the capital assets (the economy’s wealth). In sustainability analysis we are required to estimate the change in wealth per capita over time. That means estimating net investment per capita, with the hope that it is positive over the period being studied (see footnote 3).

The word “investment” could be thought to embody a sense of robust activism, but that’s only because national income statisticians have traditionally limited the term’s use to the accumulation of produced capital. When the government invests in roads, the picture that’s drawn is one of bulldozers leveling the ground and tarmac being laid by men in hard hats. The theory I am reporting extends the notion of capital beyond produced assets to include human capital and natural capital. So we are obliged to stretch the notion of “investment” also, which now includes the growth of renewable natural resources such as ecosystems. To leave a forest unmolested so as to enable it to grow would be to invest in the forest. To allow a fishery to restock under natural conditions would be to invest in the fishery. That suggests investment amounts to deferred consumption. But the matter is subtler. To provide food

---

3To formalize, assume time is discrete. Let \( s \) and \( t \) denote time \( (s \geq t) \) and let \( L(s) \) be the population size at \( s \). A common formulation of intergenerational well-being averaged across the generations is this:

Let \( U(s) \) be an aggregate measure of the well-being of all who are alive at \( s \). Intergenerational well-being at \( t \), averaged across the generations, is taken to be

\[
V(t) = \frac{1}{\delta} \left( \sum_{s \geq t} U(s) / (1 + \delta)^{s-t} \right) \left( \sum_{s \geq t} L(s) / (1 + \delta)^{s-t} \right), \quad \delta \geq 0. \tag{F2}
\]

(Compare equation (F1) in the previous footnote with equation (F2).)

Let \( k(t) \) denote the per capita stock of capital asset \( i \) at \( t \) and \( p_i(t) \) its shadow price. Wealth at \( t \) is then

\[
W(t) = \sum p_i(t) k_i(t). \tag{F3}
\]

Denote a small change in any variable by \( \Delta \). Then the fundamental proposition concerning the normative basis of national accounts is that \( V(t) \) in equation (F2) and \( W(t) \) in equation (F3) are related by the proposition

\[
\Delta V(t) \geq 0 \text{ if and only if } \sum p_i(t) \Delta k_i(t) \geq 0.
\]
to the undernourished not only increases their current well-being, but enables
them to be more productive in the future; and the latter feature makes even
consumption among the poor an investment. This suggests investment is any
decision that enhances future consumption. No doubt all that sounds odd, but
theory and empirics taken together should determine our usage of technical
terms, not preconception or customary usage. So, by “net investment” in an
asset we should mean the value of the rate of change in the stock of that asset.
At its core, economic evaluation involves studying asset management.

Unraveling the productive base

Classifying the components of an economy’s productive base is more complex
than could appear at first blush. Capital assets are patently a part of what one
loosely means by a “productive base,” but there are intangible objects such
as institutions that would not fit the intuitive meaning of “assets” easily and
yet are a society’s backbone. Let us classify them.

Capital assets

“Durability” and “tangibility” would appear to be the defining characteristics
of capital assets. That assets are durable objects is almost implicit in the term.
Tangibility is a less obvious requirement; but it facilitates measurement, and
measurement is a prime requirement of national accounts. Certainly, those
were basic considerations among quantitative economists of the past. Modern
economists have followed a more difficult trail. The theory I am describing
here extends their framework.

Capital assets can be consumer durables (houses, television sets) or
producer durables (agricultural land, factories, forests). The notion of assets
has, however, become more elastic in recent decades. It has become stan-
dard practice to regard “knowledge” as a form of capital (but note that it’s
intangible), education and health as forms of “human capital” (they too are
intangible), and “institutional,” “social,” and “religious” capital (which are not
only intangible, but to many scholars quite beyond the remit of the language
of “assets”—Arrow 2000; Solow 2000).

What should be included on the list of capital assets is up to a point a
matter of convenience. It depends on the purpose to which the notion is put.
The purpose here is to determine the form of national accounts most useful
for sustainability and policy analysis. That is what has shaped the classifica-
tion that follows:

(1) Produced capital (roads, buildings, ports, machinery, and equip-
ment). Contemporary national accounts divide an economy’s investment
in produced capital into sectoral investments (manufactures, mining, for-
estry, agriculture, and so on). Other things being equal, raising the rate of
investment in reproducible capital increases the size of an economy’s future productive base. In this limited sense investment is deferred consumption. The standard practice is to identify shadow prices of produced capital with their market prices. But there are exceptions. A previous literature on social cost–benefit analysis found shadow prices of produced capital to be different from market prices because of market imperfections. Attempts were made to construct simple formulas for estimating shadow prices in the presence of canonical forms of market imperfections (Dasgupta, Marglin, and Sen 1972; Little and Mirrlees 1974).

(2) Human capital (population, education, health). When economists discuss human capital, they usually mean education. But health is another aspect of human capital. Since both education and health are embodied in people, they together make up human capital. Of course, people qua people are themselves assets. So we partition human capital into three categories:

(i) Population (size and demographic profile). Conceptually, this category raises the deepest problems because people are the reason well-being should occupy the ethical core of sustainability and policy analyses and are as well the means to the realization of well-being. Demographic changes are both the cause and effect of changes in the other components of an economy’s productive base.

Each person should be regarded as a separate asset, endowed at any point in time with knowledge, skills, and health. So the population profile matters, including as it does both age and sex distributions. Arrow, Dasgupta, and Mäler (2003) showed that estimating the value of population when population movements are acknowledged to be related to movements in, say, GDP per head poses special problems. In the absence of a demographic theory, population growth is frequently taken to be exogenously given. Often too, analysts work with population averages. Dasgupta (2004) showed it is a reasonable approximation in that case to incorporate population by expressing the quantities of the other capital assets in per capita terms.

(ii) Education. The kind of knowledge included in human capital is in many ways tacit, something a person acquires and is private to him. But as teachers are painfully aware, knowledge isn’t costlessly transferable from one person to another; so, education involves costs and should be seen as investment in people.

Knowledge and skills are both ends and means. Reading is a pleasurable activity and thus a form of consumption. The direct benefits of literacy are then reflected in the social value of consumption. But being able to read is also of indirect value: literacy enhances a person’s productivity, implying higher earnings and greater access to consumption goods. Education therefore offers joint benefits.

The above are the private benefits of education. But when someone learns to read and write and becomes numerate, not only is it she who enjoys
a benefit, others do too because they can now communicate with her. The acquisition of education thereby confers external benefits (externalities). Total benefits of education are the sum of private and external benefits.

(iii) Health. Health brings both direct and indirect benefits. An absence of pain is one of the defining characteristics of well-being; and, other things being equal, healthier people are also more productive. Disability lowers productivity, which can be measured in principle and often is measured in practice. At an extreme, malnourished people lack energy and stamina and are at the bottom of the earnings schedule. Improvements in health over time can thus be measured in terms of increased productivity.

Investments over time in a person’s health are complementary, in that damage to one’s health at an early stage of life cannot easily be compensated for in later years. Although the ability to use knowledge for one’s ends is often thought to be “innate” (“in the genes,” as the saying goes today), recent work in developmental biology has found that what goes by the name “innate ability” is in part acquired and can be traced to a person’s early nutritional and disease history. The epidemiological environment in which an infant grows influences the development of his innate abilities.4

I have dwelt on health as means to the enjoyment of a good life. But good health also brings direct benefits. By bringing relief from pain, medication improves the quality of consumption and hence an improvement in the quality of life. Medication would thus appear to be a form of consumption.

It is an intriguing and fortunate fact that expenditures that bring relief from pain, or avoid disability, or immunize someone against infectious diseases also help to extend their life. Medication, immunization, and medical treatment thus give rise to joint products: relief from pain and discomfort, greater productivity, and a longer life.

So far I have classified the benefits an individual enjoys from an improvement in her health. But when someone is immunized against an infectious disease, not only is it she who enjoys benefits, others do too, because they are now less likely to catch the disease. So, immunization confers externalities. Those external benefits should be added to private benefits if we are to estimate the total benefits of health improvements.

(3) Natural capital (local ecosystems, biomes, sub-soil resources). Natural capital is of direct use in consumption (fisheries), of indirect use as inputs in production (oil and natural gas; the wide array of ecosystems), and of use in both (air and water). The value of natural capital may be utilitarian (as a source of food, or as a keystone species)—many people call this its use-value; it may be aesthetic (places of scenic beauty); it may be intrinsic (primates); or it may be all these things (biodiversity). Their worth to us could be from

---

4 For a recent overview of the subject, see Dasgupta (2012).
extraction (timber, gum, honey, leaves, and bark) or from their presence as a stock (forest cover), or from both (watershed). The stock could be an index of quality (air quality) or quantity. Quantity is sometimes expressed as a pure number (population size); in various other cases it is, respectively, (bio) mass, area, volume, depth. Even quality indexes are often based on quantity indexes, as in “parts per cubic centimeter” for measuring atmospheric haze.

The classification offered above is useful in economic evaluation because it is based on the reasons we value Nature. For understanding the changing landscape in contemporary economies, however, the classification in the Millennium Ecosystem Assessment (MEA) (2005a–d) is more useful. There natural capital is classified in terms of the kinds of services they provide.5

There are three further categories of goods that are components of an economy’s productive base but that should not be included on the list of capital assets when computing an economy’s wealth.

Knowledge

Science and technology and the arts and humanities are commonly regarded as capital assets. They are the seat of discoveries and inventions: Nature’s laws, abstract theorems, formulas, algorithms; and historical and cultural narratives. The products are durable, intangible public goods. By calling knowledge a “durable public good,” we mean that a piece of knowledge can in principle be used by anyone who wishes to make use of it, and can do so repeatedly. A mathematical theorem knows no cultural or national boundaries.

Many regard scientific and technological knowledge as the most significant feature of an economy’s productive base. Note though that for someone to make use of a piece of knowledge three conditions must be met: (a) a recognition by the person that the knowledge exists; (b) the person should have the relevant skills (human capital); and (c) there should be no legal or social barrier (patents and copyrights) to applying the knowledge. Research and development involves investments in codified knowledge. Frequently, though, knowledge can be obtained freely from external sources (for example, once patents have expired). These aspects of knowledge suggest that its instrumental role lies in conferring value to capital assets. Before the advent of the iron age, for example, iron ore was of little value. But once the art of extracting the ore and converting it into usable iron was discovered, the ore acquired value. Its shadow price rose.

5 Theirs is a four-way classification of ecosystem services: (i) provisioning services (food, fiber, fuel, fresh water); (ii) regulating services (protection against natural hazards such as storms; the climate system); (iii) supporting services (nutrient cycling, soil production); and (iv) cultural services (recreation, cultural landscapes, aesthetic or spiritual experiences). Notice that cultural services and a variety of regulating services (such as disease regulation) contribute directly to human well-being, whereas others (soil production) contribute indirectly (by providing the means of growing food crops). Bateman et al. (2011) provides a fine discussion of the valuation of ecosystem services.
Some types of knowledge are also consumption goods. Contemplating a beautiful idea enhances well-being directly, as does reading literature. They have direct value.

Institutions

When scholars write on the progress or regress of societies, it is not uncommon today for them to refer to “religious capital,” “social capital,” “cultural capital,” and more broadly “institutional capital.” Nevertheless, as in the case of knowledge, it is best not to regard institutions as capital assets. Let us see why.

Institutions include the myriad formal and informal arrangements to which people belong. Those arrangements influence the allocation of resources both at a moment in time and through time. Households, firms, communities, and the State are obvious examples of institutions. But institutions also include a broader kind of “enabling assets,” such as the rule of law, social norms of behavior, and habitual social practices. They form the social infrastructure within which goods and services are allocated.

What role do institutions play in economic evaluation? They influence the value of what we are here calling capital assets. Suppose the State apparatus in a country is corrupt, in particular the judicial system is unreliable. Because people find it difficult to protect their property rights, the value of the assets they own is small, other things being equal. Corruption reduces the social worth of capital assets.7

The above example shows that the key to the efficacy of institutions is the extent to which they elicit trust among people. Consider as an extreme example two islands that are identical in all respects but for the fact that in island 1 households trust one another completely whereas in island 2 households do not trust one another at all. Despite their identical configurations of capital assets, the structure of shadow prices would be wholly different. The reason is that households in island 1 would begin to trade with one another, whereas in island 2 they would remain autarkic. In short, institutions aren’t enough. If societies are to progress, people must trust one another and have confidence in their institutions.

Institutions influence the composition of consumption, saving, and the character of future institutions. There are institutions that foster progress by having in place a structure of incentives that encourage people to allocate goods and services to their most productive uses. Well-developed competitive markets, tight social norms and codes of conduct, and good governance can combine to help create and maintain trust and confidence. But mutual trust and the cooperation that can result from it can’t be guaranteed even under sound

---

6 Sociologists speak of “symbolic capital” as well. See Bourdieu and Passeran (1990).
7 No doubt the purpose of corrupt practice is to enhance the market worth of assets owned by the practitioner. In the text I am referring to the social worth of assets.
institutions. Opportunistic behavior can beget opportunistic behavior. Institutions can founder under a cascade of opportunism unless there are checks and balances in place by virtue of the presence of countervailing institutions.8

A prime example of the mutual influence of institutions and economic development is provided by financial institutions. Expansion of financial institutions helped initiate economic development in late medieval Europe. By the same token, the technological changes that accompanied economic growth over the centuries encouraged the creation of new financial institutions. (But not necessarily for the better: ideal financial markets would not perform in the way they are currently doing in many countries.)

Financial institutions create and legitimize financial capital. But financial capital does not form a separate category of assets, it consists of claims to capital assets. Someone who owns shares in a firm has a claim to a share of the market value of the firm. His shares reflect the market value of the assets to which he has claim. The same holds for currency and bonds. They are claims to whatever bundle of goods and services the holder wishes to purchase, subject only to the condition that the market value of the bundle does not exceed the value of the currency and bonds he holds. To include financial capital in the wealth of nations would be to make a mess of economic evaluation.

Time

A society’s productive base can change for the better or worse simply by the passage of time. Consider an oil-exporting country. Imagine that the ground rent on oil increases over time. Our country therefore enjoys capital gains on its reserves without having to engage in any form of capital accumulation. Its productive base expands.9 Similarly, the hypothetical country enjoys an expansion of its knowledge base by deploying scientific advances elsewhere without paying for that privilege. In both circumstances our country enjoys an expansion of its productive base simply by waiting. Its wealth increases correspondingly. Previously I noted that exogenous population growth is another example.

These examples tell us that the value of capital assets can depend explicitly on time. In short, the calendar date can matter.

Shadow prices

A commodity’s shadow price is its social scarcity value, a notion familiar in social cost–benefit analysis. Three pieces of information are required for estimating shadow prices:

---

8 On the role of trust in sustainable development, see Dasgupta (2011).
9 The same logic says that, other things being equal, an oil-importing country would find its productive base shrinking as the international price of oil rose over time.
(i) A descriptive model of the economy moving through time, including not only technological possibilities and ecological processes, but also preferences, tastes, personal and social values, and policies.

(ii) The size and distribution of the economy’s capital assets at the date at which the economic evaluation is undertaken.

(iii) A conception of intergenerational well-being.

Because they combine ethical values (item (iii)) with forecasts of future economic possibilities (item (i)), shadow prices have long proved to be contentious—for example, in social cost–benefit analysis. Problems are compounded in sustainability analysis, because in estimating wealth one is obliged to impute prices to stocks of capital assets. That explains why shadow prices are most often estimated in a piecemeal way.

Definition

Consider first consumption goods. If durable (clothing, shelter), they embody consumption services over time. If non-durable (food), they vanish on consumption. So we may as well use the term “consumption” as it occurs in everyday language, where reference is made to consumption rates. We will take it that consumption is a flow, and that it affects the flow of societal well-being at that date and that date alone.11 The shadow price of a consumption good is the contribution an additional unit of it would make to the flow of societal well-being. Provided the consumption goods are private (i.e., their use isn’t associated with significant externalities), their market prices are frequently reasonable approximations of their shadow prices.

For capital assets, the notion of shadow prices is trickier. Assets are durable and, except for consumer durables, they are inputs in the production of other goods. An asset’s shadow price is the contribution an additional unit would make to well-being across the generations, other things being equal.

The factors determining intergenerational well-being are the consumption goods and services enjoyed by contemporaries and by people across the generations. It follows that an asset’s shadow price is the social worth of the change in the consumption stream an additional unit of it would give rise to, other things being equal.

Estimation

There is an acute shortage of usable data pertaining to the valuation of ecosystem services. Apart from sub-soil resources and land (in those places where there is an active land market), the natural capital that has been studied most

---

10 Continuing debates over the magnitudes of “consumption discount factors,” which are shadow prices of future consumption relative to current consumption, illustrate this.

11 Habits, like memories, are stocks, not flows.
by environmental economists is carbon in the atmosphere. In Arrow et al. (2012a), natural capital was restricted to land, forests as stocks of timber, sub-soil resources, and carbon concentration in the atmosphere. On the other hand, the authors expanded the notion of what is generally understood to be human capital, by including longevity. Their most significant finding was that health dominates the wealth of nations, by an order of magnitude greater than all other forms of capital assets (see Table 1 below). Since this is a striking finding and as the methods deployed for estimating the shadow price of health aren’t widely known, I provide an account below. Here I sketch the methods they deployed for valuing atmospheric carbon and education, respectively.

Regional shadow price of carbon

The concentration of carbon in the atmosphere is a global capital asset, but at current levels (390 parts per million) it is acknowledged to have a negative global shadow price. That price was estimated by Stern (2006) to be minus $85 per ton, and by Nordhaus (2008) to be minus $8 per ton. The difference is almost entirely attributable to differences in the authors’ choice of the rate of time discount ($\delta$ in equation (F1); see footnote 2). In a well-known work, Tol (2009) estimated the global shadow price of carbon to be minus $50 per ton.

Being a public asset, the global shadow price of carbon is the sum of country-specific shadow prices. The latter should be expected to differ across regions if for no other reason than that the effects of global climate change will differ across regions. Nordhaus and Boyer (2000) decomposed the global price of carbon into regional prices. In their work on sustainability and the wealth of nations, Arrow et al. (2012a) used the Nordhaus–Boyer estimates.

Despite the search for international agreements, the atmosphere remains an “open access” resource. Let $K(t)$ denote the concentration of carbon at date $t$. If $X(t)$ is net global carbon emission at $t$, the increase in concentration over a brief interval of time $\Delta t$ is $X(t)\Delta t$. That amounts to the change in carbon concentration in the atmosphere during that interval. If $P_n$ (a negative number) is the shadow price of carbon concentration in country $n$, the shadow value to country $n$ of the increase in concentration is $P_n(t)X(t)\Delta t$. That is the contribution of carbon concentration to the change in the country’s wealth.

Valuing education

Assume investment in education earns a social rate of return, $r$. Being a social rate, $r$ includes the benefits others enjoy when someone acquires education. Let $T$ be the average number of years of educational attainment in the country. Assuming a steady state as a first approximation, the amount of education per worker is then proportional to $e^{rT}$. The constant of proportionality is obtained from the wages of unskilled workers, for whom $T = 0$. The stock of education capital is then obtained by multiplying education capital per worker by the number of workers, the quantity being adjusted for mortality during the working life.

If we now assume that the labor market is sufficiently competitive to imply that the marginal productivity of education capital equals the real wage, the shadow price of education capital would equal the total real wage bill divided by the stock of education capital. The value of wealth in the form of education would then be the product of the shadow price and the stock of education capital.

Valuing health

Health is both a means and an end. Someone who is healthy is more productive than he would have been otherwise (an “indirect benefit” of an improvement in health); he also experiences a greater sense of well-being (the “direct benefit”). The idea then is to decompose improvements in health into their indirect and direct components. The former would be reflected in increases in output and consumption, data for which are available, albeit in very approximate forms, in national accounts. So identifying the indirect benefits doesn’t pose special conceptual difficulties; it’s the direct benefits that require unearthing.

The direct benefits of improvements in a person’s health are twofold: (a) relief from (alternatively, avoidance of) discomforts and disabilities, and (b) extension of life. That (a) and (b) are conjoined is a deep fact; (a) and (b) are joint products of improvements in early nourishment and avoidance of infectious diseases.

I first illustrate (a) by means of a single-period model and then (b) by means of a two-period model (a generalization of the model to many periods is sketched in footnote 15). In order to estimate the social benefits of health programs, one should add spillovers of the benefits of such programs across individuals (the externalities) to the private benefits studied here.

Health as consumption

Consumption is denoted by a numerical index $C$ and the flow of well-being by $U$. $C$ is a determinant of $U$, but the effectiveness of $C$ as a source of a person’s well-being depends on his state of health. Health, a capital asset, is denoted by the numerical index $H$. Assume $U$ is an increasing function of $H$ and $C$. 

H and C are in part complements, in part substitutes. It is simplest then to multiply H and C and so construct a composite commodity, HC. In that case the person’s well-being function is $U(HC)$. The product, HC, transforms consumption, measured in its own units, into consumption measured in “efficiency” units. Put another way, measured in its own units, consumption contributes more to personal well-being if health improves. A marginal improvement in the person’s health yields the benefit $CU'(HC)$.

The analysis can be readily extended if we assume that H is maintained over the person’s lifetime of, say, T years. If r is the social discount rate, the shadow price of health capital, $P_H$, is

$$P_H = CU'(HC)\left[1 - e^{-rT}\right]/r.$$  

Suppose it requires continual expenditure (health care) E to maintain H. Then the present value of health costs would be $E(1 - e^{-rT})/r$. Further expenditure on health should be advocated if $CU'(HC) > E$.

**Health as extension of life**

The above model assumes life expectancy is fixed. I now study ways to value increases in life expectancy.\(^1\)

Broadly speaking there are two ways to estimate those benefits. One is to seek people’s “willingness-to-pay” for improvements in health (the “stated preference” approach); the other is to infer the value of health from “revealed preference.” Here I discuss how improvements in life expectancy can be valued by a study of revealed preference. At the basis of this approach is the “value of a statistical life” (VSL), the most common method for estimating which is to uncover differences in wages for jobs involving differential risks. To illustrate, imagine that someone is willing to pay up to US$85 annually to reduce a mortal risk to himself from 1/10,000 to 1/30,000. Denote the reduction as Δp. In our example, $\Delta p = 1/10,000 - 1/30,000 = 1/15,000$. The value of a statistical life, $V_L$, is defined by $V_L \Delta p = $85, or $V_L = 420/\Delta p =$1.3 million (which is the figure in use for India).

Contrary to what is widely thought, $V_L$ is not the value of life, it is the value of a statistical life. In the present example it means 15,000 identical people would be willing to spend up to $1.3 million collectively to reduce the expected number of deaths due to this particular risk by 1. Across countries $V_L$ has been found to vary positively with GDP per capita. As choices in so many other spheres of life reveal, the tradeoffs that the poor are obliged to make differ cruelly from those that can be accommodated by the rich. In a study of

---

\(^1\) $U'(HC)$ is the derivative of $U$ with respect to the composite good HC.

\(^1\) The material is taken from Arrow et al. (2012b).
cross-country estimates of $V_L$, Viscusi and Aldy (2003) found that $V_L = b Y^{0.6}$, where $Y$ is GDP per head and $b$ is a positive constant.

To illustrate how estimates of VSL can be used to value life expectancy, imagine someone who is alive in period 1 and will survive to period 2 with probability $\rho$. We are interested in the case where expenditure in health can raise $\rho$. Suppose the price that must be incurred to obtain $\rho$ is $H(\rho)$. We assume $H$ is an increasing (and, plausibly, a convex) function of $\rho$, and that $H(0) = 0$. For simplicity we now imagine that health expenditures yield no benefits to the person other than an increase in the probability of survival.

Well-being in any period is a function of consumption in that period. We write that as $U(C)$. Imagine the person faces competitive markets for consumption in period 1 (labeled as $C(1)$) and for contingent consumption in period 2 (labeled as $C(2)$). The market price of $C(2)$ relative to $C(1)$ is $P$. By “contingent consumption” we mean consumption that would be available to the person should he survive. He pays $P$ in period 1 for the right to a unit of period-2 contingent consumption.

The person begins life with total wealth $W$, expressed in period-1 consumption. His expected lifetime well-being, which we write as $V$, is assumed to be

$$V = U(C(1)) + \rho U(C(2)). \quad (1)$$

To have a meaningful problem we have to assume in addition that there exists a consumption level $C^*$, such that $C^* > 0$ and $U(C^*) = 0$. $C^*$ has been called the “welfare subsistence rate” (Meade 1955). That means $U(C) < 0$ for low values of $C$ and $U(C) > 0$ for high values of $C$. From the perspective of preparing national accounts, this requirement, which is forced upon us if we consider variations in life expectancy, makes health an unusual capital asset.

