"There is no author whose books I look forward to more than Vaclav Smil."

—BILL GATES

GROWTH

FROM MICROORGANISMS TO MEGACITIES

VACLAV SMIL

Growth

From Microorganisms to Megacities

Vaclav Smil

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Smil, Vaclav, author.
Title: Growth / Vaclav Smil.
Description: Cambridge, MA : The MIT Press, 2019. | Includes bibliographical references and index.
Identifiers: LCCN 2018059356 | ISBN 9780262042833 (hardcover : alk. paper)
Subjects: LCSH: Civilization, Modern—21st century. | Technology and civilization. | Growth. | Human ecology. | Population. | Energy development. | Economic development. | Cities and towns—Growth. | Urban ecology (Sociology).
Classification: LCC CB428 .S625 2019 | DDC 909.82—dc23
LC record available at https://lccn.loc.gov/2018059356

10 9 8 7 6 5 4 3 2 1

Contents

Preface

Growth is an omnipresent protean reality of our lives: a marker of evolution, of an increase in size and capabilities of our bodies as we reach adulthood, of gains in our collective capacities to exploit the Earth's resources and to organize our societies in order to secure a higher quality of life. Growth has been both an unspoken and an explicit aim of individual and collective striving throughout the evolution of our species and its still short recorded history. Its progress governs the lives of microorganisms as well as of galaxies. Growth determines the extent of oceanic crust and utility of all artifacts designed to improve our lives as well as the degree of damage any abnormally developing cells can do inside our bodies. And growth shapes the capabilities of our extraordinarily large brains as well as the fortunes of our economies. Because of its ubiquity, growth can be studied on levels ranging from subcellular and cellular (to reveal its metabolic and regulatory requirements and processes) to tracing long-term trajectories of complex systems, be they geotectonic upheavals, national or global populations, cities, economies or empires.

Terraforming growth—geotectonic forces that create the oceanic and continental crust, volcanoes, and mountain ranges, and that shape watersheds, plains, and coasts—proceeds very slowly. Its prime mover, the formation of new oceanic crust at mid-ocean ridges, advances mostly at rates of less than 55 mm/year, while exceptionally fast new sea-floor creation can reach about 20 cm/year (Schwartz et al. 2005). As for the annual increments of continental crust, Reymer and Schubert (1984) calculated the addition rate of 1.65 km^3 and with the total subduction rate (as the old crust is recycled into the mantle) of 0.59 km^3 that yields a net growth rate of 1.06 km^3.

That is a minuscule annual increment when considering that the continents cover nearly 150 Gm^2 and that the continental crust is mostly 35–40 km thick, but such growth has continued during the entire Phanerozoic eon, that is for the past 542 million years. And one more, this time

Figure 0.1
Slow but persistent geotectonic growth. The Himalayas were created by the collision of Indian and Eurasian plates that began more than 50 million year ago and whose continuation now makes the mountain chain grow by as much as 1 cm/year. Photo from the International Space Station (looking south from above the Tibetan Plateau) taken in January 2004. Image available at https://www.nasa.gov/multimedia/imagegallery/image_feature_152.html.

vertical, example of inevitably slow tectonic speeds: the uplift of the Himalayas, the planet's most imposing mountain range, amounts to about 10 mm/year (Burchfiel and Wang 2008; figure 0.1). Tectonic growth fundamentally constrains the Earth's climate (as it affects global atmospheric circulation and the distribution of pressure cells) and ecosystemic productivity (as it affects temperature and precipitation) and hence also human habitation and economic activity. But there is nothing we can do about its timing, location, and pace, nor can we harness it directly for our benefit and hence it will not get more attention in this book.

Organismic growth, the quintessential expression of life, encompasses all processes by which elements and compounds are transformed over time into new living mass (biomass). Human evolution has been existentially dependent on this natural growth, first just for foraged and hunted food, later for fuel and raw materials, and eventually for cultivated food and feed plants and for large-scale exploitation of forest phytomass as well as for the capture of marine species. This growing human interference in the biosphere has brought a large-scale transformation of many ecosystems, above all the conversion of forests and wetlands to croplands and extensive use of grassland for grazing animals (Smil 2013a).

Growth is also a sign of progress and an embodiment of hope in human affairs. Growth of technical capabilities has harnessed new energy sources, raised the level and reliability of food supply, and created new materials and new industries. Economic growth has brought tangible material gains with the accumulation of private possessions that enrich our brief lives, and it creates intangible values of accomplishment and satisfaction. But growth also brings anxieties, concerns, and fears. People—be it children marking their increasing height on a door frame, countless chief economists preparing dubious forecasts of output and trade performance, or radiologists looking at magnetic resonance images—worry about it in myriads of different ways.

Growth is commonly seen as too slow or as too excessive; it raises concerns about the limits of adaptation, and fears about personal consequences and major social dislocations. In response, people strive to manage the growth they can control by altering its pace (to accelerate it, moderate it, or end it) and dream about, and strive, to extend these controls to additional realms. These attempts often fail even as they succeed (and seemingly permanent mastery may turn out to be only a temporary success) but they never end: we can see them pursued at both extreme ends of the size spectrum as scientist try to create new forms of life by expanding the genetic code and including synthetic DNA in new organisms (Malyshev et al. 2014)—as well as proposing to control the Earth's climate through geoengineering interventions (Keith 2013).

Organismic growth is a product of long evolutionary process and modern science has come to understand its preconditions, pathways, and outcomes and to identify its trajectories that conform, more or less closely, to specific functions, overwhelmingly to S-shaped (sigmoid) curves. Finding common traits and making useful generalizations regarding natural growth is challenging but quantifying it is relatively straightforward. So is measuring the growth of many man-made artifacts (tools, machines, productive systems) by tracing their increase in capacity, performance, efficiency, or complexity. In all of these cases, we deal with basic physical units (length, mass, time, electric current, temperature, amount of substance, luminous intensity) and their numerous derivatives, ranging from volume and speed to energy and power.

Measuring the growth phenomena involving human judgment, expectations, and peaceful or violent interactions with others is much more challenging. Some complex aggregate processes are impossible to measure without first arbitrarily delimiting the scope of an inquiry and without resorting to more or less questionable concepts: measuring the growth of economies by relying on such variables as gross domestic product or national income are perfect examples of these difficulties and indeterminacies. But even when many attributes of what might be called social growth are readily measurable (examples range from the average living space per family and possession of household appliances to destructive power of stockpiled missiles and the total area controlled by an imperial power), their true trajectories are still open to diverse interpretations as these quantifications hide significant qualitative differences.

Accumulation of material possessions is a particularly fascinating aspect of growth as it stems from a combination of a laudable quest to improve quality of life, an understandable but less rational response to position oneself in a broader social milieu, and a rather atavistic impulse to possess, even to hoard. There are those few who remain indifferent to growth and need, India's loinclothed or entirely naked *sadhus* and monks belonging to sects that espouse austere simplicity. At the other extreme, we have compulsive collectors (however refined their tastes may be) and mentally sick hoarders who turn their abodes into garbage dumps. But in between, in any population with rising standards of living, we have less dramatic quotidian addictions as most people want to see more growth, be in material terms or in intangibles that go under those elusive labels of satisfaction with life or personal happiness achieved through amassing fortunes or having extraordinarily unique experiences.

The speeds and scales of these pursuits make it clear how modern is this pervasive experience and how justified is this growing concern about growth. A doubling of average sizes has become a common experience during a single lifetime: the mean area of US houses has grown 2.5-fold since 1950 (USBC 1975; USCB 2013), the volume of the United Kingdom's wine glasses has doubled since 1970 (Zupan et al. 2017), typical mass of European cars had more than doubled since the post–World War II models (Citroen 2 CV, Fiat Topolino) weighing less than 600 kg to about 1,200kg by 2002 (Smil 2014b). Many artifacts and achievements have seen far larger increases during the same time: the modal area of television screens grew about 15-fold, from the post–World War II standard of 30 cm diagonal to the average US size of about 120 cm by 2015, with an increasing share of sales taken by TVs with diagonals in excess of 150 cm. And even that impressive increase has been dwarfed by the rise of the largest individual fortunes: in 2017 the world had 2,043 billionaires (Forbes 2017). Relative

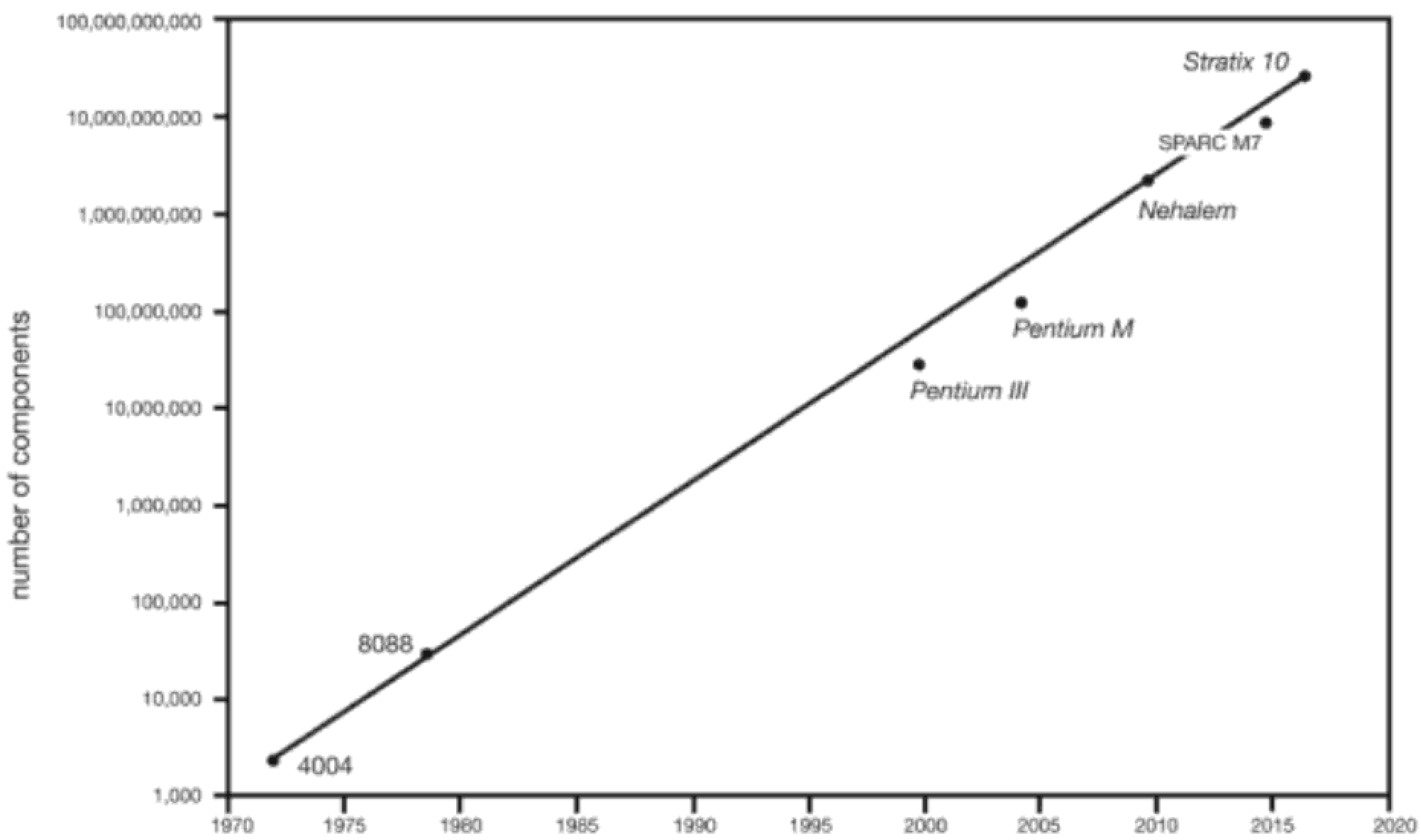


Figure 0.2
A quintessential marker of modern growth: Moore's law, 1971–2018. Semi-logarithmic graph shows steady exponential increase from 10^3 to 10^{10} components per microchip (Smil 2017a; IBM 2018b).

2018; Graphcore 2018). As in all cases of exponential growth (see chapter 1), when these gains are plotted on a linear graph they produce a steeply ascending curve, while a plot on a semilogarithmic graph transforms them into a straight line (figure 0.2).

This progress has led to almost unbounded expectations of still greater advances to come, and the recent rapid diffusion of assorted electronic devices (and applications they use) has particularly mesmerized those uncritical commentators who see omnipresent signs of accelerated growth. To give just one memorable recent example, a report prepared by Oxford Martin School and published by Citi claims the following time spans were needed to reach 50 million users: telephone 75 years, radio 38 years, TV 13 years, Internet four years, and Angry Birds 35 days (Frey and Osborne 2015). These claims are attributed to Citi Digital Strategy Team—but the team failed to do its homework and ignored common sense.

Are these numbers referring to global or American diffusions? The report does not say, but the total of 50 million clearly refers to the United States where that number of telephones was reached in 1953 (1878 + 75 years): but the number of telephones does not equal the total number of their users, which, given the average size of families and the ubiquity of phones in places of work, had to be considerably higher. TV broadcasting did not have just one but a number of beginnings: American transmission, and sales of first sets, began in 1928, but 13 years later, in 1941, TV ownership was still minimal, and the total number of TV sets (again: devices, not users) reached 50 million only in 1963. The same error is repeated with the Internet, to which millions of users had access for many years at universities, schools, and workplaces before they got a home connection; besides, what was the Internet's "first" year?

All that is just sloppy data gathering, and an uninformed rush to make an impression, but more important is an indefensible categorical error made by comparing a complex system based on a new and extensive infrastructure with an entertaining software. Telephony of the late 19th century was a pioneering system of direct personal communication whose realization required the first large-scale electrification of society (from fuel extraction to thermal generation to transmission, with large parts of rural America having no good connections even during the 1920s), installation of extensive wired infrastructure, and sales of (initially separate) receivers and speakers.

In contrast, Angry Birds or any other inane app can spread in a viral fashion because we have spent more than a century putting in place the successive components of a physical system that has made such a diffusion possible: its growth began during the 1880s with electricity generation and transmission and it has culminated with the post-2000 wave of designing and manufacturing billions of mobile phones and installing dense networks of cell towers. Concurrently the increasing reliability of its operation makes rapid diffusion feats unremarkable. Any number of analogies can be offered to illustrate that comparative fallacy. For example, instead of telephones think of the diffusion of microwave ovens and instead of an app think of mass-produced microwavable popcorn: obviously, diffusion rates of the most popular brand of the latter will be faster than were the adoption rates of the former. In fact, in the US it took about three decades for countertop microwave ovens, introduced in 1967, to reach 90% of all households.

The growth of information has proved equally mesmerizing. There is nothing new about its ascent. The invention of movable type (in 1450) began an exponential rise in book publishing, from about 200,000 volumes during the 16th century to about 1 million volumes during the 18th century, while recent global annual rate (led by China, the US, and the United Kingdom) has surpassed 2 million titles (UNESCO 2018). Add to this pictorial information whose growth was affordably enabled first by lithography, then by rotogravure, and now is dominated by electronic displays on mobile devices. Sound recordings began with Edison's fragile phonograph in 1878 (Smil 2018a; figure 0.3) and their enormous selection is now

Figure 0.3
Thomas A. Edison with his phonograph photographed by Mathew Brady in April 1878. Photograph from Brady-Handy Collection of the Library of Congress.

effortlessly accessible to billions of mobile phone users. And information flow in all these categories is surpassed by imagery incessantly gathered by entire fleets of spy, meteorological, and Earth observation satellites. Not surprisingly, aggregate growth of information has resembled the hyperbolic expansion trajectory of pre-1960 global population growth.

Recently it has been possible to claim that 90% or more of all the extant information in the world has been generated over the preceding two years. Seagate (2017) put total information created worldwide at 0.1 zettabytes (ZB, 10^{21}) in 2005, at 2 ZB in 2010, 16.1 ZB in 2016, and it expected that the annual increment will reach 163 ZB by 2025. A year later it raised its estimate of the global datasphere to 175 ZB by 2025—and expected that the total will keep on accelerating (Reinsel et al. 2018). But as soon as one considers the major components of this new data flood, those accelerating claims are hardly impressive. Highly centralized new data inflows include the incessant movement of electronic cash and investments among major banks and investment houses, as well as sweeping monitoring of telephone and internet communications by government agencies.

At the same time, billions of mobile phone users participating in social media voluntarily surrender their privacy so data miners can, without asking anybody a single question, follow their messages and their web-clicking, analyzing the individual personal preferences and foibles they reveal, comparing them to those of their peers, and packaging them to be bought by advertisers in order to sell more unneeded junk—and to keep economic growth intact. And, of course, streams of data are produced incessantly simply by people carrying GPS-enabled mobile phones. Add to this the flood of inane images, including myriads of selfies and cat videos (even stills consume bytes rapidly: smartphone photos take up commonly 2–3 MB, that is 2–3 times more than the typescript of this book)—and the unprecedented growth of "information" appears more pitiable than admirable.

And this is one of the most consequential undesirable consequences of this information flood: time spent per adult user per day with digital media doubled between 2008 and 2015 to 5.5 hours (eMarketer 2017), creating new life forms of screen zombies. But the rapid diffusion of electronics and software are trivial matters compared to the expected ultimate achievements of accelerated growth—and nobody has expressed them more expansively than Ray Kurzweil, since 2012 the director of engineering at Google and long before that the inventor of such electronic devices as the charged-couple flat-bed scanner, the first commercial text-to-speech synthesizer, and the first omnifont optical character recognition.

In 2001 he formulated his law of accelerating returns (Kurzweil 2001, 1):

> An analysis of the history of technology shows that technological change is exponential, contrary to the common-sense "intuitive linear" view. So we won't experience 100 years of progress in the 21st century—it will be more like 20,000 years of progress (at today's rate). The "returns," such as chip speed and cost-effectiveness, also increase exponentially. There's even exponential growth in the rate of exponential growth. Within a few decades, machine intelligence will surpass human intelligence, leading to The Singularity—technological change so rapid and profound it represents a rupture in the fabric of human history. The implications include the merger of biological and nonbiological intelligence, immortal software-based humans, and ultra-high levels of intelligence that expand outward in the universe at the speed of light.

In 2005 Kurzweil published *The Singularity Is Near*—it is to come in 2045, to be exact—and ever since he has been promoting these views on his website, Kurzweil Accelerating Intelligence (Kurzweil 2005, 2017). There is no doubt, no hesitation, no humility in Kurzweil's categorical grand pronouncements because according to him the state of the biosphere, whose functioning is a product of billions of years of evolution, has no role in our futures, which are to be completely molded by the surpassing mastery of machine intelligence. But as different as our civilization may be when compared to any of its predecessors, it works within the same constraint: it is nothing but a subset of the biosphere, that relatively very thin and both highly resilient and highly fragile envelope within which carbon-based living organisms can survive (Vernadsky 1929; Smil 2002). Inevitably, their growth, and for higher organisms also their cognitive and behavioral advances, are fundamentally limited by the biosphere's physical conditions and (wide as it may seem by comparing its extremes) by the restricted range of metabolic possibilities.

Studies of Growth

Even when limited to our planet, the scope of growth studies—from ephemeral cells to a civilization supposedly racing toward the singularity—is too vast to allow a truly comprehensive single-volume treatment. Not surprisingly, the published syntheses and overviews of growth processes and of their outcomes have been restricted to major disciplines or topics. The great classic of growth literature, D'Arcy Wentworth Thompson's *On Growth and Form* (whose original edition came out it in 1917 and whose revised and much expanded form appeared in 1942) is concerned almost solely with cells and tissues and with many parts (skeletons, shell, horns, teeth, tusks)

There will be nothing about the growth (the inflationary expansion) of the universe, galaxies, supernovas, or stars. I have already acknowledged inherently slow growth rates of terraforming processes that are primarily governed by the creation of new oceanic crust with spreading rates ranging between less than two and no more than about 20 cm/year. And while some short-lived and spatially limited catastrophic events (volcanic eruptions, massive landslides, tsunami waves, enormous floods) can result in rapid and substantial mass and energy transfers in short periods of time, ongoing geomorphic activities (erosion and its counterpart, sedimentary deposition) are as slow or considerably slower than the geotectonic processes: erosion in the Himalayas can advance by as much as 1 cm/year, but the denudation of the British Isles proceeds at just 2–10 cm in every 1,000 years (Smil 2008). There will be no further examination of these terraforming growth rates in this book.

