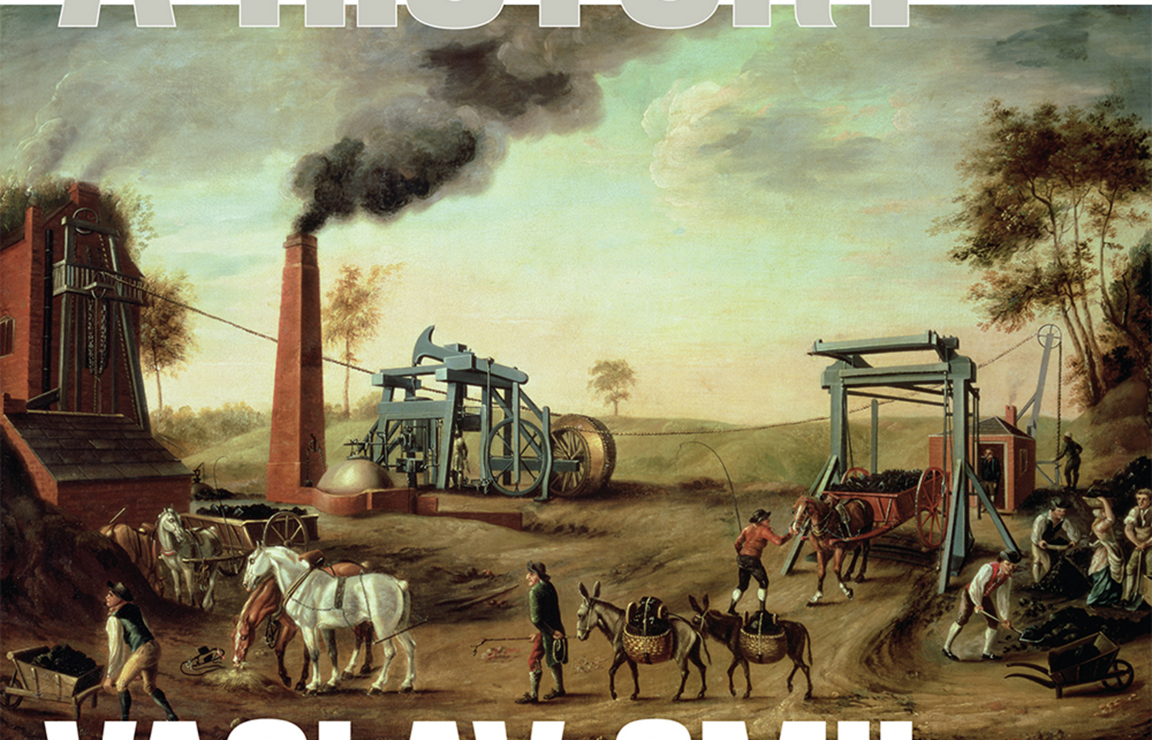
ENERGY AND CIVILIZATION

A HISTORY



VACLAV SMIL

6 Fossil-Fueled Civilization

The contrast is clear. Preindustrial societies tapped virtually instantaneous solar energy flows, converting only a negligible fraction of practically inexhaustible radiation income. Modern civilization depends on extracting prodigious energy stores, depleting finite fossil fuel deposits that cannot be replenished even on time scales orders of magnitude longer than the existence of our species. Reliance on nuclear fission and the harnessing of renewable energies (adding wind- and photovoltaic-generated electricity to more than 130-year-old hydrogeneration, and turning to new ways of converting phytomass to fuels) have been increasing, but by 2015 fossil fuels still accounted for 86% of the world's primary energy, just 4% less than a generation ago, in 1990 (BP 2016).

By turning to these rich stores we have created societies that transform unprecedented amounts of energy. This transformation brought enormous advances in agricultural productivity and crop yields; it has resulted first in rapid industrialization and urbanization, in the expansion and acceleration of transportation, and in an even more impressive growth of our information and communication capabilities; and all of these developments have combined to produce long periods of high rates of economic growth that have created a great deal of real affluence, raised the average quality of life for most of the world's population, and eventually produced new, high-energy service economies.

But the use of this unprecedented power has had many worrisome consequences and has resulted in changes whose continuation might imperil the very foundations of modern civilization. Urbanization has been a leading source of inventiveness, technical advances, gains in the standard of living, expanded information, and instantaneous communication, but it has also been a key factor behind deteriorating environmental quality and worrisome income inequality. The political implications of an uneven distribution of energy resources have both intra- and international consequences ranging from regional disparities to the perpetuation of corrupt, and often intolerant or outright violent, regimes.

Modern high-energy weapons have raised the destructive powers of nations by many orders of magnitude when compared to preindustrial capacities, and hence modern armed conflicts have seen commensurate increases not only in military but also in civilian casualties. Above all, the development of nuclear weapons has created, for the first time in history, the possibility of, if not destroying, then greatly crippling the entire civilization. At the same time, some of the most intractable means of modern aggression and warfare do not require superior command of concentrated energies as they rely on time-tested ways of individual terrorism. But even if modern civilization were guaranteed to avoid a large-scale thermonuclear conflict, it would still face profound uncertainties. Certainly the most worrisome challenge is the widespread environmental degradation. This rapid change arises from the extraction and conversion of both fossil fuels and nonfossil energies, as well as from industrial production, rapid urbanization, economic globalization, deforestation, and improper practices in crop cultivation and animal husbandry.