To simplify our calculations, we assume that the population consists of a large number of identical people and that the survival probabilities are independent of one another. In period 1 the individual chooses $C(1)$, $C(2)$, and $\rho$ so as to maximize (1) subject to his budget constraint

$$C(1) + PC(2) + H(\rho) = W. \quad (2)$$

Before proceeding to the optimization exercise, I note that (trivially) the rate of increase in $\rho U(C(2))$ when $\rho$ is increased marginally is $U(C(2))$. Equation (1) then says that the benefit to the individual of a marginal increase in survival probability $\rho$ is $U(C(2))$. I noted above that health economists express $U(C(2))$ in terms of consumption goods and then estimate it by uncovering a person’s revealed willingness to pay for a small increase in the probability of survival. This is the value of a statistical life (VSL).

---

14 Notice that the person faces a single budget constraint. If he survives to period 2, he consumes the $C(2)$ he purchased in period 1. Obviously he does not receive consumption in period 2 if he does not survive to that point.
$H'(\rho)$ in contrast is the VSL when the latter is measured in terms of the rate at which the person is able to transform wealth into an increase in the probability of survival (the budget constraint, (2)). At a personal optimum the two are of course the same (see equation (3c) below). Assuming that the individual’s optimum $\rho$ is positive, the first-order conditions of his optimization exercise are

\begin{align}
U'(C(1)) &= \mu, \quad (3a) \\
\rho U'(C(2)) &= \mu P, \quad (3b) \\
U(C(2)) &= \mu H'(\rho), \quad (3c)
\end{align}

where $\mu (> 0)$ is the multiplier associated with the budget constraint (2). Equation (3a) says that $\mu$ is the value of period-1 consumption in terms of period-1 well-being.

Recall that the person is one among a large population of identical individuals; moreover, mortality risks are independent of one another. It follows that in market equilibrium, $P$ equals the individual’s optimum choice of $\rho$. So $P$ is the actuarially fair price.

Equations (3a)–(3b) and the fact that $P = \rho$ imply

\begin{equation}
C(1) = C(2) = C \text{ (say)}. \quad (4)
\end{equation}

Equations (3a) and (4) allow us to re-express equation (3c) in consumption units as

\begin{equation}
U(C)/U'(C) = H'(\rho). \quad (5)
\end{equation}

In competitive equilibrium either side of equation (5) can serve as the VSL. The equation also confirms that $U(C) > 0$ in equilibrium.

I have sketched a complete model with perfect markets only to illustrate the connection between people’s revealed willingness to pay for reducing the risk of death and the costs incurred in bringing about that reduction. In fact of course there are few annuity markets in poor countries. The VSL in those countries won’t be related to the cost of raising life expectancy in the manner of equations (3a–c). But that doesn’t matter for the efficacy of the use of VSL estimates in valuing life expectancy. That is demonstrated in the following footnote.\(^{15}\)

\(^{15}\)To formalize the basis of VSL, let time, a discrete variable, be denoted by $s$ and $t$, respectively, with $s \geq t$. Let $\rho(s)$ be the probability of surviving to age $s$ conditional on being alive at age $(s-1)$. Consider someone aged $t$. Suppose his well-being at age $s$, should he survive to that age, is forecast to be $U(s)$. Assume next that the probability he will survive beyond age $T$ is zero (thus $t < T$). Let $\delta$ be the rate at which the person discounts his future well-being. ($\delta$ could be interpreted as his rate of impatience.) At age $t$ his expected lifetime well-being would be

\begin{equation}
V(t) = U(t) + \sum_{s=t}^{T} \Pi \left[ \rho(t + s) / (1 + \delta)^{s-t} \right]. \quad (F4)
\end{equation}

If well-being is taken to be the unit of account, VSL at age $t$ is $\partial V(t)/\partial \rho(t+1)$. From equation (1) it follows that

\begin{equation}
\text{VSL}(t) = U(t+1) + \sum_{s=t+1}^{T} \Pi \left[ \rho(t + s) / (1 + \delta)^{s-t} \right]. \quad (F5)
\end{equation}
Wealth estimates for India

To illustrate how estimates of VSL can be used to value health capital for an individual, imagine that the latter is the expected discounted years of life remaining multiplied by the value of an additional life year, assumed independent of age. To aggregate across people in a country, one can then use country-specific population and mortality data along with the estimated VSL to calculate the value of an additional year of life.

In their work on the wealth of nations, Arrow et al. (2012a) estimated changes in wealth per capita over the period 1995–2000 in five countries (Brazil, China, India, United States, and Venezuela). Their choice of countries was in part designed to reflect different stages of economic development and in part to focus on particular resource bases. Their surprising finding is that in recent years the contribution of improvements in health to well-being has far outweighed the contribution of the accumulation of produced capital and education. Table 1 gives their breakdown of wealth in India in 1995. It indicates that 98 percent of wealth in India is in its health capital, a figure so striking as to beggar belief. So let us search for weaknesses in the procedure adopted by Arrow et al. (2012a).

An obvious weakness is that the figure for VSL has been assumed to be independent of age. It could be conjectured VSL declines as one gets very old. The evidence from the United States, however, suggests that that may not be so: a good proportion of the expenditure on health in the US is incurred among the old. Moreover, the age distribution in India is still tilted toward the young, implying that the contribution to health capital by the old is relatively low. We should bear in mind also that the estimate of health capital in Table 1 doesn’t include the value of health as a consumption good (as discussed above). The direction of error in the figure isn’t therefore clear.

The relatively small figure for wealth per capita in the form of natural capital ($2,300) was surprising to me when I first read the estimate. But it is easy to explain: vast swathes of natural capital are missing from the estimate. Official ignorance of the state of an economy’s stock of natural capital assets should now be a matter of extreme embarrassment to governments.

<table>
<thead>
<tr>
<th>TABLE 1 Composition of wealth per capita in India in 1995 (US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Produced capital</td>
</tr>
<tr>
<td>Natural capital</td>
</tr>
<tr>
<td>Human capital (education)</td>
</tr>
<tr>
<td>Human capital (health)</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

SOURCE: Arrow et al. (2012a), Table 5 (rounded figures).
In a review of the empirical literature on forest services (carbon storage, ecotourism, hydrological flows, pollination, health, and non-timber forest products), Ferraro et al. (2012) have found little that can be reliably used in wealth estimates.

Table 2 reproduces estimates of the percentage change in per capita wealth in each category over the five-year period 1995–2000. Wealth per capita in the form of natural capital declined, albeit by a small percentage (0.8 percent). It will be noticed that the gains in produced capital (42.3 percent) and education are an order of magnitude greater than the gain in health capital. That’s because of the initial size of health capital, which is vastly larger than the size of the other forms of capital.

To assess the sustainability of Indian economic development over the five-year period, I calculate the percentage change in wealth per capita. I sum the net changes that occurred in the values (per capita) of the capital assets and divide the sum by the value of aggregate wealth (per capita) in year 1995. That figure is 0.9 percent. Converted into an annual figure, the growth rate is 0.18 percent. Though small, it is a positive number, which implies that subject to the very many qualifications I have noted above, economic development in India followed a sustained pattern at the end of the last century.

**Economic growth as growth in wealth**

The central finding in the literature I have reviewed here has been that by “economic growth” we should mean growth in an economy’s wealth (per capita), not growth in GDP. As an adjunct, the finding says that when we worry about the distribution of well-being in a society, we should worry about the distribution of wealth, not income. The point of economic policy wouldn’t be to maximize the rate of wealth accumulation, it would be to determine optimum accumulation rates, which could well be to aim at a gradual increase in wealth.
Of course, in a hugely distorted economy a government may be able to have its proverbial cake and eat it too. It could be that by a judicious choice of policy it is possible for a developing country today to aspire to the accumulation of wealth per capita and enjoy modest increases in GDP per head as well for a while. Only further work in wealth accounting will enable decisionmakers to know whether that is indeed possible.

Wealth is an intuitively congenial economic object. We commonly use the term “wealthy” to describe rich people and nations and on occasion recall that Adam Smith’s great treatise was on the wealth of nations. Our discussion has, however, revealed that behind what is a commonplace term lie unexpected complexities. Many of the complexities are conceptual while others concern the nuts and bolts of practical national accounting.

Without demand, supply rarely comes forth. Because there has been no official demand for environmental accounts, there are few empirical estimates of the value of environmental services, a lacuna in urgent need of repair. When and if that is done, human health and natural capital can be expected to be the dominant components of national wealth. I conjecture that the pair together will be found to swamp produced capital, even education, in importance.

These are early days in the preparation of wealth accounts. But it is sobering to realize that 60 years ago estimates of national incomes were subject to uncertainties of a magnitude people are likely to think no longer exists in current estimates. That said, we take contemporary estimates of national income too much at face value. Official estimates are silent on the proportion of incomes that go unrecorded. Estimates of transactions falling outside the market system or operating within a black market system suggest that the errors in official estimates of national income are substantial.

However, even if figures for natural resource stocks were available, the deep problem of imputing values to them would remain. Market prices may be hard facts, but shadow prices are soft. The issue isn’t merely one of uncertainty about the role natural capital plays in production and consumption possibilities, it is also a matter arising from the fact that people differ in their ethical values. Estimating the sensitivity of wealth to shadow prices should become a routine exercise in national accounts. We should expect wealth estimates to be presented as bands, not exact figures.

That people may never agree on the wealth of nations is, however, no reason for abandoning wealth as the object of interest in sustainability analysis. Our ignorance of the economic worth of natural capital remains the greatest barrier to understanding the history of economic development. Until that ignorance is lifted, policy analysis will remain crippled and sustainability will continue to be a notion we admire but cannot put into operation.
Note

Paul Demeny has been an inspiring editor even for those of us outside the mainstream of demographic research. He has encouraged us to think hard on demographic matters and has published our work even when we were unsure of its worth. And he has welcomed diverse entry points to economic demography. Readers will recognize that this essay has been fashioned with that virtue of his in mind.

References


The Last Eighty Years: Continuities and Change

Vaclav Smil

The first time Paul Demeny asked me to write for *PDR* was in 1985 in a taxi in (as it was) West Berlin, going to a Dahlem Conference on Resources and World Development. At that time I was still spending about half of my time working on China, and so I wrote an article on the country’s much-improved food supply. During the nearly three decades that followed, it was Paul’s repeated nudging that led me to ask “what next?” And so I wrote (always given a much-appreciated free range) about diet and heart disease, planetary warming, how many people the Earth can feed, environmental services, the nitrogen cycle and global population, eating meat, global catastrophic events, national socioeconomic and strategic trends, and harvesting the biosphere.

After so many different topics for *PDR*, I was not immediately sure what the present essay should be about. Eventually I proposed a topic, but then I decided to do something different, something obvious: to look back at the last eight decades in order to appraise the true nature of human advances and to emphasize our continuing inability to foresee fundamental changes.

Looking 80 years ahead did not tempt me at all because all truly long-range predictions are nothing but fairy tales, and hence must inevitably draw on an imagination that invariably tends to be either too feeble or too ridiculous. Looking back is actually one of the best ways to make clear the futility of looking far ahead: when such efforts are seen in retrospect, they seem forced, naïve, awkward, and, above all, wide of the mark in their inevitably prejudiced selectivity and (often risible) bias reflecting the changing concerns of the day (as global warming displaces acid rain and as economic crisis displaces global warming) and the herd instinct of learned prognosticators.

Growing frequency of forecasting and scenario-writing (by self-appointed experts, by think tanks of every leaning and consultancies, and by governments and corporations) might be seen as a proof that such products offer real insights and valuable guidance. But even a brief reflection dispels that impression. Evolution of human societies is marked by continually and gradually un-
folding processes whose time-specific outcomes can be usefully approximated, and sometimes even more narrowly quantified. But the persistent impossibility of making long-range realistic forecasts several generations into the future is inevitable for three principal reasons.

First, while it is possible to forecast continuation (and intensification or attenuation) of some notable unfolding trends, spatial and temporal accuracy of such forecasts remains elusive. Second, it is most unlikely that even a near-perfect quantitative forecast will embrace all the qualities that make up a new reality, that it will capture the complexities whose interactions will produce subtly or profoundly altered wholes. Third, all those predictable gradual, evolutionary processes are often diverted, strongly modified, and even reversed by profound discontinuities and saltations, by surprises whose occurrences are utterly unpredictable, whose multifaceted impacts cannot be fully foreseen even after they begin to unfold.

As for the general trends, in 1932 it was clear that many unfolding changes affecting populations, economies, and societies would not only continue but would most likely intensify during the coming generations. Indeed (to choose just a few notable examples), expectations of declining fertility, rising per capita GDP, decolonization, and increasing participation of women in the labor force have all become realities. But I cannot imagine that in 1932 any forecaster (or any prescient committee) would have identified fertile Catholic Southern Europe (where even divorce, not to mention birth control and abortion, was a dogmatic anathema) as the region that would experience the continent’s fastest decline in fertility to rates deep below the replacement level and foresee that the Protestant Nordics would be Europe’s most diligent long-term procreators.

Similarly, it is most unlikely that anyone in 1932 would have identified Nigeria, Britain’s richest African colony, as a paragon of post-colonial Africa’s failed performance, corruption, and economic decrepitude (despite the country’s enormous natural endowment and its discovery of rich oil and gas resources); or would have foreseen that the southern part of Korea, in 1932 a poor and exploited Japanese colony, would become the world’s 15th largest economy in 2012, with per capita GDP almost the same as its former colonizer (despite the country’s near complete lack of resources and the fact that by 1954 it was reduced to rubble). And as for women’s participation in the labor force, who would have said in 1932 that among the affluent countries it would be the US, at that time already the world’s richest economy, where nearly 60 percent of all women would hold jobs, while in Japan, a country where virtually all women had worked in traditional agrarian society, the rate would be among the lowest in the rich world?

To illustrate the second point, imagine that in 1980 a brilliant forecaster is immune to impressions left by the unfolding crises of the day—world crude oil prices rising close to $100/barrel in today’s monies; Western car companies
downsizing in a panic; consumers demanding small, energy-efficient models; newly ascendant mullahs getting comfortable in Tehran; and the Soviets taking over Afghanistan—and is able to pinpoint the continuing expansion of America’s vehicle fleet during the next generation by putting its total at nearly 250 million cars and trucks in 2005.

But would he also (amidst all those sudden changes and fears that were engendered by the second oil price crisis in eight years and in the total absence of anything called an SUV) predict that in 25 years the best-selling passenger vehicles would be massive truck-like designs, some of them direct derivations of armored military assault machines, that would be sold under an incongruous label of sports utility? Would he predict that the average performance of America’s new passenger cars would double between 1975 and 1985, but then remain stagnant for the next quarter-century? Would he conclude that by 2005 two of the country’s three domestic automakers, despite their near-record production run, would be very close to bankruptcy? And yet it was precisely the combination of these new realities, rather than the total count of vehicles on the road, that had the most important consequences for the US automotive industry and for the country’s economy.

As for the third reality, the last 80 years have been repeatedly punctuated by such unpredictable leaps (mostly social and political rather than, as commonly but wrongly believed, those of a technical nature), by advances whose rapid adoption brought multifaceted changes (mass jet-powered airline travel is an excellent example, with impacts ranging from unprecedented ease of intercontinental travel to endless worries about airborne terror), and by reversals that had often no less profound effects (the US pivoting from the world’s largest creditor to the greatest debtor nation). At the same time, this retrospective exercise will try to demonstrate that the most fundamental changes in the human condition during the past 80 years have not arisen from those surprising post-1930 discoveries and developments but rather from remarkable scaling up of techniques, machines, and processes that were already commercialized or commonly deployed by 1932 and from no less remarkable improvements in their qualities (that is in efficiency, convenience of use, reliability, and durability).

**Changing populations**

In order to appreciate how far we have traveled, it is necessary to list the key contrasts. In 1932 global population stood at just over 2 billion compared to 7.1 billion in 2012. The United States had 125 million people (about 40 percent of its 2012 total), China (at that time in disunited turmoil) was below 700 million (1.35 billion now), India (trying to slip from British rule) had just passed 280 million (nearly 1.2 billion now), and even relatively slow-growing European countries have grown considerably: for example, France added
about 21 million people during the past eight decades to reach today’s roughly 63 million. Total fertility rates (TFRs) were well above the replacement level even in the richest countries and very high throughout large parts of Asia and Africa, while by 2012 the demographic transition had run its course everywhere in the rich world, and TFRs are well below the replacement level not only in Europe and North America but also in China (at about 1.6 China’s rate is virtually identical with the European mean), and the global average has been reduced to less than 2.5.

Eight decades ago life expectancies at birth were just above 60 years in the richest nations (where they are now close to or above 80) and they were below 40 in many parts of Africa and Asia, but even China’s mean is now close to 75 years. In rich countries the pattern of dying was dominated by infectious diseases, there were no antibiotics and no polio vaccine, while in poor countries the combination of widespread malnutrition, communicable diseases, and unclean water kept infant mortality very high, commonly in excess of 200 per 1,000 live births. In contrast, today’s lowest infant mortality rates are less than 3 per 1,000 (Sweden, Japan), the global mean is about 40, and China’s rate is below 20. Today cardiovascular diseases and cancers are the main cause of adult death throughout the affluent world, a pattern now shared by megacities of Asia.

Growing populations, falling death rates, and rising longevity were made possible by improvements in overall quality of life that resulted from a nearly uninterrupted growth of GDP, from widespread industrialization followed by the rise of the service sector, from the expansion of international trade, from relentless urbanization, and from a much-increased mobility of populations. In turn, these epochal changes were primarily the function of rising energy consumption that led to adequate nutrition, improved water supply, better housing and sanitation, and greatly reduced mortality from communicable diseases (thanks to preventive inoculation) and bacterial infections (thanks to antibiotics). All of these changes, of course, now afford lives of material affluence for hundreds of millions of people.

Higher energy use and improved nutrition

In 1932 most of humanity (hundreds of millions of peasants in Asia, Africa, and Latin America) continued to live, as in preceding millennia, in a wooden age as they burned fuel wood (directly or used it to make charcoal), crop residues (mostly cereal straws), and dried dung. In global terms these biofuels supplied more than a third of all primary energy, and more than 90 percent in China and India. Global per capita supplies of fossil and biomass fuels (with the latter supplying at least a third of the total) averaged about 35 GJ/year (an equivalent of two tonnes of dry wood or 1.5 tonnes of good steam coal), while the national rates were only 20 GJ/capita in Japan but, even after falling some 30 percent between 1929 and 1932, about 150 GJ (six
tonnes of excellent coal) in the United States. Eighty years later the world was consuming more than 500 EJ of primary energy, with traditional biofuels supplying less than 10 percent of the total, and per capita means were more than 300 GJ (equivalent of roughly 7 tonnes of crude oil) in the US, about 170 GJ in Germany and Japan, and, after a rapid rise since the early 1980s, about 80 GJ in China.

In the early 1930s only the industrialized countries of Europe and North America and Soviet Russia were large producers and consumers of coal whose combustion dominated their energy use (more than 80 percent of all fossil fuel). And only the US, with its exceptionally high car ownership (60 percent of households had a car before the Depression), was a large crude oil consumer (also by far the largest oil producer). No oil was produced in the Arabian Peninsula, and Soviet Russia relied only on the old Baku fields. Eighty years later crude oil was the largest component of global energy use (about 30 percent of the total), with coal still slightly ahead of natural gas. Russia, Saudi Arabia, and the US were the largest producers of crude oil, and two additional forms of primary energy became important since the early 1930s: hydro and nuclear electricity, each supplying about 5 percent of the world's primary energy (and a nearly four times higher share of electricity).

By 1930 the United States had just accomplished its industrial transition from steam to electricity (America's large factory manufacturing became almost fully electrified by 1929). But although some 70 percent of households were connected to the grid, electricity use was still fairly limited. Before the Depression, Americans were buying more light bulbs and more radios (50 percent ownership by 1930), but refrigerators were uncommon (in just 10 percent of households), TV was not yet commercialized, and there was no household air conditioning. Eight decades later America's average per capita consumption of electricity was about 13 times the 1932 mean as more than 90 percent of households had color TVs, telephones, and refrigerators and more than 80 percent owned clothes washers and dryers, electric ranges, microwave ovens, and an assortment of electronic devices. And electrification is now advancing rapidly in Asia, especially in China where it powers the country's fast-paced urbanization, its massive export-oriented manufacturing, and its new high-speed trains.

Eight decades ago food production was dominated by traditional, low-yielding cultivars. New short-stalked, high-yielding varieties of rice and wheat and hybrid corn, advances achieved thanks to the application of highly rewarding classical plant breeding techniques, changed all that: hybrid corn was introduced during the 1930s; much-improved rice and wheat cultivars were widely adopted during the 1960s. In 1932 American corn yields (at less than 2 t/ha) were unchanged since the 1880s; now the mean is over 10 t/ha. Similarly, European wheat harvests were generally less than 2 t/ha, while yields of 8 t/ha are now common. In 1930 soybeans, cultivated for millennia in East Asia, were a less important crop in the US than sweet potatoes; now
soybeans are the country’s second largest crop, yielding more than 80 Mt/year, and soybean meal (the residue after expressing oil for cooking) is the main source of protein in animal feeding. In 1932 draft animals were still indispensable not only in Asia but also in Europe and even in parts of the US and they, together with dairy cattle, pigs, and poultry, produced valuable manure whose recycling was the leading source of soil nutrients.

In the early 1930s most of humanity subsisted on overwhelmingly vegetarian and barely sufficient diets produced by traditional agricultures with almost no inputs beyond the closed, and fully solar, system of cultivation and animal husbandry. Dietary transition from traditional intakes dominated by carbohydrates (staple cereals and tubers) to a more varied diet with a higher share of animal foodstuff was underway only in the industrialized countries, but meat was still expensive as its small-scale production on mixed farms was inefficient: it took a year or more of feeding (any available biomass) before pigs were ready for slaughter, and chickens were not ready for at least three to four months.

Eight decades later sub-Saharan Africa is the world’s only region with minimal external inputs into farming. In all affluent countries, as well as in the most productive regions of Asia and Latin America, food production is critically dependent on high (and in many places still rising) inputs of fossil energies, directly as fuel for field machinery (including irrigation pumps) and indirectly for the production of nitrogenous fertilizers, pesticides, herbicides, agricultural machinery, and trucks. Obviously, these external energy inputs are high in Iowa or Holland; but because intensive cultivation of rice demands high rates of nitrogen fertilizer applications (more than 500 kg N/ha a year), China is now even more dependent on fossil fuels for food production than is the US: without Haber–Bosch nitrogen, it could not feed some two-thirds of its population.

Pork and chicken for Western consumers, as well as for inhabitants of megacities in Asia and Latin America, come overwhelmingly from centralized animal feeding operations where the animals receive balanced rations of carbohydrates (mainly corn) and protein (mainly soybeans) supplemented by micronutrients; pigs now reach a slaughter weight of more than 100 kg in less than six months, and broilers are ready for the market in just 35–42 days. As a result, traditionally vegetarian India is now the only populous country with extremely low meat consumption: Annual meat supply (all rates in carcass weight/capita) is now around 50 kg in China, 80 kg in Brazil, and 120 kg in the US.

**Gradual gains vs. fundamental breakthroughs**

What has driven these profound changes in energy use and food production? Perhaps the clearest conclusion derived from looking back is that gradual
development of established devices, processes, and systems has been cumulatively more important than even the most notable fundamental scientific and technical breakthroughs achieved since 1930. This may seem counterintuitive: after all, we are constantly told that all innovation has been accelerating and that we live in an era of unprecedented knowledge breakthroughs. But the fact is that when seen from a truly historical perspective (longue durée), the unparalleled era of scientific and engineering progress occurred between 1865 and 1914 when an amazing concatenation of advances created most of the fundamental technical and economic realities of the twentieth century.

The briefest list of those veritable pre–World War I saltations—qualitative and quantitative leaps in our capabilities that have come to define modern economies—would be limited to four defining components of modern civilization: energies, prime movers, materials, and information flows. The three epochal energy advances were the development of oil and gas industries, the invention and rapid maturation of electric systems, and the introduction of new mechanical prime movers. Coal combustion was nothing new by the mid-nineteenth century, but the pre–World War I decades saw the rise of hydrocarbon (oil and natural gas) industries, and the post-1930 decades saw the rise of these two fuels to become the leading energizers of modern world. Just a few decades after the commercial introduction of electricity, widespread adoption of electric lights, motors, and electrochemical processes revolutionized industrial production, eased or eliminated many household chores, and opened the means of inexpensive instant telecommunication.

The new prime movers of unprecedented power and efficiency included steam turbogenerators, large hydraulic turbines, and combustion engines (Otto cycle burning gasoline and Diesel cycle burning heavier fuels) deployed in all forms of transportation as well as in agriculture (tractors, combines, irrigation pumps). Inexpensive methods of large-scale steelmaking and aluminum smelting began to produce the two most ubiquitous metals of modern civilization, found in infrastructures ranging from roads to buildings and in machines ranging from the largest ship to the largest airplanes while chemical syntheses of ammonia, drugs, and plastics changed the ways we produce food, treat diseases, and manufacture industrial and consumer products. Finally, invention of the first electronic components launched a new era of wireless telecommunication.