And as the book's major focus is on the growth of organisms, artifacts, and complex systems, there will be also nothing about growth on subcellular level. The enormous intensification of life science research has produced major advances in our understanding of cellular growth in general and cancerous growth in particular. The multidisciplinary nature, the growing extent, and accelerating pace of these advances means that new findings are now reported overwhelmingly in electronic publications and that writing summary or review books in these fields are exercises in near-instant obsolescence. Still, among the recent books, those by Macieira-Coelho (2005), Gewirtz et al. (2007), Kimura (2008), and Kraikivski (2013) offer surveys of normal and abnormal cellular growth and death.

Consequently, there will be no systematic treatment of fundamental genetics, epigenetics and biochemistry of growth, and I will deal with cellular growth only when describing the growth trajectories of unicellular organisms and the lives of microbial assemblies whose presence constitutes significant, or even dominant, shares of biomass in some ecosystems. Similarly, the focus with plants, animals, and humans will not be on biochemical specificities and complexities of growth at subcellular, cellular, and organ level—there are fascinating studies of brain (Brazier 1975; Kretschmann 1986; Schneider 2014; Lagercrantz 2016) or heart (Rosenthal and Harvey 2010; Bruneau 2012) development—but on entire organisms, including the environmental settings and outcomes of growth, and I will also note some key environmental factors (ranging from micronutrients to infections) that often limit or derail organismic growth.

Human physical growth will be covered in some detail with focus both on individual (and sex-specific) growth trajectories of height and weight (as well as on the undesirable rise of obesity) and on the collective growth of populations. I will present long-term historical perspectives of population growth, evaluate current growth patterns, and examine possible future global, and some national, trajectories. But there will be nothing on psychosocial growth (developmental stages, personality, aspirations, self-actualization) or on the growth of consciousness: psychological and sociological literature covers that abundantly.

Before proceeding with systematic coverage of growth in nature and society, I will provide a brief introduction into the measures and varieties of growth trajectories. These trajectories include erratic advances with no easily discernible patterns (often seen in stock market valuations); simple linear gains (an hourglass adds the same amount of falling sand to the bottom pile every second); growth that is, temporarily, exponential (commonly exhibited by such diverse phenomena as organisms in their infancy, the most intensive phases in the adoption of technical innovation, and the creation of stock market bubbles); and gains that conform to assorted confined (restrained) growth curves (as do body sizes of all organisms) whose shape can be captured by mathematical functions.

Most growth processes—be they of organisms, artifacts, or complex systems—follow closely one of these S-shaped (sigmoid) growth curves conforming to the logistic (Verhulst) function (Verhulst 1838, 1845, 1847), to its precursor (Gompertz 1825), or to one of their derivatives, most commonly those formulated by von Bertalanffy (1938, 1957), Richards (1959), Blumberg (1968), and Turner et al. (1976). But natural variability as well as unexpected interferences often lead to substantial deviations from a predicted course. That is why the students of growth are best advised to start with an actual more or less completed progression and see which available growth function comes closest to replicating it.

Proceeding the other way—taking a few early points of an unfolding growth trajectory and using them to construct an orderly growth curve conforming to a specifically selected growth function—has a high probability of success only when one tries to predict the growth that is very likely to follow a known pattern that has been repeatedly demonstrated, for example, by many species of coniferous trees or freshwater fish. But selecting a random S-curve as the predictor of growth for an organism that does not belong to one of those well-studied groups is a questionable enterprise because a specific function may not be a very sensitive predictive tool for phenomena seen only in their earliest stage of growth.

The Book's Structure and Goals

The text follows a natural, evolutionary, sequence, from nature to society, from simple, directly observable growth attributes (numbers of multiplying cells, diameter of trees, mass of animal bodies, progression of human statures) to more complex measures marking the development and advances of societies and economies (population dynamics, destructive powers, creation of wealth). But the sequence cannot be exclusively linear as there are ubiquitous linkages, interdependencies, and feedbacks and these realities necessitate some returns and detours, some repetitions to emphasize connections seen from other (energetic, demographic, economic) perspectives.

My systematic inquiry into growth will start with organisms whose mature sizes range from microbes (tiny as individual cells, massive in their biospheric presence) to lofty coniferous trees and enormous whales. I will take closer looks at the growth of some disease-causing microbes, at the cultivation of staple crops, and at human growth from infancy to adulthood. Then will come inquiries into the growth of energy conversions and man-made objects that enable food production and all other economic activities. I will also look how this growth changed numerous performances, efficiencies, and reliabilities because these developments have been essential for creating our civilization.

Finally, I will focus on the growth of complex systems. I will start with the growth of human populations and proceed to the growth of cities, the most obvious concentrated expressions of human material and social advancement, and economies. I will end these systematic examinations by noting the challenges of appraising growth trajectories of empires and civilizations, ending with our global variety characterized by its peculiar amalgam of planetary and parochial concerns, affluent and impoverished lives, and confident and uncertain perspectives. The book will close with reviewing what comes after growth. When dealing with organisms, the outcomes range from the death of individuals to the perpetuation of species across evolutionary time spans. When dealing with societies and economies, the outcomes range from decline (gradual to rapid) and demise to sometimes remarkable renewal. The trajectory of the modern civilization, coping with contradictory imperatives of material growth and biospheric limits, remains uncertain.

My aim is to illuminate varieties of growth in evolutionary and historical perspectives and hence to appreciate both the accomplishments and the limits of growth in modern civilization. This requires quantitative treatment throughout because real understanding can be gained only by charting actual growth trajectories, appreciating common and exceptional growth rates, and setting accomplished gains and performance improvements (often so large that they have spanned several orders of magnitude!) into proper (historical and comparative) contexts. Biologists have studied the growth of numerous organisms and I review scores of such results for species ranging from bacteria to birds and from algae to staple crops. Similarly, details of human growth from infancy to maturity are readily available.

In contrast to the studies of organismic growth, quantifications of long-term growth trajectories of human artifacts (ranging from simple tools to complex machines) and complex systems (ranging from cities to civilizations) are much less systematic and much less common. Merely to review published growth patterns would not suffice to provide revealing treatments of these growth categories. That is why, in order to uncover the best-fitting patterns of many kinds of anthropogenic growth, I have assembled the longest possible records from the best available sources and subjected them to quantitative analyses. Every one of more than 100 original growth graphs was prepared in this way, and their range makes up, I believe, a unique collection. Given the commonalities of growth patterns, this is an unavoidably repetitive process but systematic presentations of specific results are indispensable in order to provide a clear understanding of realities (commonalities and exceptions), limits, and future possibilities.

Systematic presentation of growth trajectories is a necessary precondition but not the final goal when examining growth. That is why I also explain the circumstances and limits of the charted growth, provide evolutionary or historical settings of analyzed phenomena, or offer critical comments on recent progression and on their prospects. I also caution about any simplistic embrace of even the best statistical fits for long-term forecasting, and the goal of this book is not to provide an extended platform for time-specific growth projections. Nevertheless, the presented analyses contain a variety of conclusions that make for realistic appraisals of what lies ahead.

In that sense, parts of the book are helpfully predictive. If a century of corn yields shows only linear growth, there is not much of a chance for exponentially rising harvests in the coming decades. If the growth efficiency of broilers has been surpassing, for generations, the performance of all other terrestrial meat animals, then it is hard to argue that pork should be the best choice to provide more protein for billions of new consumers. If unit capacities, production (extraction or generation) rates, and diffusion of every energy conversion display logistic progress, then we have very solid ground to conclude that the coming transition from fossil fuels to renewables will not be an exceptionally speedy affair. If the world's population is getting inexorably urbanized, its energetic (food, fuels, electricity) and material needs will be shaped by these restrictive realities dictating the need for incessant and reliable, mass-scale flows that are impossible to satisfy from local or nearby sources.

Simply put, this book deals in realities as it sets the growth of everything into long-term evolutionary and historical perspectives and does so in rigorous quantitative terms. Documented, historically embedded facts come first—cautious conclusions afterward. This is, of course, in contradistinction to many recent ahistoric forecasts and claims that ignore long-term trajectories of growth (that is, the requisite energetic and material needs of unprecedented scaling processes) and invoke the fashionable mantra of disruptive innovation that will change the world at accelerating speed. Such examples abound, ranging from all of the world's entire car fleet (of more than 1 billion vehicles) becoming electric by 2025 to terraforming Mars starting in the year 2022, from designer plants and animals (synthetic biology rules) making the strictures of organismic evolution irrelevant to anticipations of artificial intelligence's imminent takeover of our civilization.

This book makes no radical claims of that kind; in fact it avoids making any but strongly justified generalizations. This is a deliberate decision resting on my respect for complex and unruly realities (and irregularities) and on the well-attested fact that grand predictions turn out to be, repeatedly, wrong. Infamous examples concerning growth range from those of unchecked expansion of the global population and unprecedented famines that were to happen during the closing decades of the 20th century to a swift takeover of the global energy supply by inexpensive nuclear power and to a fundamentally mistaken belief that the growth rate underlying Moore's law (doubling every two years) can be readily realized through innovation in other fields of human endeavor.

The book is intended to work on several planes. The key intent is to provide a fairly comprehensive analytical survey of growth trajectories in nature and in society: in the biosphere, where growth is the result of not just evolution but, increasingly, of human intervention; and in the manmade world, where growth has been a key factor in the history of populations and economies and in the advancement of technical capabilities. Given this scope, the book could be also read selectively as a combination of specific parts, by focusing on living organisms (be they plants, animals, humans, or populations) or on human designs (be they tools, energy converters, or transportation machinery). And, undoubtedly, some readers will

1 Trajectories: or common patterns of growth

Growth attracts adjectives. The most common ones have been (alphabetically) anemic, arithmetic, cancerous, chaotic, delayed, disappointing, erratic, explosive, exponential, fast, geometric, healthy, interrupted, linear, logistic, low, malignant, moderate, poor, rapid, runaway, slow, S-shaped, strong, sudden, tepid, unexpected, vigorous. Most recently, we should also add sustainable and unsustainable. Sustainable growth is, of course, a clear *contradictio in adjecto* as far as any truly long-run material growth is concerned (I am ignoring any possibilities of migrating to other planets after exhausting the Earth's resources) and it is highly doubtful that we can keep on improving such intangibles as happiness or satisfaction. Most of the adjectives used to describe growth are qualifiers of its rate: often it is not the growth per se that we worry about but rather its rate, either too fast or too slow.

Even a casual news reader knows about the constant worries of assorted chief economists, forecasters and government officials about securing "vigorous" or "healthy" growth of the gross domestic product (GDP). This clamoring for high growth rates is based on the most simplistic expectation of repeating past experiences—as if the intervening growth of GDP had nothing to do with the expected future rate. Put another way, economists have an implicit expectation of endless, and preferably fairly fast, exponential growth.

But they choose an inappropriate metric when comparing the outcomes. For example, during the first half of the 1950s the US GDP growth averaged nearly 5% a year and that performance translated roughly to additional $3,500 per capita (for about 160 million people) during those five years. In contrast, the "slow" GDP growth between 2011 and 2015 (averaging just 2%/year) added about $4,800/capita (for about 317 million people) during those five years, or nearly 40% more than 60 years ago (all totals are in constant-value monies to eliminate the effect of inflation). Consequently, in terms of actual average individual betterment, the recent 2% growth has been quite superior to the former, 2.5 times higher, rate. This is simple algebra, but it is repeatedly ignored by all those bewailing the "low" post-2000 growth of the US or European Union (EU) economies.

Results of the British referendum of June 23, 2016, about remaining in the EU or leaving it, provided another perfect illustration of how the rate of change matters more than the outcome. In 94% of the areas where the foreign-born population increased by more than 200% between 2001 and 2014, people voted to leave the European Union—even though the share of migrants in those regions had remained comparatively low, mostly less than 20%. In contrast, most regions where the foreign-born population was more than 30% voted to remain. As *The Economist* concluded, "High numbers of migrants don't bother Britons; high rates of change do" (Economist 2016).

Other adjectives used to qualify growth are precisely defined terms describing its specific trajectories that conform (sometimes almost perfectly, often fairly closely) to various mathematical functions. Those close, even perfect, fits are possible because most growth processes are remarkably regular affairs as their progress follows a limited array of patterns. Naturally, those trajectories have many individual and inter- and intraspecific variations for organisms, and are marked by historically, technically, and economically conditioned departures for engineered systems, economies, and societies. The three basic trajectories encompass linear growth, exponential growth, and various finite growth patterns. Linear growth is trivial to grasp and easy to calculate. Exponential growth is easy to understand but the best way to calculate it is to use the base of natural logarithms, a mystery to many. The principle of finite growth patterns, including logistic, Gompertz and confined exponential growth functions, is, again, easy to understand, but their mathematical solutions require differential calculus.

But before taking a closer look at individual growth functions, their solutions and resulting growth curves, I will devote two brief sections to time spans and to the figures of merit involved in growth studies. In their short surveys, I will note both common and less frequently encountered variables in whose growth we are interested, be it as parents, employees, or taxpayers, as scientists, engineers, and economists, or as historians, politicians, and planners. These measures include such universal concerns as the weight and height of growing babies and children, and the growth of national economies. And there are also such infrequent but scary concerns as the diffusion of potentially pandemic infections made worse by mass-scale air travel.

Time Spans

Growth is always a function of time and during the course of modern scientific and engineering studies their authors have traced its trajectories in countless graphs with time usually plotted on the abscissa (horizontal or x axis) and the growing variable measured on the ordinate (vertical or y axis). Of course, we can (and we do) trace growth of physical or immaterial phenomena against the change of other such variables—we plot the changing height of growing children against their weight or rising disposable income against the growth of GDP—but most of the growth curves (and, in simpler instance, lines) are what James C. Maxwell defined as diagrams of displacement and what Thompson called time-diagrams: "Each has a beginning and an end; and one and the same curve may illustrate the life of a man, the economic history of a kingdom … It depicts a 'mechanism' at work, and helps us to see analogous mechanisms in different fields; for Nature rings her many changes on a few simple themes" (Thompson 1942, 139).

Growth of ocean floor or of mountain ranges, whose outcomes are driven by geotectonic forces and whose examination is outside of this book's already large scope, unfolds across tens to hundreds of millions of years. When dealing with organisms, the length of time span under consideration is a function of specific growth rates determined by long periods of evolution and, in the case of domesticated plant and animal species, often accelerated or enhanced by traditional breeding and, most recently, also by transgenic interventions. When dealing with the growth of devices, machines, structures or any other human artifacts, time spans under study depend both on their longevity and on their suitability to be deployed in new, enhanced versions under changed circumstances.

As a result, growth of some artifacts that were in use since the antiquity is now merely of historical interest. Sails are a good example of this reality, as their development and deployment (excepting those designed and used for fast yacht racing) ended fairly abruptly during the second half of the 19th century, just a few decades after the introduction of steam engines, and after more than five millennia of improving designs. But other ancient designs have seen spectacular advances in order to meet the requirements of the industrial age: construction cranes and dockyard cranes are perhaps the best example of this continued evolution. These ancient machines have seen enormous growth in their capacities during the past two centuries in order to build taller structures and to handle cargo of increasingly more voluminous ships.

Microbes, fungi, and insects make up most of the biosphere's organisms, and common time spans of interest in microbiology and invertebrate biology are minutes, days, and weeks. Bacterial generations are often shorter than one hour. Coccolithophores, single-celled calcifying marine algae that dominate oceanic phytomass, reach maximum cell density in nitrogen-limited environments in one week (Perrin et al. 2016). Commercially cultivated white mushrooms grow to maturity just 15–25 days after the growing medium (straw or other organic matter) is filled with mycelium. Butterflies usually spend no more than a week as eggs, two to five weeks as caterpillars (larval stage), and one to two weeks as chrysalis from which they emerge as fully grown adults.

In annual plants, days, weeks, and months are time spans of interest. The fastest growing crops (green onions, lettuces, radishes) may be harvested less than a month after seeding; the shortest period to produce mature staple grain is about 90 days (spring wheat, also barley and oats), but winter wheat needs more than 200 days to reach maturity, and a new vineyard will start producing only during the third year after its establishment. In trees, the annual deposition of new wood in rings (secondary growth originating in two cambial lateral meristems) marks an easily identifiable natural progression: fast-growing plantation species (eucalypts, poplars, pines) may be harvested after a decade of growth (or even sooner), but in natural settings growth can continue for many decades and in most tree species it can be actually indeterminate.

Gestation growth of larger vertebrates lasts for many months (from 270 days in humans to 645 days for African elephants), while months, or even just days, are of interest during the fastest spells of postnatal growth. That is particularly the case when meat-producing poultry, pigs, and cattle are fed optimized diets in order to maximize daily weight gain and to raise their mass to expected slaughter weight in the shortest possible time. Months and then years are the normal span of interest when monitoring growth of infants and children, and a pediatrician will compare age- and sex-specific charts of expected growth with actual individual growth to determine if a baby or a toddler is meeting its growth milestones or if it is failing to thrive fully.

Although the growth of some artifacts—be they sailing ships or construction cranes—must be traced across millennia, most of the advances have been concentrated in relatively brief growth spurts separated by long periods of no growth or marginal gains. Energy converters (engines, turbines, motors), machines, and devices characteristic of modern industrial civilization have much shorter life histories. Growth of steam engines lasted 200 years, from the early 18th to the early 20th century. Growth of steam turbines (and electric motors) has been going on since the 1880s, that of gas turbines only since the late 1930s. Growth of modern solid-state electronics began with the first commercial applications of the 1950s but it really took off only with microprocessor-based designs starting in the 1970s.

Studying the collective growth of our species in its evolutionary entirety would take us back some 200,000 years but our ability to reconstruct the growth of the global population with a fair degree of accuracy goes back only to the early modern era (1500–1800), and totals with small ranges of uncertainty have been available only for the past century. In a few countries with a history of censuses (however incomplete, with the counts often restricted to only adult males) or with availability of other documentary evidence (birth certificates maintained by parishes), we can recreate revealing population growth trajectories going back to the medieval period.

In economic affairs the unfolding growth (of GDP, employment, productivity, output of specific items) is often followed in quarterly intervals, but statistical compendia report nearly all variables in terms of their annual totals or gains. Calendar year is the standard choice of time span but the two most common instances of such departures are fiscal years and crop years (starting at various months) used to report annual harvests and yields. Some studies have tried to reconstruct national economic growth going back for centuries, even for millennia, but (as I will emphasize later) they belong more appropriately to the class of qualitative impressions rather than to the category of true quantitative appraisals. Reliable historical reconstructions for societies with adequate statistical services go back only 150–200 years.

Growth rates capture the change of a variable during a specified time span, with percent per year being the most common metric. Unfortunately, these frequently cited values are often misleading. No caveats are needed only if these rates refer to linear growth, that is to adding identical quantity during every specified period. But when these rates refer to periods of exponential growth they could be properly assessed only when it is understood that they are temporary values, while the most common varieties of growth encountered in nature and throughout civilization—those following various S-shaped patterns—are changing their growth rate constantly, from very low rates to a peak and back to very low rates as the growth process approaches its end.

the volume of 42 US gallons (or roughly 159.997 liters) was adopted by the US Bureau of the Census in 1872 to measure crude oil output, and barrel remains the standard output measure in the oil industry—but converting this volume variable to its mass equivalent requires the knowledge of specific densities.