The cumulative effects of these changes have already gone well beyond local and regional problems to the destabilizing effects of global biospheric change, above all the many unwelcome consequences of relatively rapid global warming. Modern civilization has engineered a veritable explosion of energy use and has extended human control over inanimate energies to previously unthinkable levels. These gains have made it fabulously liberating and admirably constructive—but also uncomfortably constraining, horribly destructive, and, in many ways, self-defeating. All these changes have brought generations of strong economic growth and expectations that this process, fed by incessant innovation, need not end anytime soon—but its continuation is by no means certain.

Unprecedented Power and Its Uses

Even though interrupted by two world wars and the world's worst economic crisis during the 1930s, the growth of global energy proceeded at unprecedented rates during the first seven decades of the twentieth century. Afterward there came a slowdown precipitated by OPEC's quintupling of oil prices between October 1973 and March 1974; the growth would have moderated even without that jolt because the absolute levels had grown too large to support rates of growth that are possible at lower aggregate levels. But (at a slower pace) enormous quantitative changes have continued, and they have been accompanied by some new and remarkable qualitative gains. The best compilations of global statistics show the sustained exponential growth of fossil fuel production since the large-scale extraction of such fuels began during the nineteenth century (Smil 2000a, 2003, 2010a; BP 2015; fig. 6.1).

Coal mining grew 100-fold, from 10 Mt to 1 Gt, between 1810 and 1910; it reached 1.53 Gt in 1950, 4.7 Gt in the year 2000, and 8.25 Gt in 2015 before it declined a bit to about 7.9 Gt in 2015 (Smil 2010c; BP 2016). Crude oil extraction rose about 300-fold, from less than 10 Mt in the late 1880s to just over 3 Gt in 1988; it was 3.6 Gt in the year 2000 and almost 4.4 Gt in 2015 (BP 2016). Natural gas production rose 1,000-fold, from less than 2 Gm^3 in the late 1880s to 2 Tm^3 by 1991; it was 2.4 Tm^3 in 2000 and 3.5 Tm^3 in 2015. During the twentieth century the global extraction of fossil energies rose 14-fold in aggregate energy terms.

But a much better way to trace this expansion is to express the growth in terms of useful energy, the actually delivered heat, light, and motion. As we have already seen, early conversions of fossil fuels were rather inefficient (<2% for incandescent light, <5% for steam locomotives, <10% for thermal generation of electricity, <20% for small coal stoves), but improvements in coal-fired boilers and stoves soon doubled those efficiencies, and still left a great potential for further gains. Liquid hydrocarbons burned in household furnaces and in industrial and power plant boilers are converted with higher efficiencies, and only gasoline-fueled internal combustion engines in passenger cars are relatively inefficient. The combustion of natural gas, be it in furnaces, boilers, or turbines, is highly efficient, commonly in excess of 90%, as are the conversions of primary electricity.

Consequently, in 1900 the average weighted efficiency of global energy use was no higher than 20%; by 1950 it was more than 35%, and by the year 2015 the global mean of converting fossil fuels and primary electricity had reached 50% of total commercial inputs: the International Energy Agency (IEA 2015a) accounts for 2013 show worldwide primary supply of 18.8 Gt of oil equivalent and a final consumption of 9.3 Gt of oil equivalnet, with the highest losses, predictably, in thermal electricity generation and transportation. Even more remarkably, in a key consumption sector, household heating, entire populations have experienced a complete efficiency transition just in a matter of a few decades (box 6.1).

While the total supply of all fossil energies was up 14-fold during the twentieth century, the steady progress of efficiencies supplied more than 30 times as much useful energy as was available in 1900. As a result, affluent nations, where fossil fuel already dominated the overall supply by 1900,