By the early 1930s all of these innovations introduced before World War I were successfully commercialized as new categories of machines and processes were gradually diffused throughout the world’s richest countries. Remarkably this adoption process was not derailed by the Great Depression, and after World War II came its second wave in newly industrializing countries of Asia and Latin America. Consequently, in so many fundamental ways the world of 2012 would be familiar to competent engineers working in the early 1930s and, with a few explanations, it would be readily comprehensible
also to men like Thomas Edison (electric systems), Charles Parsons (steam turbogenerators), Nikola Tesla (electric motors), Fritz Haber and Carl Bosch (synthesis of ammonia), or Lee de Forest (inventor of the triode) whose work preceded World War I.

At the same time, engineers working during the early 1930s would be impressed by the twinned changes of scale and efficiency that have been achieved during the past 80 years. Just a few key examples illustrate this progress. In 1932 the largest commonly deployed steam turbogenerators had capacities of less than 100 MW, while by 2012 units in excess of 1,000 MW are common; and the efficiencies of the best large coal-fired power stations had nearly doubled to more than 40 percent. The best lights now convert electricity to visible radiation with efficiencies that are an order of magnitude higher than those of 80 years ago. Smelting of pig iron in blast furnaces now takes only half as much energy per tonne of hot metal as was needed in the early 1930s, while the global output of pig iron is an order of magnitude greater than even its pre-Depression peak in 1929. Energy needed to synthesize a tonne of ammonia has been reduced by nearly two-thirds, while the global output rose more than 60-fold, from about 2 to about 130 Mt/year. And even in the US, new high-efficiency cars now travel nearly three times farther on a unit of gasoline, while global car sales in excess of 60 million vehicles are more than 50 times the 1932 level.

Concentrating on the fundamentals also makes it clear that electronic computing, nonexistent in 1932, has not been the most significant innovation in recent history. This claim is now often made because of the ubiquity of electronic hardware and software, with microchips controlling everything from cars to cellphones, and with complex programs watching over electric grids and air traffic as much as they do over electronic telecommunications and bank transactions. Those who make that claim betray their uninformed bias by the fact that their use of “technology” is limited to only computer-related electronic devices and processes, as if securing adequate energy supplies and growing enough food does not entail any technical skills. And they also argue a clear cart-before-the-horse proposition since all electronic computing and allied telecommunications would cease immediately without a massive and extraordinarily reliable (and largely fossil fuel–generated) electricity supply: as amazing and as potent as they are, they are still a class of secondary innovations dependent on a system whose fundamental concept emerged during the 1880s and whose global coverage has been created by decades of cumulative development.

Obviously, computerization has made many products and processes cheaper, and most of them also more reliable and more convenient to use. But it has not changed the physical and biophysical fundamentals of energy and food production, the two great pillars of any civilization. High quality of life is perfectly possible without ubiquitous microchips: after all, we had a rich and (by all rational measures) affluent civilization before 1970, the year...
when the first integrated circuit was designed by Intel. But high quality of life is unthinkable with persistent malnutrition, contaminated water, unheated homes, and material poverty. Technical advances that eliminated those deprivations and shortcomings had predated electronic computing by decades and generations. Indeed, if the qualification “most significant” is taken, logically, as making the greatest positive difference for the largest number of people, then none of the great innovations introduced after 1930—whether electronic computing, nuclear electricity, genomics, or medical diagnostics—qualifies.

Unprecedented growth of human population during the past 80 years could have happened without any solid-state electronics—but not without fossil fuels to power industrialization and urbanization, the two transformations that have employed and sheltered billions, and not without growing enough food for an additional 5 billion people, an achievement fundamentally predicated on an adequate supply of plant nutrients. We could do without packing more and more transistors on a chip, but we cannot do without providing sufficient quantities of essential amino acids to growing infant brains and adolescent muscles: nitrogen in proteins, not silicon in chips, has been (and will always be) the indispensable element of quality.

I have called the development of nuclear electricity generation a successful failure. The industry now generates about 13 percent of global electricity, but its expansion in the US stopped three decades ago, in the EU it was in retreat even before the Fukushima disaster, and, remarkably, traditional biomass fuels (wood, charcoal, and crop residues) still contain more energy than the fission-based generation. And great strides in lowering infant mortality and extending longevity throughout Asia and Latin America had nothing to do with the sequencing of the human genome and hardly anything to do with the number of magnetic resonance imaging (MRI) or computerized axial tomography (CAT) machines per 100,000 people, but plenty to do with properly cased new wells, old-fashioned sewage treatment plants, and more animal protein from milk and chicken in children’s diets.

Failures of imagination and reluctant commitments

The preceding argument was not about denigrating the role of innovation, merely about separating fundamentals from the less important factors responsible for the great changes of the past 80 years. Those decades brought plenty of impressive and diverse innovations that truly matter, and perhaps their most surprising shared attribute is how unanticipated they were. In 1830, with science being still a restricted domain of curious (and often well-off) individuals, with no industrial research (indeed with no centralized, highly productive industries), with steam-driven transportation in its infancy, and with no means of effective telecommunication, it was exceedingly difficult to anticipate any important technical advances that would change the world
over the next one or two generations. By 1930 science had become a substantial enterprise, industrial research was an essential part of modern production, news traveled rapidly on radio waves, and looking into the future was both a popular literary pastime and a scientific quest.

And yet the historical record reveals several constancies demonstrating that our capacities to anticipate future technical advances did not improve. First is the continuing lack of bold but realistic imagination. There is never any shortage of irrelevant speculation about colonizing Mars, flying to work in small private planes, or eating food entirely encapsulated in pills. But my searches have not uncovered any previsions of the US Midwest planted to soybeans for animal feed or a widespread use of horizontal drilling in oil and gas production. Similar examples could be chosen for every decade and for every class of innovations, and when looking for a combination of importance and variety I chose the year 1938 when the discovery of polytetrafluoroethylene (Du Pont’s now ubiquitous Teflon) was as much a surprise as the first major discovery of crude oil in Saudi Arabia and Frank Whittle’s turbojet engine.

In the first instance it was, as with so many inventions, pure serendipity as Roy Plunkett accidentally pumped freon into a cylinder that was left overnight in cold storage and found that the resulting white powder had not only an extraordinary slipperiness but also high resistance to cold, heat, and acids. The other two cases illustrate a much more common reality of institutional incredulity, aversion to risk taking, and lack of imagination. The Anglo–Persian Oil Company (active in southwestern Iran, just across the Gulf, since the early 1900s) believed throughout the 1920s that no oil would be found in Saudi Arabia. Standard Oil of California secured a concession only in 1933 and the first production well flowed in 1938: how different might world history have been had Saudi oil been discovered in the early 1920s?

A similar question can be asked about Frank Whittle’s jet engine. The young pilot-engineer patented his idea of a jet engine (gas turbine) in 1930; but even after years of trying, his enthusiasm about this superior mode of air propulsion was not shared by the British Air Ministry, which in 1935 refused to pay the trifling £5 fee for his patent’s renewal. The approach of World War II eventually changed those calculations, but the accelerated post-1939 development of jet engines and fighter planes came too late to make a difference to the length and the outcome of the war. But what difference might those machines have made if the Royal Air Force had begun to manufacture jet fighters three or five years ahead of the Luftwaffe?

There is yet another constancy regarding risk aversion and lack of imagination: even a cursory look back shows many instances in which the organizations that invented new products or processes were reluctant, even quite unwilling, pioneers of these innovations, choosing not to pursue them further and not to convert working prototypes into commercial products. I know of no more astounding example of this behavior than the way Xerox treated the
innovations that came out of its Palo Alto Research Center (PARC) during the 1970s. PARC created most of the indispensable components of user-friendly desktop computing: it was the first place in the world where one could find a standalone machine operated with the aid of a mouse, WYSIWYG editing, graphical user interface with icons and pop-up menus, laser printing, spell checkers and thesaurus, text editing and graphics, and access to file servers and printers with point-and-click actions.

But Xerox never attempted to commercialize its remarkable system, and the company nearly disintegrated. Meanwhile PARC’s accomplishments were exploited by Steve Jobs and Steve Wozniak to launch the first successful commercial PC in 1977 and then by IBM and Microsoft to introduce PARC-derived Windows 1.0. Again, another fascinating consideration for alternative history: what if Xerox’s aggressive development of PARC-based personal computing had preempted the rise of Apple and Microsoft?

Xerox has not been the only major corporation failing to grasp an early chance: Intel’s first microchip designed in 1970 did not go into a consumer product made by an electronic giant, but into a small calculator made by Busi-com, a Japanese company that folded just a few years later. And, to choose just one more example from a very different industry, production of steel in basic oxygen furnaces (now the principal method of smelting the world’s most important metal) was not commercialized by any of the world’s leading steelmakers but in an unlikely place, at Vereinigte Österreichische Eisen- und Stahlwerke AG in Linz, Austria.

But the current obsession with innovation assumes that things will be different as entrepreneurs pursue development of specific solutions rather than waiting for a natural winnowing and eventual emergence of truly distinguished contributions. So far, this deliberate quest has had a very limited success even in those cases where large sums were spent on new solutions. After all, the world is not running solely on nuclear electricity (although the industry has received tens of billions in subsidies since the 1950s), fuel cell–powered cars are nowhere to be seen, electric car sales account for less than 0.1 percent of the market, algae do not produce gasoline (Exxon Mobil alone spent $600 million on the quest before abandoning the effort), and new renewables cover a tiny share of the global energy demand. Is the situation any better as far as our foresight regarding major social, economic, and strategic trends is concerned? Do the past eight decades reveal some real progress toward a more successful anticipation of transformational socioeconomic changes? The answers are uniformly negative.

Politics, economics, and social transformations

Many political shifts and economic discontinuities that result in major social changes and sudden strategic reversals are the result of hard-to-miss trends, but that does not make it any easier to anticipate their outcomes in the only...
way that would make a useful difference: by realizing the totality of their eventual impacts. Once Hitler attacked the Soviet Union in June 1941 and once the US entered the war in December of that year, the ultimate Nazi defeat was only a question of time required for mobilization of America’s enormous natural and industrial resources and for the reassertion of Russia’s strategic strength. Predicting an Allied victory in December 1941 was thus an unremarkable feat, but nobody could have predicted the war’s most important long-term outcomes: that less than four years later the destruction of two Japanese cities would launch a new era of a decades-long, extremely expensive and perilous nuclear arms race; that less than five years later the two great powers would be adversaries in the Cold War that would eventually involve countries as far-flung as Angola and Vietnam; and that less than nine years later Europe would be partitioned over the next four decades and that China would be under Communist Party control.

Observations of elusive outcomes can be made for every one of the past eight decades and for all major global actors, and I will limit myself to just four more cases of epochal events in recent history: China’s transformation after Mao Zedong’s death; the collapse of the Soviet Union; Japan’s great turnaround; and America’s economic retreat. All of them were easy to predict in theory (after all, every man has to die, no dictatorship is eternal, no economy can grow at rapid rates forever, and no great power is immune to internal rot), but neither their specific modalities nor their complex consequences could readily have been predicted. Hence their arrivals came as a surprise, and the world lives with their enormous consequences.

There was a high likelihood that many changes would take place after Mao’s death in 1976, but who could have predicted that just two years later his old comrade Deng Xiaoping would begin a slow reversal of some of the key Maoist policies, that this process would accelerate after the Tian’anmen killings of 1989, that in just one generation China would become the workshop of the world, and that just three decades later the Chinese Communist Party would be a de facto owner of one trillion dollars of US government debt?

Would such predictions have sounded more credible than saying in the early 1980s—when Ronald Reagan was launching an expensive arms race with the Soviet Union and when the CIA was still making worrisome assessments of enormous and growing Soviet power—that a decade later there would be no Soviet Union and that its dissolution would be inexplicably rapid and, even more surprisingly, almost completely non-violent?

Also during the 1980s America’s political and corporate elites convinced themselves that Japan’s rise was unstoppable as the Nikkei index had quadrupled in five years to nearly 40,000 by December 1989, and as Japan was widely perceived as the premiere candidate for economic and technical supremacy in the twenty-first century. But by 2012 the Nikkei index was below 9,000 (the equivalent of the Dow Jones being just above 3,000 in 2012 instead
of trading above 13,000), and Japan (to be sure, still rich in global comparison) was looking back at two decades of lost direction, economic stagnation, and political drift while contemplating the future of de-industrialization, rapid population aging, and declining population size.

Lastly, in 1980, as America began to worry about Japan, it was still (as it had been for the previous eight decades) the world’s largest creditor nation; but in September 1985 the US Department of Commerce revealed that the US had become a net debtor. Who would have said by the end of that year, as the deficit reached trifling $28 billion, that 20 years later America’s annual current account would be nearly $700 billion (a sum larger than the total GDP of Saudi Arabia or Switzerland) in debt and that most of the goods sold by Walmart, its and the world’s largest retailer, would be made in China? How many things, and how many prospects, would be different if America had remained a large net creditor nation with a manufacturing sector almost as strong as it was in the mid-1980s?

**Humanity in control?**

The past matters in so many obvious genetic, cultural, and social terms, but in modern consciousness it is being reduced to ever briefer intervals between the incoming e-messages (one billion people incessantly checking and re-checking trivial Facebook ephemera, yet unable to place even epochal events within the correct century). Modern societies are animated by promises of a better future, a future designed to their preferred specifications, and some of our past successes have deluded many people into thinking that we can do progressively better, that we are finally close to being in control. That is why innovation has now become the mantra of affluent Western societies. Too many experts and too many influential opinion-makers are now trying to persuade us that the accelerating pace of technical fixes will solve our most complex problems: renewable energy conversions will soon displace fossil fuels, transgenic crops will feed the world, carbon capture and sequestration will prevent global warming, nanomaterials and additive manufacturing will produce anything cheaply and rapidly, and personal genomics will have us all living as long as Methuselah.

Answers are promised for everything, fate and chance are risible terms that have no place in the modern discourse of the future—as if the world (everything else being equal) would not have been different if the trench gas had killed rather than just temporarily blinded Hitler in October 1918, if Stalin had his fatal stroke in March 1943 rather than in March 1953, if in 1945 Truman decided to support Mao rather than Chiang, or if the Soviet hardliners had prevailed in Moscow and Cuba in October 1962. This is hardly a new argument, but one repeated with a much greater insistence: this time it is really different, this time we are really at a threshold of a truly new world because
infinitely intelligent machines will soon make humans superfluous (one of Raymond Kurzweil’s scenarios has that scheduled for as early as 2045) and will run everything on our behalf with wondrous perfection. Similarly, Diamondis and Kotler promise an imminent unfolding of “the most transformative and thrilling period in human history” whose miraculous advances will bring the age of abundance, of unprecedented material plenty and a pristine environment enjoyed by everyone.

Although the claims of accelerating change appear to be well supported by the relentless rise of computing power and by the shrinking intervals between the releases of new models of electronic gadgets, a closer look reveals the fallacy of this claim. It is a perfect case of *pars pro toto* error or, as I called this specific instance, Moore’s curse, a mistaken belief that the progress of all innovation resembles the biennial doubling of transistors crowded on a microchip, the rule formulated by Gordon Moore two generations ago. But this rule does not apply to the ways we produce our foods, energize our societies, transport our resources and goods, build our cities, and extend our lifespans: these activities have seen many improvements but no radical shifts and no steadily accelerating advances.

In this essay I have tried to demonstrate that the changes of the past 80 years were primarily due to unprecedented quantities, scales, and magnitudes rather than to radical shifts in kind, and although I hesitate to make any long-range forecasts I would expect that the next 80 years will not be much dissimilar where those existential fundamentals are concerned. Obviously, I do not foresee (*pace* Kurzweil) the world run by the eternal wisdom of self-generated software, nor do I take seriously the fairy tales of a superabundant era that is almost upon us.

But what if the next 80 years will see another period of epochal advances akin to those fundamental changes that took place between 1865 and 1914? How can we be sure that another burst of such creativity does not lie ahead? Of course, we cannot, but it is highly improbable because such discontinuities have been exceedingly rare during human evolution. Limiting the span to the last million years leaves us with the mastery of fire (as early as 790,000 years ago), hunting with projectile weapons (spears, some 400,000 years ago), the appearance of modern *Homo sapiens* (about 200,000 years ago), domestication of plants and animals (6,000–11,000 years ago), and the rise of the cities (3,000–5,000 years ago).

Preindustrial mechanical prime movers (water wheels, wind mills) were fairly common only in some regions, and even there they did not change the overwhelming reliance on animate primate movers (human and animal muscles); and while some countries were able to secure more reliable food supplies, famines remained common well into the early modern era and crop yields remained insufficient well into the nineteenth century. Unless one sub-
scribes to the unproven notion of accelerating evolution, anything resembling the great saltation of 1865–1914 is most unlikely to be reprised just a century or two later. But I do foresee events strongly shaped by contingencies and chance, and I hope that we will benefit from some important innovations that will be as unanticipated, and in their overall impact as surprising, as the jet engines or integrated circuits that helped to change the world of the past four generations.

And what of the planetary threat of global warming? We cannot be sure whether the eventual outcome will be tolerable (average global temperature increase of less than 2°C), highly worrisome (2°–4°C), or catastrophic (>4°C). In any case, there does not seem to be any resolve to confront the challenge, and it may be helpful to think of the alternative. The relative stability of climate that has prevailed during the past half-millennium will have to end, sooner rather than later: after all, we live in just another interglacial era. Would we prefer another period of global cooling akin to the last episode, which left a large chunk of mid-latitudes—regions where most of the world’s economic product and most of the planet’s food surplus now originate—under thick layers of ice or that converted those landscapes into tundra or cool steppe (Canada’s Prairies, the Corn Belt, France and Germany)? Would it be easier to adapt to a pronounced cooling than to a substantial warming?

But I would confidently predict that the most consequential impacts of the coming social, political, strategic, and environmental shifts will not be those that now dominate our concerns. Looking back reminds us about the enduring power of unruly realities, and to think that the coming decades will be profoundly different because we will be, finally, in control is only the latest addition to a long history of hubris and wishful thinking.

References

Curious readers may consult the following publications for relevant details.


POPULATION THEORY
AND MEASUREMENT
Demographic Metabolism: A Predictive Theory of Socioeconomic Change

WOLFGANG LUTZ

This essay introduces a general theory of how societies change as a consequence of the changing composition of their members with respect to certain relevant and measurable characteristics. These characteristics can either change over the life course of individuals or from one generation to the next. While the former changes can be analytically identified and described by certain age- and duration-specific transition schedules, the latter changes resulting from cohort replacement can be modeled and projected using standard models of population dynamics.

Building on earlier qualitative work by Karl Mannheim and Norman Ryder, this new theory applies the quantitative tools of multi-dimensional mathematical demography to forecast the future composition of a population according to relevant characteristics. In the case of persistent characteristics (such as highest educational attainment) that typically do not change from young adulthood until the end of life, quantitative predictions about the distributions of such characteristics in the population can readily be made for several decades into the future. For other characteristics that tend to change over the life course (such as labor force participation), standard age/duration-specific patterns can be assumed. Hence, unlike other models that are called “theories” but cannot be used to make explicit quantitative statements about the future, this theory of socioeconomic change can make such statements in a way that can potentially be falsified. It can therefore be called a theory with predictive power according to Karl Popper’s criteria (Popper 1959).¹

This is a theory predicting aggregate-level change rather than individual behavior. It is a macro-level theory focusing on the changing composition of a population and hence has no micro-level analogue. It can be called a “demographic” theory of socioeconomic change, implying that its inspiration and approach are demographic though its purpose is not. It is not primarily intended to explain and forecast demographic variables (such as population size, birth and death rates, migration, and the like); rather the goal is to predict
socioeconomic change in a broader sense (ranging from values and religions
to skills and productivity of the labor force) using a demographic paradigm.

In this essay I first discuss the writings of Karl Mannheim and Norman Ryder, both of whom suggested that the process of social change can be analytically captured through the process of younger cohorts replacing older ones. This is followed by a presentation and discussion of the basic propositions of the theory and a description of the multi-dimensional demographic model of the changing composition of the population over time. Finally, the possible empirical applications of this theory are illustrated through two examples. The first reconstructs and projects the changing composition of the population by the highest level of educational attainment. This important individual characteristic can be unambiguously measured and is persistent over the life course after a certain age. The second example addresses a “soft” variable, namely the prevalence of European identity in addition to national identity. Forecasting the prevalence of European identity undoubtedly has great significance for the future of the European Union. I conclude with a brief discussion of further possible applications and key implications of this theory of socioeconomic change with predictive power.

Antecedents: Karl Mannheim and Norman Ryder

In its general form the idea that societies change as new generations take
over is as old as human reflection, and one finds early writings on this topic
by the pre-Socratic philosophers and in Confucian philosophy. It is such a
plausible concept corresponding so well to everyday experiences of most fami-
lies, institutions and companies across all cultures and times that it is indeed
surprising how little systematic scientific effort has been made to formally
describe this force for socioeconomic change. While this is true even for the
systematic analysis of historical evolutions, it is even more the case when it
comes to forecasting.

Over the past two centuries, the strain of writing that comes closest to
the idea of projecting along cohort lines (the approach proposed here) is the
literature on the succession of generations that was prominent in the late
nineteenth and early twentieth centuries for explaining the sequence of his-
torical epochs. In this view, the process was driven by the replacement of old
generations by new ones with new views of the world, new priorities, and
new styles. A comprehensive synthesis of this approach was offered by the
sociologist Karl Mannheim in his 1928 essay “The problem of generations”
(Mannheim 1952 [1928]). More than a generation later the demographer
Norman Ryder published a seminal article titled “The cohort as a concept in
the study of social change” (Ryder 1965). Ryder’s article remains the key ref-
eree for anyone dealing with cohort analysis. Given Ryder’s demographic
background it is surprising that this article is entirely qualitative in nature,
without a single table. And at no point in his article does Ryder refer to the
great potential of this cohort approach not only for studying the past but also
for forecasting future social change. Because these writings of Mannheim and
Ryder are the most influential milestones in explaining social change through
cohort replacement and hence form the historical basis on which this new
theory will build intellectually, I discuss them in some detail.

Karl Mannheim (1893–1947) was Professor of Sociology in Frankfurt
and later taught at London University. In “The problem of generations,”
Mannheim contrasts two opposite views on generations. One of these he calls
“positivist,” which is focused on measurement and the average periods of time
required for the older generation to be superseded by the new in public life
(1952, p. 278). He refers to this approach as having a biological perspective
that we could accurately translate as meaning demographic. The opposite
approach, which he calls “romantic-historical,” is distinctly non-quantitative
and associated with the writings of the German historian Dilthey. Here the
central notion is “entelechy,” one generation’s expression of the “inner aim”
or its “inborn way of experiencing life and the world.” Mannheim seems
clearly more amenable to the first view, but his focus was only on looking
backward, trying to understand the forces driving history, rather than look-
ing toward the future.

Mannheim defines a generation as being determined by its “social loca-
tion” (soziale Lagerung), something shared by all members of a generation. He
compares it to the way one is a member of a specific social class, viewed not
cross-sectionally but over time: both generation and class “endow the indi-
viduals sharing in them with a common location in the social and historical
process, and thereby limit them to a specific range of potential experience,
predisposing them for a certain characteristic mode of thought and experi-
ence, and a characteristic type of historically relevant action” (p. 291). He
discusses what produces generational units and under what conditions a new
group of people growing up is sufficiently different from the previous one in
order to be called a new generation. In this sense Mannheim’s sociological
approach, which to some extent seeks to identify the qualitative inner spirit
of a generation, is quite different from the formal cohort approach proposed
here, where inner values (entelechy) may be a consequence but not a defin-
ing criterion for membership in a generation.

One other decisive difference between my proposed approach and
Mannheim’s is that members of a generation (cohort in my case) are not all
required to have the same social location, that is, to be similar in key aspects.
Quite the opposite, I consider cohorts as being composed of groups of people
with clearly distinguishable properties (which may define social locations)
such as speaking different languages or having different levels of educational
attainment or different national/European identities. Hence, the basic idea is
not that generations are homogeneous but rather that they are heterogeneous
in measurable ways, while their characteristics are persistent along cohort lines and the composition of the properties in the population changes as a consequence of the changing proportions of cohorts that possess the relevant characteristics.

More than a generation after Mannheim, Norman Ryder took up the topic from a demographic perspective. His seminal essay on the cohort and social change was published in 1965, but the paper originated in his 1951 doctoral dissertation at Princeton (eventually published as *The Cohort Approach* (Ryder 1980)). Ryder starts from the central but (then and now) rarely used notion of “demographic metabolism,” which he defines as the “massive process of personnel replacement” driven by the births, lives, and deaths of individuals (1965, p. 843). While individuals die, societies become immortal if reproduction is sufficient to offset mortality. This view of society is combined with the notion that the appearance of new individuals provides an “opportunity for social change.” While for individuals the flexibility to change over their lifetime tends to be limited, “the continual emergence of new participants in the social process and the continual withdrawal of their predecessors compensate the society for limited individual flexibility” (p. 844). Given the assumed inflexibility of individuals over their lifetimes, he arrives at the conclusion that “the society whose members were immortal would resemble a stagnant pond” (ibid.). An important additional thought is that “metabolism may make change likely, or at least possible, but it does not guarantee that the change will be beneficial” (ibid.).