Just over six barrels of heavy crude oil (commonly extracted in the Middle East) are needed to make up one tonne of crude oil, but the total may be as high as 8.5 barrels for the lightest crudes produced in Algeria and Malaysia, with 7.33 barrels per tonne being a commonly used global average. Similarly, converting volumes of wood to mass equivalents requires the knowledge of specific wood density. Even for commonly used species, densities differ by up to a factor of two, from light pines (400 kg/m^3) to heavy white ash (800 kg/m^3), and the extreme wood densities range from less than 200 kg/m^3 for balsa to more than 1.2 t/m^3 for ebony (USDA 2010).

The history of ubiquitous artifacts illustrates two opposite mass trends: miniaturization of commonly used components and devices on one hand (a trend enabled to an unprecedented degree by the diffusion of solid-state electronics), and a substantial increase in the average mass of the two largest investments modern families make, cars and houses, on the other. The declining mass of computers is, obviously, just an inverse of their growing capability to handle information per unit of weight. In August 1969, the Apollo 11 computer used to land the manned capsule on the Moon weighed 32 kg and had merely 2 kB of random access memory (RAM), or about 62 bytes per kg of mass (Hall 1996). Twelve years later, IBM's first personal computer weighed 11.3 kg and 16 kB RAM, that is 1.416 kB/kg. In 2018 the Dell laptop used to write this book weighed 2.83 kg and had 4 GB RAM, or 1.41 GB/kg. Leaving the Apollo machine aside (one-of-a-kind, noncommercial design), personal computers have seen a millionfold growth of memory/mass ratio since 1981!

As electronics (except for wall-size televisions) got smaller, houses and cars got bigger. People think about houses primarily in terms of habitable area but its substantial increase—in the US from 91 m^2 of finished area (99 m^2 total) in 1950 to about 240 m^2 by 2015 (Alexander 2000; USCB 2017)—has resulted in an even faster growth rate for materials used to build and to furnish them. A new 240 m^2 house will need at least 35 tonnes of wood, roughly split between framing lumber and other wood products, including plywood, glulam, and veneer (Smil 2014b). In contrast, a simple 90 m^2 house could be built with no more than 12 tonnes of wood, a threefold difference.

Moreover, modern American houses contain more furniture and they have more, and larger, major appliances (refrigerators, dishwashers, washing machines, clothes dryers): while in 1950 only about 20% of households had washing machines, less than 10% owned clothes dryers and less than 5% had air conditioning, now standard even in the northernmost states. In addition, heavier materials are used in more expensive finishes, including tiles and stone for flooring and bathrooms, stone kitchen counters and large fireplaces. As a result, new houses built in 2015 are about 2.6 times larger than was the 1950 average, but for many of them the mass of materials required to build them is four times as large.

The increasing mass of American passenger cars has resulted from a combination of desirable improvements and wasteful changes (figure 1.2). The world's first mass-produced car, Ford's famous Model T released in October 1908, weighed just 540 kg. Weight gains after World War I (WWI) were due to fully enclosed all-metal bodies, heavier engines, and better seats: by 1938 the mass of Ford's Model 74 reached 1,090 kg, almost exactly twice that of the Model T (Smil 2014b). These trends (larger cars, heavier engines, more accessories) continued after World War II (WWII) and, after a brief pause and retreat brought by the oil price rises by the Organization of the Petroleum Exporting Countries (OPEC) in the 1970s, intensified after the mid-1980s with the introduction of sport-utility vehicles (SUVs, accounting for half of new US vehicle sales in 2019) and the growing popularity of pick-up trucks and vans.

In 1981 the average mass of American cars and light trucks was 1,452 kg; by the year 2000 it had reached 1,733 kg; and by 2008 it was 1,852 kg (and had hardly changed by 2015), a 3.4-fold increase of average vehicle mass in 100 years (USEPA 2016b). Average car mass growth in Europe and Asia has

Figure 1.2
The bestselling American car in 1908 was Ford Model T weighing 540 kg. The bestselling vehicle in 2018 was not a car but a truck, Ford's F-150 weighing 2,000 kg. Images from Ford Motor Company catalogue for 1909 and from Trucktrend.

been somewhat smaller in absolute terms but the growth rates have been similar to the US rise. And while the worldwide car sales were less than 100,000 vehicles in 1908, they were more than 73 million in 2017, roughly a 700-fold increase. This means that the total mass of new automobiles sold globally every year is now about 2,500 larger than it was a century ago.

Time is the third ubiquitous basic unit. Time is used to quantify growth directly (from increased human longevity to the duration of the longest flights, or as time elapsed between product failures that informs us about the durability and reliability of devices). More importantly, time is used as the denominator to express such ubiquitous rates as speed (length/time, m/s), power (energy/time, J/s), average earnings (money/time, $/hour), or national annual gross domestic product (total value of goods and services/time, $/year). Rising temperatures are encountered less frequently in growth studies, but they mark the still improving performance of turbogenerators, while growing total luminosity of illumination informs about the widespread, and intensifying, problem of light pollution (Falchi et al. 2016).

Modern societies have been increasingly concerned about immaterial variables whose growth trajectories describe changing levels of economic performance, affluence, and quality of life. Common variables that the economists want to see growing include the total industrial output, GDP, disposable income, labor productivity, exports, trade surplus, labor force participation, and total employment. Affluence (GDP, gross earnings, disposable income, accumulated wealth) is commonly measured in per capita terms, while the quality of life is assessed by combinations of socioeconomic variables. For example, the Human Development Index (HDI, developed and annually recalculated by the United Nations Development Programme) is composed of three indices quantifying life expectancy, educational level, and income (UNDP 2016).

And in 2017 the World Economic Forum introduced a new Inclusive Development Index (IDI) based on a set of key performance indicators that allow a multidimensional assessment of living standards not only according to their current level of development but also taking into account the recent performance over five years (World Economic Forum 2017). There is a great deal of overlap between HDI for 2016 and IDI for 2017: their rankings share six among the top 10 countries (Norway, Switzerland, Iceland, Denmark, Netherlands, Australia). Perhaps the most interesting addition to this new accounting has been the quantifications of happiness or satisfaction with life.

Small Himalayan Bhutan made news in 1972 when Jigme Singye Wangchuck, the nation's fourth king, proposed to measure the kingdom's progress by using the index of Gross National Happiness (GNH Centre 2016).

Turning this appealing concept into an indicator that could be monitored periodically is a different matter. In any case, for the post-WWII US we have a fairly convincing proof that happiness has not been a growth variable. Gallup pollsters have been asking Americans irregularly how happy they feel since 1948 (Carroll 2007). In that year 43% of Americans felt very happy. The measure's peak, at 55%, was in 2004, the low point came after 9/11 at 37%, but by 2006 it was 49%, hardly any change compared to more than half a century ago (47% in 1952)!

Satisfaction with life is closely connected with a number of qualitative gains that are not easily captured by resorting to simple, and the most commonly available, quantitative measures. Nutrition and housing are certainly the two best examples of this reality. As important as it may be, tracing the growth of average daily per capita availability of food energy may deliver a misleadingly reassuring message. Dietary improvements have lifted food supply far above the necessary energy needs: they may have delivered a more than adequate amount of carbohydrates and lipids and may have satisfied the minimum levels of high-quality protein—but could still be short of essential micronutrients (vitamins and minerals). Most notably, low intakes of fruit and vegetables (the key sources of micronutrients) have been identified as a leading risk factor for chronic disease, but Siegel et al. (2014) showed that in most countries their supply falls below recommended levels. In 2009 the global shortfall was 22% with median supply/need ratios being just 0.42 in low-income and 1.02 in affluent countries.

During the early modern era, the rise of scientific methods of inquiry and the invention and deployment of new, powerful mathematical and analytical tools (calculus during the mid-17th century, advances in theoretical physics and chemistry and the foundations of modern economic and demographic studies during the 19th century) made it eventually possible to analyze growth in purely quantitative terms and to use relevant growth formulas in order to predict long-term trajectories of studied phenomena. Robert Malthus (1766–1834), a pioneer of demographic and economic studies, caused a great of concern with his conclusion contrasting the means of subsistence that grow only at a linear rate with the growth of populations that proceeds at exponential rates (Malthus 1798).

Unlike Malthus, Pierre-François Verhulst (1804–1849), a Belgian mathematician, is now known only to historians of science, statisticians, demographers, and biologists. But four decades after Malthus's essay, Verhulst made a fundamental contribution to our understanding of growth when he published the first realistic formulas devised explicitly to express the progress of confined (bounded) growth (Verhulst 1838, 1845, 1847). Such growth governs not only the development of all organisms but also the improving performance of new techniques, diffusion of many innovations and adoption of many consumer products. Before starting my topical coverage of growth phenomena and their trajectories (in chapter 2), I will provide brief, but fairly comprehensive, introductions into the nature of these formal growth patterns and resulting growth curves.

Linear and Exponential Growth

These are two common but very different forms of growth whose trajectories are captured by simple equations. Relatively slow and steady would be the best qualitative description of the former, and increasingly rapid and eventually soaring the best of the latter. Anything subject to linear growth increases by the same amount during every identical period and hence the equation for linear growth is simple:

$$N_t = N_0 + kt$$

where a quantity at time t (N_t) is calculated by enlarging the initial value (N_0) by the addition of a constant value k per unit of time, t.

Analysis of a large number of stalagmites shows that these tapering columns of calcium salts created on cave floors by dripping water often grow for millennia in a near-linear fashion (White and Culver 2012). Even a relatively fast growth of 0.1 mm/year would mean that a stalagmite 1 meter tall would grow just 10 cm in thousand years ($1{,}000\ \text{mm} + 1{,}000 \times 0.1$). The plotted outcome of its linear growth shows a monotonously ascending line (figure 1.3). This, of course, means that the growth rate as the share of the total stalagmite height will be constantly declining. In a stalagmite growing at 0.1 mm/year for 1,000 years it would be 0.01% during the first year but only 0.009% a millennium later.

In contrast, in all cases of exponential growth the quantity increases by the same rate during every identical period. The basic functional dependence is

$$N_t = N_0 (1 + r)^t$$

where r is the rate of growth expressed as a fraction of unity growth per unit time, for example, for a 7% increase per unit of time, $r = 0.07$.

This exponential growth can be also expressed—after a trivial multiplicative unit-of-timekeeping adjustment—as

$$N_t = N_0 e^{rt}$$

where e ($e = 2.7183$, the base of natural logarithms) is raised to the power of rt, an easy operation to do with any scientific hand-calculator. We can of machines have been often linear, including the average power of American cars since the Ford Model T in 1908, maximum thrust and bypass ratio of jet engines since their origin, maximum train speed and boiler pressure of steam locomotives (since the beginning of regular service in 1830), and maximum ship displacements.

And sometimes simple linear growth is an outcome of complex interactions. Between 1945 and 1978, US gasoline consumption had followed an almost perfectly linear course—and after a brief four-year dip it resumed a slower linear growth in 1983 that continued until 2007 (USEIA 2017b). The two linear trajectories resulted from an interplay of nonlinear changes as vehicle ownership soared, increasing more than seven times between 1945 and 2015, while average fuel-using efficiency of automotive engines remained stagnant until 1977, then improved significantly between 1978 and 1985 before becoming, once again, stagnant for the next 25 years (USEPA 2015).

Some organisms, including bacteria cultivated in the laboratory and young children, experience periods of linear growth, adding the same number of new cells or the same height or the same mass increment, during specific periods of time. Bacteria follow that path when provided with a limited but constant supply of a critical nutrient. Children have spells of linear growth both for weight and height. For example, American boys have brief periods of linear weight growth between 21 and 36 months of age (Kuczmarski et al. 2002), and the Child Growth Standards of the World Health Organization (WHO) indicate a perfectly linear growth of height with age for boys between three and five years, and an almost-linear trajectory for girls between the same ages (WHO 2006; figure 1.4).

Exponential Growth

Exponential growth, with its gradual takeoff followed by a steep rise, attracts attention. Properties of this growth, formerly known as geometric ratio or geometric progression, have been illustrated for hundreds of years—perhaps for millennia, although the first written instance comes only from the year 1256—by referring to the request of a man who invented chess and asked his ruler-patron to reward him by doubling the number of grains of rice (or wheat?) laid on every square. The total of 128 grains (2^7) is still trivial at the end of the first row; there are about 2.1 billion grains (2^{31}) when reaching the end of the middle, fourth, row; and at the end, it amounts to about 9.2 quintillion (9.2×10^{18}) grains.

The key characteristic of advanced exponential growth are the soaring additions that entirely overwhelm the preceding totals: additions to the last row of the chessboard are 256 times larger than the total accumulated height (cm)

age (years)

age (years)

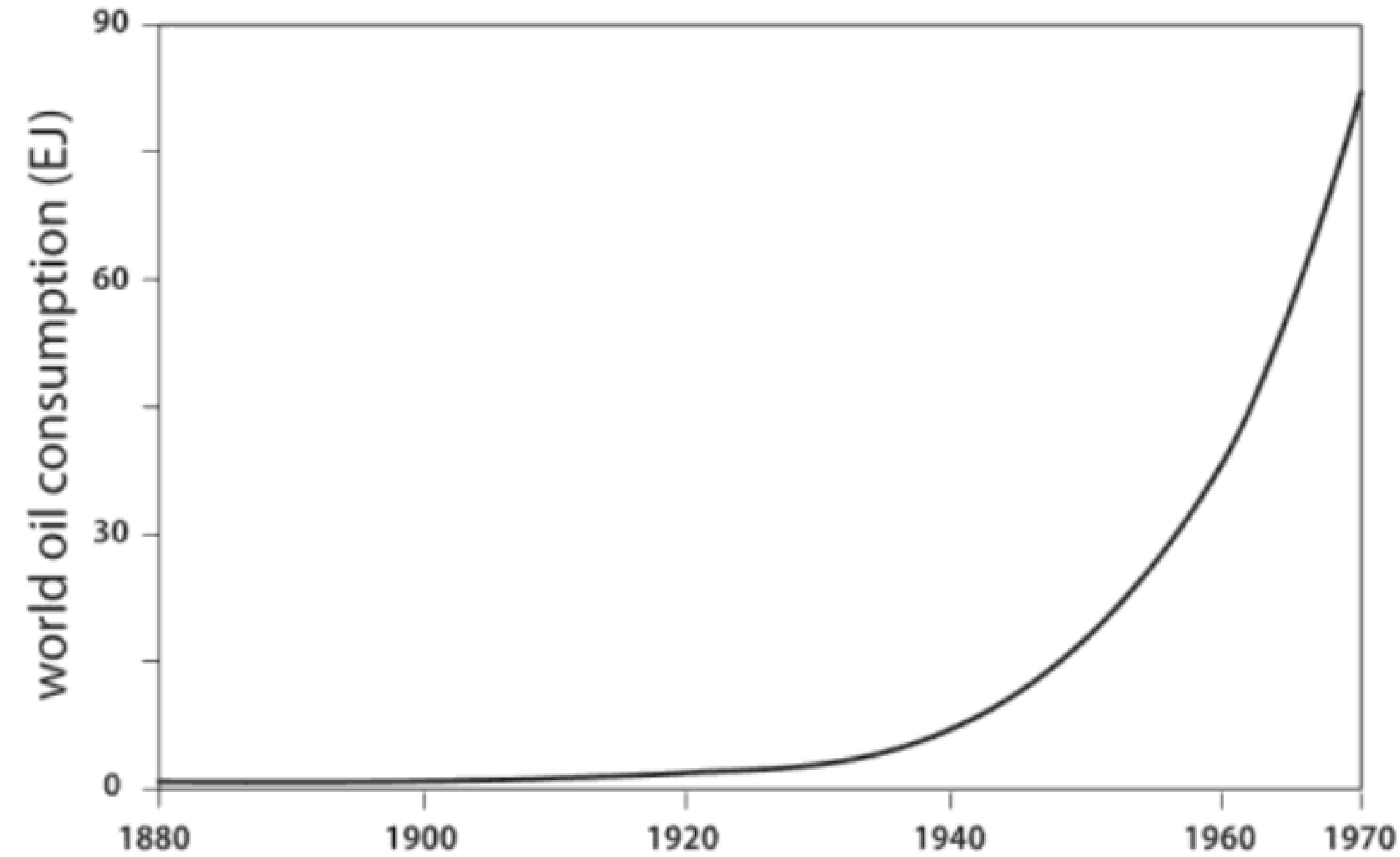
Figure 1.4
Graphs of expected height-for-age growth (averages and values within two standard deviations) for boys and girls 2–5 years old. Simplified from WHO (2006).

at the end of the penultimate row, and they represent 99.61% of all added grains. Obviously, undesirable exponential growth may be arrested, with various degrees of effort, in its early stages, but the task may quickly become unmanageable as the growth continues. When assuming an average rice grain mass of 25 milligrams, the grand total (all too obviously not able to fit any chessboard) would equal about 230 Gt of rice, nearly 500 times more than the grain's global annual harvest—which was just short of 500 Mt in 2015.

Over long periods even minuscule growth rates will produce impossible outcomes and there is no need to invoke any cosmic time spans—referring back to antiquity will do. When imperial Rome reached its apogee (in the second century of the common era), it needed to harvest about 12 Mt of grain (much of it grown in Egypt and shipped to Italy) in order to sustain its population of some 60 million people (Garnsey 1988; Erdkamp 2005; Smil 2010c). When assuming that Rome would have endured and that its grain harvest would have grown at a mere 0.5% a year its total would have now reached about 160 Gt, or more than 60 times the world's grain harvest of 2.5 Gt in 2015 used to feed more than 7 billion people.

Linear scale is a poor choice for charting exponential growth whose complete trajectory often encompasses many orders of magnitude. In order to accommodate the entire range on a linear y axis it becomes impossible to make out any actual values except for the largest order of magnitude, and the result is always a J-curve that has a nearly linear section of relatively slow gains followed by a more or less steep ascent. In contrast, plotting constant exponential growth on a semilogarithmic graph (with linear x axis for time and logarithmic y axis for the growing quantity) produces a perfectly straight line and actual values can be easily read off the y axis even when the entire growth range spans many orders of magnitude. Making a semilog plot is thus an easy graphic way of identifying if a given set of data has been a result of exponential growth. Figure 1.5 compares the two plots for such a phenomenon: it charts the growth of one of the key foundations of modern civilization, the almost perfectly exponential rise of global crude oil consumption between 1880 and 1970.

The fuel's commercial production began on a negligible scale in only three countries, in Russia (starting in 1846) and in Canada and the US (starting in 1858 and 1859). By 1875 it was still only about 2 Mt and then, as US and Russian extraction expanded and as other producers entered the market (Romania, Indonesia, Burma, Iran), the output grew exponentially to about 170 Mt by 1930. The industry was briefly slowed down by the economic crisis of the 1930s, but its exponential rise resumed in 1945 and,



Fi[illegible]

Growth of annual global crude oil consumption, 1880–1970: exponential growth plotted on linear and semilog scales. Data from Smil (2017b).

propelled by new huge Middle Eastern and Russian discoveries, by the mid-1970s the output was three orders of magnitude (slightly more than 1,000 times) higher than 100 years previous.

Temporary periods of exponential growth have not been uncommon in modern economies, where they have marked the rise of domestic product in such rapidly developing nations as Japan, South Korea, and post-1985 China, and where they characterized annual sales of electronic consumer goods whose mass appeal created new global markets. And, of course, fraudulent investing schemes (Ponzi pyramids) are built on the allure of the temporary exponential rise of make-believe earnings: arresting exponential growth in its early stages can be done in manageable manner, sudden collapse of Ponzi-like growth will always have undesirable consequences. Progress of technical advances has been also often marked by distinct exponential spells, but when the exponential growth (and its perils) became a major topic of public discourse for the first time it was in relation to rising sizes of populations (Malthus 1798).