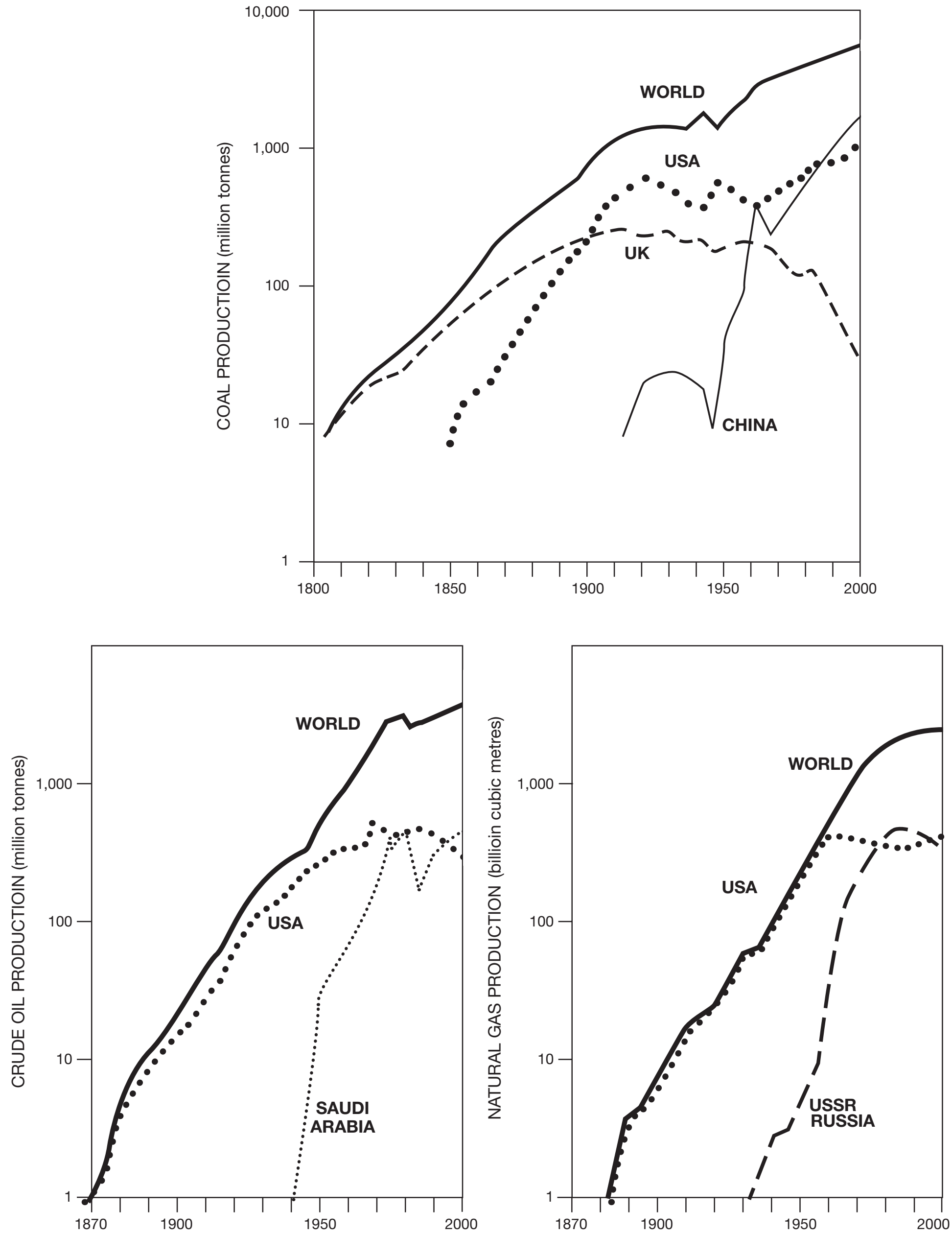


Figure 6.1

Production of the three principal fossil fuels: global totals and annual outputs for the largest producers. Plotted from data in United Nations Organization (1956), Smil (2010a), and BP (2015).

Box 6.1
Efficiency of household heating

In less than 50 years I have lived in homes heated by four different fuels and have seen the conversion efficiency of this key energy service tripled (Smil 2003). During the late 1950s, living in a village surrounded by forests close to the Czech-Bavarian border, we heated our house, as did most of our neighbors, with wood. My father ordered pre-cut logs of spruce or fir, and it was my summer duty to chop them into ready-to-stoke pieces of wood (and also into some finer pieces for kindling) and then to stack them in a sheltered place to air-dry. The efficiency of our woodstoves was not more than 30–35%. When I studied in Prague, all energy services—space heating, cooking, electricity generation—depended on lignite, and the coal stove I had in my room, in a thick-walled former monastery, had an efficiency of about 45%. After moving to the United States we rented the upper floor of a suburban house that was heated by fuel oil delivered by truck and burned in a furnace with no more than 60% efficiency. Our first Canadian house had a natural gas furnace rated at 65%, and when I designed a new super-efficient home, I installed a natural gas furnace rated at 94%—and have since replaced it with one rated at 97%.

now derive more than twice or even three times as much useful energy per unit of primary supply than they did a century ago, and because traditional biomass energies were converted with very low efficiencies (<1% for light, <10% for heat), those low-income nations where modern energies became dominant only during the latter half of the twentieth century now derive commonly five to ten times as much useful energy per unit of primary supply than they did a century ago. In per capita terms—with the global population at 1.65 billion in 1900 and 6.12 billion in the year 2000—the global increase in useful energy supply was more than eightfold, but this mean hides large national differences (more on this topic later in the chapter in the discussion of economic growth and the standard of living).

Another way to appreciate the aggregate size of modern energy flows is to compare them with traditional uses, in both absolute and relative terms. Best estimates show the worldwide total of biomass fuel consumption rising from around 700 Mt in 1700 to about 2.5 Gt in the year 2000. This would be from about 280 Mt to 1 Gt in terms of oil equivalent, less than quadrupling in three centuries (Smil 2010a). During the same time, the extraction of fossil fuels rose from less than 10 Mt to about 8.1 Gt of oil equivalent, about an 800-fold expansion (fig. 6.2). In gross energy terms, the global supply of biofuels and fossil fuels was about the same in 1900