Ryder’s definition of a cohort has since become the standard in demography: “A cohort may be defined as the aggregate of individuals (within some population definition) who experienced the same event within the same interval” (p. 845). In most cases, birth is taken as the defining event, but this is only a special case of the more general approach. Cohorts differ from synthetic cross-sections (comparing people of different ages at the same point in time) because “time and age change pari passu for any cohort” (ibid.) Unlike Mannheim, Ryder allows for heterogeneity within generations/cohorts. This is also an important feature of the new theory introduced here, in which members of one birth cohort are subdivided into different categories such as groups with different educational attainment. Hence, under this definition a cohort is by its very nature an aggregate measure, that is, it refers to groups of people. In Ryder’s words: “Thus a cohort experiences demographic transformation in ways that have no meaning at the individual level of analysis, because its composition is modified not only by status changes of the components, but also by selective changes of membership” (ibid.).

But in contrast to Ryder, the present theory does not necessarily depend on the assumption of complete cohort determinism. The tools of multi-dimensional population dynamics described below also allow for changes over the life course of cohorts that may be a function either of age or of
external period changes (such as the introduction of new technologies that affect all cohorts) and interactions between the two. Hence, the possibility of lifelong learning and changing of cohorts can in itself provide an opportunity for socioeconomic change, and immortality would not necessarily result in Ryder’s “stagnant pond.” To what degree cohort effects dominate age and period effects might differ depending on the specific characteristic studied. In some cases (such as highest educational attainment after a certain age) it is persistent by definition; in other cases this is a matter of empirical analysis for the past and of corresponding assumptions for the future.

Inspired by Ryder, I have adopted the term “demographic metabolism” as the tentative name of this theory of socioeconomic change because it addresses the changing nature of a population through the replacement of individuals with certain characteristics by individuals with other characteristics. However, unlike in Ryder’s usage, not only does metabolism affect the population in its entirety, but the notion is also generalized to refer to the changing composition of certain sub-categories such as the labor force or the young adult population. This generalization of Ryder’s idea is discussed further in the concluding section.

A macro-level theory based on people as units

Before introducing the quantitative forecasting model and offering numerical illustrations of this theory, I specify and discuss some basic definitions and assumptions on which the approach rests. These foundations of the theory are presented in the form of four propositions.

Proposition 1: People—individual humans—are the primary building blocks of every society and the primary agents in any economy. Hence, they form the basic elements of any theory of social and economic change.

In the terminology of the pre-Socratic philosophers, people are the atoms of society. Theoretically, one could delve into sub-atomic structures: modern brain research shows that any decision at the individual level is the result of complex interactions among different parts of our brains. But for the present purpose, it is sufficient to assume that decisions happen at the level of individuals, who then interact with other individuals.

Proposition 2: For any population, members can be sub-divided into disjoint groups (states) according to clearly specified and measurable individual characteristics (in addition to age and sex) for any given point in time.

In principle, any sub-division of people that satisfies this criterion is legitimate and allows the application of the theory described below. Age is a special characteristic for forecasting because it automatically changes in tandem with chronological time. Hence, the applications and illustrations of the theory are designed to divide the entire population by the particular charac-
teristics of interest and then further to sub-divide all people in this category by age and sex. Over time, people stay in their categories and simply become one year older every calendar year unless they die or move to another category.

**Proposition 3:** At any point in time, members of a sub-population (state) defined by certain characteristics can move to another state (associated with different characteristics), and these individual transitions can be mathematically described by a set of age- and sex-specific transition rates.

Transitions may occur not only to another state inside the system but also to an absorbing state (death) or to a state outside the system (out-migration). New individuals arriving (through birth or in-migration) will be instantly allocated to one state within the system. Not all transitions among a given set of states may be possible. Sometimes transitions are only possible in one direction, such as from lower to higher educational categories or from the single to the married state, from which people may move to the divorced or widowed state.

**Proposition 4:** If any given population consists of sub-groups that are significantly different from each other with respect to relevant characteristics, then a change over time in the relative size of these sub-groups will result in a change in the overall distribution of these characteristics in the population and hence in socioeconomic change.

The historical record offers many examples of significant changes of entire cultures, economies, and technologies as a consequence of the changing composition of populations. Most radical have been changes in the dominant characteristics among populations resulting from mass migrations combined with differential mortality. One may think of the end of the Roman Empire or the Spanish conquest of the Americas: new people brought with them their very different ways of thinking, acting, and using technologies (characteristics) and replaced the previously dominant characteristics, thus causing radical social change. In such major transitions, one may assume that most individuals did not actually change their characteristics over their life course. Rather the bearers of the old characteristics disappeared (through being killed, succumbing to new diseases, or out-migration) and were replaced by new arrivals and their offspring, a radical form of demographic metabolism. In some cases the surviving children of the bearers of the old characteristics were also socialized to display the new characteristics, thus creating a new and different society with the prevailing new characteristics through inter-generational change. Early efforts to assimilate aborigines in Australia were based on this strategy. On the other hand, existing characteristics can also be lost over generations without such major discontinuities. History is filled with examples of this sort that can also be viewed as a consequence of the changing characteristics of the populations, which also include the abilities to use certain technologies.
The choice of a characteristic that is worth studying with respect to its changing distribution in a society is necessarily context-specific. Hence, the definition of relevant sub-groups depends on the questions asked. In the second part of this essay I give several empirical examples of using this approach to address specific questions concerning future socioeconomic change. Before doing so, I briefly describe the methodological tool to make such forecasts.

The multi-dimensional cohort-component model

Most demographic methods deal with the transitions of people from one state to another over a certain time interval and are in one way or another based on the life table. In its most fundamental form the two states are being alive and being dead, and the life table was constructed on the basis of age-specific mortality rates to determine the probability of surviving to each age and the remaining years of life expectancy at any given age. These tables were originally calculated separately for men and women because observed age-specific mortality rates tended to differ substantially by sex. Aside from this differentiation by age and sex, conventional demography still considers populations to be largely homogeneous—for example, assuming that all men aged 50–54 are exposed to the same risk of death. In the multi-state case this restriction is relaxed and mortality rates can differ for sub-groups as defined by further characteristics.

The cohort-component model has become the dominant method of projecting populations. While a simple exponential growth model considers changing population size solely as a function of the growth rate, the cohort-component model can deal with irregular age distributions and the differential impacts of fertility, mortality, and migration patterns on future age structures. The one-dimensional cohort-component model starts with a population stratified by age and sex and then projects cohorts by advancing their age $t$ years over every time interval of $t$ years while simultaneously exposing them to assumed age- and sex-specific schedules of mortality and migration and exposing the female cohorts to age-specific fertility rates. Because fertility, mortality, and migration are the three fundamental components of population change, this method of projecting along cohort lines is called the cohort-component model. In the multi-dimensional case the population is further sub-divided along additional characteristics. Since the origins of the multi-state model lie in regional population studies, these additional divisions were initially sub-regions, for example, provinces or states of one country. This explicitly heterogeneous population is then projected into the future by simultaneously considering different fertility, mortality, and migration schedules for the individual provinces as well as transitions (migrations) among the provinces. More generally, provinces can also be defined by other criteria such as marital status, labor force participation, and highest educational attainment.
The simple cohort-component model goes back to Cannan (1895). Later it was formalized and widely applied by Whelpton (1936) and Leslie (1945), after whom the Leslie Matrix is named. The detailed formal presentation of this model is given in any standard demography textbook (e.g., Preston, Heuveline, and Guillot 2001). The multi-state model is based on the generalization of the simple life table (single decrement table) to multiple-decrement and increment-decrement life tables (Rogers 1975; Schoen and Nelson 1974; Keyfitz 1985).Essentially such tables describe movements of people that can go back and forth between more than two states. These methodological advances, made in the 1970s largely by scholars affiliated with the International Institute for Applied Systems Analysis (IIASA), led naturally to the multi-state population projection model that can simultaneously project the populations of different categories (states, regions) with different fertility, mortality, and migration patterns as well as movements among the categories. Again, the methodology is extensively documented elsewhere (Rogers 1975; Keyfitz 1985; Keyfitz and Caswell 2005; Rogers and Willekens 1986). Here it suffices to say that the methodology is based on the Markovian assumption that the probability of transition to another state for a specific period of time is defined only by being present in the current state in addition to age and sex. In other words, all persons of the same age and sex in the same state have identical risks of dying or of moving to another state irrespective of their earlier transition histories.

After the initial focus on regionally defined states, applications were expanded to the analysis of marital status transitions (Schoen and Nelson 1974), the analysis of labor force participation and working life tables (Willekens 1978), the analysis of health and morbidity (Manton 1988), and family status cross-classifying marital status and number of children (Lutz 1989). The application of the multi-state projection model to education was to my knowledge used for the first time in a study of population–development–environment interactions on the island of Mauritius (Lutz 1994).

In the model developed for the study of Mauritius, the entire adult population is sub-divided into six states as defined by the cross-classification of three education categories (primary or less, secondary, and tertiary) and labor force participation (see Figure 1). The children and in-school population constitute a seventh state. The figure also depicts the essential transitions of the multi-state model: new births enter the category of child population, which advances one year in age every calendar year and is simultaneously exposed to assumed sets of mortality and migration rates. At a given age, then, all children become members of the adult population not (yet) in the labor force and according to their highest educational attainment. In adulthood, people can move back and forth between being out of the labor force and in the labor force, and they age one year every year within their respective categories, being always exposed to age-specific mortality and migration risks.
which are status-specific. The female population of reproductive age is also exposed to assumed age-specific fertility rates—which differ by status—resulting in the births that then enter the model.

While Figure 1 gives the structure of one specific model, it also typifies the general design of a multi-dimensional demographic forecasting model. The two following illustrations of other applications differ from it only in their specific definitions of the state space and the possible transitions among states.

**Modeling the dynamics of changing educational attainment distributions by age and sex**

The multi-state model described above requires the fewest assumptions for studying characteristics that are persistent over the individual life cycle. One such characteristic, highest educational attainment, is typically attained at younger ages and then maintained throughout life. While this is almost entirely the case for primary and junior secondary education, the completion of
tertiary education can also occur later in the life cycle. Because comparable data exist for many countries on the timing of transitions to higher levels of education, the age pattern of education transitions can be appropriately modeled and taken into account.

This persistence of educational attainment over time becomes particularly visible when looking at the distributions by age for countries that recently experienced significant increases in schooling. Figure 2 shows the education and age pyramids for South Korea as reconstructed and projected using the model of multi-dimensional population dynamics. Here, the education classification has been added to the conventional age pyramids through color (shading) with the following four educational categories: never been to school, some primary education, completed junior secondary education, and completed first-level tertiary. These 60 years of improvements in Korea’s human capital reflect probably the most rapid expansion in education in history. While in 1960 the majority of the female population above age 20 had never been to school, 30 years later the population below age 50 already has nearly universal secondary education. By 2020 most of the uneducated older cohorts will have died, while young women in Korea will be among the world’s most educated with more than half of them having completed college. By then, the demographic metabolism at the population level will have fully run its course and Korea will be a different society and economy than it was in 1960 in almost every imaginable dimension.

The data plotted in Figure 2 come from an ongoing project carried out by the Wittgenstein Centre for Demography and Global Human Capital (Vienna) to reconstruct and project educational attainment distributions by age and sex for more than 170 countries (Lutz et al. 2007; KC et al. 2010; Lutz and KC 2011). While the reconstruction uses the past age and sex distributions as given by the UN Population Division and therefore only needs to consider differential mortality (and where possible migration) by level of education, the projection also includes education-specific fertility rates.

This reconstruction of the educational attainment distributions by age and sex also produced a new data set that is superior to previous data sets (Barro and Lee 1996; Dela Fuente and Domenech 2006; Cohen and Soto 2007; Benhabib and Spiegel 1994) in at least four respects: (1) its detail (four educational categories for 5-year age groups of men and women) for most countries of the world; (2) the consideration of differential mortality by level of education; (3) the strict consistency of the definition of educational categories over time (a major problem in historical data sets, in which the underlying educational definitions often change); and (4) its natural extension to forecasting.

These new data have already formed the basis for a reassessment of the global empirical evidence on returns to education (Lutz and KC 2011). While there is little doubt that education has an important positive effect on
**FIGURE 2** Age and education pyramid for South Korea 1960–2020 distinguished according to four educational attainment categories

**1960**
- Total population: 24.8 million

**1990**
- Total population: 43.0 million

**2020**
- Total population: 49.5 million

NOTE: No educational classification given for population below age 15.
individual earnings, the evidence for its effect on aggregate-level economic growth has been more ambiguous (Becker 1993; Barro and Sala-i-Martin 2003; Benhabib and Spiegel 1994; Pritchett 2001). However, these earlier studies were based typically on the mean years of schooling of the entire adult population above age 25. As Figure 2 clearly illustrates for Korea, in 1990 this crude indicator averaged the human capital of the elderly uneducated cohorts with that of the highly educated young ones. As a result, the indicator has little statistical signal in its change over time and does not correlate strongly with the increase in economic growth, which in Korea accelerated when the better-educated cohorts entered the young adult ages. A recent study estimating economic growth regressions using the new age-specific human capital data found consistently positive and significant effects of educational attainment on economic growth (Lutz, Crespo Cuaresma, and Sanderson 2008). Because it also included the full educational attainment distributions (and not just mean years of schooling), the study also showed that for the poorest countries it is essential to complement primary education with broad-based secondary education in order to boost economic growth. This approach also has a stronger effect on poverty eradication than focusing on tertiary education in an otherwise largely uneducated population.

Beyond economic growth, these new data were also used to reassess the aggregate-level returns to education on a broad range of outcomes from population growth and health to quality of institutions. For instance, we have shown that the entry of large cohorts of better-educated men and women into the young adult ages appears to play a key role in the transitions of societies to modern democracies, as assessed through an analysis of global time series (Lutz, Crespo Cuaresma, and Abbasi-Shavazi 2010).

As shown in the case of South Korea, the multi-dimensional demographic approach with a focus on age, sex, and level of education can provide a portrait of key dimensions—if not the single most important dimension—of the rapid socioeconomic change that Korea has experienced over the past several decades. Starting in the 1940s and 1950s in a situation of widespread poverty—worse than many Eastern African countries at the time—and a setting in which more than 80 percent of the population received no schooling, a massive expansion of schooling at the primary and secondary level (complemented by an effective family planning program) resulted in a sharp fertility decline and was a necessary condition for the remarkable rates of economic growth that followed. While human capital alone may be not sufficient and effective governance and the right macro-economic strategies are also important, it is hard to see how Korea and the other Asian tigers could have had this economic success without the underlying changes in human capital. As Figure 2 illustrates, this change happened along cohort lines where demographic metabolism was the major engine of change in each relevant age group. Plausibly, with this cohort perspective, an informed observer of
developments in Korea around 1970 could have predicted the subsequent course of socioeconomic development.

Moreover, the particularly rapid improvement in the education of younger women brought fundamental changes to Korean family structure and society in general. The total fertility rate declined from around 6.0 in the late 1950s to around 1.2 today. This change is also visible in the changing age structure as depicted in Figure 2. Korea is facing rapid population aging. It is not yet clear to what extent this will pose a problem in light of the fact that high levels of education among Korea’s younger generations will not only yield higher productivity but will also likely result in much better health at older ages and a later age at retirement. While these specific economic consequences of aging cannot be readily forecast by this theory (partly because they have no precedent in human history), the fundamental underlying forces of aging and changing characteristics along cohort lines can be forecast with high certainty. These forecasts are likely to be more accurate and more relevant than those of most macro-economic models over a similar time horizon.

**Predicting “soft” variables such as preferences and identities**

My second example of the theory presented here comes from a field that is both highly contentious and highly political and is usually considered beyond the reach of quantitative modeling, not to mention forecasting: predicting the future evolution of European identity among EU citizens. The following description draws on the study by Lutz, Kritzinger, and Skirbekk (2006).

Many political observers of the current economic crisis and its effect on the Euro and the future of the European Union expect a revival of nationalism in many member countries and as a result a possible dissolution of the EU. It is often argued that existing economic interdependencies (particularly in the banking sector) are the main force that still holds the EU together. These observers overlook the changing opinions and identities of European citizens, the people who make up Europe.

Political scientists tend to think that the question of identity is crucial for the legitimacy of any political system (Easton 1965). In this view the future of the European Union depends primarily on whether European citizens also maintain a European identity. Furthermore, recent work stresses that people usually have multiple identities that do not necessarily compete with each other. In this view European identity complements but does not displace national and regional identities (Risse 2000). The European Commission is concerned with this issue and collects information on it in the Eurobarometer surveys. The relevant question on European identity asks: “In the near future, do you see yourself as [nationality] only, as [nationality] and European, as European and [nationality] or European only?” Since 1996 this question has been asked
with identical wording more than a dozen times in the EU-15 (members of the EU as of 1995) with national samples of around 1,000 in each round.

In 2004, 42 percent of the adult population of the EU-15 above age 18 said that they felt themselves to be solely nationals of their own country, while 58 percent gave an answer that reflected multiple identities including a European identity. Figure 3 shows a clear decline in multiple identity with age. In other words, the older the respondent, the greater the chance that he or she will feel only a national identity. While for younger age groups those with only national identities are a minority, for the population above age 60 they constitute a majority.

Does this pronounced age pattern imply that as people get older they tend to assume a stronger national identity and abandon multiply identities they might have held earlier? In this case of a dominating age effect, the significant population aging that will occur over the coming decades would imply a decline in the proportion of citizens with multiple identities. Yet, the same pattern could also be explained in terms of a cohort effect: young cohorts are being socialized in a way that produces a higher prevalence of multiple identities than found among the previous cohorts which they then maintain throughout their lives. This effect would result in significant increases in future European identity through demographic metabolism, with the younger, more European-minded cohorts replacing the older ones. Both of these contrasting interpretations are possible, and based on one cross-sectional survey alone their validity cannot be assessed empirically. Only panel data, providing age profiles at different points in time, allow us to discriminate between age and cohort effects.

Lutz, Kritzinger, and Skirbekk (2006) conducted statistical analyses showing a highly significant positive cohort effect. The study found that for cohorts born one year later, the proportion with multiple identities is on average half a percentage point higher. This confirms the view that the

![FIGURE 3 Proportion of the European population with multiple identities (including European identity) by age, 1996, 2004, and forecast for 2030](image)
trend toward a greater prevalence of multiple identities in the European Union largely happens along cohort lines. In other words, cohorts born more recently are socialized in such a way that they adopt fewer solely national identities and more multiple identities. They then largely maintain these identities throughout their lives.

Because expressed national versus European identity can be viewed as a characteristic of individuals meeting the criteria set out above, its changing prevalence in the population can also be forecast using the new model. Doing so, Lutz, Kritzinger, and Skirbekk (2006) forecast future trends in the degree of multiple identities in Europe under the assumption that the estimated cohort effect remains relevant over time. In other words, we incorporate the assumption that in each subsequent cohort the proportion with multiple identities increases by 0.5 percentage points and remains stable along cohort lines. The top line in Figure 3 gives the predicted age-specific proportions with multiple identities in 2030 and shows a marked upward shift. In terms of absolute numbers the results predict that in 2030, only 104 million adult European Union citizens (EU-15) will have strictly national identities, while 226 million will have multiple identities. This study also suggests that the relentless forces of cohort replacement by which the older, more nationally oriented cohorts will die out and the younger, more European-minded cohorts will take their place are likely to produce significant changes in the pattern of European identity. These predicted cohort shifts are likely to have significant long-term implications for fundamental political and economic developments in Europe, even though short-term politics and market reactions are likely to remain volatile. With respect to the current political debate on the legitimacy of Europe-wide institutions, this study suggests that most commentators and analysts ignore one of the most important forces that can likely lead to greater European integration—namely, the demographic metabolism that will replace predominantly nationalistic cohorts with younger ones that predominantly identify with Europe.

If the theory of socioeconomic change introduced in this essay can produce quantitative forecasts for issues that are as soft and qualitative as the notion of national/European identity, it can also be applied to forecasting the future prevalence of a large array of personal views, opinions, preferences, and values. The preconditions for the applicability of this theory are the availability of data that allow a clear distinction between the different characteristics studied and the estimation of the existence and strength of cohort effects as well as age/duration-specific transition rates.

Discussion and outlook

Virtually all human populations can be subdivided into groups whose members differ from each other according to measurable characteristics that also influence their behavior. If membership in these groups is stable over the
life course after a certain age (such as ethnicity, native language, highest educational attainment, or stable values and identities) or shows patterns of transition that can be modeled on the basis of plausible assumptions (such as marital status, labor force participation, income level, or health status), then the multi-dimensional demographic approach allows us to model how societies change over time according to the changing relative sizes of these sub-groups. Perhaps no other theory in the social sciences has potentially greater predictive power over decades into the future.

Many questions remain. Some of them I will mention in this concluding section, others will be elaborated in subsequent research. One question that requires significant further elaboration is how to deal with uncertainty in the context of prediction. Over the past two decades an extensive literature on probabilistic population projections has appeared (e.g., Lutz and Goldstein 2004; Lee 1998). These approaches, which replace point forecasts with ranges of explicitly stated uncertainty, can, in principle, also be applied to the kinds of multi-dimensional forecasts presented here. The main problem is that the uncertainty distributions need to be established for a much larger number of input parameters, and there often is scant empirical basis for grounding these assumptions. A more serious theoretical issue relates to the possibility of falsification of probabilistic statements. Under a probabilistic view, the probability of actually realizing a precise point forecast is practically zero: the real outcome almost certainly will be higher or lower. (This same problem arises with most predictive theories in the natural sciences: the question of falsification of probabilistic statements is one of the issues that are still largely unresolved in the theory of science (Pearl 2000).)

Space limitations have allowed only the description of two examples of predictions based on the theory of demographic metabolism. But these two have shown that this concept can be operationalized in a practical and meaningful way, resulting in new and relevant information about the future. The examples given also illustrate that the theory is applicable to almost any other meaningful and measurable sub-division of the population and even for cases in which the persistence along cohort lines is more questionable and becomes a matter of empirical assessment.

This demographic theory can possibly bring innovation not only to sociology but also to economics. Demographers tend to use many sociological and economic concepts in their work. Why should a genuinely demographic concept, applied far beyond the traditional realm of demography, not also advance thinking in those other disciplines? While for sociology and the study of social change the applicability of this theory is clear, there are many applications in economics as well. Wherever in economics the almost ubiquitous assumption of strictly homogeneous human agents is considered as too strong, this model offers a quantitative way of explicitly addressing heterogeneity. This can range from distinguishing between groups of people who have dif-
different sets of indifference curves underlying their choices to groups of people with different discount rates in their assessment of utilities or to people who have different degrees of rationality in their behaviors.

Finally, I should emphasize that the innovative aspect of this essay does not lie in the methodologies of multi-dimensional (multi-state) demography or in highlighting the importance of changes along cohort lines. Both have been well established in the literature for the past 40–50 years, although not given a high level of attention. The new feature of this essay is the combination of these elements in a way that forms a comprehensive theory of socioeconomic change with quantitative predictive power. If accepted and further corroborated, this theory could significantly alter the way socioeconomic change is viewed, analyzed, and forecast, with direct implications for policy decisions and choices at the level of individuals, organizations, and societies.

Notes

Funding for this work was made possible by an Advanced Grant of the European Research Council, “Forecasting Societies’ Adaptive Capacities to Climate Change,” grant agreement ERC-2008-AdG 230195-FutureSoc.

1 To my knowledge, Karl Popper has never explicitly written about demographic theories, but thanks to his desire to eat his favorite dish (Wiener Erdäpfelgulasch, or Viennese potato goulash), I had the opportunity to have a lengthy discussion with him about this topic in 1984, when I stayed in London for an internship with the World Fertility Survey and Popper’s private doctor in Vienna asked me to deliver a large number of packages with his favorite dish to him. In return for this service, he talked to me at his private residence, patiently answering the questions of a young demographer with an interest in philosophy of science. In essence, he told me that, with respect to the validity of theories, demography is not different from all the other sciences. But through its quantitative nature it should be actually easier than in many other social sciences to define hypotheses and theories that are specific enough so that they can be tested, i.e. potentially falsified. He also asked what would be the most important theories with predictive power in demography. After some hard thinking I could only come up with the theory of demographic transition, which predicts that all societies starting from pre-modern conditions will experience a fertility decline following the mortality decline. Despite its lack of precision about the time lags involved and the precise course of fertility decline, this theory has predictive power and as a consequence underlies all the population projections for developing countries where further fertility declines to at least replacement levels are assumed. After listening to my explanations with interest, he asked whether demographic models could be applied to forecast other social trends as well, in particular utilizing the great inertia of population changes over time. Since I had never thought about it in this way, he encouraged me to do so. And it took me over 25 years to finally take up the challenge.

I remember well his last words as we waited for a taxi outside his house: “As scientists we have to be like bats, sending out signals to an unknown world and based on the echoes we receive, build an image (theory) which is the basis for predictions where we can safely fly. And as we fly, we have to listen very carefully to be able to update our predictions.”