That famous work—*An Essay on the Principle of Population*—by Thomas Robert Malthus had precedents in the work of Leonhard Euler, a leading scientist of the 18th century who left Switzerland to work in Russia and Prussia (Bacaër 2011). In Berlin, after his return from Russia, Euler published—in Latin, at that time still the standard language of scientific writing—*Introduction to Analysis of the Infinite* (Euler 1748). Among the problems addressed in the book was one inspired by Berlin's 1747 population census which counted more than 100,000 people. Euler wanted to know how large such a population, growing annually by one thirtieth (3.33% a year), would be in 100 years. His answer, determined by the use of logarithms, was that it could grow more than 25 times in a century: as $P_n = P_0\,(1 + r)^n$, the total in 100 years will be $100{,}000 \times (1 + 1/30)^{100}$ or 2,654,874. Euler then proceeded to show how to calculate the annual rate of population increase and the doubling periods.

But it was Malthus who elevated the powers of exponential growth to a major concern of the new disciplines of demography and political economy. His much-repeated conclusion was that "the power of population is indefinitely greater than the power in the earth to produce subsistence for man" because the unchecked population would be rising exponentially while its means of subsistence would be growing linearly (Malthus 1798, 8):

> Taking the population of the world at any number, a thousand millions, for instance, the human species would increase in the ratio of—1, 2, 4, 8, 16, 32, 64, 128, 256, 512, etc. and subsistence as—1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc. In two centuries and a quarter, the population would be to the means of subsistence as 512 to 10: in three centuries as 4096 to 13, and in two thousand years the difference would be almost incalculable, though the produce in that time would have increased to an immense extent.

Charles Darwin illustrated the process with references to Malthus and Linnaeus and with his own calculation of the consequences of unchecked elephant breeding (Darwin 1861, 63):

> There is no exception to the rule that every organic being increases at so high a rate, that if not destroyed, the earth would soon be covered by the progeny of a single pair. Even slow-breeding man has doubled in twenty-five years, and at

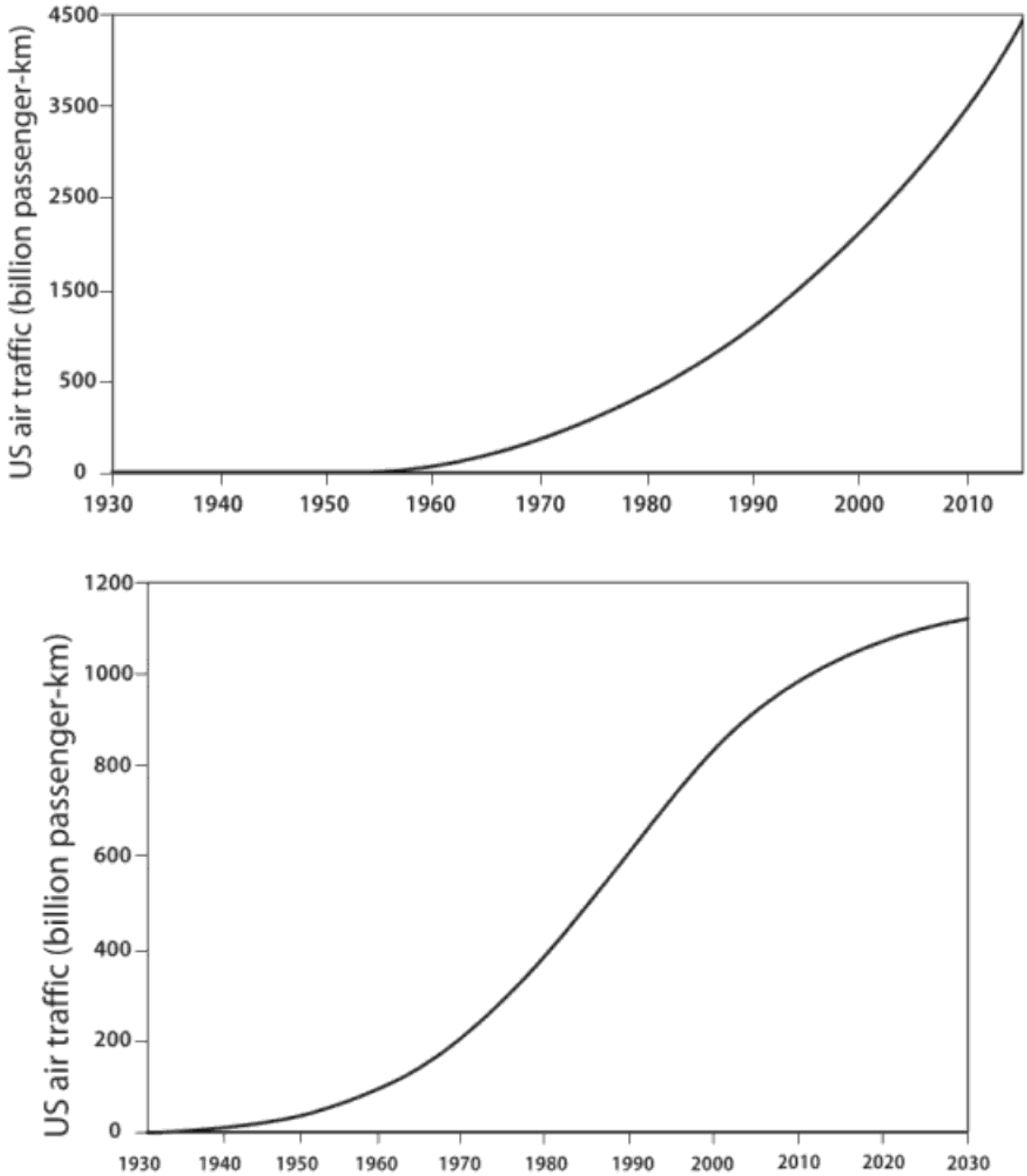


Figure 1.6

Predictions of growth of US air travel (in billions of passenger-kilometers) based on the period 1930–1980 (top, the best fit is quartic regression) and 1930–2015 (bottom, the best fit is a logistic curve with the inflection year in 1987). Data from various annual reports by the International Civil Aviation Organization.

built. This example of a sobering contrast between early rapid advances of a technical innovation followed by inevitable formation of sigmoid curves should be recalled whenever you see news reports about all cars becoming electric by 2025 or new batteries having impressively higher energy densities by 2030.

But the final, inescapable power of this reality may seem inapplicable in those cases where exponential growth has been underway for an extended period of time and when it keeps setting new record levels. More than a few normally rational people have been able to convince themselves—by repeating the mantra "this time it is different"—that performances will keep on multiplying for a long time to come. The best examples of these, often collective, delusions come from the history of stock market bubbles and I will describe in some detail just two most notable recent events, Japan's pre-1990 rise and America's New Economy of the 1990s.

Japan's economic rise during the 1980s provides one of the best examples of people who should know better getting carried away by the power of exponential growth. After growing 2.6 times during the 1970s, Nikkei 225 (Japan's leading stock market index and the country's equivalent of America's Dow Jones Industrial) increased by 184% between January 1981 and 1986, added 43% in 1986, nearly 13% in 1987, almost 43% in 1988, and a further 29% in 1989 (Nikkei 225 2017). Between January 1981 and December 1989, Nikkei 225 had more than quintupled, the performance corresponding to average annual exponential growth of 17% for the decade and 24% for its second half. Concurrently, Japan's GDP kept on growing at an annual rate surpassing 4%, as the yen's exchange rate strengthened from ¥239/US$ in January 1980 to ¥143/US$ by December 1989.

A sobering denouement had to come, and in chapter 6 I will trace that swift post-1989 unfolding. But exponential growth is a potent delusion-maker, and in 1999, 10 years after the Nikkei's peak, I was thinking about the Japanese experience as we were waiting to claim our rental car at San Francisco airport. Silicon Valley was years into its first dotcom bubble, and even with advance reservations people had to wait for the just-returned cars to get serviced and released again into the halting traffic on the clogged Bayshore freeway. Mindful of the Japanese experience, I was thinking that every year after 1995 might be the last spell of what Alan Greenspan famously called irrational exuberance, but it was not in 1996 or 1997 or 1998. And even more so than a decade earlier, there were many economists ready to assure American investors that this spell of exponential growth was really different, that the old rules do not apply in the New Economy where endless rapid growth will readily continue.

During the 1990s, the Dow Jones Industrial Average—driven by America's supposedly New Economy—posted the biggest decadal gain in history as it rose from 2,810 at the beginning of January 1990 to 11,497 at the end of December 1999 (FedPrimeRate 2017). The performance corresponded to annual exponential growth of 14% during the decade, with the peak gains of 33% in 1995 and 25% in 1996. Continuation of that growth pointed to a level around 30,000 by 2010. And the Nasdaq Composite Index—reflecting the rising computing and communication capabilities and, above all, the soaring performance of speculation-driven Silicon Valley companies—did even better during the 1990s: its exponential growth averaged almost 26% annually between April 1991, when it reached 5,000 points, and March 9, 2000, when it peaked at 5,046 points (Nasdaq 2017).

Even some normally cautious observers got swept away by this. Jeremy Siegel, at the Wharton School of Business, marveled: "It's amazing. Every year we say it can't be another year of 20 percent-plus (gains)—and then every year it's 20 percent-plus. I still maintain we have to get used to lower, more normal returns, but who knows when this streak is going to end?" (Bebar 1999). And the boosters made money by wholesaling the impossible: one bestseller saw an early arrival of Dow Jones at 40,000 (Elias 2000), another forecast the unstoppable coming of Dow 100,000 (Kadlec and Acampora 1999). But the end came and, again, it was fairly swift. By September of 2002, Dow Jones was down to 9,945, a nearly 40% decline from its 1999 peak (FedPrimeRate 2017), and by May 2002 Nasdaq Composite fell nearly 77% from its March 2000 peak (Nasdaq 2017).

Exponential growth has been common in many cases of technical advances and, as I will show in chapter 3, in some instances it has persisted for decades. The maximum power of steam turbines is a perfect example of this long-lasting exponential growth. Charles Algernon Parsons patented the first design in 1884 and almost immediately built a small machine—which can be seen in the lobby of the Parsons Building at Trinity College in Dublin—with power of just 7.5 kW, but the first commercial turbine was 10 times larger as the 75 kW machine began generating electricity in 1890 (Parsons 1936).

The subsequent rapid rise brought the first 1 MW turbine by 1899, a 2 MW machine just three years later, the first 5 MW design in 1907, and before WWI the maximum capacity reached 25 MW with the turbine installed at the Fisk Street station of the Commonwealth Edison Co. in Chicago (Parsons 1911). Between the year of the first commercial 75 kW model in 1890 and the 25 MW machine of 1912, maximum capacities of Parsons steam turbines were thus growing at an annual compounded exponential rate of more than 26%, doubling in less than three years. That was considerably faster than the growth of early steam engine capacities during the 18th century, or the rated power of water turbines since the 1830s when Benoît Fourneyron commercialized his first designs.

And some performances advance exponentially not by a constant improvement of the original technique but by a series of innovations, with the next innovation stage taking off where the old technique reached its limits: individual growth trajectories are unmistakably S-shaped but their envelope charts an exponential ascent. The history of vacuum tubes, briefly reviewed in chapter 4, is an excellent example of such an exponential envelope spanning nearly a century of advances. In chapter 4 (on the growth of artifacts), I will also look in detail at perhaps the most famous case of modern exponential growth that has been sustained for 50 years, that of the crowding of transistors on a silicon microchip described by Moore's law that has doubled the number of components every two years.

And before leaving the topic of exponential growth, this is an apposite place to note a simple rule for calculating the doubling period of quantities subject to it, be they cancerous cells, bank accounts or the processing capacity of computers, or, in reverse, to calculate a growth rate by using the known doubling time. Exact results are obtained by dividing the natural logarithm of 2 (that is 0.693) by the prevailing growth rate (expressed as a fraction of one, e.g. 0.1 for 10%), but a quite good approximation is dividing 70 by the growth rate expressed in percent. When the Chinese economy was growing at 10% a year, its doubling period was seven years ; conversely, the doubling of components on a microchip in two years implies an annual exponential growth rate of about 35%.

Hyperbolic Growth

Unbounded, and hence on Earth always only temporary, exponential growth should not be mistaken (as it sometimes is) for hyperbolic growth. While exponential progress is characterized by an increasing absolute rate of growth, it remains a function of time as it approaches infinity; in contrast, hyperbolic growth culminates in an absurdity as a quantity grows toward infinity in a finite time interval (figure 1.7). This terminal event is, of course, impossible within any finite confines and a moderating feedback will eventually exert a damping effect and terminate the hyperbolic progress. But when starting at low rates, hyperbolic trajectories may be sustained for relatively long periods of time before their progression stops and another form of growth (or decline) takes over.

Cailleux (1951) was the first to note what he called the *surexpansion*, the fact that the global population was growing at an ever-increasing rate, the process made possible by an accelerated evolution of civilizations: "Ainsi est-il normal de lier la surexpansion humaine à la présence de l'Esprit" (Cailleux 1951, 70). This process fits a quasi-hyperbolic equation: $P=a/(D-t)M$ where a, D and M are constants. Meyer and Vallee (1975, 290) thus concluded that the growth of human population, "far from tending 'naturally' toward an equilibrium state…exhibits a unique characteristic, that of self-acceleration."

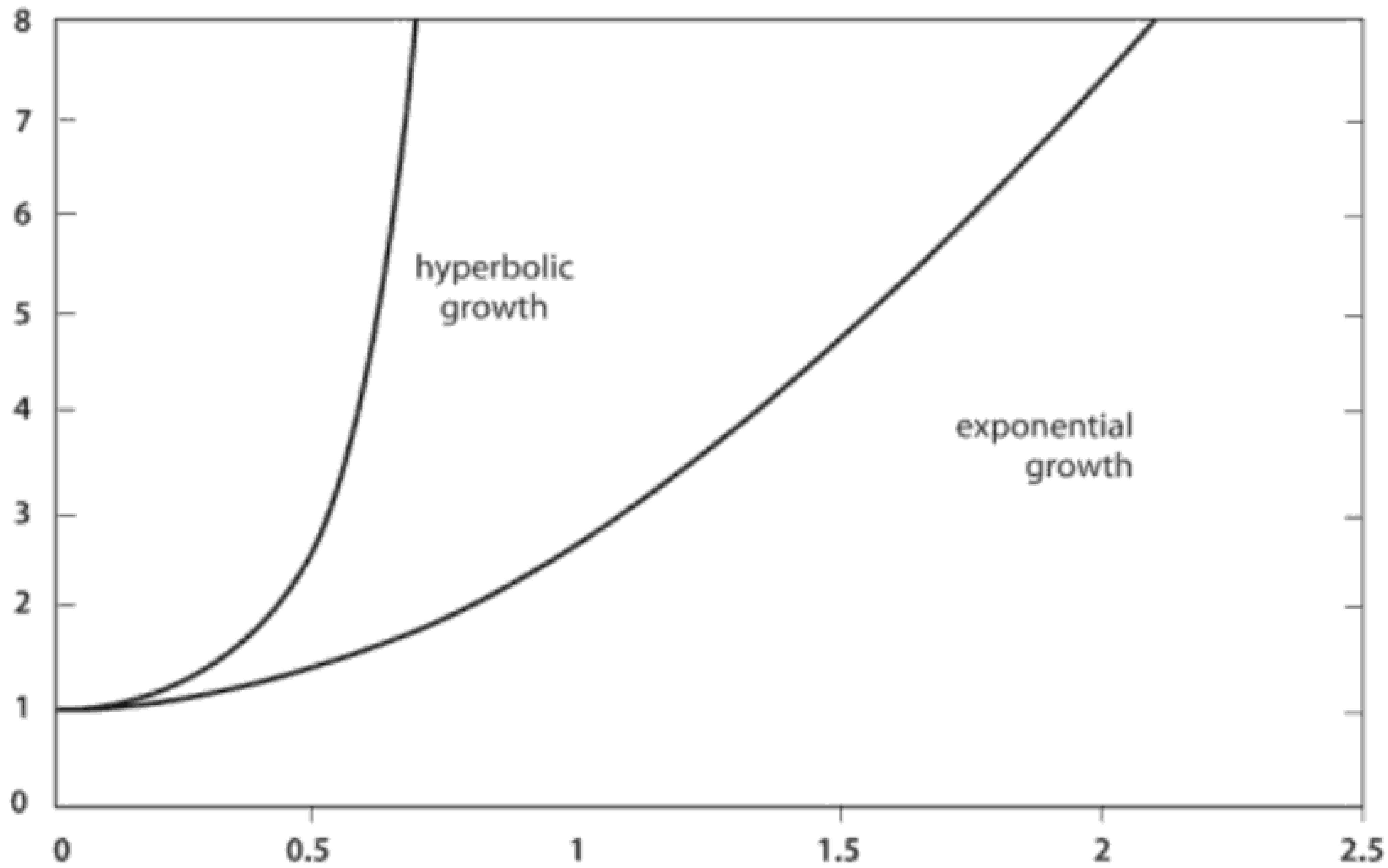


Figure 1.7
Hyperbolic growth curve in comparison with exponential growth.

But only temporarily, as the projection of this growth would eventually lead to infinite population. Von Foerster et al. (1960, 1291) had actually calculated that "Friday, 13 November, A.D. 2026" will be the doomsday when the "human population will approach infinity if it grows as it has grown in the last two millennia." Obviously, that could never happen, and just a few years after von Foerster et al. published their paper the annual growth of global population peaked and a transition to a new trajectory began.

Even so, Hern (1999) argued that global population growth had shown striking parallels with malignant growth because some cancers also display decreasing doubling times of cell proliferation during their most invasive phase. Starting the count 3 million years ago, he calculated that by 1998 the human population had undergone 32.5 doublings, with the 33rd (reaching 8.59 billion) to be completed in the early 21st century. When adding biomass of domestic animals to the anthropomass, the 33rd doubling was already completed. Some malignant tumors cause death of the host organism after 37–40 doublings, and (assuming the trend continues) the 37th doubling of the human population will be reached in a few hundred years.

Nielsen's (2015) analysis of world population growth showed that there were actually three approximately determined episodes of hyperbolic growth during the past 12,000 years: the first one between 10,000 and 500 BCE, the second one between 500 and 1200 CE, and the third one between 1400 and 1950, in total accounting for about 89% of total growth over the past 12 millennia. The first two transitions (500 BC to 500 CE and 1200–1400) were

And, obviously, even if the best seeds are planted and when the crop receives optimum nutrients, moisture, and protection against weeds and pests, the maximum yield remains limited by light intensity, length of the growing period, minimum temperature tolerated by the species, and the vulnerability to many kinds of natural catastrophes. As I will show in chapter 2 (in the section on crop growth), many regions of previously rising productivities now have diminished returns with intensive inputs of fertilizers and enhanced irrigation and their yield trajectory has been one of minimal gains or outright stagnation. Clearly, there is no universal, super-exponential progression toward superior harvests. Human ingenuity has brought more impressive gains when it did not have to reckon with the complexities of organisms whose life cycles are determined by assorted environmental constraints. Technical advances provide the best examples of self-accelerating development following hyperbolic growth trajectories, and the maximum unit power of prime movers and top travel speeds offer accurately documented illustrations.

The maximum unit power of modern prime movers (primary sources of mechanical power) shifted first from less than 1,000 W for steam engines in the early 17th century, to water turbines (between 1850 and 1900), and then to steam turbines, whose record ratings now surpass 1 GW (figure 1.8).

Figure 1.8
Relay growth of the largest stationary prime mover capacities (Smil 2017b). Overlapping logistic growth of unit ratings of steam engines, water turbines, and steam turbines produces a temporary hyperbolic growth trend with nearly seven-order-magnitude gain in 300 years.

Further extension could be made by including rocket engines deployed only for brief periods of time: power of the Saturn C 5 rocket that launched the Apollo mission to the Moon was about 2.6 GW (Tate 2012). Similarly, maximum travel speeds increased from endurance running (10–12 km/h sustained by messengers) and riders on good horses (average speed 13–16 km/h) to fast sailing ships (mid-19th century clippers averaging around 20 km/h with top speeds above 30 km/h), trains (maxima of around 100 km/h reached before 1900), to commercial airplanes powered by reciprocating engines (speeds rising from 160 km/h in 1919 to 550 km/h in 1945) and, finally, to jetliners (more than 900 km/h since the late 1950s).