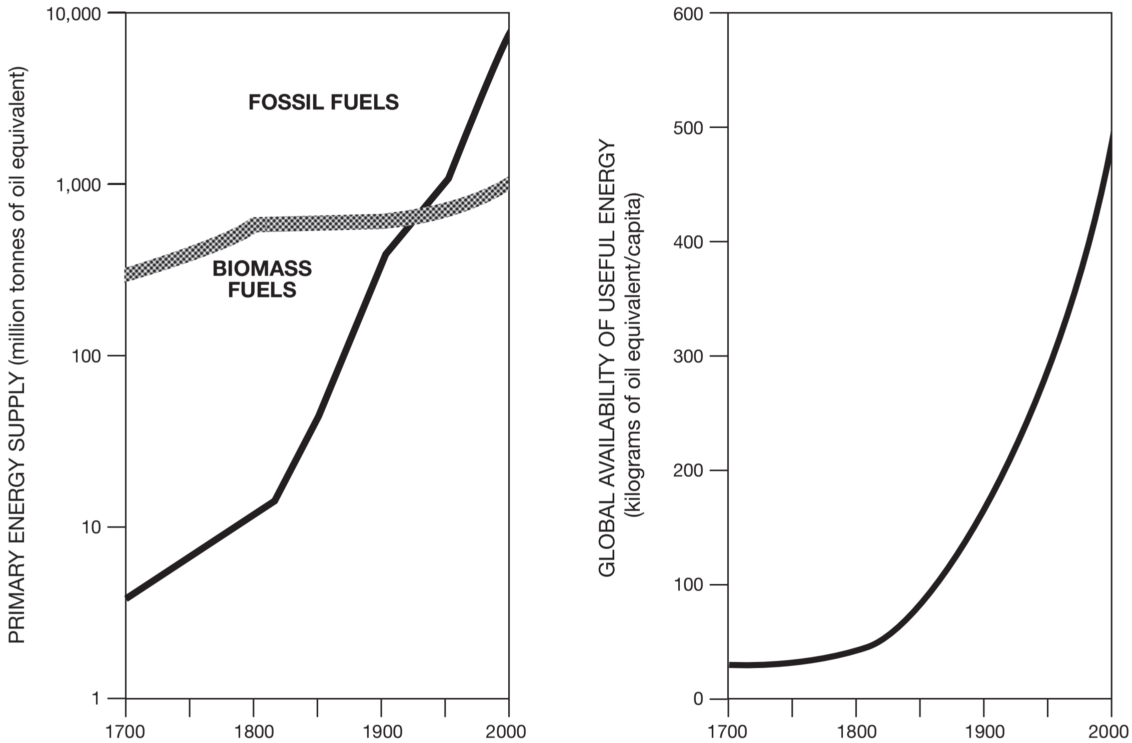


Figure 6.2
Global output of fossil fuels had surpassed the total supply of traditional biomass energies just before the end of the nineteenth century (left). Increase of useful energy was more than twice as high as the increase of the total primary supply (right). Plotted from data in United Nations Organization (1956) and Smil (1983, 2010a).

(both at roughly 22 EJ); by 1950 fossil fuels supplied nearly three times as much energy as wood, crop residues, and dung; and by the year 2000 the difference was nearly eightfold. But when adjusted for actually delivered, useful energy, the difference in the year 2000 was nearly 20-fold.

Surges in energy use raised average per capita consumption levels to unprecedented heights (fig. 6.3). The energy needs of foraging societies were dominated by the provision of food, and their annual consumption averages did not go above 5–7 GJ/capita. Ancient high cultures added slowly rising energy use for better shelters and clothing, for transportation (energized by food, feed, and wind), and for a variety of manufactures (with charcoal prominent). New Kingdom Egypt averaged no more than 10–12 GJ/capita, and my best estimate for the early Roman Empire is about 18 GJ/capita (Smil 2010c). Early industrial societies easily doubled

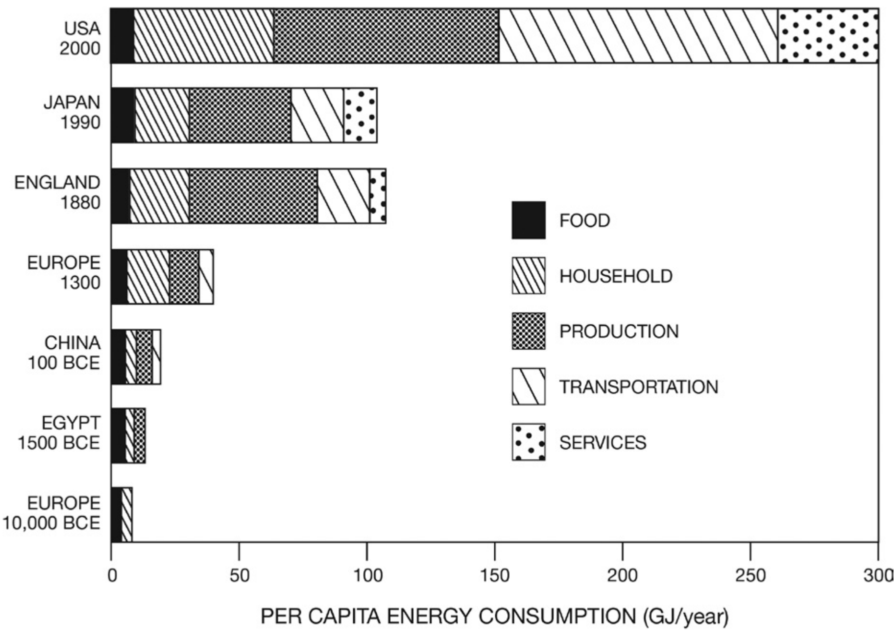


Figure 6.3
Comparisons of typical annual per capita energy consumption during different stages of human evolution. Large increases in absolute consumption have been accompanied by growing shares of energy used by households, industries, and transportation. Pre-nineteenth-century values are only approximations based on Smil (1994, 2010c) and Malanima (2013a); later figures are taken from specific national statistical sources.