2 If considered relevant, this model could easily be modified to also allow for adults to go back to school for continued education before reentering the labor market.
References


Cross-Sections Are History

Richard A. Easterlin

International cross-section regressions on real gross domestic product (GDP) per capita are widespread in the social sciences. Findings of significant associations between GDP per capita and a multitude of economic, social, and political variables are commonplace, and these results are often read as demonstrating the effect of economic growth on the variables under study (Easterlin and Angelescu 2012). For variables integral to production and consumption, such as material living standards or the rural–urban distribution of employment, such inferences are plausible. But economic growth is often viewed also as the main force responsible for such outcomes as the expansion of schooling, improved health, increased life expectancy, fertility decline, women’s empowerment, the extension of political and civil rights, and the like.

Moreover, these cross-section relationships are often taken to be predictive of time-series change, of what is likely to happen as a result of economic growth. Studies of the historical experience of individual countries, however, frequently fail to confirm expectations based on cross-section relationships (Easterlin and Angelescu 2012; Easterly 1999; Thornton 2005). As one moves outside the purely economic realm to social and political variables, this lack of confirmation of an association with GDP per capita is especially apparent. Why, then, do we often find a significant cross-section relationship if the implied causal connection is not confirmed by time-series analysis? The answer suggested here is that cross-sections register the results of history, not insights into likely experience.

The head start hypothesis

To focus the discussion, take as a demographic example the international cross-section regression of life expectancy at birth ($e_0$) on GDP per capita illustrated in Figure 1. This relationship is sometimes thought to demonstrate the causal impact of economic growth on life expectancy, as proclaimed, for example, in the title of Pritchett and Summers’s (1996) econometric analysis of international data, “Wealthier is healthier.”

©2013 The Population Council, Inc.
Thanks to the seminal work of Samuel Preston (1975), demographers for many years have, instead, viewed the life expectancy differences between rich and poor countries in Figure 1 as the outcome of two component factors—the first, a movement along a given e0/GDP function, reflecting the positive impact of improved living levels attributable to economic growth (the Pritchett–Summers effect); and the second, an upward shift in the function, the result of advances in health knowledge and technology, whose application leads to increased life expectancy at given levels of GDP per capita. Preston’s empirical analysis identified the principal cause of life expectancy improvement post-1900 as the upward shift of the function; indeed, advances in health technology were estimated to account for 75 to 90 percent of the overall improvement.

Unlike the common “black box” treatment of production technology in economics, Preston offered numerous specific examples of advances in health technology. Other scholars have provided similar evidence (e.g., Durand 1960; Easterlin 2004, chapter 7; Cutler, Deaton, and Lleras-Muney 2006 and the references therein). The major breakthroughs in health technology occurred roughly in three stages. The first was new methods of preventing the transmission of disease, starting around the middle of the nineteenth

![Figure 1: Life expectancy and GDP per capita, 185 countries, 2009](image)

NOTE: The fitted regression is \( y = 6.10x + 15.68 \), where \( y \) = life expectancy in years, \( x \) = log GDP per capita. Adj \( R^2 = 0.6741 \) (t-stats in paren). \((19.16) \quad (5.58)\)

century. The so-called sanitation revolution at that time aimed at cleaning up cities through purer water supplies, better sewage disposal, paved streets, education in personal hygiene, and the like. The second stage, starting around the 1890s, was the introduction of vaccines aimed at preventing certain infectious diseases. The third, beginning in the late 1930s, was the development of new drugs (antimicrobials) to cure infectious disease. These advances, it should be noted, are quite different from the new technologies of the “industrial revolutions” responsible for the growth of GDP per capita (cf. Easterlin 1996).

A case could be made that all of the observed improvement in life expectancy is the result of advances in health technology. Preston’s 10 to 25 percent share of causality allotted to higher GDP per capita is based on a positively sloped $e_0$/GDP function. But if one were to go back to the mid-nineteenth century, before the major breakthroughs in health knowledge and technology mentioned above, it is by no means clear that the slope of the function was positive at that time. True, higher GDP per capita led to better food, clothing, and (perhaps) shelter, hence more resistance to disease. Counteracting this trend, however, was the fact that higher GDP per capita also led to increased urban concentration and consequently greater exposure to disease. The reaction of the nineteenth-century economist Nassau Senior to the horrifying descriptions of Britain’s “great towns” in the 1838 Poor Law Reports tells the story:

What other result [besides high urban mortality] can be expected, when any man who can purchase or hire a plot of ground is allowed to cover it with such buildings as he may think fit, where there is no power to enforce drainage or sewerage, or to regulate the width of the streets, or to prevent houses from being packed back to back, and separated in front by mere alleys and courts, or their being filled with as many inmates as their walls can contain, or the accumulation within and without, of all the impurities which arise in a crowded population? (as quoted in Flinn 1965)

It seems plausible that the positive slope of Preston’s $e_0$/GDP function was itself the product of new knowledge, because sanitary reform and similar advances gradually eliminated the sizable excess of urban over rural mortality. Thus, the negative effect on life expectancy that growing GDP per capita induced via increased urban concentration of the population was removed, leaving only the positive effect of improved living standards.

By this reasoning, all of the modern improvement in life expectancy is due to advances in health technology, not to higher GDP per capita. But if this is so, how can one account for the positive association of life expectancy and GDP per capita in Figure 1?

One argument is that the higher income resulting from modern economic growth is essential to financing increased private and government
expenditures associated with improved health technology. But the measures necessary to implement advances in health technology do not seem to have required, on average, particularly high levels of income. If they did, then less developed countries (LDCs) would have been hard put to implement public health programs in the twentieth century without substantial external aid. In fact, such aid was trivial; an assessment published in 1980 concluded that “total external health aid received by LDCs is less than 3% of their total health expenditures” (Preston 1980). Clearly, despite their low levels of income, LDCs were able almost entirely on their own to fund applications of new health technology. Higher GDP per capita was not essential to implement these advances.

An alternative explanation—what might be called the “head start hypothesis”—is suggested here for the significant positive cross-section association between life expectancy and GDP per capita. Assume that the historical trajectories of GDP per capita and life expectancy are independent of each other and that each is governed by advances in its own underlying technology. In the case of GDP per capita, call it “production technology”; in the case of life expectancy, “health technology” (cf. Easterlin 1996). Assume also that the initial levels of GDP per capita and life expectancy are low and fairly similar among countries, and that the takeoff dates for advances in production technology and health technology in different countries may differ from each other. Suppose that the countries with a head start in production technology and thus in the growth of GDP per capita are the same as those with a head start in health technology, and thus in the improvement of life expectancy. Then the result will be that at any given date the head start countries will have both higher GDP per capita and higher life expectancy than the countries that followed, even though GDP and life expectancy have no causal relation to each other. The positive cross-section relationship between life expectancy and GDP per capita is the result of the same countries having a head start in the technology underlying each variable.

To illustrate numerically, consider the takeoff dates in economic growth and life expectancy improvement (reflecting, respectively, the introduction of

<table>
<thead>
<tr>
<th>Region</th>
<th>Date of takeoff</th>
<th>Years from takeoff to 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) Economic</td>
<td>(2) Life</td>
</tr>
<tr>
<td></td>
<td>growth</td>
<td>expectancy</td>
</tr>
<tr>
<td>Western Europe</td>
<td>1820</td>
<td>1890</td>
</tr>
<tr>
<td>East Asia, excluding Japan</td>
<td>1950</td>
<td>1950</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>2000</td>
<td>1970</td>
</tr>
</tbody>
</table>
modern production technology and health technology) listed for each of the three regions in Table 1 (columns 1 and 2). Although not precise, the dates are roughly in keeping with historical experience. Western Europe (and its overseas descendants) leads the way in both technologies, and sub-Saharan African brings up the rear. If one calculates years since takeoff in both life expectancy and GDP per capita, Western Europe has the greatest number of years and sub-Saharan Africa the fewest (columns 3 and 4). If one then plots years since takeoff for the three regions, as is done in Figure 2, we reproduce the positive slope of the regression line in Figure 1. The interpretation of Figure 1 suggested by Figure 2 is that Western Europe has the highest life expectancy and GDP per capita because it had a substantial head start in the implementation of both economic and health technologies. Correspondingly, sub-Saharan Africa has the lowest life expectancy and GDP per capita because it was the last to implement both technologies. The high values on life expectancy and GDP per capita for Western Europe and the low values for sub-Saharan Africa do not indicate a causal relation—by assumption they are independent of each other. They simply indicate that Western Europe was the leader in both technologies, and sub-Saharan Africa was a follower. Although the timing and rate of spread differs between the two technologies, the geographic pattern of diffusion is the same, and the result is a statistically significant international cross-section.

Discussion

The geographic diffusion of both the industrial revolution (i.e., modern economic growth) and the mortality revolution (i.e., the modern improvement
in life expectancy) is marked by the same leader–follower sequence, with Western European countries and their overseas descendants in the vanguard and sub-Saharan Africa bringing up the rear. The result is that any point-of-time comparison of countries in recent decades will show a significant positive relation between life expectancy and GDP per capita, with Western Europe and its descendants high on both variables and sub-Saharan Africa low. Although some analysts take this association as support for a causal connection running from economic growth to life expectancy, this inference is belied by two empirical observations. The first is the substantial difference in the timing of the onset of each phenomenon. While the industrial revolution dates from the early nineteenth century, the mortality revolution does not take off until over half a century later. The second is the marked difference in the rapidity of spread of the two revolutions. Although the mortality revolution starts later, it spreads much faster. By the latter part of the twentieth century, the mortality revolution had clearly reached sub-Saharan Africa; whether the industrial revolution has taken hold there yet remains arguable. These empirical observations suggest that each revolution must be analyzed in its own right in order to understand its underlying causes and spread. The presumption based on international cross-section regressions that growth is the motive force—that “wealthier is healthier”—is misleading. Cross-section regressions merely record that one set of countries got an early start on both revolutions and another set, a late start.

These conclusions, illustrated here for the industrial and mortality revolutions, are more generally applicable. In the past several centuries, Western Europe has been in the forefront of a number of revolutionary changes—the development of universal schooling, the rise of political democracy, the adoption of deliberate fertility control and consequent decline in childbearing, the empowerment of women, the welfare state, and so forth. Each of these is characterized by a particular pattern of onset and spread. For example, while the expansion of public schooling and the rise of political democracy both antedate by different lengths of time the onset of modern economic growth, universal primary schooling is today widespread throughout the world, while the diffusion of political democracy remains quite limited. Understanding these patterns calls for much more than simple cross-section regressions.

There remains, of course, the question why Western Europe in recent centuries has been the birthplace of what are viewed by many as major advances in the human condition. A plausible starting point is the scientific revolution. The new empirical and experimental mode of inquiry it initiated became the key to systematic knowledge of the natural world and, in time, the social world as well. The benefits of this knowledge have resulted in unprecedented improvements in human well-being, the evidence for which is documented, but not explained, by cross-section regressions.
Note

I am grateful for valuable assistance to Robson Morgan and Malgorzata Switek, and to the University of Southern California for financial assistance.

References

Population Change among the Elderly: International Patterns

ANDREW STOKES
SAMUEL H. PRESTON

Modern demography is, above all, about demographic transition. This is to be understood in two distinct, but interrelated senses. First, there is description in the spirit of the narrow, literal definition of demography—a description of quantitative aspects of human populations…. On a more ambitious level, there is the task of explanation.

—Demeny (1968: 502)

In his brilliant editorship of Population and Development Review, Paul Demeny encouraged both types of research on demographic change. This essay is in the spirit of the first type of study. Rather than dealing with the entire population, however, it restricts its attention to ages above 65, where population changes reflect, amplify, and reshape the broad patterns associated with the demographic transition. In this study, we develop and apply a demographic accounting model adapted to the segment of the population above age 65. Through an assessment of the terms in this model, we estimate the contribution of contemporaneous factors and their historical antecedents to rates of growth of the older population.

Background

In a population with fixed age-specific rates of fertility, mortality, and migration, all age groups grow at the same rate (Sharpe and Lotka 1911). Such a demographic regime characterized most of human history over broad sweeps of time and space. Starting in the eighteenth and nineteenth centuries, relatively steady declines in mortality began in Western countries. Such declines typically increased growth rates at ages under five years and above 60, the ages of greatest vulnerability to death (Coale 1972; Coale and Demeny 1983). These declines were followed after some decades by declines in fertility. Initially, fertility declines reduce only the growth rate at age zero, but eventually such reductions become visible throughout the age structure by reducing the number of
Population change at ages 65+ draws from the same demographic sources as population change at other ages, but adds its own filters and lags. Changes in fertility and in infant mortality will affect the population aged 65+ only after a delay of 65 years. Changes in mortality can be registered instantaneously in the death rate and thus in the growth rate at ages 65+, but they may also affect the rate of entrance into the 65+ population by virtue of past improvements in survivorship at ages below 65. Likewise, migration rates at older ages will immediately influence the growth rate at ages 65+, but they can also influence the rate of arrival at age 65 through lifetime migration patterns at younger ages. These factors are formalized in the next section.

Methods

We begin with a demographic accounting framework provided by the balancing equation of population growth as adapted to an age segment of the population rather than applied to the entire population. Through decomposition of the terms in this accounting identity, we estimate the contribution of contemporaneous factors and their historical antecedents to the growth rate of the older population.

The balancing equation of population growth expresses population growth over a time interval in terms of flows into the population via births and in-migration and flows out of the population via deaths and out-migration (Preston, Heuveline, and Guillot 2001). These terms are converted to rates of change by dividing the number of entrances and exits by person-years lived in the interval. When applying the accounting identity to the population aged \( x \) and above, the analog to the number of births is the number of \( x \)th birthdays that occur in the population during the interval (Brass 1975). The counting of deaths, in-migrants, out-migrants, and person-years of exposure is limited to ages \( x \) and above. The population growth rate above age \( x \) can be expressed as

\[
r_{x+} = b_{x+} - d_{x+} + m_{x+}
\]

In this equation \( r_{x+} \) is the growth rate of the population, \( d_{x+} \) is the death rate, and \( m_{x+} \) is the net rate of migration. All rates refer to the population aged \( x \) and above over a specified time interval. For expositional purposes, we assume that the time interval is from \( t \) to \( t+1 \). The remaining term, \( b_{x+} \), is the rate of new arrivals at age \( x \) by virtue of reaching one’s \( x \)th birthday. It is sometimes termed the “birthday rate” by direct analogy to the birth rate, but we use the broader term “arrival rate.” Each term on the right-hand side of the equation has a number of events during the specified period in the nu-
merator and the number of person-years lived above age $x$ during the same period in the denominator. Equation 1 connects the growth rate of the older population to contemporaneous demographic processes of births, deaths, and net migration, permitting evaluation of their respective contributions to the growth rate of the older population.

The number of $x$th birthdays during the period is in turn a product of the number of births into the cohort $x$ years earlier—i.e., births in the interval $t-x$ to $t-x+1$ for a one-year period—and any change in the size of the cohort resulting from mortality or migration. Thus the $b_{x \cdot}$ term in Equation 1 can be expressed as the birth rate $x$ years earlier, times the probability of survival between ages 0 and $x$ for that cohort, times a factor by which the cohort’s size is modified through migration between ages 0 and $x$. This decomposition connects the growth rate of the older population to past demographic events, thereby providing insight into future aging processes before they are revealed by the passage of time.

Further decomposition of Equation 1 is possible through elaboration of the $d_{x \cdot}$ term into components representing the intensity of mortality and the effects of age composition on the death rate. (See Appendix.)

The intensity of mortality is the age-standardized death rate above age $x$, i.e., the death rate that would prevail if the age composition were that of the standard population. The age-compositional effects on the crude death rate at ages $x+$ are represented by the component of the death rate above age $x$ that is a result of age-distributional differences between the actual population and the standard population. This term will have positive values where the population above age 65 is exceptionally old.

**Data**

In our primary analysis, we focus on growth rates above age 65, a common age of retirement and entitlement in many countries. We examine growth in the older population and its sources between 2005 and 2010 in a sample of 13 high-income countries drawn from North America, Western, Northern, and Southern Europe, and Asia. Countries were selected on the basis of the availability of detailed historical data series and are broadly representative of variation in demographic conditions among high-income countries. In a second stage, we explore growth at ages 65+ in the aggregates of more developed (MDCs) and less developed (LDCs) countries.

The growth rate of the population aged 65 and above, the arrival rate at age 65, and the death rate above age 65 are calculated using population and death counts by five-year age groups drawn from the United Nations (2011). The arrival rate at age 65 is calculated using data on population size in adjacent five-year age groups centered around 65 at the beginning and end of the time interval. In particular, the number of 65th birthdays occurring in the interval 2005–2010 is estimated by observing the number of people aged...
60–64 in 2005 and the number aged 65–69 in 2010. Linear interpolation between these numbers provides an estimate of the number of people who experienced their 65th birthday between 2005 and 2010. The remaining term in Equation 1, the net migration rate, is derived as a residual.

The three historical components of the arrival rate at age 65 are calculated using country-level data on the number of births, deaths, and exposure-to-risk from the Human Mortality Database (HMD) (2011). The relevant birth cohort for persons experiencing their 65th birthday in the interval 2005–2010 is persons born in 1940–45. The “historical birth rate” term, $b_{x}$, is calculated using the number of births in this period divided by person-years lived at ages 65+ during 2005–2010. For details on the construction of the series, see the Appendix in Preston and Stokes (2012).

To estimate the probability of survival between birth and age $x$, we combine age-specific death rates for the relevant birth cohort from each consecutive five-year period and age group. For example, for analysis of the population 65+, we combine deaths and exposure at ages 0–4 during 1940–44 with deaths and exposure at ages 5–9 during 1945–49 and so on until ages 60–64 during 2000–2005. We then apply standard life-table techniques to the cohort age-specific mortality rates to estimate the proportion of original cohort members who live to exact age 65 (using the ratio of life-table values $l_{65}/l_{0}$) (Preston, Heuveline, and Guillot 2001).

For decomposition of the death rate above age $x$ into components representing the intensity of mortality and the effect of age composition, the standard population used combines the age distribution of countries in Europe (excluding Eastern Europe) and North America as well as Australia, New Zealand, and Japan (standard population values available upon request).

For analysis of the growth of the older population and its sources in MDCs and LDCs, we use the UN classification system, under which MDCs are defined to include countries of Europe, North America, Australia, New Zealand, and Japan and LDCs comprise the remainder. We remove Eastern Europe from the set of MDCs because its demographic history is both atypical and hard to recover from available data. We use different procedures to identify the contributions of historical births and cohort survival to arrival rates between the two regions because of differences in data availability. The methods used for constructing demographic series for MDCs and LDCs are described in the Appendix of Preston and Stokes (2012).

**Results**

**High-income countries**

Table 1 presents growth rates of the 65+ population and the sources of growth in 13 high-income countries for 2005–2010. The mean and variance of each column are presented at the bottom of the table.
By far the fastest growth of the older population (2.6 percent annually) occurred in Japan and Australia. Different demographic factors are responsible for the rapid growth of the older population in these two countries. The proximate cause of the more rapid growth of the older population of Australia was its high rate of arrival into the 65+ population combined with slightly better-than-average mortality in 2005–2010. As shown in Table 2, the very high arrival rate in Australia is a product of an extraordinarily high rate of immigration at ages below 65 in the past. The size of birth cohorts reaching age 65 in Australia between 2005 and 2010 was inflated by approximately 61 percent as a result of a huge inflow of migrants at younger ages.

The exceptional rate of growth in Japan’s older population in 2005–2010, on the other hand, is principally attributable to the unusually low rate of mortality in older ages in that country. The death rate in Japan is only 78 percent of the mean value of the 13 countries. Table 3 shows that none of the exceptionally low mortality of the 65+ population in Japan is attributable to an unusual

### TABLE 1 Demographic components of growth in the 65+ population, 2005–2010

<table>
<thead>
<tr>
<th>Country</th>
<th>Growth rate of the population at ages 65+</th>
<th>Rate of arrival at age 65</th>
<th>Death rate at ages 65+</th>
<th>Net migration rate at ages 65+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>0.0255</td>
<td>0.0668</td>
<td>0.0412</td>
<td>0.0000</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.0075</td>
<td>0.0531</td>
<td>0.0460</td>
<td>0.0004</td>
</tr>
<tr>
<td>Finland</td>
<td>0.0205</td>
<td>0.0642</td>
<td>0.0448</td>
<td>0.0011</td>
</tr>
<tr>
<td>France</td>
<td>0.0100</td>
<td>0.0519</td>
<td>0.0420</td>
<td>0.0001</td>
</tr>
<tr>
<td>Italy</td>
<td>0.0136</td>
<td>0.0543</td>
<td>0.0424</td>
<td>0.0016</td>
</tr>
<tr>
<td>Japan</td>
<td>0.0269</td>
<td>0.0621</td>
<td>0.0350</td>
<td>−0.0002</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.0210</td>
<td>0.0676</td>
<td>0.0465</td>
<td>0.0000</td>
</tr>
<tr>
<td>Norway</td>
<td>0.0106</td>
<td>0.0639</td>
<td>0.0508</td>
<td>−0.0025</td>
</tr>
<tr>
<td>Spain</td>
<td>0.0144</td>
<td>0.0555</td>
<td>0.0445</td>
<td>0.0035</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.0186</td>
<td>0.0678</td>
<td>0.0495</td>
<td>0.0003</td>
</tr>
<tr>
<td>Switzerland</td>
<td>0.0171</td>
<td>0.0637</td>
<td>0.0431</td>
<td>−0.0035</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.0135</td>
<td>0.0617</td>
<td>0.0486</td>
<td>0.0005</td>
</tr>
<tr>
<td>United States</td>
<td>0.0203</td>
<td>0.0666</td>
<td>0.0484</td>
<td>0.0022</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0169</td>
<td>0.0615</td>
<td>0.0448</td>
<td>0.0003</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0035</td>
<td>0.0033</td>
<td>0.0018</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

**Region**

<table>
<thead>
<tr>
<th></th>
<th>Growth rate of the population at ages 65+</th>
<th>Rate of arrival at age 65</th>
<th>Death rate at ages 65+</th>
<th>Net migration rate at ages 65+</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDCs</td>
<td>0.0179</td>
<td>0.0611</td>
<td>0.0441</td>
<td>0.0010</td>
</tr>
<tr>
<td>LDCs</td>
<td>0.0250</td>
<td>0.0845</td>
<td>0.0595</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**NOTE:** Means reported under each column are unweighted. Variances are multiplied by a factor of 100 to reduce the number of reported decimals. MDCs are defined to include countries of Europe, North America, Australia, New Zealand, and Japan and LDCs comprise the remainder. Eastern Europe is removed from MDCs for the above calculations. Description of the methods used to construct historical demographic series for MDCs and LDCs are included in the Appendix to Preston and Stokes (2012).

**SOURCES:** United Nations (2011); Human Mortality Database (2011).
age distribution. If Japan had the same death rate at ages 65+ as the mean of the 13 countries (44.8 per 1000 instead of 35.0 per 1000), its growth rate would actually be slightly below the mean rather than the highest. Japan is a world leader in longevity (National Research Council 2011); one consequence is an exceptionally rapid growth of the 65+ population.

Countries with the lowest rates of growth in the 65+ population between 2005 and 2010 are France and Belgium. These countries had exceptionally low arrival rates at age 65 in 2005–2010. Table 2 shows that the low arrival rate reflects very low birth rates during World War II (Dublin 1945). Italy and Spain had the third and fourth lowest arrival rates, but their position is not a result of a deficiency of births but rather of relatively low survivorship to age 65 combined with relatively low lifetime immigration into

### TABLE 2 Demographic components of the rate of arrival at age 65, 2005–2010

<table>
<thead>
<tr>
<th>Country</th>
<th>Rate of arrival at age 65</th>
<th>Ratio of births in 1940–45 to person-years lived above age 65</th>
<th>Probability of survival from birth to age 65 for the 1940–45 birth cohort</th>
<th>Factor by which cohort born 1940–45 has changed through migration by age 65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>0.0668</td>
<td>0.0507</td>
<td>0.8201</td>
<td>1.6083</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.0531</td>
<td>0.0628</td>
<td>0.7521</td>
<td>1.1235</td>
</tr>
<tr>
<td>Finland</td>
<td>0.0642</td>
<td>0.0861</td>
<td>0.7481</td>
<td>0.9972</td>
</tr>
<tr>
<td>France</td>
<td>0.0519</td>
<td>0.0544</td>
<td>0.7484</td>
<td>1.2744</td>
</tr>
<tr>
<td>Italy</td>
<td>0.0543</td>
<td>0.0778</td>
<td>0.7288</td>
<td>0.9574</td>
</tr>
<tr>
<td>Japan</td>
<td>0.0621</td>
<td>0.0841</td>
<td>0.7258</td>
<td>1.0174</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.0676</td>
<td>0.0830</td>
<td>0.8064</td>
<td>1.0092</td>
</tr>
<tr>
<td>Norway</td>
<td>0.0639</td>
<td>0.0774</td>
<td>0.8138</td>
<td>1.0145</td>
</tr>
<tr>
<td>Spain</td>
<td>0.0555</td>
<td>0.0771</td>
<td>0.7025</td>
<td>1.0240</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.0678</td>
<td>0.0708</td>
<td>0.8297</td>
<td>1.1549</td>
</tr>
<tr>
<td>Switzerland</td>
<td>0.0637</td>
<td>0.0633</td>
<td>0.8197</td>
<td>1.2276</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.0617</td>
<td>0.0784</td>
<td>0.7866</td>
<td>1.0005</td>
</tr>
<tr>
<td>United States</td>
<td>0.0666</td>
<td>0.0753</td>
<td>0.7694</td>
<td>1.1495</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0615</td>
<td>0.0724</td>
<td>0.7732</td>
<td>1.1199</td>
</tr>
<tr>
<td>Variance</td>
<td>0.9386</td>
<td>2.8192</td>
<td>0.3011</td>
<td>2.0418</td>
</tr>
</tbody>
</table>

#### Region

<table>
<thead>
<tr>
<th>Region</th>
<th>Rate of arrival at age 65</th>
<th>Ratio of births in 1940–45 to person-years lived above age 65</th>
<th>Probability of survival from birth to age 65 for the 1940–45 birth cohort</th>
<th>Factor by which cohort born 1940–45 has changed through migration by age 65</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDCs</td>
<td>0.0611</td>
<td>0.0761</td>
<td>0.7574</td>
<td>1.0607</td>
</tr>
<tr>
<td>LDCs</td>
<td>0.0845</td>
<td>0.2225</td>
<td>0.3799</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

* By assumption

NOTE: Means reported under each column are unweighted. Variances are based on the logarithm of the data since the terms are multiplicative and are multiplied by a factor of 100 to reduce the number of reported decimals. See Table 1 for definition of MDCs and LDCs.

the cohorts reaching age 65. In Italy’s case, 4.3 percent of the births during 1940–45 left the country before reaching age 65, the greatest loss suffered by any country.