In both instances, the accelerating growth has been achieved by the relay phenomenon as the overlapping logistic (self-limiting) curves produce an impressively ascending envelope. Obviously, this relay cannot continue because it would eventually produce impossibly high growth rates, be it of unit power or speed.. As has been the case with global population growth, a temporary hyperbolic envelope will be eventually transformed to a logistic trajectory. In fact, this has already been the case when technical advances are considered in practical, realistic terms and not as sequences of maximum performances.

Obviously, constructing the maximum speed envelope by overlapping logistic curves of speeds for horses, sailing ships, trains, cars, airplanes, and rockets shows a progression of transportation modes that is not sequentially substitutable. High-volume urban transport progressed from horse-drawn vehicles to motorized road vehicles and subways but it will not progress to jet-powered flight. The opposite is true, as the average speed of urban traffic has declined in almost every major city since the 1960s, and just its doubling would be impossible even if every vehicle was part of a synchronized, automated urban system (unless all crossroads were eliminated, infrastructurally an impossible transformation in existing cities). The average speed of rapid trains has increased only marginally since their first deployment in 1964 and, once again, it is a very safe bet that billions of people traveling by train will not do so in a decade or two in a supersonic hyperloop fashion.

The typical speed of large container ships (30–40 km/h) is not radically higher than the typical speed of the 19th-century clippers; of course, their cargo capacities are orders of magnitude apart, but there has been no hyperbolic growth of speeds in ocean shipping, and there is no realistic prospect that this fundamental mode of transport that enabled modern economic globalization will enter a new age of radically increased speeds. The cruising speed of the latest Boeing 787 (913 km/h) is nearly 7% lower than the cruising speed of the company's first commercial jetliner, Boeing 707, in 1958

(977 km/h). Again, there is no realistic prospect that billions of passengers will be soon casually traveling at supersonic speeds. The seemingly hyperbolic envelope of maximum performances tells us little about the actual trajectories of speeds that have created modern economies by moving billions of people and billions of tonnes of raw materials, foodstuffs, and finished products.

The same is, inevitably, true about other envelopes of growing technical capabilities. The largest rockets may produce gigawatts of power during very brief periods of takeoff, but that is irrelevant as far as actual capacities of myriads of machines energizing modern civilization are concerned. Most electric motors in our appliances have power that is smaller than that delivered by a well-harnessed horse: washing machines need 500 W, a well-fed horse could easily sustain 800 W. The typical, or modal, capacity of steam turbines in large electricity-generating stations has been fairly stable since the 1960s, with units of 200–600 MW dominant in new plants fired by coal or natural gas, and with turbogenerators larger than 1 GW reserved mostly for the largest nuclear stations. And the power of typical road vehicles has gone up slightly only because they got heavier, not because they need to be more powerful to go from red light to red light or to cruise within posted speed limits on highways, for which a motive power of ~11 kW/t of vehicle weight is sufficient for 100 km/h travel on level roads (Besselink et al. 2011). Again, a synthetic rising trajectory is composed of disparate progressions that do not imply any unified rising trend of ever-ascending substitutions.

And there is no shortage of historical examples of technical advances that do not show any automatic, tightly sequenced acceleration of performance. Steelmakers continued to rely on open-hearth furnaces for nearly a century after they perfected their use, and the hard-wired rotary-dial telephone changed little between its adoption during the 1920s and the introduction of the push-button design in 1963 (Smil 2005 and 2006b). And there is no doubt about the long-term trajectory of hyperbolic growth on the Earth: it must either collapse or it must morph into a confined progression which might become a part of a homeostatic coexistence of humanity and the biosphere including an eventual upper limit on the information content in the external memory (Dolgonosov 2010).

Confined Growth Patterns

These are, above all, the trajectories of life: the biosphere's mass of recyclable nutrients allows for an enormous variety of specific genetic expressions and mutations but it puts fundamental limits on the performance of primary production (photosynthesis) and hence on the accumulation of secondary production (heterotrophic metabolism of organisms ranging from microbes to the most massive mammals). These limits unfold through intra- and interspecific competition of microorganisms, plants and animals for resources, through predation and viral, bacterial and fungal infections, and all multicellular organisms are subject to intrinsic growth limits imposed by apoptosis, programmed cell death (Green 2011).

No tree grows to heaven but neither does any artifact, structure or process, and confined (or constrained) growth patterns characterize the development of machines and technical capabilities as much as they describe the growth of populations and expansion of empires. And, inevitably, all diffusion and adoption processes must conform to that general pattern: no matter if their early trajectory shows rapid or slow progress, it is eventually followed by a substantial slowdown in growth rate as the process asymptotically approaches saturation and often reaches it (sometimes after many decades of diffusion) only a few percent, even only a fraction of a percent, short of the maximum. No households had electricity in 1880 but how many urban dwellings in Western cities are not connected to the grid today?

Given the ubiquity of phenomena exhibiting confined growth, it is not surprising that many investigators sought to fit them into a variety of mathematical functions. The two basic classes of trajectories of bounded growth are those of S-shaped (sigmoid) growth and of confined exponential growth. Scores of papers describe original derivations and subsequent modifications of these curves. There are also their extensive reviews (Banks 1994; Tsoularis 2001) and perhaps the best overview is table S1 in Myhrvold (2013) that systematically compares equations and constraints for more than 70 nonlinear growth functions.

S-shaped Growth

S-shaped functions describe many natural growth processes as well as the adoption and diffusion of innovations, be they new industrial techniques or new consumer items. Initially slow growth accelerates at the J-bend and it is followed by a rapid ascent whose rate of increase eventually slows down, forming the second bend that is followed by a slowing ascent as the growth becomes minimal and the total approaches the highest achievable limit of a specific parameter or a complete saturation of use or ownership. By far the best known, and the most often used function of the S-shaped trajectory is the one expressing logistic growth.

Unlike with exponential (unbounded) growth, whose rate of increase is proportional to the growing quantity, relative increments of logistic (limited) growth decrease as the growing quantity approaches its maximum possible level that in ecological studies is commonly called carrying capacity. Such growth seems to be intuitively normal:

> A typical population grows slowly from an asymptotic minimum; it multiplies quickly; it draws slowly to an ill-defined and asymptotic maximum. The two ends of the population-curve define, in a general way, the whole curve between; for so beginning and so ending the curve must pass through a point of inflection, it *must* be an S-shaped curve. (Thompson 1942, 145)

The origins of the formally defined logistic function go back to 1835 when Adolphe Quetelet (1796–1874; figure 1.9), Belgian astronomer and at that time Europe's leading statistician, published his pioneering analysis *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale* in

Figure 1.9
Adolphe Quetelet and Pierre-François Verhulst. Steel engravings from the author's collection of 19th-century images.

In his 1845 paper, Verhulst assumed that the checks to further population growth would be proportional to the size of the excess population (*population surabondante*), and when he used this growth function to derive the ultimate limits of Belgian and French populations he set them, respectively, at about 6.6 and 40 million to be reached before the end of the 20th century. But in his last paper on population growth he concluded that the barriers to population growth are proportional to the ratio between the excess population and the total population (Verhulst 1847). That change resulted in a larger ultimate population or, as the asymptotic value became eventually known, in higher carrying capacities (Schtickzelle 1981).

Essentially, Verhulst's equation captures the shifting dominance between two feedback loops: a positive feedback loop (FBL) initiates growth that is eventually slowed down and brought into balance by a negative feedback that reflects the limits to growth prevailing in the finite world. As Kunsch (2006, 35) put it, logistic growth "is described as combining exponential growth embodied in (+) FBL, and goal-seeking growth, embodied in a (–) FBL." In that sense, Verhulst's function, with two feedback loops competing for dominance, can be seen as the very foundation of feedback-based systems dynamics developed by Jay Forrester at the Massachusetts Institute of Technology during the 1950s and 1960s (Forrester 1971) and applied by the study supported by the Club of Rome on the global *Limits to Growth* (Meadows et al. 1972).

This key systemic concept of constrained growth (a high density of organisms being the proximate constraining factor and resource availability being the complex causative driver) has been very useful when conceptualizing many natural, social, and economic developments involving series of feedbacks but its mechanistic application can result in substantial errors. Verhulst's original population forecasts are the earliest illustrations of such errors because population maxima are not preordained by any specific growth function but depend on changing a nation's, and ultimately the planet's, productive potential through scientific, technical, and economic development. How long such higher, evolving, maxima can be sustained is another matter. Verhulst eventually raised his Belgian population maximum by the year 2000 from 6.6 to 9.5 million—but by the end of the 20th century the Belgian and French populations were, respectively, at 10.25 and 60.91 million: for Belgium that was about 8% higher than Verhulst's adjusted maximum, but for France the error was 52%.

Although the second half of the 19th century saw an explosion of demographic and economic studies, Verhulst's work was ignored and it was rediscovered only during the 1920s and became influential only during the

1960s (Cramer 2003; Kint et al. 2006; Bacaër 2011). This was not the only instance of such forgetting: Gregor Mendel's fundamental experiments in plant genetics done during the 1860s were also ignored for nearly half a century (Henig 2001). Could the neglect of Verhulst's work be ascribed to Quetelet's reservations about his pupil's contributions published in the older man's eulogy after the younger man's premature death in 1849? Udny Yule had a better explanation: "Probably owing to the fact that Verhulst was greatly in advance of his time, and that the then existing data were quite inadequate to form any effective test to his views, his memories fell into oblivion: but they are classics on their subject" (Yule 1925a, 4).

The next appearance of logistic function (without using that name) was to quantify the progress of autocatalytic reactions in chemistry. While catalysis denotes the increasing rate of a chemical reaction caused by the presence of an additional element (notably, one of heavy metals) or a compound (often in minute quantities), autocatalysis describes a reaction that is catalyzed by its own products. Autocatalytic processes—reactions showing rate acceleration as a function of time followed by eventual saturation—are essential for the growth and maintenance of living systems and without them abiotic chemistry could not have given rise to replication, metabolism, and evolution (Plasson et al. 2011 Virgo et al. 2014).

After Wilhelm Ostwald (1853–1932, a leading chemist of the pre-WWI era) introduced the concept in 1890 (Ostwald 1890) it was quickly realized that the progress of the process follows a logistic function: the concentration of one reagent rises from its initial level, first slowly then more rapidly, but then, limited by the supply of the other reagent, it slows down while the concentration of the second reagent declines to zero. In 1908 T. Brailsford Robertson (1884–1930), an Australian physiologist at the University of California, noted that comparing the curve for monomolecular autocatalytic reaction with the increase of body weight of male white rats, "the resemblance between the curve of growth and that of an autocatalysed reaction is at once obvious" (figure 1.12)—but comparing the curve for autocatalyzed monomolecular reaction with the one showing the increase in body weight of a man showed that the latter trajectory has two superimposed curves (Robertson 1908, 586).

Both are sigmoid curves but Robertson did not mention Verhulst. Three years later, McKendrick and Kesava Pai (1911) used the function, again without naming Verhulst, to chart the growth of microorganisms, and in 1919 Reed and Holland (1919) made a reference to Robertson (1908) but did use the term logistic in their growth curve for the sunflower. That example of plant growth became later widely cited in biological literature on growth.

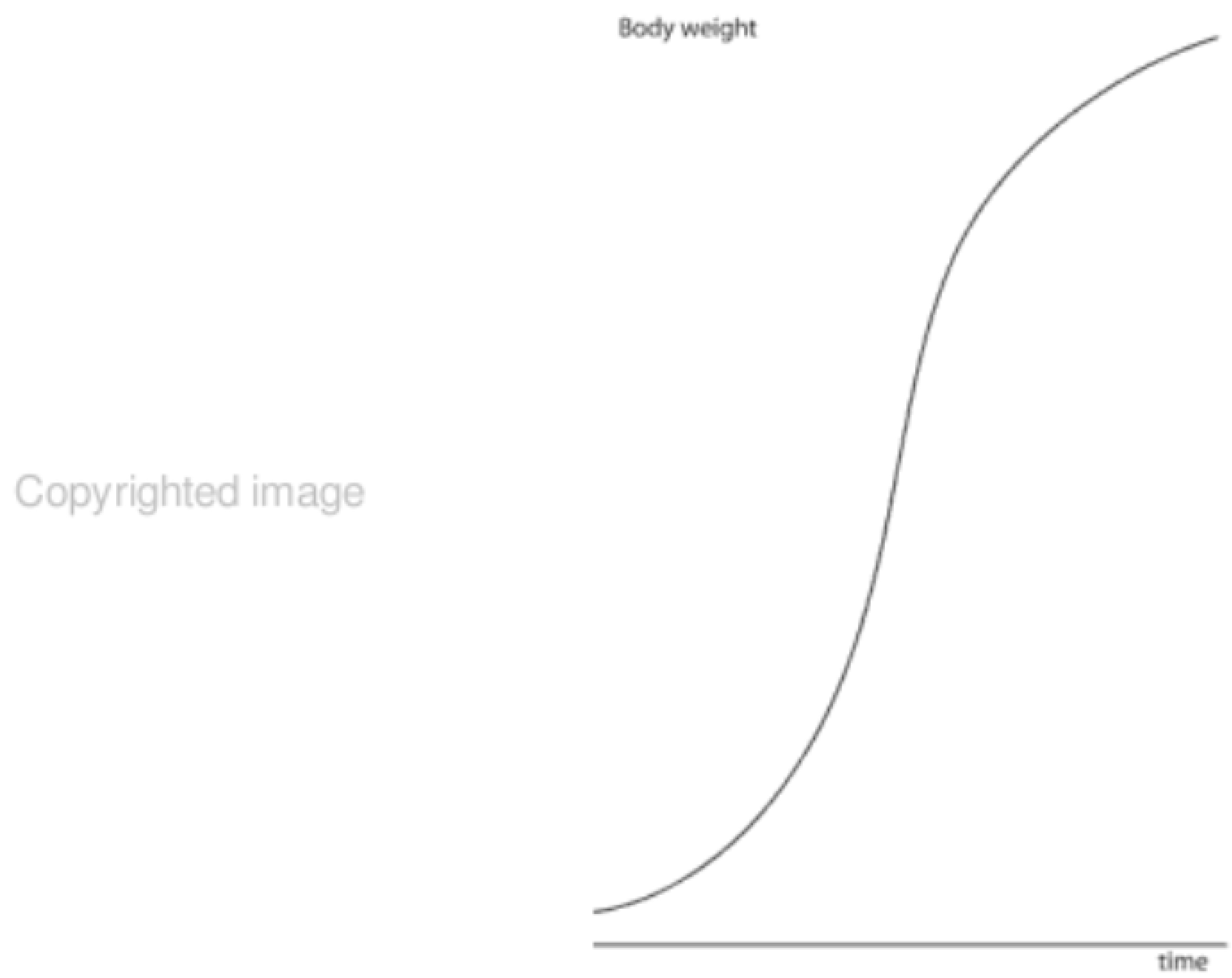


Figure 1.12
Robertson's (1908) comparison of the progress of an autocatalytic reaction with body weight increase of male white rats.

Observed growth of *Helianthus* height between planting and the 84th day follows closely a four-parameter logistic function with the inflection point falling on the 37th day (figure 1.13).

In 1920 the logistic function reappeared in demography when Raymond Pearl and Lowell Reed, professors at Johns Hopkins University, published a paper on the growth of the US population (Pearl and Reed 1920), but only two years later they briefly acknowledged Verhulst's priority (Pearl and Reed 1922). Much like Verhulst in the mid-1840s, Pearl and Reed used the logistic function to find the maximum population of the US supportable by the country's agricultural resources (Pearl and Reed 1920, 285):

> The upper asymptote … has the value 197,274,000 roughly. This means that … the maximum population which continental United States, as now areally limited, will ever have will be roughly twice the present population. We fear that some will condemn at once the whole theory because this number is not sufficiently imposing. It is so easy, and most writers on population have been so prone, to extrapolate population by geometric series, or by a parabola or some such purely empirical curve, and arrive at stupendous figures, that calm consideration of real probabilities is most difficult to obtain.

And as was the case with Verhulst's maxima for Belgium and France, Pearl and Reed also underestimated the supportable maximum of the US population. By 2018 its total had surpassed 325 million, nearly 65% above

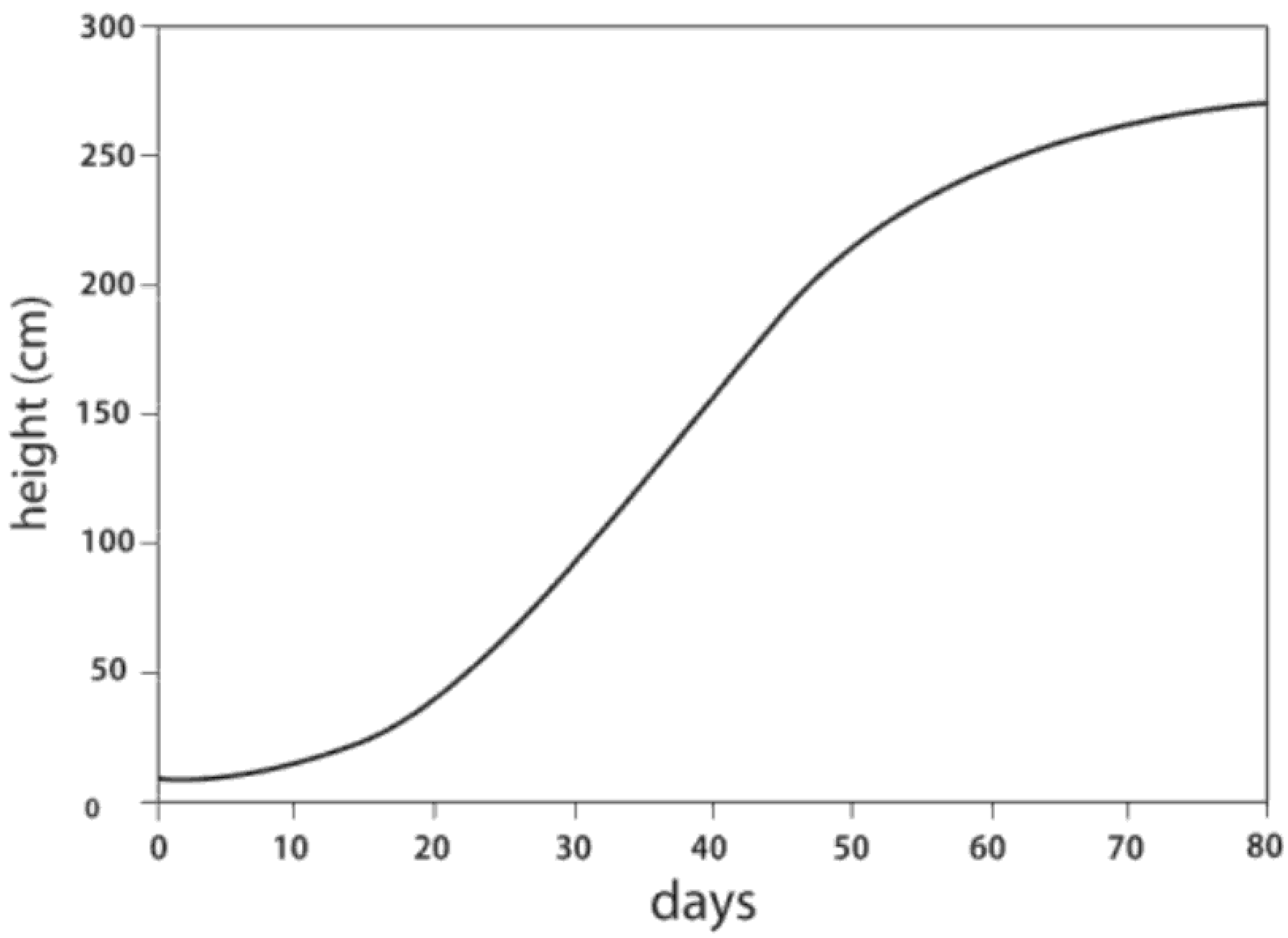


Figure 1.13
Logistic growth (inflection point at 37.1 days, asymptote at 292.9 cm) of a sunflower plant plotted by Reed and Holland (1919).

their calculation of the maximum carrying capacity (figure 1.14)—even as the country has been diverting 40% of its largest crop into corn-based fermentation of ethanol and still remains the world's largest food exporter. But Pearl had little doubt about the predictive power of his equation: in 1924 he compared the curve "in a modest way" with Kepler's law of planetary motion and with Boyle's law of gases (Pearl 1924, 585).