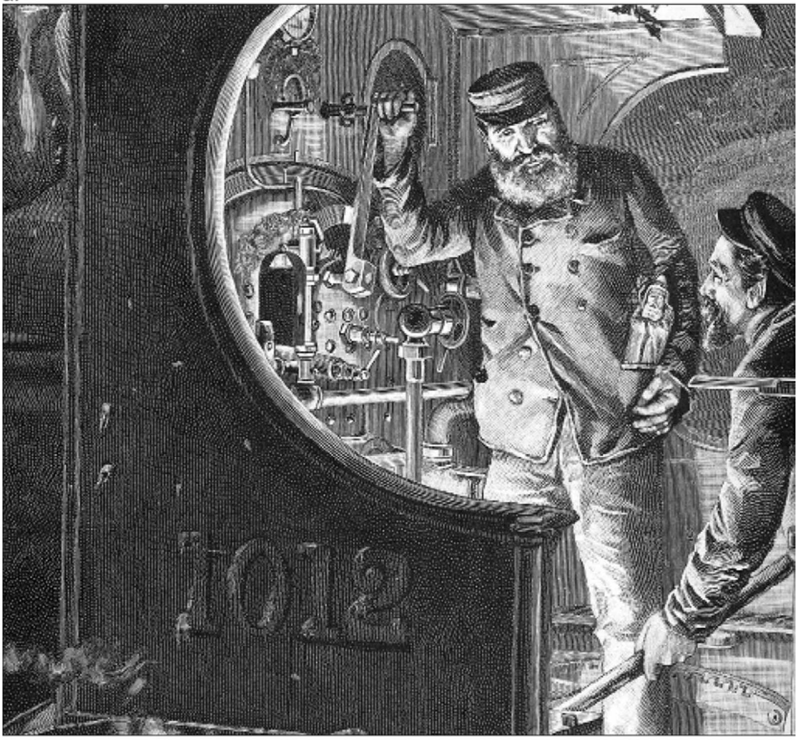
the traditional per capita energy use. Most of that increase went into coal-fueled manufactures and transportation. Malanima (2013b) put the European averages at about 22 GJ/t in 1500, followed by a stagnation at 16.6–18.1 GJ/t until 1800.

Afterward came a pronounced differentiation among the industrializing nations and countries and those whose economies remained largely agrarian. Kander (2013) puts the mean for England and Wales rising from 60 GJ/capita in 1820 to 153 GJ/capita by 1910, while during the same period the German rate more than quintupled (from 18 to 86 GJ/capita) and the French rate tripled (18 to 54 GJ/capita), but the Italian rates rose by only 20% (from 10 to just 22 GJ/capita). For comparison, the average U.S. rate rose from less than 70 GJ to about 150 GJ between 1820 and 1910 (Schurr and Netschert 1960). A century later all richer European countries were above 150 GJ/capita, the United States was above 300 GJ/capita, and as the average consumption rates rose, their composition changed (fig. 6.3).

In foraging societies food was the only energizer; my estimate puts food and fodder at about 45% of all energy in the early Roman Empire (Smil 2010c). In preindustrial Europe food and fodder ranged from 20% to 60%, but by 1820 the mean was no more than about 30%; by 1900 it was less than 10% in the UK and Germany. By the 1960s fodder energy had declined to a negligible level and food had become no more than 3% and even less than 2% of all energy supply in the most affluent societies, whose consumption became dominated by industrial, transportation, and household uses of fuels and electricity (fig. 6.3). Per capita deliveries of electricity have risen by two orders of magnitude in high-income economies, by 2010 becoming around 7 MWh/year in Western Europe and about 13 MWh/year in the United States. Contrasts between energy flows controlled directly by individuals are no less impressive.

When in 1900 a Great Plains farmer held the reins of six large horses while plowing his wheat field, he controlled—with considerable physical exertion, perched on a steel seat, and often enveloped in dust—no more than 5 kW of animate power. A century later his great-grandson, sitting high above the ground in the air-conditioned comfort of his tractor cabin, controlled effortlessly more than 250 kW of diesel engine power. In 1900 an engineer operating a coal-fired locomotive pulling a transcontinental train at close to 100 km/h commanded about 1 MW of steam power, the maximum performance permitted by manual stoking of coal (Bruce 1952; fig. 6.4). By the year 2000, pilots of a Boeing 747 retracing the transcontinental route 11 km aloft could choose an auto-mode for a large part of the journey as four gas turbines developed up to about 120 MW and the plane flew at 900 km/h (Smil 2000a).

a.



b.

Figure 6.4

Stoking a late-nineteenth-century steam locomotive (top) and piloting a Boeing jetliner (bottom). The two pilots control two orders of magnitude more power than did the stoker and the engineer in their locomotive. Locomotive from VS archive; Boeing cockpit from http://wallpapersdesk.net/wp-content/uploads/2015/08/2931_boeing_747.jpg.

This concentration of power also demands much greater safety precautions because of the inevitable consequences of errors in control. Drivers sitting atop coaches used on inter-city journeys until the nineteenth century usually controlled steady power of no more than 3 kW (four harnessed horses) deployed to transport 4–8 people; pilots of intercity jetliners control 30 MW developed by jet engines to fly 150–200 passengers. Temporary inattention or an error of judgment will obviously have vastly different consequences when an operator is in control of 3 kW or 30 MW, a four-orders-of-magnitude difference. An obvious way to cut such risks is to deploy electronic controls.

The world's safest public transportation system ever—Japan's *shinkansen* between Tokyo and Osaka, which celebrated 50 years of accident-free operation on October 1, 2014 (Smil 2014b)—centralized electronic controls from its very beginning: an automatic train control keeps a proper distance between trains and instantly engages the brakes if the speed exceeds the indicated maximum; a centralized traffic control carries out route control; and the earthquake detection system senses the very first seismic waves reaching Earth's surface and can halt or slow trains before the main shock arrives (Noguchi and Fujii 2000). Modern jetliners have been highly automated for decades, and advance controls are now becoming common in passenger cars. Electronic controls and continuous monitoring—whose penetration now ranges from room thermostats to the operation of large blast furnaces, and from anti-lock braking in cars to ubiquitous CCTV in cities—have emerged, together with the mass adoption of computers and portable electronics, as a major new category of electricity demand.