Among the three contemporaneous demographic processes shown in Table 1, the greatest contributor to international differences in the 65+ growth rate is the arrival rate. The variance in the arrival rate is roughly double that of the death rate, while contemporary migration into the 65+ population makes only a small contribution to international differences. Spain is an exception with a relatively high immigration rate into the elderly population, but this is probably a temporary phenomenon related to an amnesty program in 2005 that applied to illegal immigrants (Preston and Wang 2007). Norway and Switzerland experience moderate out-migration among the elderly.

The arrival rate at a given age is a function of the relative size of the original birth cohort, the proportion of those births who survived to age 65, and the rate of migration into the birth cohort before reaching age 65. The

<table>
<thead>
<tr>
<th>Country</th>
<th>Death rate at ages 65+</th>
<th>Age-standardized death rate in the 65+ population</th>
<th>Age-compositional effects on the crude death rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>0.0412</td>
<td>0.0409</td>
<td>0.0003</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.0460</td>
<td>0.0473</td>
<td>−0.0013</td>
</tr>
<tr>
<td>Finland</td>
<td>0.0448</td>
<td>0.0473</td>
<td>−0.0025</td>
</tr>
<tr>
<td>France</td>
<td>0.0420</td>
<td>0.0392</td>
<td>0.0028</td>
</tr>
<tr>
<td>Italy</td>
<td>0.0424</td>
<td>0.0424</td>
<td>0.0000</td>
</tr>
<tr>
<td>Japan</td>
<td>0.0350</td>
<td>0.0350</td>
<td>0.0000</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.0465</td>
<td>0.0489</td>
<td>−0.0024</td>
</tr>
<tr>
<td>Norway</td>
<td>0.0508</td>
<td>0.0464</td>
<td>0.0045</td>
</tr>
<tr>
<td>Spain</td>
<td>0.0445</td>
<td>0.0446</td>
<td>−0.0001</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.0495</td>
<td>0.0462</td>
<td>0.0033</td>
</tr>
<tr>
<td>Switzerland</td>
<td>0.0431</td>
<td>0.0408</td>
<td>0.0022</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.0486</td>
<td>0.0476</td>
<td>0.0010</td>
</tr>
<tr>
<td>United States</td>
<td>0.0484</td>
<td>0.0464</td>
<td>0.0020</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0448</td>
<td>0.0441</td>
<td>0.0008</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0018</td>
<td>0.0017</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>Death rate at ages 65+</th>
<th>Age-standardized death rate in the 65+ population</th>
<th>Age-compositional effects on the crude death rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDCs</td>
<td>0.0441</td>
<td>0.0441</td>
<td>0.0000</td>
</tr>
<tr>
<td>LDCs</td>
<td>0.0595</td>
<td>0.0764</td>
<td>−0.0168</td>
</tr>
</tbody>
</table>

Table 3 Demographic components of the death rate at ages 65+, 2005–2010

<table>
<thead>
<tr>
<th>Country</th>
<th>Death rate at ages 65+</th>
<th>Age-standardized death rate in the 65+ population</th>
<th>Age-compositional effects on the crude death rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>0.0412</td>
<td>0.0409</td>
<td>0.0003</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.0460</td>
<td>0.0473</td>
<td>−0.0013</td>
</tr>
<tr>
<td>Finland</td>
<td>0.0448</td>
<td>0.0473</td>
<td>−0.0025</td>
</tr>
<tr>
<td>France</td>
<td>0.0420</td>
<td>0.0392</td>
<td>0.0028</td>
</tr>
<tr>
<td>Italy</td>
<td>0.0424</td>
<td>0.0424</td>
<td>0.0000</td>
</tr>
<tr>
<td>Japan</td>
<td>0.0350</td>
<td>0.0350</td>
<td>0.0000</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.0465</td>
<td>0.0489</td>
<td>−0.0024</td>
</tr>
<tr>
<td>Norway</td>
<td>0.0508</td>
<td>0.0464</td>
<td>0.0045</td>
</tr>
<tr>
<td>Spain</td>
<td>0.0445</td>
<td>0.0446</td>
<td>−0.0001</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.0495</td>
<td>0.0462</td>
<td>0.0033</td>
</tr>
<tr>
<td>Switzerland</td>
<td>0.0431</td>
<td>0.0408</td>
<td>0.0022</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.0486</td>
<td>0.0476</td>
<td>0.0010</td>
</tr>
<tr>
<td>United States</td>
<td>0.0484</td>
<td>0.0464</td>
<td>0.0020</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0448</td>
<td>0.0441</td>
<td>0.0008</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0018</td>
<td>0.0017</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>Death rate at ages 65+</th>
<th>Age-standardized death rate in the 65+ population</th>
<th>Age-compositional effects on the crude death rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDCs</td>
<td>0.0441</td>
<td>0.0441</td>
<td>0.0000</td>
</tr>
<tr>
<td>LDCs</td>
<td>0.0595</td>
<td>0.0764</td>
<td>−0.0168</td>
</tr>
</tbody>
</table>

NOTE: Means reported under each column are unweighted. Variances are multiplied by a factor of 100 to reduce the number of reported decimals. The standard population is the 65+ population of more developed countries in 2005–2010 (according to UN classification) excluding Eastern Europe. See Table 1 for definition of MDCs and LDCs.

relative size of the original birth cohort is the largest source of variance in the arrival rate, although the migration factor is also powerful, in part because Australia is such an outlier (Table 2). France also experienced very rapid immigration, which swelled the size of the original birth cohort by 27 percent by age 65. Cohort survival probabilities also varied, ranging from a low of 70 percent in Spain to a high of 83 percent in Sweden. These rankings reflect comparative mortality levels among children and young adults in the period 1940–60. For example, 13 percent of births died before reaching age 10 in the Spanish life table of 1950, compared to 3 percent of Swedish births in the same period (Keyfitz and Flieger 1968: 456, 504).

Table 3 shows, perhaps surprisingly, that the age structure of the 65+ population has a relatively minor influence on the death rate of the 65+ population. The largest effects are in Sweden and Norway, where relatively old age structures raise the death rate by 3–5 per 1000.

In terms of the components of demographic growth among the elderly, the United States stands out for its normalcy. Its values are not exceptional on any of the eight indexes in Tables 1–3. That circumstance is about to change as baby boomers reach age 65. How results vary with the initial age is considered in the next section.

Extension of results to other ages

In our primary analysis, we have focused on age 65 because of its social significance. Had we chosen another age, the ranking of growth rates across countries would be different. Figure 1 (a) shows growth rates in the five largest developed countries during 2005–2010 at age $x+$, where $x$ varies from 60 to 80. The patterns in this figure are erratic, with lines that connect the growth rates of particular countries often intersecting. For example, the growth rate above age 60 in US is well above that of France and Italy and on par with Japan, while above age 80, the US growth rate is well below that of Japan, France, and Italy.

The fluctuation in growth rate rankings is explained primarily by variability in the arrival rates, reflecting erratic births series in the period from 1940 to 1960. As an illustration of this point, in Figure 1 (b) we plot the growth rate above age $x$ in the same five countries included in Figure 1 (a). However, in this case we follow a particular set of birth cohorts through time, i.e., those aged 60+ in 1985–90, 65+ in 1990–95,…80+ in 2005–10. Because observations for a country pertain to the same set of birth cohorts, fluctuations in the number of births cannot influence the series. And it is obvious in Figure 1 (b) that when such fluctuations are removed, the series become much more orderly than in Figure 1 (a), with only one cross-over occurring. The growth rate series rise smoothly with age for a particular cohort because mortality improvements are occurring more or less continuously as the cohort ages (Preston and Stokes 2012). This demonstration helps to confirm the
evidence in Table 1 that the arrival rate is the principal source of variation in growth rates at older ages.

World regions

We now examine the growth rate of the older population for the aggregate of more developed and less developed countries, using the United Nations’ classification of these groups. Tables 1–3 present the growth rate of the 65+ population and its decomposition for the two regions. Note that the elements of growth for MDCs ins Tables 1–3 are quite similar to the mean of the 13 developed countries.
The older population in LDCs is growing more rapidly than that in MDCs. This difference is entirely a result of higher arrival rates in LDCs. The death rate of the older population in LDCs is much higher than that in MDCs (59.5 per 1000 compared to 44.1 per 1000) despite the fact that its 65+ population is much younger than that in MDCs. As shown in Table 3, once age-distributional differences are controlled, the 65+ death rate in LDCs exceeds that in MDCs by 32.3 per 1000 rather than by 15.4 per 1000.

The decomposition of the arrival rate shows pronounced differences between MDCs and LDCs (Table 2). The term reflecting birth rates in the period 1940–45 is three times higher in LDCs than in MDCs. However, the probability of surviving from birth to age 65 for this birth cohort was only half as great in LDCs (38 percent) as in MDCs (76 percent). If LDCs had maintained the same cohort survival rates as MDCs and all other elements of the growth rate equation remained the same, the annual growth rate of the 65+ population in LDCs would have been 10.9 percent, more than four times its value for 2005–2010. While this is an implausible scenario, it is nevertheless evident that a history of extremely high mortality in LDCs has left a massive imprint on contemporary demographic processes at older ages.

Thus, despite much higher historic birth rates in LDCs, their significantly lower cohort survival, together with the high contemporary death rate of its 65+ population, has produced a growth rate of the 65+ population in LDCs that is only 0.007 higher than that of MDCs.

The future

Our findings on the determinants of older population growth at the beginning of the twenty-first century primarily reflect twentieth-century demographic history. The distinction between arrival rates and death rates is also useful for clarifying elements of the future growth of the 65+ population. Figures 2 (a) and (b) show the arrival rates, death rates, and growth rates that are implied by UN medium-variant population and mortality projections for more developed and less developed countries. In both cases, the death rate declines until about 2030–35, where it levels off. Because projected mortality levels continue to improve, the leveling of death rates is a consequence of an increasingly older age structure among the 65+ population.

Arrival rates also show a declining trend in both sets of countries, but the trend is interrupted during the present decade. In the case of more developed countries, the interruption is clearly a result of the postwar baby boomers reaching age 65. In less developed countries, on the other hand, the impending rapid increase in arrival rates appears to be primarily a consequence of large declines in infant and child mortality. Growth rates at ages under 10 accelerated rapidly from 1940–45 to 1950–55, coinciding with the widespread introduction of anti-malarial programs and other public health efforts, including the use of antibiotics. The accelerated growth rates were maintained as these cohorts
aged and are now beginning to affect growth rates at ages 65+ (Preston and Stokes 2012).

The sudden burst in arrival rates in both sets of countries will have ended by around 2015. After that date, arrival rates are projected to decline, echoing earlier declines in birth rates, and to do so faster than declines in death rates. Accordingly, growth rates of the older population in both sets of countries are projected to show a persistent pattern of decline after 2015.

The classic graphical representation of the demographic transition shows death rates initially declining, birth rates declining later, the differ-

**FIGURE 2**  Growth, arrival, and death rates, population 65+, 1995–2000 to 2050–2055 in (a) more developed countries and (b) less developed countries

NOTE: Population and mortality data after 2010 are based on UN medium-variant projections. Observations are located at the beginning of the 5-year interval over which each rate is calculated. Projections for MDCs are based on all countries classified as more developed under UN classification, including countries of Eastern Europe.

ence between birth rates and death rates first widening and then narrowing, and growth ultimately declining to low levels or ceasing altogether. Figure 2 shows such a pattern in both developed and developing countries for the population aged 65+. While in MDCs this pattern is a somewhat coincidental result of the timing of the baby boom, the projected pattern of growth in LDCs at ages 65+ is an authentic echo of the earlier demographic transition pertaining to all ages combined. And with annual growth rates at ages 65+ still as high as 2 percent in 2050, it is clear that this age-distributional tail will be hosting its own version of a demographic transition for most of the twenty-first century.

Appendix

For a discrete time interval $\tau$, the population growth rate above age $x$, $r_{sx}(\tau)$, can be expressed as

$$r_{sx}(\tau) = b_{sx}(\tau) - d_{sx}(\tau) + m_{sx}(\tau)$$  \hspace{1cm} (A1)

where $d_{sx}(\tau)$ is the death rate, $m_{sx}(\tau)$ is the net rate of migration, and $b_{sx}(\tau)$ is the rate of new arrivals reaching their $x$th birthday (the arrival rate). Each term on the right-hand side of the equation has a number of events during the period $\tau$ in the numerator and the number of person-years lived above age $x$ during period $\tau$ in the denominator.

The number of $x$th birthdays during the period $\tau$ is in turn a product of the number of births into the cohort over the period $\tau - x$ and any change in the size of the cohort resulting from mortality or migration. Thus the $b_{sx}(\tau)$ term in Equation 1 can be expressed as

$$b_{sx}(\tau) = b_0(\tau - x) p_0^{x-x} x f_0^{x-x}$$  \hspace{1cm} (A2)

where $b_0(\tau - x)$ is the number of births in the population over the interval $\tau - x$ divided by person-years lived above age $x$ during period $\tau$; $p_0^{x-x}$ is the probability of survival between ages 0 and $x$ for a member of the birth cohort born during the interval $\tau - x$; and $f_0^{x-x}$ is the factor by which the birth cohort’s size is modified through migration between ages 0 and $x$. The term $b_0(\tau - x)$ is related to the actual birth rate in the interval $\tau - x$ by a factor equal to the ratio of the population aged $x$ and above in period $\tau$ to the total population in the interval $\tau - x$.

Further decomposition of Equation A1 is possible through elaboration of the $d_{sx}(\tau)$ term into components representing the intensity of mortality and the effects of age composition on the death rate. The death rate above age $x$ can be expressed as

$$d_{sx}(\tau) = \sum_x [c_a^d(\tau) m_a(\tau) + \sum_x [c_a^d(\tau) m_a(\tau) - \sum_x c_{a+1}^d m_a(\tau)]$$

$$= \sum_x [c_a^d(\tau) m_a(\tau) + \sum_x [c_a^d(\tau) - c_{a+1}^d] m_a(\tau)]$$  \hspace{1cm} (A3)
where \( \phi_{a}^{(\tau)} \) is the proportion of the population aged \( a \) to \( a+1 \) in the interval \( \tau \); \( m_{a}^{(\tau)} \) is the death rate between age \( a \) and \( a+1 \) during the period; and \( \phi_{a}^{(S)} \) is the proportion of the population ages \( a \) to \( a+1 \) in an arbitrarily chosen standard population. Each product is summed between ages \( x \) and \( \beta \), the highest age attained. The first term in Equation A3 is the age-standardized death rate above age \( x \), representing the intensity of mortality. The second term is the component of the death rate above age \( x \) that is a result of age-distributional differences between the actual population and the standard population.

### Note

This project was funded through support from the National Institutes of Health–National Institute on Aging, award number P30 AG012836, the Boettner Center for Pensions and Retirement Security; and National Institutes of Health–National Institute of Child Health and Development Population Research Infrastructure Program R24 award number HD-044964, both at the University of Pennsylvania.

### References


The French School of Demography: Contextualizing Demographic Analysis

VÉRONIQUE PETIT
YVES CHARBIT

Demographic analysis, according to Pressat’s *Dictionary of Demography*, is “a form of statistical analysis which employs, for the most part, a modest array of mathematical and statistical techniques to deal with the data produced by censuses, surveys and vital registration systems” (1985: 52). In this essay we explore the methodological advances in this field between the 1960s and the late 1990s. We argue, against the grain of the standard definition of the subject, that these efforts were directed at refining demographic analysis in a quest for scientific purity, to the exclusion of other objectives. As Greenhalgh puts it, despite what alternative histories of the field might suggest, “the conventional rhetoric of steady scientific progress continues to dominate discussions of the creation of population thought” (2012: 122). Demography, we believe, has paid a high price for this narrowing of focus. Epistemological questions such as the conceptualization of the individual and the treatment of culture have been eluded; ideological biases have entered. We argue that an alternative paradigm, more open to interdisciplinarity, should be explored.

Our attention is primarily directed to France and to the so-called French school of demography, which the National Institute for Demographic Studies (INED) very early claimed to personify. As is known, under the leadership of Louis Henry the department of historical demography at INED achieved major methodological progress with regard to the knowledge of populations in the pre-census era; the department of psycho-sociology, successively headed by Jean Stoetzel, Alain Girard, and Henri Léridon, completed large socio-demographic surveys. However, another influential part of that school proclaimed the uniqueness of its contribution to the heart of the discipline, “pure” demographic analysis. It was defined as the study of population dynamics, paying
minimum attention to epistemological, conceptual, and methodological issues and problems that underlie what is now termed population studies. The following discussion goes beyond the case of France, as the discipline evolved in other countries as well.

The quest for demographic purity

Among INED senior staff, Roland Pressat was one of the founders of the French school of demography, while Daniel Courgeau was one of the leaders of its second generation. Courgeau divides the development of demographic analysis into four distinct stages: cross-sectional, cohort, biographical, and multilevel. The progression traces an increasing emphasis on the plurality of levels of observation and on the complexity of interactions in the analysis of individual behavior (Courgeau 2004).

Cross-sectional analysis and the independent variables

From the eighteenth century to the mid-twentieth, cross-sectional analysis was the dominant approach in demography. Following the path of political arithmetic, this approach began by making widespread use of censuses and vital records, then in the nineteenth and twentieth centuries adding a range of demographic indicators. Cross-sectional analysis, says Courgeau, removes “all thickness from human life,” insofar as it assumes demographic phenomena to be determined by the characteristics of the examined population, at the moment they occur (2004: 39). In Ryder’s words, “aggregate analysis destroys individual sequences, and diverts attention from process. By implying that the past is irrelevant, cross-sectional analysis inhibits dynamic inquiry and fosters the illusion of immutable structure” (1965: 859). We would add that contextual explanatory factors that are introduced in work based on cross-sectional data are not strongly related to individual demographic behavior. Economic, political, religious, and social changes observed at the societal level are governed by their own causality independently of the micro level of the individual.

A brief backward glance to the 1920s and 1930s can help account for this situation. The “methodological toolkit” of demography was largely formed during this period, one of extreme conservatism in the discipline. The need “to create abstract measures of population processes and to treat these measures as if they comprise the best, or even the only, way to conceive of the issues at hand” resulted in a reification of demographic thought and delayed the emergence of newer methodologies (Wilson and Oeppen 2003: 113). Despite the prevalence of heterogeneity as an empirical fact, demography remained firmly—even perversely—attached to the implicit hypothesis of homogeneity. It has searched for average behavior, as if driven by Quetelet’s
obsessive quest for the “average man.” Its measures and practices emphasized central tendency at the expense of variance, the plurality of experiences, and heterogeneity, all of which are characteristics crucial for analyzing and understanding social phenomena. The focus on homogeneity governed the choice of the specific social and economic determinants used in the analysis of demographic processes. Thus in Britain, the 1911 Census of Fertility set the parameters for British demography throughout the twentieth century. Following Szreter (1996), Wilson and Oeppen pointed to a notable consequence of this professional practice: indicators repeatedly transmitted became natural constructions in demographic research, becoming so familiar that their nature and underlying assumptions were forgotten (2003: 120). Demographers, they remark, are often more fascinated by their models than by the actual results of the processes observed in their research. Yet it is precisely those results that are drawn on in designing public policy in population and health. A consequence of this “methodological lock-in,” clearly in evidence when comparing demography to such other disciplines as statistics, economics, and the natural sciences, has been that “many of the analytical and explanatory frameworks used by demographers show a remarkable record of survival” (De Gans 1999: 124).

**Cohort analysis and homogeneity**

After World War II, in response to criticisms leveled against cross-sectional analysis, demographers embarked on new research based on a second methodology, cohort analysis. In contrast to the fictive cohort posited in cross-sectional analysis, cohort analysis follows “the destiny of actual cohorts…. What matters is the time spent in a given state [by] groupings of subpopulations observed at every age and drawn from a given generation” (Courgeau 2004: 48). In the late 1990s an advocate of traditional demographic analysis, Chantal Blayo, another senior research fellow at INED, triggered a methodological debate when she defended the importance of longitudinal analysis for “isolating various demographic phenomena in their purity, in order to rid them of the effect of disruptive phenomena and to draw comparisons between countries and periods.” Demographers, she held, should study a phenomenon throughout the life of a single generation or cohort in a population “that preserves all of its features and the same features as long as the phenomenon is observed” (Blayo 1995: 1504). However, the application of cohort analysis requires that the population be treated as a homogeneous entity and that disruptive phenomena be independent of the examined phenomenon. Contrary to Blayo, Courgeau argues (2004: 48–49) that cohort analysis, like cross-sectional analysis, necessitates a holistic approach, but for a reason that is, in our view, equally problematic: cohort analysis is based on a hypothesis of homogeneity, whereas in reality heterogeneity prevails.
Biographical analysis: Methodological individualism

In France, as elsewhere, the limitations of both cross-sectional and cohort analysis led in the early 1980s to the development of a third methodology, biographical analysis. This development was said to represent a revolution in demographic analysis, since it amounted to a renunciation of the attempt to isolate every phenomenon in order to see, instead, “how a given event can subsequently affect the life of the individual and how certain characteristics may compel an individual to behave differently from another individual” (Courgeau and Lelièvre 1997; Courgeau 2004: 13). The new approach required collection of a new kind of demographic data, through so-called biographical surveys, since traditional population registers, censuses, and surveys did not provide the required degree of specificity on individuals. The focus of research thus shifted from homogeneous subpopulations to individual trajectories. The unit of analysis was now the individual biography, conceived as a stochastic process. The anticipated future trajectory of an individual at a given moment in time can be assumed to be a function of his or her previous experience. To its proponents biographical analysis was a resolutely individualistic approach based on methodological individualism, and as such it was clearly in opposition to cross-sectional analysis. But these proponents added a crucial assumption: “the behavior of a man needs to be connected with his past life history, without searching in society for the reasons for his actions” (Courgeau 2004: 70; emphasis added). That assumption, we would note, is open to a major criticism. While the biographical approach takes into account the individual and the past motivations for his or her present actions, it neglects the fact that the attitudes and behaviors of individuals are in important respects subject to social control. The individualized approach therefore suffers from a decontextualized vision of the individual, reduced to a free agent and isolated from the social environment. Ryder had earlier made this point in noting that the broader historical context shapes the behavior of demographic cohorts (1965: 848–851).

Multilevel analysis: A step toward contextualization

A fourth type of demographic analysis was an effort to remedy this deficiency. It entailed combining contextual and multilevel analysis. This new approach sought to examine not only the events of individual trajectories, as in biographical analysis, but also the characteristics of the environments in which individuals live. The approach identifies the relevant levels of analysis and privileges the most important ones. While it clearly represents a methodological innovation, allowing for a greater consideration of social complexity, its implementation depends on the existence of appropriate data. This constraint explains why applications of multilevel analysis have developed only slowly (see Golaz and Bringé 2009).
Courgeau presents the four types of demographic analysis as successive stages of progress in the explanatory power of demography. However, it would be wrong to suppose that the demographic community simply proceeded from one type of analysis to the next. The types are not mutually exclusive; choice among them depends on the available data and on the specific research objectives. INED, as a notable practitioner of demographic analysis, provides a case in point. Every year since 1969 its authoritative research unit, Département de conjoncture, produces a report to Parliament on demographic trends in the French population, along with European and wider international comparisons, and provides in-depth treatment of a selected topic (e.g., abortion, divorce, new family forms). Cross-sectional as well as cohort data are presented, often simultaneously, depending on their availability and relevance to the topic considered. In the late 1960s, French demographers, most of them employed at INED, debated the implications of the observed downturn in the country’s total fertility rate after its 1964 peak of 2.9. Was there an actual change in the fertility of more recent cohorts or simply a change in the timing of births? (On this debate, see for instance Blayo 1972, 1974; Blayo and Festy 1975; Bourgeois-Pichat 1972; Calot, Hémery, and Piro 1969; Pressat 1973a, 1973b; Marchal and Rabut 1972). Events soon settled the debate: it turned out that high fertility was a phenomenon of the past. After reaching a maximum in the cohort born in 1931, fertility declined to 2.2 by the 1943 cohort, the most recent one for which a reliable estimate could then be made.