Applications of logistic growth function began to spread. Robertson used information about the growth of dairy cows, domestic fowl, frogs, annual plants and fruits in his voluminous survey of *The Chemical Basis of Growth and Senescence* (Robertson 1923). A year later, Spillman and Lang (1924) published a detailed treatment of *The Law of the Diminishing Returns* with many quantifications of bounded growth rates. Reed and Berkson (1929) applied the logistic function to several bimolecular reactions and to the proteolysis of gelation by pancreatin, and Bliss (1935) used it to calculate a dosage-mortality curve. And during the two decades before WWII, Pearl and his collaborators applied the logistic curve "to almost any living population from fruit flies to the human population of the French colonies in North Africa as well as the growth of cantaloupes" (Cramer 2003, 6).

In 1945 Hart published a comprehensive examination of logistic social trends with scores of examples classified as series reflecting the

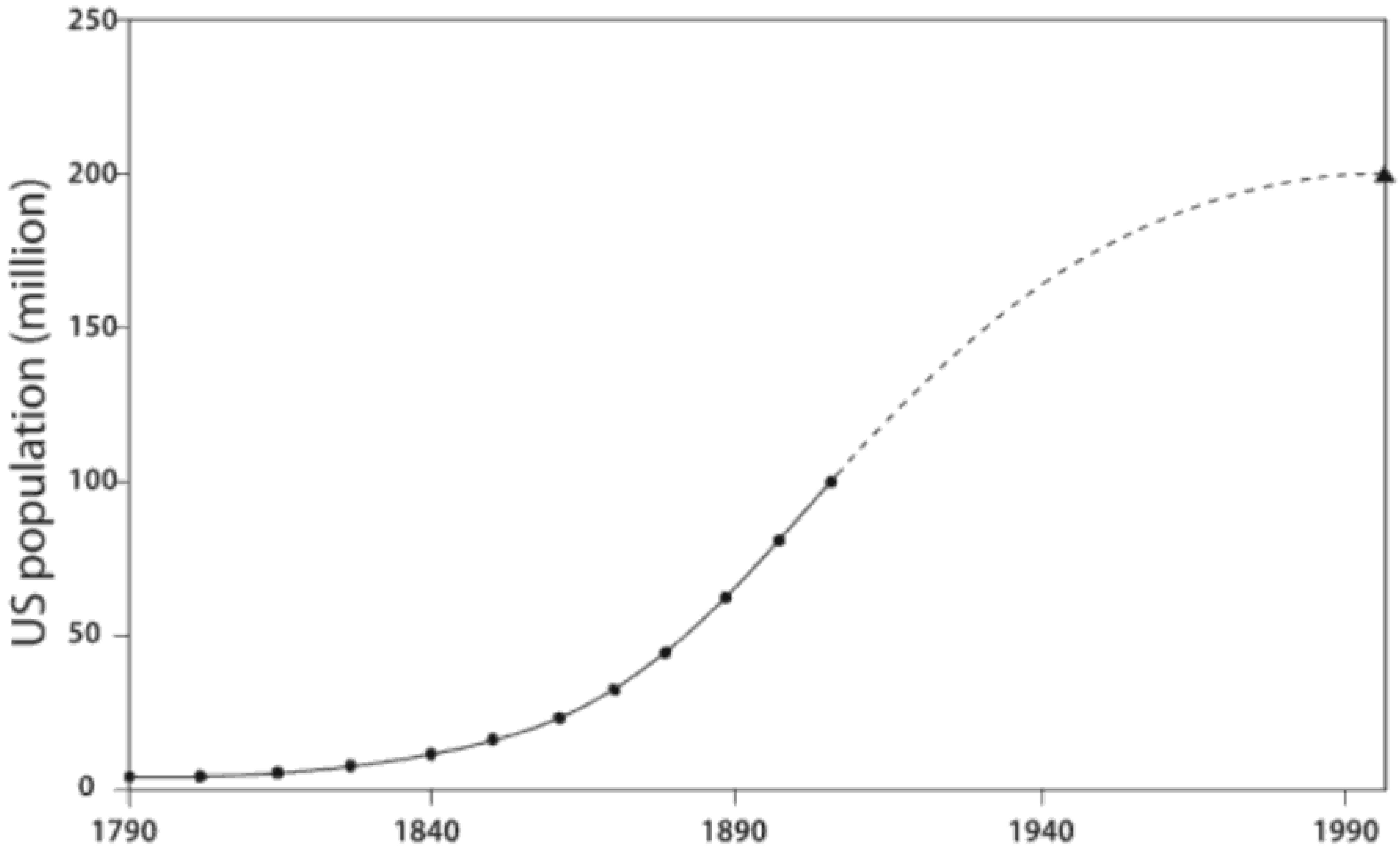


Figure 1.14
Forecast of US population growth based on the logistic curve (inflection point in 1919, asymptote at 197.3 million) fitted to decennial census data between 1790 and 1910 (Pearl and Reed 1920).

growth of specific social units (populations, cities, crop yields, output and consumption of industrial products, inventions measured by patenting, length of railways), the diffusion of specific cultural traits (school enrolments, car ownership, social and civic movements), and what he called indices of social efficiency, including life expectancy, speed records, and per capita incomes (Hart 1945). Two decades of rapid post-WWII population and economic growth driven by technical expansion were dominated by numerous instances of exponential growth, but the logistic function regained a greater prominence with the rise of modern ecological consciousness during the late 1960s and 1970s. Not surprisingly, there are many publications describing how to fit a logistic curve to data (Cavallini 1993; Meyer et al. 1999; Arnold 2002; Kahm et al. 2010; Conder 2016).

There is another fairly commonly used growth model, the Gompertz curve, whose origins are even older than the Verhulst function. The model was originally proposed in 1825 by Benjamin Gompertz (1779–1865), a British mathematician, in order to estimate the progress of human mortality (Gompertz 1825). It shares three constants, the asymptote and a fixed degree of skewness with the logistic function but, as already noted, the logistic function has the inflection point exactly halfway between the two still-growing family of sigmoid curves have been a new growth equation developed by Birch (1999) and, as already noted, a generalized logistic function by Tsoularis (2001). Birch (1999) modified Richards's equation to make it more suitable for generic simulation models, particularly for representing the growth of various plant species within mixed vegetation, while Tsoularis (2001) proposed a generalized logistic growth equation incorporating all previously used functions as special cases.

Logistic Curves in Forecasting

Logistic curves have been a favorite tool of forecasters because of their ability to capture, often very closely, growth trajectories of both living organisms and anthropogenic artifacts and processes. Undoubtedly, their use can provide valuable insights but, at the same time, I must caution against any overenthusiastic reliance on logistic curves as failure-proof forecasting tools. Noël Bonneuil's (2005, 267) verdict recalled the "golden age of the logistic curve, when Pearl enthusiastically applied the same function to any case of growth he could find, from the length of tails of rats to census data of the United States" and dismissed the claims of strikingly accurate applications of this model to historical data sets by tagging these "triumphs as shallow: most constrained growth processes do resemble the logistic, but to say so adds little understanding to dynamics in history…Curve fitting is too often an exercise that misleads in two fronts: not only should it not be taken as probative, but it can also conceal important detail."

Obviously, using these curves for long-range forecasting is no guarantee of success. Their application may be revealing and it can provide useful indications about coming limits, and throughout this book I will introduce retrospective fittings that are remarkably accurate and that may offer reliable indications of near-term growth. But in other cases, even highly accurate logistic fits of past trajectories may provide highly misleading conclusions about the coming advances and the forecasting errors may go far beyond those expected and acceptable ±10–25% deviations over a period of 10–20 years.

In one of the earliest surveys of logistic trends published at the end of WWII, Hart (1945) included speed records of airplanes between 1903 and 1938: that trajectory produces a very good logistic fit with the inflection point in 1932 and the maximum speed of close to 350 km/h—but technical innovation invalidated that conclusion twice within a dozen years. First, improvements in the performance of reciprocating engines (required to power wartime aircraft) brought their output to practical limits and they were soon adopted for commercial aviation. Lockheed L-1049 Super

Constellation, first flown in 1951, had a cruising speed of 489 km/h and maximum speed of 531 km/h, about 50% higher than the forecast of Hart's logistic ceiling.

Super Constellation became the fastest transatlantic airliner but its dominance was short-lived. The ill-fated British de Havilland Comet flew for the first time in January 1951 but was withdrawn in 1954, and the first scheduled jet-powered flight by an American company was Pan Am's Boeing 707 in October 1958 (Smil 2010b; figure 1.15). Turbojets, the first gas turbines in commercial flight, had more than doubled the pre-WWII cruising speeds of passenger aircraft (with the first service in 1919) and generated a new logistic curve with the inflection point in 1945 and asymptote around 900 km/h (figure 1.16). More powerful and more efficient turbofan engines, first introduced during the 1960s, enabled large aircraft and lower fuel consumption, but maximum cruising speeds have remained basically unchanged (Smil 2010b).

During the 1970s, it appeared that that the air speed trajectory might be raised yet again by supersonic airplanes, but Concorde (cruising at 2,150 km/h, 2.4 times faster than wide-body jetliners) remained an expensive exception until it was finally abandoned in 2003 (Glancey 2016. By 2018 several companies (Spark Aerospace and Aerion Corporation for Airbus, Lockheed Martin, and Boom Technology in Colorado) were working on designs of new supersonic aircraft and although any expectations of an early large-scale commercial operation would be highly premature, another eventual doubling of (at least some) cruising speeds cannot be excluded later in the 21st century.

One of the best extensive illustrations of excessive logistic enthusiasm is a book on predictions whose subtitle—*Society's Telltale Signature Reveal the Past and Forecasts the Future*—indicates the author's belief in the predictive efficacy of logistic fits. Modis (1992) used logistic curves to forecast trajectories of many modern techniques (ranging from the share of cars with catalytic converters to the performance of jet engines) and assorted economic and social phenomena (ranging from the growth of oil and gas pipelines to passenger air traffic). One of the agreements between data and curve that he singled out was the growth of world air traffic: he predicted that by the late 1990s it will reach 90% of the estimated ceiling. In reality, by 2017 air freight was 80% higher than in the year 2000, and the number of passengers carried annually had more than doubled (World Bank 2018).

In addition, Modis presented a long table of predicted saturation levels taken from Grübler (1990). Less than 30 years later some of these forecasts have become spectacularly wrong. A notable example of these failures is the prediction of the worldwide total of cars: their count was to reach 90% of

The plane that raised a logistic growth ceiling of the cruising speed: Boeing 707. Image from wikimedia.

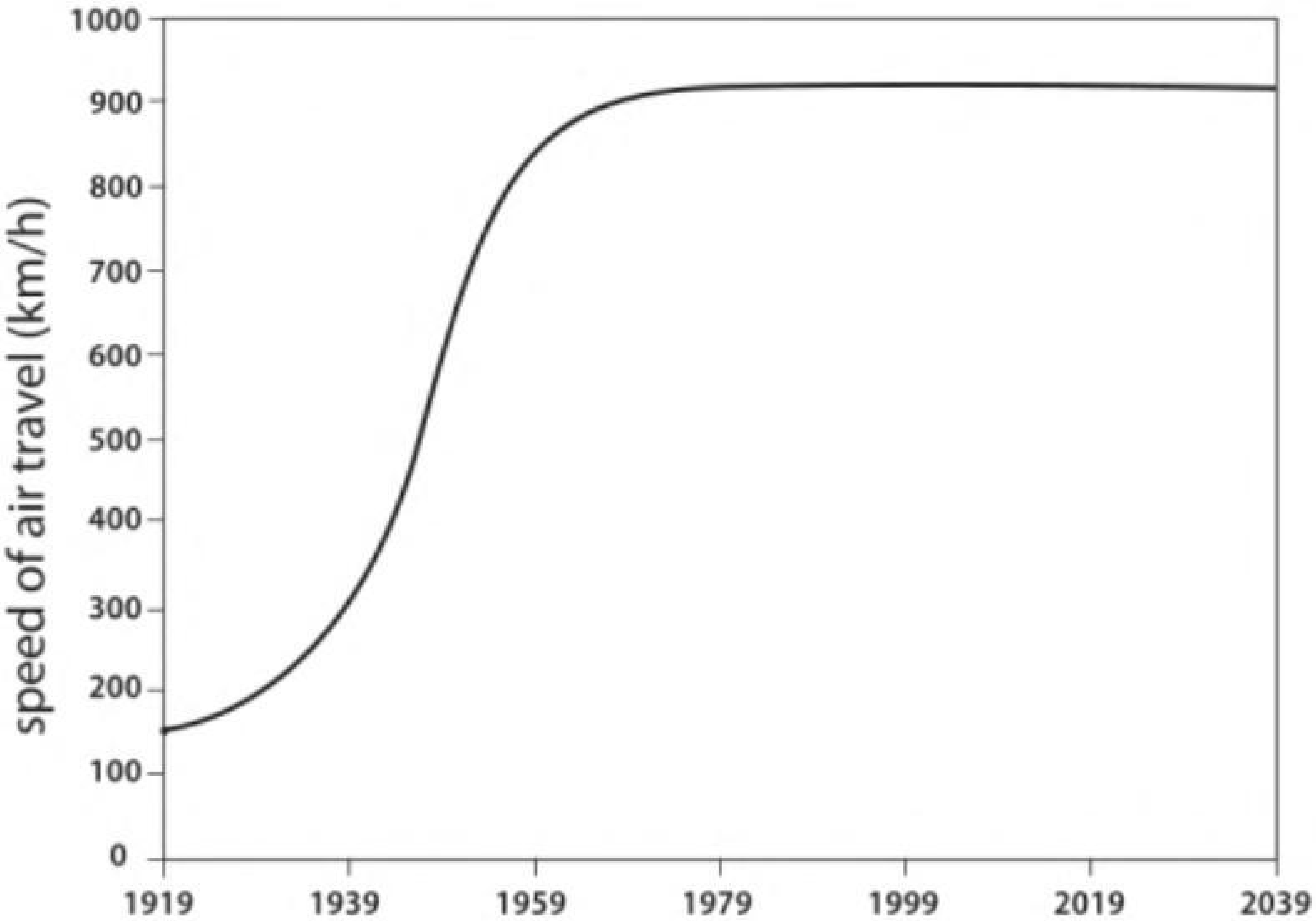


Figure 1.16
Logistic curve tracing the growth of cruising speed of commercial airliners 1919–2039 (inflection point in 1945, asymptotic cruising speed of 930.8 km/h). Plotted from data on speeds of specific airplanes, starting with KLM's de Havilland DH-16 in 1919 and ending with Boeing 787 in 2009.

the saturation level by 1988. At that time there were about 425 million car registrations, implying the eventual saturation at some 475 million cars—but one billion cars were registered by 2017, more than twice the supposed maximum, and their global count still keeps rising (Davis et al. 2018).

Marchetti (1985 and 1986b) brought the dictate of logistic growth "into one of the most defended strongholds of human ego, that of freedom, and in particular freedom in his creative acts" by concluding that "each of us has some sort of internal program regulating his output until death ... and people die when they have exhausted 90–95% of their potential" (Marchetti 1986b, figure 42). After analyzing Mozart's cumulative output, he concluded that when the composer died at 35 "he had already said what he had to say" (Marchetti 1985, 4). Modis (1992) enthusiastically followed this belief but he carried it even further.

After fitting the cumulative number of Mozart's compositions into an S-curve, Modis (1992, 75–76) claimed not only that "Mozart was composing from the moment he was born, but his first eighteen compositions were never recorded due to the fact the he could neither write nor speak well enough to dictate them to his father." And he asserted, with accuracy on the order of 1%, that this logistic fit also indicates the total potential for 644 compositions and hence Mozart's creativity was 91% exhausted when he died and, echoing Marchetti, there was "very little left for Mozart to do. His work in this world has been practically accomplished."

I wonder what Bonneuil would have to say about these verdicts! I did my own fittings, using the enduring Köchel catalogue of 626 compositions listed between 1761 and 1791 (Giegling et al. 1964). When plotting the totals in five-year intervals, a symmetrical logistic curve with the inflection point in 1780 was the best fit ($R^2=0.995$): its saturation level was at 784 compositions and it predicted the total of 759 of them by 1806 when Mozart would have turned 50 (figure 1.17a). When I entered cumulative totals for every one of Mozart's productive years, I found that the best-fitting curve ($R^2=0.9982$) was an asymmetrical (five-parameter) sigmoid that predicted the total of 955 compositions by 1806 (figure 1.17b).

But quadratic regression (second order polynomial) is also a great fit for Mozart's three decades of productivity, as is quartic (fourth order polynomial) regression (both with R^2 of 0.99) and they would predict, respectively, just over 1,200 and more than 1,300 compositions completed by 1806 (figures 1.17c and 1.17d). The verdict is clear: various curves could be found to fit Mozart's composing trajectory, but none of them should be seen as revealing anything credible about the creativity of which Mozart was deprived by his early death (or, *pace* Modis, which he would have been

Figure 1.17
Fitting Mozart's oeuvre into growth curves: symmetrical (a) and asymmetrical (b) logistic functions and quadratic (c) and quartic (d) regression have all high degrees of fit ($R^2 = 0.99$) but predict substantially different long-term outcomes for the year 1806 when Mozart (who died in 1791) would have been 50 years old. Compositions by date listed in Giegling et al. (1964).

unable to realize even if he had lived much longer). Besides, all of this misses the most obvious point of such curve-fitting exercises based on cumulative numbers of creative acts (compositions, novels, or paintings): those analyzed numbers are mere quantities devoid of any qualitative content and they do not reveal anything about the course of a creative process or about the appeal and attractiveness of individual creations.

Marchetti has been also an enthusiastic user of logistic curves in forecasting technical developments in general and composition of global primary energy demand in particular. In his studies of energy transitions, he adopted a technique developed by Fisher and Pry (1971). Originally used to study the market penetration of new techniques, it assumes that the advances are essentially competitive substitutions which will proceed to completion (that is, to capturing most of the market or all of it) in such a way that the rate of fractional substitution is proportional to the remainder that is yet to be substituted.

before the year 2000—but by 2015 more than 2.5 billion people used them daily for cooking and heating; in absolute terms, the annual demand for these fuels is nearly twice as large as a century ago; and in 2015 they supplied at least 8% of all primary energy (Smil 2017a).

Curiously, Marchetti's original analysis of primary energy shares excluded hydroelectricity: in 2015 it delivered 55% more electricity than nuclear fission. But he included a rapid ascent of a new "solar/fusion" category whose contribution was to surpass coal's share around 2020—but in 2019 there is no commercially generated fusion electricity (indeed no fusion electricity and no prospects for any early breakthrough), while in 2018 solar photovoltaics produced an equivalent of less than 0.5% of the world's primary energy supply. Obviously, the unerring internal clock has failed and all of Marchetti's supposedly immutable growth trajectories departed substantially from their predetermined schedules.