The twentieth-century growth of global electricity output was even faster than the expansion of fossil fuel extraction, whose annual average was about 3% (fig. 6.5). Less than 2% of all fuel was converted to electricity in 1900; the share was still less than 10% in 1945, but by the century's end it had risen to about 25%. In addition, new hydroelectric stations (on a large scale after World War I) and new nuclear capacities (since 1956) further expanded electricity generation. As a result, the global electricity supply went up by about 11% a year between 1900 and 1935, and by more than 9% annually thereafter until the early 1970s. For the remainder of the century the growth of electricity generation declined to about 3.5% annually, largely because of lower demand in high-income economies and higher conversion efficiencies. New ways of electricity generation from renewable sources such as solar energy and wind have shown notable advances only since the late 1980s.

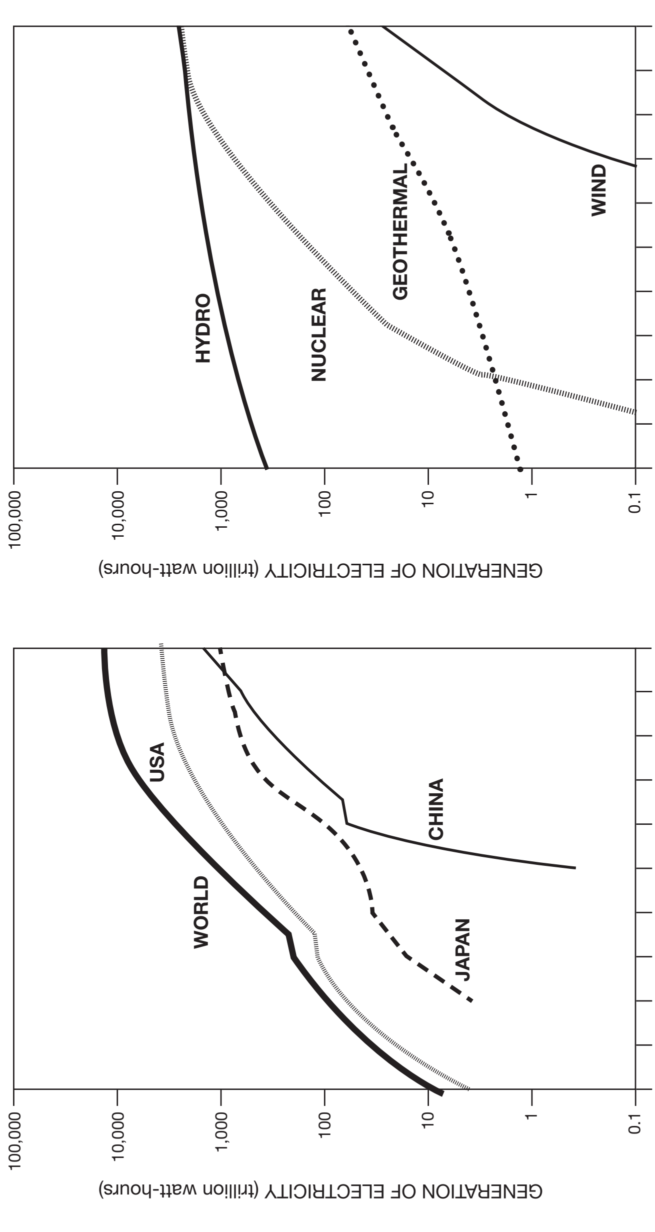


Figure 6.5
Global electricity generation has been growing considerably faster than the supply of fossil fuels. The largest economies have been always the leading producers, and thermal generation (now largely coal- and natural gas–based) continues to dominate the global output (left). Hydroelectricity and nuclear generation remain, respectively, in the second and third places, while wind and solar electricity have seen rapid post-2000 gains (right). Plotted from data in United Nations Organization (1956), Palgrave Macmillan (2013), and BP (2015).

No gain enabled by this new power has been more fundamental than the substantial rise in global food production, which has made it possible to provide adequate nutrition to nearly 90% of the world's population (FAO 2015b). No change has molded modern societies more than the process of industrialization, and no new developments have contributed more to the emergence of an interdependent global civilization than the evolution of mass transportation and the enormous expansion of our capacities for amassing information and engaging in communication with a frequency and an intensity that have no historical precedents. But these impressive gains have not been shared equally, and I will examine the extent to which the benefits of global economic growth have gone disproportionately to a minority of the world's population, and will also note considerable intranational inequities. Even so, there have been many universal improvements.

Energy in Agriculture

Fossil fuels and electricity have become indispensable inputs in modern farming. They are used directly to power machines and indirectly to build them, to extract mineral fertilizers, to synthesize nitrogenous compounds and a still expanding variety of protective agrochemicals (pesticides, fungicides, herbicides), to develop new crop varieties, and most recently to energize the electronics used in many functions that now support precision farming. Fossil fuels and electricity have brought higher and more reliable yields. They have displaced virtually all draft animals in all rich countries and greatly reduced their importance in the poor ones, and the replacement of muscles by internal combustion engines and electric motors sustained the reduction of labor started by preindustrial farming advances.

Indirect fossil fuel subsidies in agriculture began already (on a very small scale) in the eighteenth century when the smelting of ion ores was converted from charcoal to coke, expanded with widespread adoption of steel machinery during the latter half of the nineteenth century, and reached new highs with the introduction of larger and more powerful field machines, irrigation pumps, and crop-processing and animal husbandry equipment during the twentieth century. But machinery's embedded energy cost is a fraction of the energy used directly to run tractors, combines, and other harvesters, to pump water, to dry grain, and to process crops. Because of their inherently high efficiency, diesel engines have come to dominate most of these uses, but gasoline and electricity are also major energy inputs.