A more general observation to be made about these analytical developments is that different conceptions of population with differing epistemological implications now coexist within the field. The general consensus around the object and methods of analysis that long characterized the demographic community appears to have collapsed in recent decades. Rather than displaying a shift from one type of analysis to the next, the chief feature of the current landscape of demography is a juxtaposition of different analytical approaches.

The statistical individual: An actor without qualities

The stages in the development of demographic analysis described above correspond to different epistemological treatments of the individual and her societal context, but they all incorporate demographic analysis into a minimal sociological framework. In this regard, demographic research in the 1980s was out of step with contemporary advances in knowledge and theory elsewhere in the social sciences (and even in natural sciences such as genetics), with their interest in socialization, learning conditions, the transmission of identity, and similar issues. The individual as conceived by demographers was, and still is, a being endowed with very few social and psychological
qualities. As Le Bras has noted, demography is the only science in which “processes exist in a pure state, with the minimum of hypotheses concerning human behavior” (2005: 421). Demography lacks significant cultural and historical thickness. There is no choice but to treat the individual as a sketchy statistical entity—otherwise it would be impossible to submit such a complex and puzzling entity to demographic analysis. It has been argued that sociology shares this defect: according to Lahire (2006: 18), “sociology appears to operate on the basis of partly outdated and partly challenged psychological knowledge, as if encysted in its theories of action and cognition.” However, demography is arguably much more resistant than sociology to the contributions of psychology or philosophy, so the shortfall in its case is greater. Demographers, largely engaged in applied research, seldom examine or question the concept of the individual. They draw uncritically on the concept in its common Western sense. Only when confronted with other societies shaped by different philosophical traditions, cultures, and religions are they forced to recognize the limitations of their discipline’s conceptual and methodological ethnocentrism. Anthropology experienced a salutary crisis in dealing with decolonization, but that political transformation had very little effect on demography: the new states that emerged with independence were keen to develop their own statistics, but they typically retained the same instruments and employed the same categories (even the same notion of ethnic group) as the former colonial powers.

Demography shares with other disciplines the problem that individuals have complex identity affiliations and relationships linking them to a range of different reference groups and networks, which are sometimes competing and even conflicting, and individuals will tend to use their many affiliations or memberships according to the circumstances in which they find themselves. The researcher must somehow avoid prejudging the relative significance of these affiliations and relationships. The relevant analytical levels vary among societies, and even among social groups within the same population. Moreover, in Western societies at least, social relations are in the process of being radically redefined—seen, for example, in the experience of disaffiliation (particularly among the poor) and individualization (among the young). Ironically or paradoxically, just as demography is seeking to capture individuals in society within the context of their various relations and networks, individuals are increasingly seeking to redefine the relations they wish to maintain with society at their own level and on their own terms.

Describing the wide range of contexts that determine demographic behaviors, Hobcraft has inferred two major consequences from the increasing complexity of demographic determinants. First, more elaborate hypotheses are needed to measure the relative influence of each level of determinant. Second, the angle of observation must be modified, as the object of analysis becomes less purely demographic and more socio-demographic or anthropo-
demographic. “The viewpoint thus shifts away from marriages, divorces, deaths, migrations and the description of the structure of households toward a perspective that encompasses couples and intimacy, parenthood, potential and well-being, the position within society and geographic space and interpersonal relations” (Hobcraft 2007: 50). A broader conception of the individual is needed. The implicit or explicit idea of a simple *homo demographicus*, like the often-derided *homo economicus*, must be abandoned (Petit 2012).

**The temptation of reductionism and the black box of culture**

Social reality is that of a complex world that resists understanding. It is far from the “population” artifact typically examined by demographers. Of course, to some extent all researchers in the social sciences work with some sort of artifact; all are confined to “approaching” social reality (to quote Gaston Bachelard). But the reductive approach to social complexity implied by demography entails an especially stark disembodiment of reality. Consider, for example, demographic transition theory. Successive refinements of this theory in the second half of the twentieth century, in the absence of serious epistemological debates within the discipline, were largely inspired by advances in areas outside demography, such as sociological theory (functionalism, modernization), anthropological theory (diffusion, culturalism, institutional approaches), and economic theory (growth theory, microeconomics). Demographers display little interest in higher theoretical issues. Indeed, they risk being subject to C. Wright Mills’s fierce warning: “theory without data is empty, data without theory are blind” (Mills 1968).

An implication of reducing the objective of demography to measurement is that demographers can work only with methods that permit measurement and on phenomena that are by nature measurable. But what is a measurable phenomenon? This question is fundamental since it is precisely along these lines that the distinction between the demographic field and the non-demographic field tends to be drawn. Over forty years ago, in his book *La Population*, Alfred Sauvy examined the qualities of individuals that called for measurement. In his view, some were obvious characteristics such as age, but he regarded as more problematic anything that pertained to culture, identity, and social processes—in short, anything that could not easily be turned into a quantitative variable (1968: 55).

To tackle complexity requires modifying the treatment of the cultural realm. Once demographers have exhausted all statistical possibilities, if the variables used in their model have failed to explain the totality of the observed object, it is not uncommon for them to say: “the rest—that which is unexplained, not to say inexplicable—is cultural.” Thus do they admit impotence when faced with this multifaceted, elusive magma that is peculiarly resis-
tant to quantification and modeling. They recognize the limitations of their discipline and seek to absolve themselves of responsibility by arguing that culture is a matter beyond the boundaries of their expertise. That is hardly an acceptable standpoint, however. When demographers deal with culture, they should be able to provide a substantive definition of what they mean and some progress toward quantification—in terms that make sense to an anthropologist (Hammel and Howell 1987; Hammel 1990). Even if convinced of the need to incorporate the cultural dimension in demographic research, demographers are faced with a number of conceptual and methodological problems that cause most practitioners to postpone the task.

Culture can be examined from two different perspectives: that of groups, defined by the key cultural features forming their identity (given that culture, like identity, is necessarily contextual and historical); or that of cultural mechanisms, such as education, transmission, socialization, or cultural integration. Demographers may seek to quantify certain aspects of culture for the purposes of their models; or they may take cultural complexity into account within their research questions by using more anthropologically oriented approaches and by refining the contextualization of demographic behaviors. The distinction between demographic and other human behaviors (social, economic, political) is becoming less and less relevant insofar as the cultural dimension appears to dominate the biological.

Examining how culture should be treated in demography raises the broader question of whether we need to measure a phenomenon in order to understand it, a much-debated issue in the social sciences. We might invert the logic and say that we need to understand a phenomenon (to identify its components, its mechanisms) in order to measure it as accurately as possible. Should understanding the object precede its measurement or vice versa? The answer has implications for the timing of quantitative and qualitative investigations when they are pursued in combination. While we do not claim to have an answer, we would argue that a particular dialectic is at work. The logical order between measurement and understanding is not as straightforward as it is generally assumed to be.

Which paradigm for demography among the social sciences?

Courgeau has argued that new methodological developments in demography, in particular multilevel analysis, because it opens demography “to the analysis of new structures” relevant across many of the human sciences, should favor a broad rapprochement among them (2004: 209–210). This seems doubtful: this mere methodological progress of demography, presented as no less than a new paradigm, at most will lead to a rapprochement of disciplines or researchers sharing this specific epistemological framework. Admittedly,
economics, quantitative sociology, and human geography appear to be direct beneficiaries of the opportunities provided by this type of analysis, but it is difficult to imagine how anthropology, history, and a more comprehensive form of sociology might also benefit. The latter disciplines will continue to emphasize subjectivity and contingency at the expense of causality. This impasse provides a useful illustration of Jean-Louis Fabiani’s remark that “the epistemological break is in itself institutionalizing: it produces demarcation” (2006: 24). In particular, the tension between statistical reasoning and historical contextualization is likely to continue.

There is another way, however, to construct an epistemological framework capable of transcending disciplinary boundaries: that is by seeking to reconcile quantitative and qualitative approaches. Although agreement can be reached to acknowledge social complexity in the development of a new epistemology, there is a risk of ending up with “a science located half-way between two scientific approaches if not a science located half-way from science” (Passeron 2006: 57). Demographers now often use quantitative and qualitative methods within the same research project, but tend to view the former as the yardstick of validity for the latter. They may pride themselves on pioneering qualitative methods, but still feel compelled to remark that their quantitative approach is “more rigorous” than their qualitative approach, thus implying that not all their results are equally valid. That the social sciences, as though affected by a nagging inferiority complex, continually seek to justify themselves in relation to the natural sciences by trying to prove that they can be just as scientific is a consequence of an outdated or inadequate knowledge of the epistemology of the natural sciences—and, more specifically, of a belief in the supremacy of causality and determinism. Finally, despite methodological progress, demographic research has left partly unsolved crucial problems that will need to be addressed in twenty-first-century demography. For example, research on the meaning of age and age-categories has been generally conducted by anthropologists, historians, and philosophers, but not demographers.

The challenge of openness

Jean Piaget classified demography (construed as applied statistics) as one of the nomothetic sciences. As such, he noted, the demographic field remained “relatively closed yet thriving, the absence of possible experimentation (in the strict sense of dissociation of factors) being compensated by the relative precision of measurements and the success of the various statistical methods relating to variances and the various functional connections accessible to calculation” (1970: 68–69). Although somewhat outdated, since it does not take account of the recent advances in demographic analysis, Piaget’s comment accurately depicts demography’s situation and its boundary problems with
related disciplines. Fortunately, for an increasing number of demographers the gap between demography and the other social sciences has narrowed. Yet the coexistence and ranking of fundamentally different epistemological frameworks remains a fundamental challenge for anyone seeking to open demography and to advance interdisciplinarity. (A proposed avenue forward, through development of a “comprehensive demography,” is set out in Charbit and Petit 2011.)

As early as 1972, Jean Bourgeois-Pichat advocated opening demography to the other social sciences, chiefly because, without a substantial theoretical base of its own, demography had no choice but to show an interest in the theoretical contributions of other disciplines. A number of sub-disciplines have emerged—economic demography, demographic anthropology, genetic demography, historical demography—but they have systematically privileged a particular vision of the relation between measurement and explanation. This was probably the only means of avoiding an even more sterile imprisonment within the confines of measurement.

The relations between demography and other disciplines are largely based on the general representations of science fostered by researchers in their chosen discipline, and especially of their alleged degree of scientificity. Other disciplines now provide demographers with methodological, conceptual, and theoretical resources and sources of inspiration to complement the advances made within demography. It is no coincidence that “most of the transformations affecting the general configuration of disciplines occur at the boundaries or margins of fields of knowledge” (Fabiani 2006: 20), even if increased openness to other disciplines may have the perverse effect of reducing the incentive to develop theories that would be strictly internal to demography.

Renewing theory, embracing interdisciplinary work, and even engaging in deconstruction, as anthropologists have done, are urgent tasks for demography—the more so, at the present time, because the second half of the twentieth century witnessed the emergence of new population problems that demography is incapable of handling alone. A brief backward glance is particularly illuminating in this respect. In the nineteenth century, fertility decline in France provoked significant concern among the French elite, who contrasted it to the dynamism of the British and German populations. The phrase “the population question” was often used, with the implication that satisfactory answers could not be provided by numbers alone or within the demographic dynamic itself. In other words, when searching for economic, social, cultural, and political explanations, they adopted a “population studies” approach avant la lettre.

As our brief account shows, demography subsequently became increasingly sophisticated (and therefore increasingly withdrawn and isolated), a characteristic of INED in the period covered by this essay (roughly early 1950s–late 1980s).
Today, there is evidence of a return to openness in population studies, with a broadening of research questions. New demographic trends—population aging, ultra-low fertility, new forms of morbidity—have gained traction, raising issues extending beyond the narrow discipline into the areas of reproductive health, population and poverty, population and the environment, international migration, and “biopolitics” (to use Foucault’s term). Even if the hybridization between demography and the other social sciences occurs mainly at the margins, it is nonetheless important in introducing new modes of thinking to demography and in fostering exchanges with related disciplines. Once characterized by a stability and solidity that seemingly would long stand the test of time, demography today has expanded to encompass a range of specializations and practices and a welcome degree of interdisciplinarity.

Note

This essay borrows from Véronique Petit, Counting Populations, Understanding Societies (Springer, 2012). The authors gratefully acknowledge comments on an earlier draft.

References


The Twilight of the Census

Dav i d C o l e m a n

The lights are going out in census offices all over Europe. The traditional central source of population information, the demographers’ rod and staff, the main source of data for planning for local authorities and central government, is being replaced by more timely alternative approaches, including population registers. In this regard, Europe is the world’s demographic laboratory: the only continent where so many countries have adopted alternatives to the census (Valente 2011, p.14). It is only recently that all countries in the world have even held a census. At the other end of the scale, that source of data needed so badly by the least developed countries is being replaced in the most developed ones: yet another in the growing list of “demographic transitions” in which the data of demography itself are now the subject matter. It seems natural to regard this as an inevitable result of the coming of the electronic age, of the digitization, rapid transfer, and easy linkage of data, heralding a real revolution in knowledge and things knowable. But if so, why do some of the most advanced countries on Earth, including the richest and most powerful, retain their census, albeit with modifications? What may be lost, as well as gained, by this transition? And will the demographic convenience of modern methods of data-gathering come with a cost to freedom, as some fear?

Background to the census: From state exploitation and control to welfare and planning

These questions are not just of interest to demographers. The census has been entwined with the fundamental concerns of states ever since states arose. Earlier approaches to the counting of the people were seldom motivated by concern for their welfare, although the survival of the people and the survival of the state were ultimately synonymous. Compulsion was, and remains, central. If there were no enumeration without taxation, or if enumeration risked draft into the army or forced labor, the temptation was, and remains, to evade it. Correspondingly severe sanctions have been applied to defaulters,
for example, in respect of those evading the old Chinese household register: “The family will be placed on the list of those liable for military service...any common people who hide from the census will be punished according to law and drafted into the army. If in their search the military come across minor officials who have suppressed the facts, those officials are to be decapitated” (cited in Cox 1970).

In time the census became less concerned with what the people could be obliged to do for the state and more concerned with what the state could do for them. By the late nineteenth century, questions on the condition of the people—housing, health, and disability—indicated the growing role of the census in social reform. Retrospective questions on children ever-born and children surviving, with a scale of social grading devised for the census as in the Census of 1911 in Great Britain, threw a spotlight on social and geographical inequality in infant mortality and (indirectly) on the onset of family limitation. As state spending increased in the developed world to between 30 percent and 60 percent of GDP, census questionnaires have grown in parallel. With the need for knowledge arising from local and national government responsibilities for schools and roads, educational qualifications, occupation, employment, and geographic mobility, the number of questions has inflated to 50 or more. And as populations have become more diverse through immigration, or as ancient minorities have asserted themselves, questions on identity have crept into the census, raising issues about the definition of populations (Kreager 1992).

The traditional census up to the 1970s

Up to the 1970s, all countries used the traditional form of enumeration: compulsory, comprehensive, total, periodical, instantaneous, household-based. Questions typically included details of the relationship of household members to one another, age, sex, marital status, education and employment, immigration and citizenship, previous residence. Fertility history, health status, journey to work, and car ownership may be asked; also racial or ethnic origin; rarely, income. Originally census questionnaires were delivered by an army of (mostly temporary) officials to each household and collected, checking that they had been completed. Now questionnaires are mostly posted out and posted back, or answered online. In the Czech Republic in 2011, 25 percent of forms returned were electronic, in Lithuania the majority, with back-up from enumerators. Post-enumeration surveys and other checks attempt to evaluate accuracy and missing data that could not be established by repeated calls. Missing data, even missing persons, are imputed by increasingly sophisticated statistical techniques (e.g., Abbott 2007). Until recently only the census could produce total population counts for the small areas that concern local authorities.
Reasons for abandoning the conventional census

Why do so many national authorities now look beyond the conventional census? Rapid population turnover, from accelerated internal and international migration, harshly highlights the inevitable delay in processing census results and their rapid obsolescence under modern conditions (*The Economist* 2010). International migration, always the least well recorded element of demographic change, is now its chief driver. Census data represent a statistical snapshot, taken once a decade, fixed in time in a world moving fast. That picture becomes distinctly sepia-toned as reality leaves it behind. And even on first appearance, the results show events as they were only in the recent past, not the present. With the most modern methods the first processed results seldom appear before a year’s delay, and detailed analyses can take two years or more. Results of the (conventional) 2011 British census will not be complete for nearly three years (ONS 2012). The chief customers of the census, the local authorities, will have to wait for their data until March–October 2013. As British politician Harold Macmillan remarked as long ago as 1956 when Chancellor of the Exchequer, “some of our statistics are too late to be as useful as they ought to be. We are always, as it were, looking up a train in last year’s Bradshaw” [timetable]. In promoting the American Community Survey, Leslie Kish, late chief of the US Census Bureau, remarked that the census “was great for its time, but not for today. Having censuses for local data once every 10 years is just not good enough anymore, and it hasn’t been for some time.” It may also be that coordinated repeat surveys organized by the EU such as SHARE, GGS, EVS, and others, despite their modest sample size and response rates, have persuaded politicians and bureaucrats that the census is less necessary.

Social and political developments have made censuses, and all inquiries by the state, less acceptable to some sections of the public. (And an incessant bombardment from unofficial “surveys” intended to extract money for charities and sales drives may have induced some fatigue.) Assurances that census information will be kept from the taxman and other official bodies may be disbelieved as trust declines while demand for greater accountability and democratic rights has increased (Warren 1999: chapters 4, 5, 6; Rose 1999). Some claim that a more geographically and socially mobile, less homogeneous society weakens trust and erodes social capital. In a supposedly “post materialist” world, the “second demographic transition” may threaten the data on which depends the evidence for its existence. In 2012 the annual Edelman Trust Barometer from 25 developed and semi-developed countries showed an unprecedented 9-point decline in the public’s trust of government. Others claim that the basic premise is wrong, especially in Europe (Roger 2012). Libertarian political sentiments have gained ground in, for example, the US and Canada, objecting to almost any state data collection, and questioning its utility.
The state, as well as the people, has also begun to defect. Influential political voices have given added weight to popular mistrust, abolishing Canada’s census long form and threatening the American Community Survey. As Joseph Chamie has observed (personal communication), “Most politicians prefer to remain ignorant. Without data, they can say crazy things and pass stupid policies and can’t be challenged.” In the UK Francis Maude, Cabinet Office minister in the 2010 Coalition government, has denounced the census as obsolete and inaccurate and hopes to save £480 million by cancelling the 2021 census—an unusually long-range policy for a politician (Daily Telegraph, 21 February 2011).

Writing about the 2010 debate on the Canadian census, Roderic Beaujot remarked, “It would appear that some of our political leaders consider that government can be based on values and good ideas, without the need for data and research. This is not unlike the view in some social sciences circles, that we should move to a post-empirical approach that is based on theory and qualitative approaches rather than empirical data” (Beaujot 2012, p. 1).

Even without considering ideological objections, social and economic change is making conventional enumeration more difficult in modern countries. Correct address lists are becoming difficult to construct. Some population groups are hard to contact or to count accurately. Young people are geographically highly mobile, especially students and young professionals. More people living by themselves complicate addresses for an occasional enumeration. An increasing number of people have two homes. Immigrants, especially illegal immigrants and overstayers, may be averse to government contact. Census questions duplicate some information already held in other official sources.

The census can be very expensive, especially if it relies on enumerators to deliver and collect forms and deal directly with respondents. Enumerators typically look after about 100 households each, so a large number of personnel must be recruited and trained. The US conducted a “shoe-leather” census until 1970. Even in 2000, 35 percent of households had to be visited in person by enumerators (the constitutional implications of the US census prohibit imputation of persons). Problems of cost and organization are exacerbated by the need to mobilize a large workforce every decade for a short time, rather than employing a smaller regular staff. And given the decadal interval, knowledge, experience, and expertise tend to be lost in the national statistical office as well. The average cost per head expected in the 2010 round of censuses in member states of the UN’s Economic Commission for Europe was $8.84 (UNECE 2012, Table 1). The US easily tops the list at $48.90 per person for the 2010 census. Generally traditional enumerations are the most expensive. The register-based or mixed censuses of Austria, Belgium, Denmark, Finland, Netherlands, Norway, Slovenia, and Switzerland all cost less than $1.30 per head, Denmark’s being cheapest of all at $0.03 (UNECE 2010, Table 4; expected costs on a purchasing power parity basis). However, costs are not easy to compare. The cost of a traditional enumeration census is one-
off, not an annual cost for maintaining a set of registers. The cost per year falls to nearly zero in the middle of the intercensal period, rising to a peak in the census year in a kind of sine-wave pattern. The cost of a census based on registers will be high when it is set up for the first time, then diminishes over time (ONS 2012, Figure 2).

Replacing the army of enumerators by postal questionnaires saves money but may reduce coverage and accuracy. Many modern censuses only achieve contact with about 90 percent of their target population, and much less than that in areas with “hard-to-reach” populations. For example in the 2001 British census, in the area with the worst “census response rate,” only 62 percent of the estimated population returned a form (ONS 2012). When the results of the census and the intercensal estimates do not tally, as often happens, there is no absolute yardstick to determine which is right, although the census used to be taken to be the gold standard. Intercensal estimates drift off-message as time passes, in recent years as a result of incorrectly estimated migration. In Germany, which until 2011 had taken no census since 1987, the population estimates were thought to be 1.3 million too high by 2010. On the other hand, the “missing million” judged to have been omitted from the UK census total of 1991 by comparison with intercensal estimates and other data (actually 1.2 million, Mitchell et al. 2002) was partly added to the total to bring it better into line with the population estimates. The results of the UK 2001 census overturned that correction, the blame now being placed on defective intercensal emigration estimates. Re-enumeration by the Census Coverage Survey and analysis and imputation of missing data and people were expected to give a compete estimate of 100 percent of the population to be covered—the so-called revision-proof “One Number Census.” Inevitably, later local authority protests showed that claim to have been somewhat hubristic. The 2011 census total for England and Wales was 476,000 more than estimates based on the 2001 census; 200,000 of that was attributed to further error in the 2001 census total (ONS 2012).

Errors of up to 0.5 percent are regarded as acceptable, although discrepancies are more serious in the inner urban areas. Where numbers determine federal or other national central government subventions, lawsuits can drag on for years. The US government distributes almost $450 billion each year on criteria based on census results, and similar weighty sums follow census results in other countries. Naturally cities and local areas contest the results when they show a decline compared with the population estimates or the previous census.

**What are the alternatives?**

As state involvement in planning and welfare has expanded, so have the numbers of questions in censuses. The longer the questionnaire the higher the level of non-response and the greater the trouble and expense. To bal-
ance detail and coverage, some census bureaus have used a two-part inquiry whereby the majority of households received a basic shorter questionnaire, while a sample—usually 10 percent or 20 percent—received a long form with all the questions on which the government wanted answers. This was tried in the UK in 1961 but not repeated. These “long forms” can be very long indeed. In recent years about 17 countries have adopted this strategy, notably the US and Canada, but only Italy (2011), South Korea (2005 only), and Russia (2002 only) in the rest of the developed world. Elsewhere, it has been adopted by Bangladesh, Brazil, China, Ethiopia, Jamaica, Mexico, Nepal, Pakistan (1998), the Philippines (2000), Puerto Rico (2000) Sudan, and Vietnam (UN 2012).

In the US the long form introduced in 1960 was sent to one in six of the population asking details of socioeconomic status and housing conditions. In 2010 the short form really was short—just ten questions. The decennial enumeration has been updated since 2000 by the American Community Survey, covering a shifting sample in monthly rounds amounting to 3 million people per year (Mather et al. 2005). This rolling minicensus was described by former head of the Census Bureau Robert Groves as “one of the most valuable additions to society’s knowledge about itself, because it’s so current” (cited by Carl Bialik, Wall Street Journal, 30 March 2012). However, its compulsory status was revoked in the House of Representatives in May 2012 and the survey then declared unconstitutional, to the consternation of business interests and statisticians (Philips 2012; Rampell 2012). Its fate in the Senate is unclear in a dispute as yet unresolved (US Census Bureau 2002).