The only correct conclusion of Marchetti's analysis is that global energy substitutions unfold slowly, but his specific timing—about 100 years to go from 1% to 50% of the market, what he called time constant of the system—has been an exception rather than a rule. Only coal has done that, going from 1% just before 1800 to 50% a century later—while crude oil's global share has never reached 50%. By 2015, more than a century after it surpassed 1% of global energy supply, natural gas was still just short of 25%, while wind- and solar-generated electricity have reached, after two decades of subsidized development, just 2% of global primary energy consumption by 2016. These lessons of failed forecasts should be kept in mind whenever I use logistic fits to indicate (not to forecast!) possible future developments: some may foretell specific levels fairly well, while others may turn out to be only rough indicators, and others yet may fail as unexpected superior solutions emerge.

But which ones will surpass our expectations? Since 1900, the maximum battery energy densities rose from 25 Wh/kg for lead-acid units to about 300 Wh/kg for the best lithium-ion designs in 2018, a 12-fold gain that fits a logistic curve predicting about 500 Wh/kg by 2050 (figure 1.20). We must hope that new discoveries will vault us onto a new logistic trajectory as even 500 Wh/kg is not enough for battery-powered machines to displace all liquid derived from crude oil: the diesel fuel used to power heavy machines, trains, and ships has energy density of 13,750 Wh/kg. In contrast, another (fairly mature) logistic curve has a much higher probability to provide useful guidance: the stock of US passenger vehicles—after growing from just 8,000 to 134 million during the 20th century and to 189 million by 2015—will most likely grow by no more than about 25% by 2050.

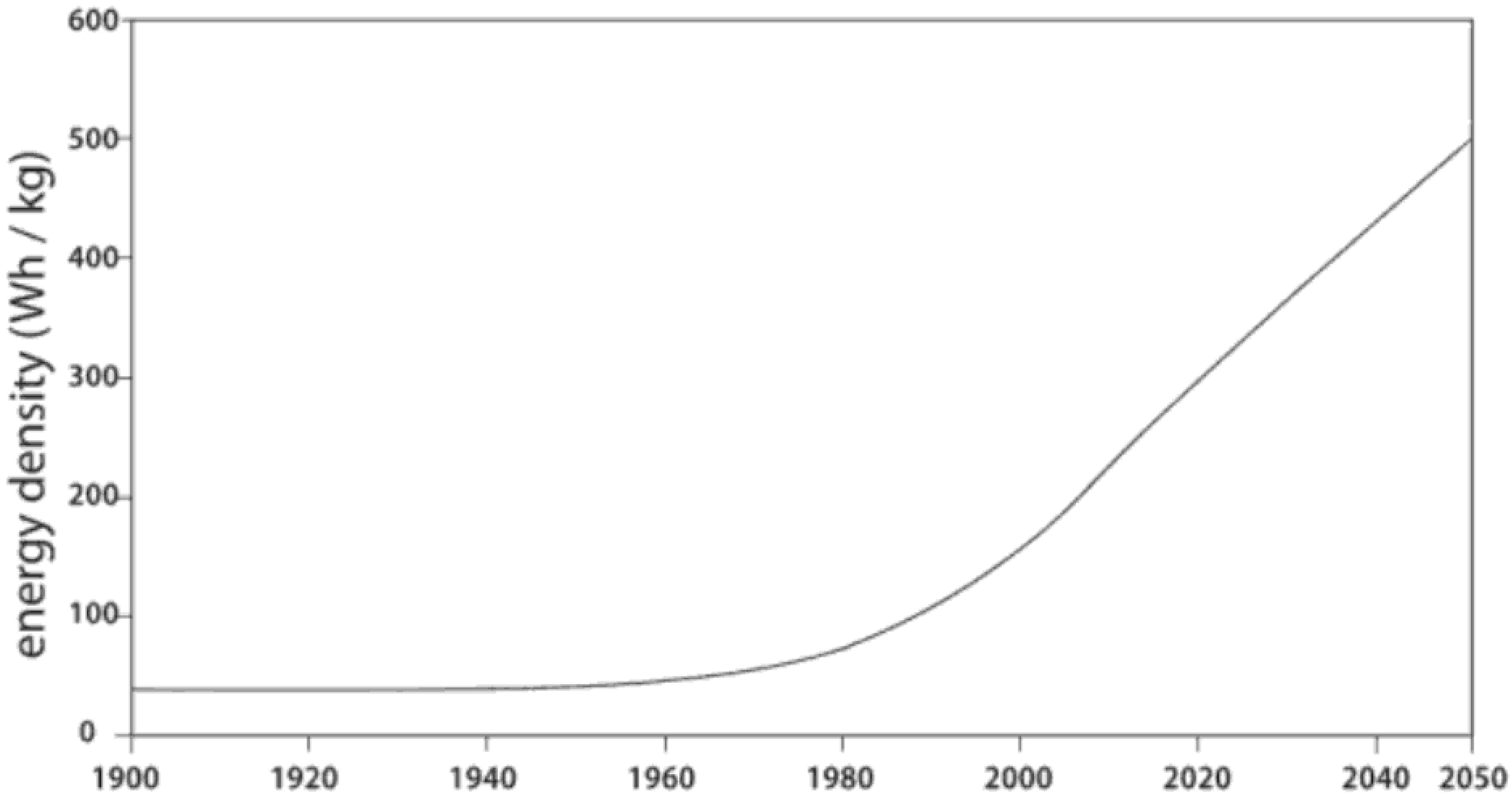


Figure 1.20
Logistic growth trajectory (inflection point in 2024, asymptote at 625.5 Wh/kg) of battery energy densities, 1900–2017. Plotted from data in Zu and Li (2011) and from subsequent news reports.

Confined Exponential Growth

Many growth phenomena do not follow S-shaped trajectories and belong to the other major class of finite growth patterns, confined exponential distributions. Unlike exponential growth, with its doubling time, these curves trace exponential decay, with its declining growth rates. Their maximum slope and curvature occur right at the beginning and hence they have no inflection point, and their concave shapes become more prominent with higher growth rates (figure 1.21). Such trajectories illustrate many phenomena of diminishing returns and are encountered with processes ranging from heat and mass transfer to tracing yield response to crop fertilization. The confined exponential function often used in these fertilizer application/crop response studies is also known as Mitscherlich equation (Banks 1994).

Confined exponential functions also capture well many diffusion processes, be it public interest in a news item, or adoption of technical innovations, often called technology transfer (Rogers 2003; Rivera and Rogers 2006; Flichy 2007). Comin and Hobijn (2004) concluded—after examining all major classes of technical innovations (including textiles, steelmaking, communications, information, transportation, and electricity) from the closing decades of the 18th century to the beginning of the 21st century—that a robust pattern of trickle-down diffusion dominates. Innovations originate

Figure 1.21
Examples of confined exponential growth curves (based on Banks 1994).

mostly in advanced economies and then get adopted elsewhere, with the quality of human capital, type of government, openness to trade, and adoption of predecessor innovations being the key factors that determine the rate of trickling down.

The spread of a technical innovation (adoption of new manufacturing processes or new prime movers), rising ownership of a new consumer product (share of families owning a microwave oven or air conditioning), or a displacement of an old product by a better version (color TV driving out black-and-white TV) are examples of diffusion processes that commonly follow a sigmoid function. But there are also instances of an immediate rapid takeoff followed by gradual slowdown, with the complete trajectory resembling a bow segment. This kind of confined exponential trajectory in innovation diffusion is also known as the Coleman model, and Sharif and Ramanathan (1981 and 1982) offered a comprehensive evaluation of binomial and polynomial innovation diffusion models.

The model applies to all situations where the population of potential adaptors (companies, customers) is both limited and constant, where all of them eventually adopt (there are no intercontinental flights powered by piston engines; there are no vacuum electronic computers) and where the diffusion proceeds independently of the number of adopters. Binomial models of confined exponential growth—limited to two variables representing the population that has already adopted an innovation and the potential adopters—have captured well such phenomena as the adoption of the fluoridation of the US water supply or the diffusion of credit card banking (Evans 2004).

Given the variety of growth processes, it is not surprising that even the two large categories of growth trajectories—S-shaped functions and confined exponential growth function—cannot subsume all variations of real-world growth. Ultimately, growth trajectories must be governed by first principles expressed through biochemical reactions, material limits, entropy change, and information decay, but actual nonlinear progressions will show irregularities and deviations from specific growth functions. As a result, some growth processes are best captured by a combination of growth functions: for example, California's post-1860 population growth followed an exponential path for 100 years until 1960 and then entered a confined exponential stage (Banks 1994). Brody (1945) found this combination useful for capturing the growth of livestock.

And the evolution of technical advances offers examples of very slow linear growth suddenly accelerating into an exponential expansion followed by confined exponential growth. And technical and economic advances get interrupted by extended performance plateaus caused by such external interventions as economic downturns or armed conflicts. Consequently, too much effort could be spent on fitting assorted growth phenomena into chosen growth models, or on seeking the "best" function for a particular growth trajectory. Doing that may have both heuristic and economic rewards—for example, a highly accurate model of body mass of an aquacultured fish species would help to optimize the consumption of relatively expensive protein feed—but this quest has been repeatedly subverted by moving the ultimate bar, that is by changing the maximum value whose level determines the trajectories of all S-shaped functions.

Staying with an aquacultural example, the growth rate of farmed salmon (produced since the late 1960s in offshore pens, now in Europe, North and South America and in New Zealand) has been doubled with the approval of AquaBounty genetically engineered fish (all sterile females) for human consumption in 2015 (AquaBounty 2017). A growth-promoting gene from Chinook salmon put into fertilized Atlantic salmon eggs makes them grow like a trout would, reaching the market weight of 2–3 kg in 18–24 months rather than in three years. The transferred gene also allows the fish to be grown in warmer waters and in complete containment.

Examples of such fundamental innovation-induced shifts in the asymptote abound, and I will cite just one more here, with many others to come in the topical chapters of this book. Waterwheels were the first inanimate prime movers to provide stationary power for tasks ranging from grain milling and water pumping and from powering blast furnace bellows to forging iron. For nearly two millennia they were wooden, and even by the early 18th century their average capacities were less than 4 kW, with only a few machines approaching 10 kW. At that point, the trajectory of waterwheel growth would have indicated future maxima of less than 100 kW—but by 1854, Lady Isabella, England's largest iron overshot wheel, reached a capacity of 427 kW (Reynolds 1970). Meanwhile, water turbines, derived from horizontal water wheels, began to make their inroads. In 1832 Benoît Fourneyron installed his first low-head (2.4 m) small capacity (38 kW) reaction turbine to power forge hammers in Fraisans, but just five years later he built two 45 kW machines with water falling more than 100 m (Smith 1980).

Other turbine designs (by James B. Francis and Lester A. Pelton) followed during the second half of the 19th century, and Viktor Kaplan patented his axial flow machine in 1920. Turbines took over from waterwheels as the prime movers in many industries but, above all, they enabled inexpensive conversion of falling water into electricity, with capacities above 1 MW by 1900, and by the 1930s, when America's largest hydro stations were built on the Columbia and Colorado Rivers, turbine capacities surpassed 100 MW. The first technical innovation, moving from wood to iron, raised the maximum power about fourfold, the second one (moving from wheels to turbines) lifted that value by an order of magnitude, and since the early 20th century it has grown by two orders of magnitude as the largest water turbines now rate 1,000 MW.

Collective Outcomes of Growth

A perceptive observer of organisms, artifacts, and achievements (be they record running speeds or average incomes) is aware that collective outcomes of their growth do not fit into a single category that could be characterized (whether almost perfectly or with satisfactory approximation) by an all-embracing mathematical function: growth of children and adolescents does not end up with the same distribution as does the growth of towns and cities. But many measured attributes do fall into two basic categories as they form either a normal distribution or as they extend over a range of values conforming (more or less closely) to one of many specific power laws. The first common category of growth outcomes includes species, objects, or properties whose distribution is dominated by a single quantity around which all individual measurements are centered and this clustering produces a typical value, and large deviations from this mean are relatively rare.

easy. But the distribution's common occurrence in nature has also led many to assume, falsely, that its fit is more universal than is justified by unruly realities.

Instances once thought to fit the normal pattern turned out to be better expressed by other distributions, and the long-standing explanation of normal distribution, invoking the central limit theorem (the sum of a numerous independent random variable tends to be normally distributed regardless of its underlying distribution) is not always (even approximately) satisfactory, while an alternative explanation, relying on the maximum entropy property, has its own problems (Lyon 2014). These caveats do not invalidate the commonality of normal distribution; they merely alert us to the fact that many distributions are more complicated than is suggested by sample averages.

After Quetelet, the normal distribution and the arithmetic mean became the norm for statistical analyses of many phenomena, but this changed once Galton (1876) and McAlister (1879) called attention to the importance of the geometric mean in vital and social statistics. Galton (1879, 367) pointed out the absurdity of applying the arithmetic mean (normal distribution) to wide deviations (as excess must be balanced by deficiency of an equal magnitude) and illustrated the point with reference to height: "the law is very correct in respect to ordinary measurements, although it asserts that the existence of giants, whose height is more than double the mean height of their race, implies the possibility of the existence of dwarfs, whose stature is less than nothing at all."

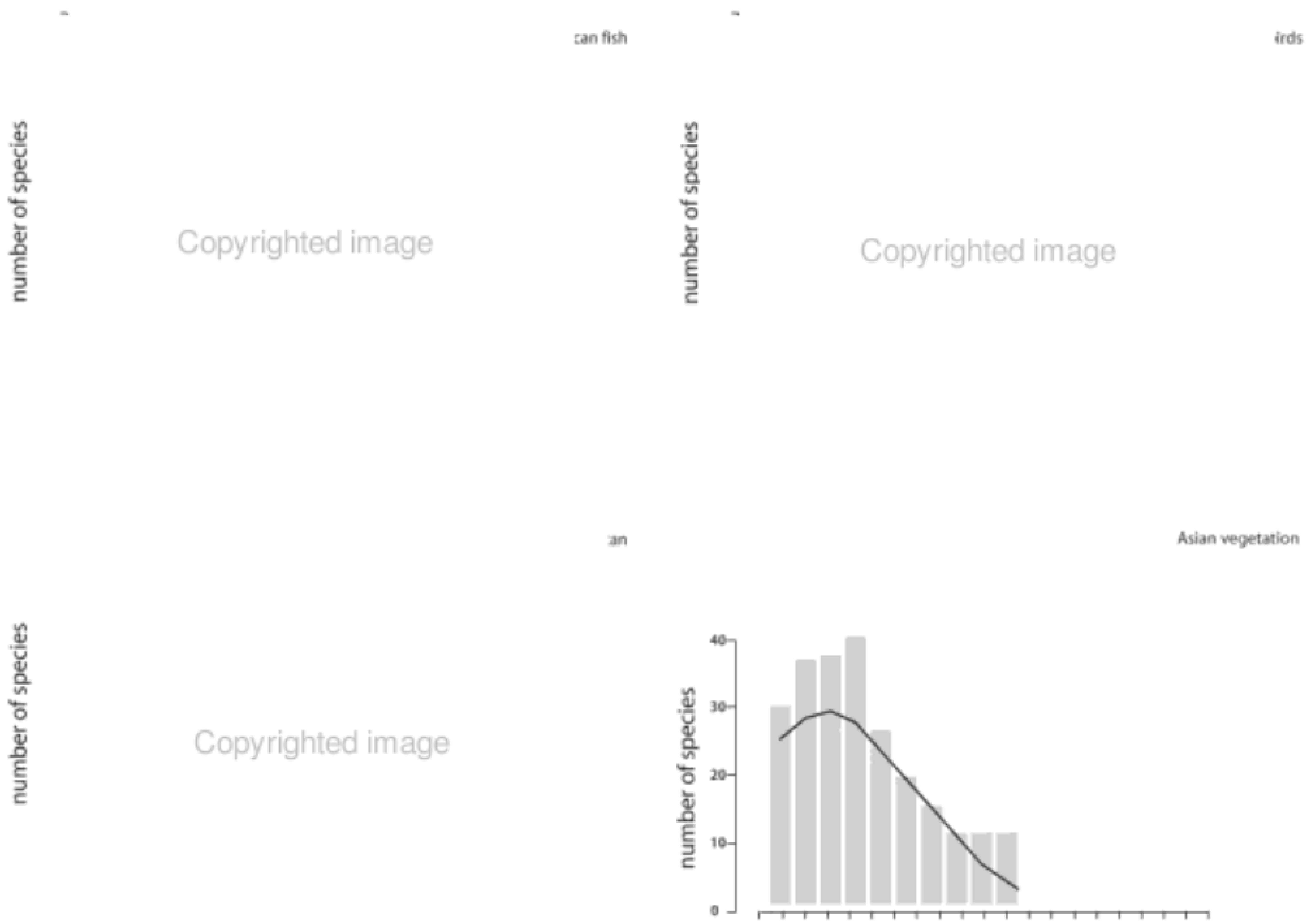
Skewed (nonnormal) distributions in nature are a common outcome of specific growth and interspecific competition. When the number of species in a community is plotted on a vertical axis and their abundance (numbers of individuals belonging to those species) are on a horizontal axis, the resulting asymmetric "hollow" curve has a long right tail—but the distribution will conform fairly closely to a normal curve when the horizontal values are expressed in decadic logarithms. Properties of this lognormal distribution have been well known since the mid-19th century: skewed to the left and characterized by its mean (or median) and a standard deviation (Limpert 2001). Lognormal distribution means that most species constituting a community will be present in moderate numbers, a few will be very rare and a few will be encountered in very high numbers.

Earlier studies of species abundance distribution (SAD) in ecosystems have identified lognormal abundance among 150 species of diatoms, hundreds of species of moths in England, Maine and Saskatchewan, and scores of species of fish and birds (Preston 1948; May 1981; Magurran 1988). Other interesting findings of lognormal SAD arising from the growth of organisms have included instances as dissimilar as airborne contamination by bacteria and fungi (Di Giorgio et al. 1996), abundance distribution of woody species in a fragment of *cerrado* forest in southeastern Brazil (Oliveira and Batalha 2005), and the length of terminal twigs on self-similar branches of Japanese elm trees (Koyama et al. 2017).

But lognormal SAD is not the norm in nature. Williamson and Gaston (2005) looked at three different distributions: the abundance of British breeding birds, the number of trees with breast-height diameter larger than 1 cm in a Panamanian forest plot, and the abundance of butterflies trapped at Jatun Sacha, Ecuador. The first two sets were complete enumerations and they showed left skew when the abundance was transformed to logarithms, while the third, incomplete, count showed right skew. They concluded that the lognormal distribution is placed uncomfortably between distributions with infinite variance and the log-binomial one, that a satisfactory species abundance distribution should have a thinner right-end tail than does the lognormal pattern, and that SAD for logarithmic abundance cannot be Gaussian.

Šizling et al. (2009) showed that lognormal-like species-abundance distributions (including the power-fraction model) cannot be universally valid because they apply only to particular scales and taxa, and the global species/range size distributions (measured in km^2) for raptors and owls are extremely right-skewed on untransformed axes, which means that when transformed they are not lognormally distributed (Gaston et al. 2005). Ulrich et al. (2010) found that completely censused terrestrial or freshwater animal communities tend to follow lognormal species abundance distributions more often than log-series or power-law types (and do so irrespective of species richness, spatial scale), but they also failed to identify a specific shape that should apply to a certain type of community and hence they strongly supported a pluralistic way of dealing with species abundances.

Baldridge et al. (2016) used rigorous statistical methods to compare different models of SAD and they found that in most cases several of the most popular choices (log-series, negative binomial, Poisson lognormal) provided roughly equivalent fits. By far the most comprehensive examination of lognormal distributions in ecosystems is by Antão et al. (2017), who analyzed 117 empirical data sets, all from intensely sampled communities, for plants, invertebrates, fish, and birds in marine, aquatic, and terrestrial habitats. They found excellent or good lognormal fits for many sets of fish, birds, and plants, but a significant share of species/abundance



Lognormal species abundance distributions (x axes in log2 classes) of North American fish and birds and less regular distributions of North American and Asian vegetation. Simplified from Antão et al. (2017).

distributions (on the order of 20%, and including both vegetation and vertebrates) can also exhibit multiple modes. This multimodality appears to increase with ecosystemic heterogeneity, that is when the examined sets include a broader spatial scale and greater taxonomic range (figure 1.24).