The use of internal combustion engines in field machinery started first in the United States, in the same decade that passenger cars finally became a mass-produced commodity (Dieffenbach and Gray 1960). The first American tractor factory was set up in 1905; power takeoff for attached implements was introduced in 1919; and power lifts, diesel engines, and rubber tires were introduced in the early 1930s. Until the 1950s mechanization proceeded much more slowly in Europe. In the populous countries of Asia and Latin America it really started only during the 1960s, and the shift is still under way in many poor countries. The mechanization of field work has been the main reason behind the rising labor productivity rise and the reduction of agricultural populations: a strong early twentieth-century Western horse worked at a rate equal to the labor of at least six men, but even early tractors had power equivalent to 15–20 heavy horses, and today's most powerful machines working on Canadian prairies rate up to 575 horsepower (Versatile 2015).

In chapter 3 I showed how rising productivity reduced average labor inputs to American wheat farming from about 30 h/t of grain in 1800 to less than 7 h/t in 1900; by the year 2000 the rate was down to about 90 minutes. Inevitably, this released labor found its way to cities, resulting in a worldwide decline of rural populations and a still continuing rise of urbanization (reviewed later in this chapter). American statistics illustrate the resulting displacements. The country's rural labor fell from more than 60% of the total workforce in 1850 to less than 40% in 1900; the share was 15% in 1950, and in 2015 it was just 1.5% (USDOL 2015). For comparison, agricultural labor in the EU is now about 5% of the total, but in China it is still around 30%.

American draft horses reached their highest number in 1915, at 21.4 million animals, but mule numbers peaked only in 1925 and 1926, at 5.9 million (USBC 1975). During the second decade of the twentieth century total draft power was about ten times as large as that of all the newly introduced tractors; by 1927 the two kinds of prime movers had equal power capacity, and the peak animal total was halved by 1940. But mechanization alone could not have released so much rural labor. Higher crop yields, brought about by new crop varieties responding to higher fertilization, applications of herbicides and pesticides, and more widespread irrigation, were also necessary.

The importance of a well-balanced supply of plant nutrients was formulated by Justus von Liebig (1803–1873) in 1843 and became widely known as Liebig's law of the minimum: the nutrient in the shortest supply will determine the yield. Of the three macronutrients (elements required in relatively large quantities)—nitrogen, phosphorus, and potassium—the latter two were not difficult to secure. In 1842 John Bennett Lawes (1814–1900) introduced the treatment of phosphate rocks by diluted sulfuric acid to produce ordinary superphosphate, and this led to discoveries of large phosphate deposits in Florida (1888) and Morocco (1913), while potash (KCl) could be mined at many sites in Europe and North America (Smil 2001).

Supplying nitrogen, the macronutrient always needed in the largest quantities per unit of cropped land, was the greatest challenge. Until the 1890s the only inorganic option was to import Chilean nitrates (discovered in 1809). Then relatively small amounts of ammonium sulfate began to be recovered from new by-product coking ovens; the expensive cyanamide process (coke reacting with lime produced calcium carbide, whose combination with pure nitrogen produced calcium cyanamide) became commercial in Germany in 1898; and at the very beginning of the twentieth century an electric arc (the Birkeland-Eyde process, 1903) was used to produce nitrogen oxide, to be converted to nitric acid and nitrates. None of these methods could supply fixed nitrogen on a mass scale, however, and the outlook for feeding the world changed fundamentally only in 1909 when Fritz Haber (1868–1934) invented a catalytic, high-pressure process to synthesize ammonia from its elements (Smil 2001; Stoltzenberg 2004).

Its rapid commercialization (by 1913) took place in the BASF plant in Ludwigshafen under the leadership of Carl Bosch (1874–1940). But the first practical use of the process was not to produce fertilizers but ammonium nitrate to make explosives during World War I. The first synthetic nitrogen fertilizers were sold in the early 1920s. The pre–World War II output remained limited, and even by 1960 more than a third of American farmers did not use any synthetic fertilizers (Schlebecker 1975). The synthesis of ammonia and subsequent conversions to liquid and solid fertilizers are energy-intensive processes, but technical advances lowered the overall energy cost and enabled the worldwide applications of nitrogenous compounds to reach an equivalent of about 100 Mt N by the year 2000, accounting for about 80% of the compound's total synthesis (box 6.2, fig. 6.6).

No other energy use offers such a payback as higher crops yields resulting from the use of synthetic nitrogen: by spending roughly 1% of global energy, it is now possible to supply about half of the nutrient used annually by the world's crops. Because about three quarters of all nitrogen in food proteins come from arable land, almost 40% of the current global food supply depends on the Haber-Bosch ammonia synthesis process.

Box 6.2
Energy costs of nitrogenous fertilizers

The energy requirements of the Haber-Bosch synthesis include fuels and electricity used in the process and energy embodied in the feedstocks. The coke-based Haber-Bosch synthesis process in BASF's first commercial plant required more than 100 GJ/t NH_3 in 1913; before World War II the rate was down to around 85 GJ/t NH_3. After 1950 natural gas–based processes lowered the overall energy cost to 50–55 GJ/t NH_3; centrifugal compressors and high-pressure reforming of steam and better catalysts lowered the requirements first to less than 40 GJ/t by the 1970s, and then to around 30 GJ/t by the year 2000, when the best plants needed only about 27 GJ/t NH_3, close to the stoichiometric energy requirement (20.8 GJ/t) for ammonia synthesis (Kongshaug 1998; Smil 2001). Typical new natural gas–based plants use 30 GJ/t NH_3, about 20% more when using heavy fuel oil, and up to about 48 GJ/t NH_3 for coal-based synthesis (Rafiqul et al. 2005; Noelker and Ruether 2011).