In Canada, where the census is held every five years, the short form was introduced in 1971, initially with 1/3 of households receiving the long form, 2/3 the short form. In recent years 20 percent of households received the long form. The short form comprised seven questions, the long form 47 plus eight on housing. By 2001 the long form had grown to 52, including questions on religion and birthplace of parents. By 2006, that had changed to eight questions on the short form and 53 on the long one. In 2011 the short form included the eight plus two more on language, still achieving a 98 percent response rate. But the long form had met its nemesis, abolished for the 2011 census by the Conservative government and replaced in 2010 by a voluntary National Household Survey (NHS) addressed to 4.5 million households. Denying that such a voluntary survey can substitute for mandatory census questions, Canada’s Chief Statistician, Munir Sheikh, resigned in protest amidst a flurry of adverse media, statistical, and commercial comment (Dillon 2010; Beaujot 2012). The response rate to the NHS has fallen to 69 percent. Now even the short form is under Conservative attack for its “compulsory intrusiveness,” and Statistics Canada is reported to be considering register-based alternatives (Globe and Mail, 5 May 2011). Politicians in North America seem to be marching in step away from comprehensive enumeration. In two major examples we see that political pressures on the census have
driven the organizers to confine detailed questions to a fraction of all forms, only for the long forms to be challenged by further opposition to enumeration and abolished, or caponized by being made voluntary.

The rolling census

In France, they do things differently. There, a novel rolling census (La Collecte Tournante) has trundled along since 2005, replacing the conventional census last held in 1999. The rolling census is supplemented by an annual survey supported by an address register in the larger municipalities, the Répertoire d’Immeubles Localisés (Cézard and Lefebvre 2008; Desplanques 2008). A full census is conducted every year in one in five of the 35,750 small municipalities (fewer than 10,000 inhabitants). In the 900 larger municipalities (over 10,000 inhabitants) a sample survey of 8 percent of dwellings is conducted every year. Thus the total population in all small municipalities and 40 percent of the population of larger municipalities is covered over the five-year cycle—about 70 percent of the national population (Valente 2010) giving national information for each successive median year of the five-year cycle. As usual, the change was a response to demands for more timely data, a more even spread of costs over time, a lighter burden on respondents, and better quality control. This method does not reduce overall costs: each year’s operations cost about one-seventh of a conventional enumeration (Pofantis 2008). But the advantages of quality and timeliness are great (Clanché 2011). No other country has adopted this approach although the UK ONS, obliged to find a census-substitute for 2021, is considering that among other options (ONS 2012, p. 10).

These developments in enumeration solve some problems but leave others untouched. Information is more timely but never covers all of the population at the same time. Individuals still have to answer questions. Problems of low response rates persist and no absolute yardstick of population size is created. This method yields no list of persons in the country who are obliged to contribute to the society or who are entitled to enjoy its privileges. Above all, it cannot cope any better than a conventional census with the powerful effects of international migration—the most dynamic of the three components of population change in most modern societies. Census-related methods can measure migration only indirectly. They make no use of the routine accumulation of administrative data, which have near-universal coverage, whereby the nation’s employment, pensions, health, tax, and other encounters with the state are quietly recorded.

Population registers

Only population registers can do that. They offer a radical supplement, and possible replacement, for the traditional enumeration-based censuses. The
utility of registers has been transformed by the huge explosion of information technology capability in the last thirty years, enabling population and administrative data to be entered, retrieved, compared, linked, and updated between corrected registers almost instantly, and population counts produced quickly. In some systems, total population counts can be provided more or less on demand. Of the larger countries of Western Europe, only France and the UK make no direct use of registers (other than address registers) in their enumerations.

The essence of a register is that it is current, being constantly kept up to date by additions and deletions through some form of record linkage. It is not merely a static list frozen at one point in time. A few countries have maintained a form of population register for centuries at the local parish or municipal level for the care of souls or for local civil administrative purposes, notably in the Lutheran Scandinavian countries (including Estonia). These were not initially in a form suitable for aggregation in a national population count and contained only the simplest information. As methods developed, as early as 1872 the International Statistics Congress recommended the general adoption of population registers (Brown 1872, pp. 447, 449).

By the beginning of the twentieth century population registers were operating in a number of modern and modernizing countries: Belgium, Chile, China, Czechoslovakia, Finland, Germany, Hungary, Italy, Japan, Korea, Liechtenstein, Luxemburg, the Netherlands, Spain, Sweden, and Switzerland. By the end of 1967 they were operating in at least 65 countries (UN 1969, table 1). Their modern (but not of course initially digitized) form took shape in the mid-twentieth century: in 1964 in Norway, 1967 in Sweden, and in some countries developed over time from simpler to more complex forms: from 1924 to 1968 in Denmark, from 1969 to 1971 in Finland, and from 1941 to 1972 in France. It was some time before these registers began to supplement, never mind replace, conventional censuses based on enumeration. Initially they were neither intended to serve nor capable of serving other than municipal-level functions, and the information they carried was elementary.

Almost all modern countries maintain other official registers of various kinds detailing the entitlements or obligations of their residents with respect to health and education services, taxes, license fees, and criminal records. These are not population registers; they are maintained for administration, not for enumeration. These “administrative sources” have come to embrace the majority, in some cases all, of the population as the supply and demand for services, welfare, and regulation involving the state has grown. Thus in the UK 95 percent of households are registered with the UK TV licensing authority; 72 percent of persons of eligible age are registered with the Drivers and Vehicles Licensing Authority. All in theory are now registered from birth with the National Health Service, and all over age 16 have a National Insurance Number. Perhaps the most interesting registers are those held by financial service companies. In the UK in 2010 there were 84.6 million debit cards in circulation, and 55.6
million credit cards in a total population of 61 million. In 2010, 90 percent of the adult population had a debit card, and 64 percent had a credit card. These facilitated 6.5 billion transactions in 2010 worth £410 billion (UK Cards Association 2011). Card companies hold substantial amounts of data about their subscribers, much more than many official registers.

Replacing the census

The Scandinavian countries of Lutheran heritage have obviously had a head start in these matters. From the eighteenth century, the Lutheran clergy were responsible for maintaining lists of households in their parishes as well as registering baptisms, marriages, and burials. Their societies are accustomed to registration systems and for the most part find their development uncontentious. All four Nordic countries now have “censuses” that are wholly register based, starting with Denmark in 1981, followed by Finland in 1990 and Norway and Sweden in 2011 (Statistics Sweden 2006; Wult 2008; UNECE 2007). The Netherlands held its last conventional census in 1971. Since then it has conducted a register-based “virtual census” supplemented by a survey (Prins 2000). Austria and Slovenia introduced register-based censuses in 2010 and 2011 respectively (Statistics Austria 2008, 2011), and 11 other European countries have adopted a mixed or hybrid system comprising register-based information supplemented by a “light” enumeration (Czech Republic, Estonia, Italy, Latvia, Lithuania, Switzerland), register data supplemented by a survey (Belgium, Iceland, Netherlands), or a combination of all three (Germany and Poland) (United Nations 2010, 2012). The former West Germany has had no census since 1987, the former East Germany not since 1981. The German 2011 combined census is evaluating the data from administrative registers (local population registers and the Federal Employment Agency register), from detailed questionnaires sent to 17.5 million owners of buildings and dwellings, and the results of a sample survey of about 10 percent of all residents (Augustyniak 2011). This unique arrangement was made necessary by the absence of any centralized registers of population or of dwellings in Germany, reflecting sensitivities to past politics. Registration in the communal (local) population register is compulsory, however. That includes, among others, the names, sex, date and place of birth, marital status, citizenship, place of residence, and address of each person. The 13,500 local authorities use standardized systems. Most respondents were not involved directly in this operation, an important consideration in Germany and in register-based operations generally.

In the 2010 round, 21 out of 40 countries in the UNECE still used the traditional form of census, mostly countries in Southern and Eastern Europe (most of the area east of Hajnal’s line; see UNECE 2010, Tables 2a, 2b; Valente 2010, map 1), plus Canada and the United States. Five countries adopted a register-based approach entirely; the remaining 11 adopted a variety of ap-
proaches. Between the 2000 and 2010 rounds, nine countries changed their form of enumeration: six to a combined method (Czech Republic, Estonia, Germany, Italy, Lithuania, Poland), two to a register-based census (Austria and Sweden 2011), and one to a rolling census (France).

**Drawbacks of register-based enumeration**

But there are major imperfections in the data-gathering capacity of register-based censuses, and their development also raises broader political, social, and technical issues. Register data are peculiar in that data supply and demand operate the wrong way round. Normally, in censuses and surveys, the organizers determine what information is needed and devise questions to elicit those data. With a register-based census, the nature of the administrative sources dictates what can be provided, and it may well have been devised for purposes that differ from the usual requirements of census takers. They are set up for administrative, not statistical purposes. The quality of registers may be uneven and may initially at least be the responsibility of different organs of local or central government. Data deemed of lesser importance for administrative purposes will receive less attention—for example on heating and water supply in the Norwegian housing register. Data on events that happened before the establishment of the register or on events occurring abroad will be absent. In most countries cohabitation is not registered except incidentally in relation to tax or custody of children.

Data that should be identical in different registers may not be, and, for linkage, individuals need to have the same person-number in the same format. Making a set of registers fit for the purpose of substituting for census can take many years. For example, the Central Population Register in Finland was set up in 1969 and the eight other registers (business, dwellings, housing, education, employment, family, household, and income) came on stream at various later times, the latest being employment in 1987. While they began to contribute to the census a few years after their creation, they were not all fit to contribute together to the first wholly register-based census until 1990. All this involves high initial set-up costs although, once up and running, costs will be low and real savings can be made.

Some phenomena of interest are typically omitted by registers. For example, German administrative data do not contain information on self-employment, education, or migration background (that is, birthplace or citizenship of the respondent’s parents). There are no complete registers of buildings and dwellings, or of people living in residential homes or other communal institutions. Separate surveys must supply that information.

The Austrian registers lack data on occupations. That is unusual. But like most registers they also lack information about the journey to work or to school or the means of transport. “Activities” as opposed to “attributes” tend not to be recorded. Information on persons available for work and on
jobseekers is also generally lacking, and on unemployed persons who are not registered as unemployed (e.g., Statistics Norway 2008; Statistics Netherlands 2012). “Households,” defined as persons living in the same dwelling with joint board, cannot easily be reconstructed using register information (Statistics Norway 2008, p. 6).

The questions in a conventional census are altered from time to time to meet emerging information needs in the society, for example on use of information technology, and to illuminate the new salience of disability and diversity. The census can respond flexibly to new needs simply by inserting new questions. It is more difficult to change registers in that way. In practice, surveys are used in some countries to supplement data missing from register-based censuses, although only on a sample basis. In the Netherlands “virtual census” of 2001 and 2011, register data were integrated with information from the regular Labour Force Survey to provide data, absent or incomplete in the registers themselves, on economic characteristics (occupation, unemployment, educational attainment).

Registers do not wholly solve the problems of keeping tabs on migration unless arrivals and departures are linked to the population register at the border. Out-migrants or emigrants may forget to cancel their registration. When people leave permanently without canceling their registration, in theory they would “live” forever. Illegal immigrants and overstayers may evade the registers. But in the countries discussed here it is almost impossible to live a normal life without registration, since public services are thereby inaccessible.

For a variety of reasons, then, a number of European countries that have abandoned the traditional census have not yet gone the whole distance to an entirely register-based survey (Statistics Netherlands 2012). In all, 13 countries adopted a mixed approach in the 2010/11 census round, combining register data either with a “light” enumeration or with survey data, whether routine (from the Labour Force Survey) or specially constructed for the event. Six of them made this switch for the first time (Czech Republic, Estonia, Germany, Italy, Lithuania, and Poland), as did two countries outside Europe (Israel and Turkey). Six others had used this method in 2000 as well (Belgium, Latvia, Netherlands, Slovenia, Spain, Switzerland; UN 2012).

Broader implications

So much for the detail. What are the broader implications of the movement away from comprehensive decennial censuses, with their 40 or 50 compulsory questions, toward a wholly or partly register-based system where an individual’s profile is on permanent and in some aspects unalterable official record? Is anything lost in this exchange? The statistician William Kruskal (1984) felt that the US decennial census, with all its publicity and its (theoretical) involvement of all residents, was a major national “ceremony” of importance, the only time
when all residents were involved simultaneously with a major official activity, asserting their presence as US residents for the record. In the US, this is given added significance by the Constitutional status of the census. In Canada, too, statisticians have risen up to defend the census from parliamentary attack, noting that “participation in statistical enquiries is an essential part of citizenship in a modern democracy” (Dillon 2010) and, echoing Kruskal, that the census has a “symbolic value, not unlike history, literature, maps and museums in defining ourselves as a country, ... evidenced in the media interest in census results” (Sheamur 2011, all cited in Beaujot 2012). In Europe, where opposition to census taking and other official inquiries has been growing, such affectionate attachment to the traditional census is difficult to imagine although it certainly has its defenders (Boyle and Dorling 2004).

Although many find census questions irksome, few now fear that a response will involve them in cost or obligation. In developed countries census results are ring-fenced from other agencies of government, especially taxation authorities. Individual questionnaires are usually kept secret for 100 years, and small-area breakdowns of published data are randomly anonymized to prevent individuals being identified. Generally speaking, the public in most countries has had faith in the privacy of the census; otherwise non-response would be much higher than it is.

Registration has different implications. Here the record is permanent. It follows individuals through time and space, logging their major life events. It is difficult to evade. In its most developed form the person-number of the population register links to tax and medical records. Big Brother may be watching you, some fear, and rather efficiently thanks to modern IT. Administrative sources provide plenty of opportunities for silent data-gathering, with little involvement or awareness by citizens, potentially more complete and comprehensive than any census. Exceptional measures are needed, in democratic countries, to ensure that information does not leak from those official agencies permitted by law to access the data (Watner and McElroy 2003). For example, the local registers of the Netherlands are connected with cut-outs, with an electronic one-way valve such that data may be sent from them to Statistics Netherlands, but the latter cannot interrogate the local registers (Prins 2000). Elaborate ring-fencing arrangements between registers provided by government agencies also exist in the Austrian and other systems. No serious problems have been reported in European countries using registers, although these are all established democracies with well-developed self-consciousness concerning human rights. But it was inauspicious that in 2011 six people were arrested for stealing and distributing the entire Israeli population registry data on 9 million persons.

Politics and culture matter greatly in determining public attitudes toward censuses and registers. Countries where registration has been a fact of social life for centuries, as in Scandinavia, and where this system has so far been
demonstrably harmless, appear to have had few problems in coping with the transformation of their traditional registers into machine-readable form and eventually integrating them to create a census substitute. But past oppression has cast a long shadow. The extension of German census questions to religious and racial origin had disastrous results for Jews and other population groups beginning in 1933 (Aly et al. 2004). But the listing of Jews and other victims was performed initially through the census data of 1933 and 1939, not through any existing register. These served to create lists maintained by the Nazis. It is the political culture that determines the problem, not the method of gathering data nor even the nature of the data collected. Furthermore, it does not require a census or register to identify people who are distinguished by appearance, surname, or area of residence. The events up to 1945 have had a lasting effect on German attitudes toward data collection, and on those of Jews everywhere except in Israel. Paradoxically, Israel has had a population register since 1948, computerized in 1965, using a person-number system in which religion, ethnic group, and nationality are recorded (National Center for Health Statistics 1980; Zadka 2008). Israel also controls, and has “frozen,” the Palestinian population register, which determines who may lawfully reside in the West Bank (Human Rights Watch 2012).

Lights in census offices continue to burn bright east of Hajnal’s line. There only the most western and westernized countries (Estonia, Slovenia) have adopted registers for the census. The English-speaking countries, Australia, Canada, Ireland, New Zealand, the United Kingdom, and the United States, all still depend on traditional censuses, with some experimentation in North America. Until recently none but the UK had considered using administrative data for census purposes, but political opposition to traditional information-gathering is pointing the way to registers in Canada too. In the UK an ambitious program to create an identity card backed by a limited register, proposed by the last Labour government to improve security and the control of immigration, met strong resistance (Watner and McElroy 2003; Enterprise Privacy Group 2005). It was scrapped by the incoming Conservative–Liberal Democrat coalition government ostensibly on grounds of costs and intrusiveness, although some denounced its removal as mere political opportunism, exploiting a difficulty for the governing party. There had been precedents: identity cards and a simple national register had been created in wartime in 1915 and 1939, the latter surviving seven years into peacetime only to be abandoned in 1952, again by an incoming Conservative government. The latter register formed the basis for food rationing and, later, the NHS number.

In the Anglosphere, elements of the libertarian and traditional Right and the liberal Left find common cause in suspicion of the census, regarding suggestions of a population register as a further intrusion by big government. Census questions have been probing deeper into identity—first ethnic origin, then religion. In the UK questions on income and on sexual orientation were
considered for 2011 but rejected. In reaction to census “mission creep,” some journalists on the Right have suggested that the UK census be “pared back to its original purpose: a simple population count” (Johnson 2011). Francis Maude, Minister responsible for the census, announced a decision to scrap the census in favor of new ways of counting the population that would be “less intrusive.” There has been no tradition of keeping population registers in the English-speaking world, which shares a Common Law tradition, not the codification of statute typical of Continental Europe. Official identity documents are regarded as an undesirable infringement on freedom, something best confined to the less happy lands of Continental Europe. Some English people take perverse pleasure in having no such documents, even though ordinary life in Britain, especially financial transactions, cannot get far without a bank card, a driving license, a national insurance number, or a diary full of passwords. Similar attitudes seem to have been inherited by the countries of the Anglosphere with their traditions of even more robust independence and individualism.

If the case for population registers were to make progress in the English-speaking countries, major questions would arise on the future of information on ethnic and religious identification and identity. The UK and all the English-speaking countries abroad have developed comprehensive sets of self-ascribed questions on ethnic identity, ancestry, religion, and language. These questions have proliferated with successive censuses, in line with policy responses to the new diversity arising from large-scale immigration. In the UK they have become pervasive for the enforcement of equality legislation and its quotas and targets—routinely posed for application to school and university, for recruitment, promotion and contracts in the public sector, and for services provided by the National Health Service and by local authorities down to planning permission for kitchen extensions. In Eastern Europe, where nation-states have only recently emerged from ancient empires, these questions apply mostly to long-standing indigenous minorities. In between, in Western Europe such questions (except, in some countries, religion) are not asked; multicultural policy has not been developed. Instead, registers allow the enumeration of “foreign background” populations derived from the birthplace / citizenship of individuals and their parents, as in Denmark, Netherlands, Norway, Sweden, and some others. In France, however, the official collection of such data is held to be incompatible with the equality of citizenship, and surveys suggest that public opinion is very uneasy about using “ethnoracial” classifications in any inquiry (Simon and Clément 2006). Most Southern European countries collect information only on birthplace and on nationality.

Any move in the English-speaking countries toward registers for enumeration would fall foul of these policies. Register data suit fixed attributes of individuals (date of birth, sex, educational attainment) or characteristics that alter relatively slowly and in an objectively verifiable manner (occupa-
tion, address), both of which are routinely recorded administratively. They are unsuited to self-ascribed characteristics divided into many different categories, or specified exclusively and at will by the respondent independently of any official list. Would the 420,000 who returned “Jedi” as their religion in the UK census in 2001 choose to enter that into a register? Ethnoracial or religious identity on a register would endow that person with a (potentially) permanent ethnic or racial label. The consequent racialization of society, making ethnic divisions in society permanent through statistics and reinforcing self-conscious separate identity, would be unlikely to promote common citizenship and social cohesion.

It is difficult to envisage such a development in a democratic society. The prospect might provoke a reassessment of ethnic statistics. Basic aggregate information could continue to be gleaned from survey questions. Registers without personal ethnic or religious labels would be a first step away from what many regard as a disturbing trend of social fragmentation, where statistical categories structure society rather than reflect it. Similar concerns have been raised elsewhere, for example regarding the proliferation of ethnic and color questions in the census of Canada (Beaujot 2012).

Where do we go from here? In Continental Europe at least, the trend toward the use of registers to supplement or replace the census is clear and probably unstoppable. Perhaps few countries will rely on registers alone. The favored approach in 2012 is to supplement register data with a survey to provide data that registers cannot reach. Whether that is just a half-way house to wholly register-based censuses remains to be seen. We may have passed the high-water mark of information-gathering about populations, and the price of further progress down the road to the register census will be fewer demands for personal information, just as the means to process it and link it grow.

As the world develops and IT capacity is globalized, countries outside Europe in the developing world are creating electronic population registers (e.g., Brazil, Malaysia, Singapore, Thailand, Turkey). Europe and the West may well be overtaken in the modernization of population data as they have been in economic growth. Not all these projects have respectable antecedents. The modern Chinese household registration system (hukou) (Chan and Buckingham 2008) is the lineal descendant of the world’s oldest. That existed as an inventory of population for state purposes and to control internal migration. It still does the latter, attempting to separate the rural and urban social and economic spaces, cutting off rural migrants to cities from urban entitlements. The United Nations noted (2005, p. 3) “with deep concern the *de facto* discrimination against internal migrants in the fields of employment, social security, health service, housing, and education that indirectly result[s], inter alia, from the restrictive national household registration system (hukou) which continues to be in place despite official announcements regarding reforms.”

The world’s biggest registration project, in democratic India, serves different ends. This astonishingly ambitious project aims to register all residents
of India by 2021, including illiterates, and provide them with an identity card with biometric data (Office of the Registrar-General India 2009). One of its chief advantages echoes the ancient benefits of the parish registers: an official record of name, existence, and parentage, therefore of entitlement (to benefits, to property, to vote) and of inheritance. Millions of impoverished Indians, lacking documents, are denied the ability to resist exploitation and exercise claims on the state. That is an instructive contrast of the divergent purposes to which registers and enumeration may be put in different kinds of political regime: some benevolent, as in this case; others malign, as in Germany in the 1930s.

Note

The author is grateful to the following colleagues for advice in the preparation of this essay: Roderic Beajot, Joseph Chamie, Richard Gisser, John Haskey, and Anna Rotkirch.

References


SELECTED PUBLICATIONS
OF PAUL DEMENY

Paul Demeny was born in Nyiregyhaza, Hungary in 1932. He graduated from the Reformed College of Debrecen in 1951 and from the University of Budapest in 1955. He attended the Institut universitaire de hautes études internationales (HEI), Geneva, in 1957 and received a PhD in economics from Princeton University in 1961—where his mentors were Frank Notestein and Ansley Coale in population studies, and William Baumol, Oscar Morgenstern, and Jacob Viner in economics and economic history. He subsequently held appointments as assistant professor of economics at Princeton University and as associate at Princeton’s Office of Population Research. He served on the faculty of the University of Michigan (associate professor and professor of economics and associate director of the Population Studies Center), and was a visiting professor in the Demography Department, University of California, Berkeley. In 1971 he was appointed professor of economics at the University of Hawaii and founding director of the East-West Population Institute at the East-West Center, Honolulu. He joined the Population Council in 1973 as vice president and director of its Demographic Division (later the Center for Policy Studies), remaining at the Council—since 1989 as Distinguished Scholar—until his retirement in 2012. The first issue of Population and Development Review, the journal he founded and has edited over nearly four decades, appeared in September 1975. He served as president of the Population Association of America in 1986. He is a Fellow of the American Association for the Advancement of Science, an external member of the Hungarian Academy of Sciences, and in 2003 was named Laureate of the International Union for the Scientific Study of Population.


“Demographic applications of computer models.” In Computer Methods in the Analysis of Large-Scale Social Systems, Cambridge, MA: Joint Center for Urban Studies, Massachusetts Institute of Technology and Harvard University, 1965, pp. 68–70.


SAMUEL H. PRESTON is Professor of Demography and Sociology, Population Studies Center, University of Pennsylvania.

LANT PRITCHETT is Professor of the Practice of International Development, Harvard University, Kennedy School of Government, and Senior Fellow, Center for Global Development, Washington, DC.

VACLAV SMIL is Distinguished Professor Emeritus, University of Manitoba, Winnipeg.

ANDREW STOKES is a doctoral student in Demography and Sociology, Population Studies Center, University of Pennsylvania.

ANN SWIDLER is Professor of Sociology, University of California, Berkeley.

JACQUES VALLIN is Directeur de recherche émérite, Institut national d’études démographiques (INED), Paris.

MARTINA VIARENGO is Assistant Professor, Department of Economics, Graduate Institute of International and Development Studies of Geneva.

PAUL E. WAGGONER is Director Emeritus, The Connecticut Agricultural Experiment Station, New Haven.

WANG FENG is Senior Fellow and Director, Brookings-Tsinghua Center for Public Policy and Professor of Sociology, University of California, Irvine.

SUSAN COTTS WATKINS is Professor of Sociology, University of Pennsylvania and Visiting Research Scientist, California Center for Population Research, University of California, Los Angeles.

IDDO K. WERNICK is Senior Research Associate, Program for the Human Environment, The Rockefeller University, New York.

CHARLES F. WESTOFF is Professor of Sociology and Demographic Studies, Emeritus, and Senior Research Demographer, Office of Population Research, Princeton University.

DOUGLAS WOLF is Gerald B. Cramer Professor of Aging Studies and Senior Associate, Center for Policy Research, Syracuse University.