Another commonly examined case of a lognormal distribution has become known as Gibrat's law (or Gibrat's rule of proportional growth), named after Robert Gibrat, a French engineer who realized that the proportional growth rate of companies within an industry is independent of their absolute size (Gibrat 1931). This produces a lognormal distribution—but an overview of about 60 published analyses (Santarelli et al. 2006) found that it is impossible either to confirm the general validity of the law or to reject it systematically. The rule appears to apply only in relation to certain sectors (particularly in the services) and to the largest size classes. This heterogeneous outcome across industries and size classes precludes seeing it, despite frequent references in economic literature, as a strictly valid law. But Eeckhout (2004) concluded that the size distribution for all US cities (based on the 2000 census) is lognormal rather than fitting the most commonly assumed power-law (Zipf) model (for more on this, see the growth of cities in chapter 5).

Asymmetrical Distributions

Asymmetrical distributions are commonly encountered when analyzing many natural and anthropogenic phenomena. Many of them are applicable to outcomes that have not been created by any gradual growth processes but rather by sudden, violent releases of energy. They include the intensity of solar flares, the size of lunar craters, the magnitude of earthquakes and volcanic eruptions, and the size of forest fires. But they also apply to the magnitude of terrorist attacks, to sudden and economically crippling losses (intensity of electricity outages), as well as to the constant flow of both numerical and verbal information, including frequencies of nine digits in assemblages of numbers ranging from logarithmic tables to newspapers and cost data, and word and surname frequencies in most languages (Clauset et al. 2009).

These, often highly asymmetrical, distributions vary over wide ranges, commonly spanning many orders of magnitude. They are a common outcome of inanimate growth processes, be it the height of mountains produced by tectonic uplift and subsequent erosion or the size of islands produced by plate tectonics, erosion, coral accretion and deposition processes. There is only one Qomolangma (Mount Everest) at 8,848 m (figure 1.25), just four mountains between 8.2 and 8.6 km, 103 mountains between 7.2 and 8.2 km, and about 500 mountains higher than 3.5 km (Scaruffi 2008). Similarly, there is only one Greenland (about 2.1 million km^2) and just three other islands larger than 500,000 km^2, more than 300 islands larger than 1,000 km^2, thousands of protuberances smaller than 100 km^2, and so on.

But highly asymmetric distribution is also a common outcome among anthropogenic growth processes. Towns have grown into cities and many cities have evolved into large metropolitan areas or conurbations in every country on every inhabited continent—but in 2018, there was only one Tokyo metro area with nearly 40 million inhabitants (figure 1.25), 31 cities had more than 10 million people, more than 500 cities had surpassed 1 million, and thousands of cities were larger than 500,000 (UN 2014 and 2016). On linear scales, plots of such distributions produce curves that are best characterized either by exponential functions or by a power-law function.

A perfect power-law function (approximating the form $f(x) = ax - k$ where a and k are constant) produces a nearly L-shaped curve on a linear plot, and when both axes are converted to decadic logarithms, it produces a straight line. Obviously, neither exponential nor power-law functions can be well characterized by their modal or average values; in the real world

Figure 1.25
Peaks of two asymmetric distributions, one natural and one anthropogenic: there is only one Qomolangma and one Tokyo. Qomolangma image is available at wikimedia and Tokyo's satellite image is from NASA's Earth Observatory collection.

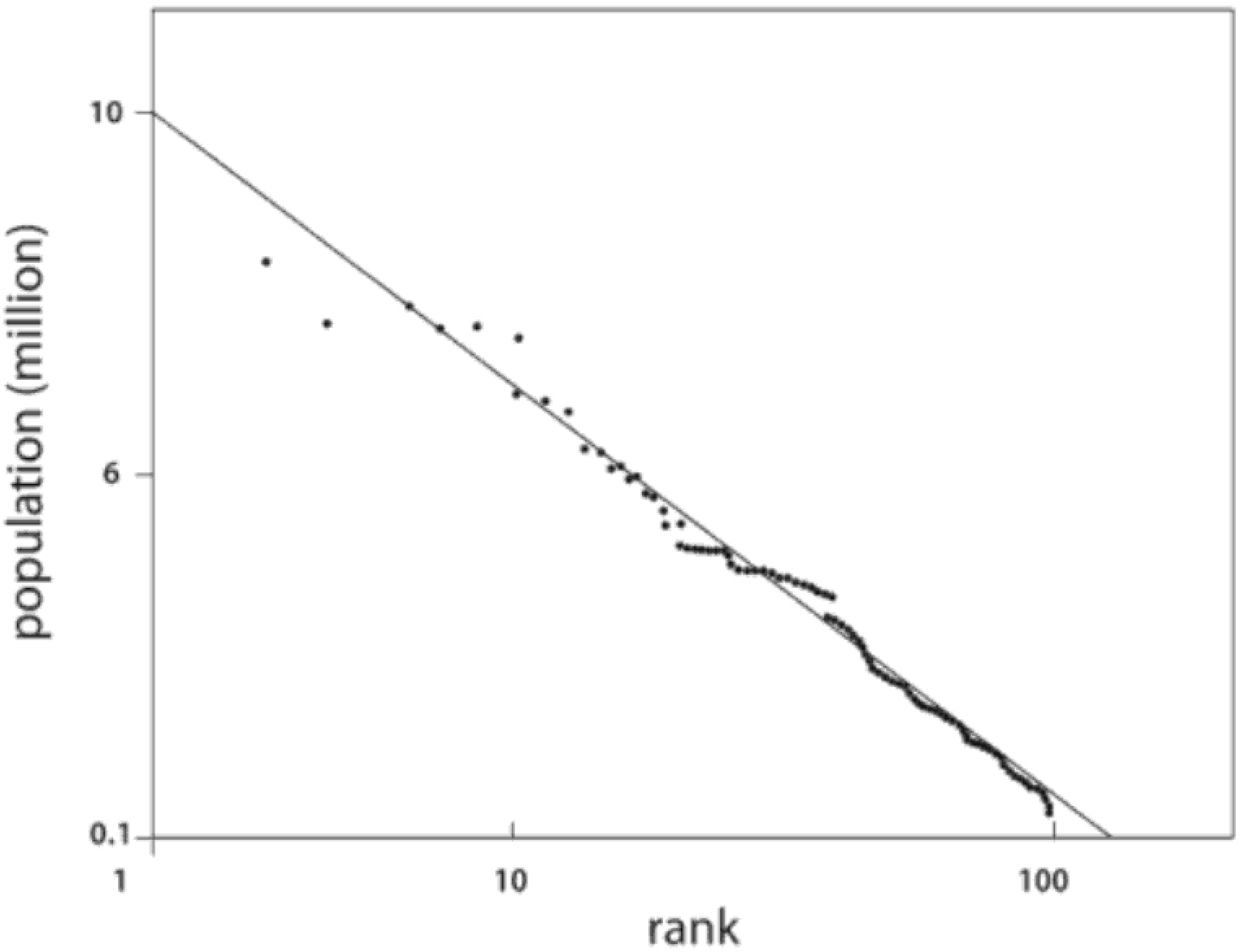


Figure 1.27 Ranking of the 100 largest US metropolitan districts based on 1940 census (Zipf 1949).

many social, economic, and physical phenomena ranging from company sizes (on national or global scales) to the characteristics of Internet traffic (Saichev et al. 2010; Pinto et al. 2012). Other inverse power laws are relatively less known in general, although they are much referred to within specific disciplines. In 1925 Udny Yule, based on the conclusions of J. C. Willis, presented almost perfect power-law frequency distribution of sizes of genera for a large family of plants (*Leguminosae*) and for two families of beetles, *Cerambycidae* and *Chrysomelidae* (Yule 1925b). In 1926 Alfred Lotka identified the inverse distribution in the frequency of scientific publications in a specific field (Lotka 1926).

In 1932 Max Kleiber, a Swiss biologist working in California, published his pioneering work on animal metabolism that challenged the nearly 50-year-old Rubner's surface law that expected animal metabolism to be proportional to two thirds of body mass (Rubner 1883; Kleiber 1932). Kleiber's law—simply stating that an animal's metabolic rate scales to the ¾ power of its mass and illustrated by the straight mouse-to-elephant line—has been one of the most important generalizations in bioenergetics. But Kleiber derived his exponent from only 13 data points (including two steers, a cow, and a sheep) and later extensive examinations have uncovered many significant departures from the ¾ power (for more, see the section on animals in chapter 2).

Jaromír Korčák called attention to the duality of statistical distribution, with the outcome of organic growth organized in normal fashion, while the distribution of the planet's physical characteristics—area and depth of lakes, size of islands, area of watersheds, length of rivers—follows inverse power law with distributions highly skewed leftward (Korčák 1938 and 1941). Korčák's law was later made better known, via Fréchet (1941), by Benoit Mandelbrot in his pioneering work on fractals (Mandelbrot 1967, 1975, 1977, 1982). But a recent reexamination of Korčák's law concluded that his ranked properties cannot be described with a single power-law exponent and hence the law is not strictly valid even for sets consisting of strictly similar fractal objects presented in his original publications (Imre and Novotný 2016).

The Gutenberg-Richter law—the second author's name is well known due to his classification system of earthquake magnitudes (Richter 1935)—relates the total number of earthquakes, *N*, to their magnitude, *M* (Gutenberg and Richter 1942). Ishimoto and Iida (1939) were the first authors to note this relationship. In the equation $N = 10^{a-bM}$ *a* indicates the activity rate (how many earthquakes of a given magnitude in a year) and *b* is usually close to 1 for interplate events but it is higher along oceanic ridges and lower for intraplate earthquakes. Quincy Wright (1942) and Lewis F. Richardson (1948) used power law to explain the variation of the frequency of fatal conflicts with their magnitude.

And Benoit Mandelbrot's pioneering studies of self-similarity and fractal structures further expanded the applications of power laws: after all, the "probability distribution of a self-similar random variable X must be of the form *Pr(X>x)=x-D*, which is commonly called *hyperbolic* or *Pareto* distribution" (Mandelbrot 1977, 320). Mandelbrot's *D*, fractal dimension, has many properties of a "dimension" but it is fractional (Mandelbrot 1967). Mandelbrot (1977) had introduced a more general power law—nearly the most general, as Gell-Mann put it—by modifying the inverse sequence, by adding a constant to the rank, and by allowing squares, cubes, square roots or any other powers of fractions (Gell-Mann 1994). Zipf's law is then just a special case with those two constants at zero. Fractal dimension equals 1 for smooth Euclidian shapes, between 1 and 2 for two-dimensional shapes—seacoast length has *D* of 1.25 (Mandelbrot 1967)—and as much as 2.9 (of possible 3) for such complex three-dimensional networks as human lungs (Turner et al. 1998).

Distributions of those collective outcomes where growth conforms to an inverse relationship have the negative exponent (constant scaling parameter) often close to one or ranging between one and three. Although power-law distributions appear to come up frequently in the studies of physical and social phenomena, it requires the deployment of standard statistical techniques to ascertain that an observed quantity does conform to a power-law distribution and that such a fit does not result merely from wishful thinking. Chen (2015) pointed out that while the inverse power function suggests a complex distribution and the negative exponential function indicates a simple distribution, a special type of the former function can be created through averaging of the latter.

And, almost always, linear fits on log-log scales do not persist across the entire range that is often spanning many orders of magnitude, but display noticeable curvatures. These heavy-tailed distributions are not exponentially bounded, and much more commonly they have a heavy right tail rather than left tail, but both tails may be heavy. Heavy-tailed scaling is obvious in the distribution of common natural events (including earthquake magnitudes, solar flux intensities, and size of wildfires), as well as information flows (distribution of computer file sizes, Web hits) and major socioeconomic phenomena resulting from population and economic growth, including the distribution of urban population and accumulated wealth (Clauset et al. 2009; Marković and Gros 2014; figure 1.28).

Jang and Jang (2012) studied the applicability of Korčák-type distribution for the size of French Polynesian islands. They found that above a certain value of island area in each sampling interval (scale), the double-log plot followed a straight line, but that it remained essentially constant below it: numbers of small islands do not vary with size. And power-law distributions are not the only ones with heavy tails: lognormal distribution and Weibull and Lévy distributions are also one-tailed, while the more complex Cauchy distribution is two-tailed. Consequently, when the sample size is limited and the data show large variance, it may be difficult to differentiate among these functions. Moreover, Laherrère and Sornette (1998) argued that a stretched exponential function (with an exponent smaller than one) provides a better fit in many commonly encountered probability distributions in nature and society and demonstrated the claim with data for French and American urban agglomerations. Tails of stretched exponential distributions are fatter than the exponential fit but much less fat than a pure power-law distribution.

Clauset et al. (2009) tested a large number of data sets describing real-world phenomena and claimed to follow power laws (figure 1.28). Their sets came from physics, earth and life sciences, computing and information sciences, engineering, and economics, with growth-related items including the numbers of distinct interaction partners in the metabolic network

Figure 1.2[illegible]
Heavy-tailed lognormal distributions of earthquake magnitudes, sizes of forest fires, cities, and the citations of academic papers. Simplified from Clauset et al. (2009).

of *Escherichia coli*, number of species per genus of mammals, and populations of US cities in the 2000 census. Their rigorous tests found that 17 of 24 data sets were consistent with power-law distribution—but, remarkably, they also concluded that the lognormal distribution could not be ruled out for any sets save one, because "it is extremely difficult to tell the difference between log-normal and power-law behavior. Indeed over realistic ranges of x the two distributions are very closely equal, so it appears unlikely that any test would be able to tell them apart unless we have an extremely large data set" (Clauset et al. 2009, 689).

Mitzenmacher (2004) came to the same conclusion as far as lognormal and power-law distributions are concerned, and Lima-Mendez and van Helden (2009) showed how an apparent power law can disappear when data are subjected to more rigorous testing. Most instances of power-law distributions do not even have strong statistical support, and any purely empirical fitting—while interesting, perhaps even remarkable—does not justify unsubstantiated suggestions of universality. Allometric scaling of metabolism is a rare exception as it finds strong statistical support across many orders of magnitude, from bacteria to whales (for details see chapter 2). And even if the law passes statistical muster, it commonly lacks a convincing generative mechanism. Long before the recent interest in complex systems and power-law scaling, Carroll (1982) listed five different classes of models that could explain city rank-size (Zipf) distributions but many of them had directly contradicted each other.

Similarly, Phillips (1999) listed 11 separate concepts to explain self-organization principles applied in earth and environmental sciences. Invariant behavior of many physical phenomena and their power-law distributions have been explained by various optimization schemes, cooperative effects, preferential attachment (most famously, rich get richer), self-similarity and fractal geometry, organized criticality, and by nonlinear dynamic behavior including multiplicative cascades (Mandelbrot 1982; Bak 1996; Pietronero et al. 2001; Yakovenko and Rosser 2009). But skepticism is in place, and on a closer examination power law does not appear to be either as ubiquitous or as fundamental as has been suggested by those who prefer to explain complex realities with simple models. Stumpf and Porter (2012, 665) considered the wide range of power-law occurrences and concluded that "the mechanistic insights are almost always too limited for the identification of power-law behavior to be scientifically useful."

But even if the statistics are convincing and even if there is empirical support for a theory explaining the generative process, "a critical question remains: What genuinely new insights have been gained by having found a robust, mechanically supported, and in-all-other-ways super power law? We believe that such insights are very rare" (Stumpf and Porter 2012, 666). Rather than demonstrating the existence of a universal principle, power laws illustrate that equifinality is common in complex open systems, as many different processes can lead to identical or very similar outcomes and hence these outcomes cannot be used to infer clear-cut causes (von Bertalanffy 1968).

What does this all mean for evaluating and understanding the distribution of many growth outcomes? Few things are incontestable: applications of power laws share the same fundamental relationship with one quantity varying as a power of another, the change being scale invariant and the direction positive or negative. Perhaps the most commonly applicable example in the first category was one of the fundamental breakthroughs in bioenergy, the discovery of metabolic scaling in animals (Kleiber 1932):

2 Nature: or growth of living matter

Perhaps the most remarkable attribute of natural growth is how much diversity is contained within the inevitable commonality dictated by fundamental genetic makeup, metabolic processes, and limits imposed by combinations of environmental factors. Trajectories of all organismic growth must assume the form of a confined curve. As already noted, many substantial variations within this broad category have led to the formulation of different growth functions devised in order to find the closest possible fits for specific families, genera or species of microbes, plants or animals or for individual species. S-shaped curves are common, but so are those conforming to confined exponential growth, and there are (both expected and surprising) differences between the growth of individuals (and their constituent parts, from cells to organs) and the growth of entire populations.

Decades-long neglect of Verhulst's pioneering growth studies postponed quantitative analyses of organismic growth until the early 20th century. Most notably, in his revolutionary book Darwin did not deal with growth in any systematic manner and did not present any growth histories of specific organisms. But he noted the importance of growth correlation—"when slight variations in any one part occur, and are accumulated through natural selection, other parts become modified" (Darwin 1861, 130)—and, quoting Goethe ("in order to spend on one side, nature is forced to economise on the other side"), he stressed a general growth principle, namely that "natural selection will always succeed in the long run in reducing and saving every part of the organization, as soon as it is rendered superfluous, without by any means causing some other part to be largely developed in a corresponding degree" (Darwin 1861, 135).

This chapter deals with the growth of organisms, with the focus on those living forms that make the greatest difference for the functioning of the biosphere and for the survival of humanity. This means that I will look at cell growth only when dealing with unicellular organisms, archaea and bacteria—but will not offer any surveys of the genetic, biochemical and bioenergetic foundations of the process (both in its normal and aberrative forms) in higher organisms. Information on such cell growth—on its genetics, controls, promoters, inhibitors, and termination—is available in many survey volumes, including those by Studzinski (2000), Hall et al. (2004), Morgan (2007), Verbelen and Vissenberg (2007), Unsicker and Krieglstein (2008) and Golitsin and Krylov (2010).

The biosphere's most numerous, oldest and simplest organisms are archaea and bacteria. These are prokaryotic organisms without a cell nucleus and without such specialized membrane-enclosed organelles as mitochondria. Most of them are microscopic but many species have much larger cells and some can form astonishingly large assemblages. Depending on the species involved and on the setting, the rapid growth of single-celled organisms may be highly desirable (a healthy human microbiome is as essential for our survival as any key body organ) or lethal. Risks arise from such diverse phenomena as the eruptions and diffusion of pathogens—be they infectious diseases affecting humans or animals, or viral, bacterial and fungal infestations of plants—or from runaway growth of marine algae. These algal blooms can kill other biota by releasing toxins, or when their eventual decay deprives shallow waters of their normal oxygen content and when anaerobic bacteria thriving in such waters release high concentrations of hydrogen sulfide (UNESCO 2016).

The second subject of this chapter, trees and forests—plant communities, ecosystems and biomes that are dominated by trees but that could not be perpetuated without many symbioses with other organisms—contain most of the world's standing biomass as well as most of its diversity. The obvious importance of forests for the functioning of the biosphere and their enormous (albeit still inadequately appreciated and hugely undervalued) contribution to economic growth and to human well-being has led to many examinations of tree growth and forest productivity. We now have a fairly good understanding of the overall dynamics and specific requirements of those growth phenomena and we can also identify many factors that interfere with them or modify their rates.

The third focus of this chapter will be on crops, plants that have been greatly modified by domestication. Their beginnings go back to 8,500 BCE in the Middle East, with the earliest domesticates being einkorn and emmer wheat, barley, lentils, peas, and chickpeas. Chinese millet and rice were first cultivated between 7,000 and 6,000 BCE and the New World's squash was grown as early as 8,000 BCE (Zohary et al. 2012). Subsequent millennia of