The average performance was about 35 GJ/t in 2015; the last rate corresponds to about 43 GJ/t N. But most farmers do not apply ammonia (a gas under normal pressure) and prefer liquids or solids, especially urea, which has the highest share of nitrogen (45%) among solid compounds that are easily applied even to small fields. Converting ammonia to urea, packaging, and transportation bring the overall energy cost to 55 GJ/t N. Using this rate as the global average means that in 2015, with about 115 Mt N used in agriculture, the synthesis of nitrogenous fertilizers claimed about 6.3 EJ of energy, or just over 1% of the global energy supply (Smil 2014b).

Stated in reverse, without Haber-Bosch synthesis the global population enjoying today's diets would have to be almost 40% smaller. Western nations, using most of their grain as feed, could easily reduce their dependence on synthetic nitrogen by lowering their high meat consumption. Populous low-income countries have more restricted options. Most notably, synthetic nitrogen provides about 70% of all nitrogen inputs in China. With over 70% of the country's protein supplied by crops, roughly half of all nitrogen in China's food comes from synthetic fertilizers. In its absence, average diets would sink to a semistarvation level—or the currently prevalent per capita food supply could be extended to only half of today's population.

The mining of potash (10 GJ/t K) and phosphates and the formulation of phosphatic fertilizers (altogether 20 GJ/t P) would add another 10% to that total. The total energy cost of other agricultural chemicals is much lower.

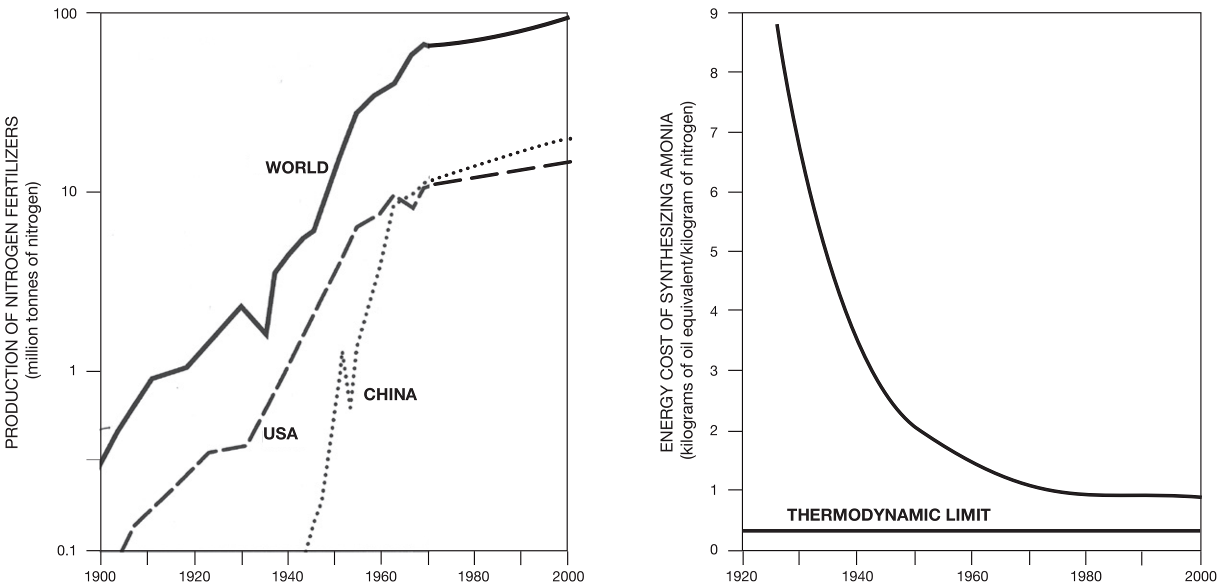


Figure 6.6
An exponential increase in the global production of nitrogenous fertilizers (left) has been accompanied by an impressive decline in the energy costs of ammonia synthesis (right). Plotted from data in Smil (2001, 2015b) and FAO (2015a).

The post–World War II growth of fertilizer applications was accompanied by the introduction and expanding use of herbicides and pesticides, chemicals that reduce weed, insect, and fungal infestations of crops. The first commercial herbicide, marketed in 1945, was 2,4-D, which kills many broadleaved plants without serious injury to crops. The first insecticide was DDT, released in 1944 (Friedman 1992). The global inventory of herbicides and pesticides now contains thousands of compounds, mostly derived from petrochemical feedstocks: their specific syntheses are much more energy-intensive than the production of ammonia (commonly >100 GJ/t, and some well above 200 GJ/t), but the quantities used per hectare are orders of magnitude lower.

The global extent of irrigated farmland roughly quintupled during the twentieth century, from less than 50 Mha to more than 250 Mha, reaching about 275 Mha by 2015 (FAO 2015a). In relative terms this means that about 18% of the world's harvested cropland is now irrigated, about half of it with water pumped mostly from wells, with about 70% of the irrigated land being in Asia. Where irrigation draws water from aquifers, the energy cost of pumping (using mostly diesel engines or electric pumps) invariably accounts for the largest share of the overall (direct and indirect) energy cost of crop cultivation. Irrigation still supplies most of the withdrawn water to furrows, but much more efficient, and more expensive, sprinklers (above all the circular center pivots) are used in many countries (Phocaides 2007).

Only approximate calculations can be made to trace the rise of the direct and indirect use of fossil fuels and electricity in modern farming. During the twentieth century, as the world population grew 3.7 times and the harvested area expanded by about 40%, anthropogenic energy subsidies soared from only about 0.1 EJ to almost 13 EJ. As a result, in 2000 an average hectare of cropland received roughly 90 times more energy subsidies than it did in 1900 (fig. 6.7). Or, skipping the numbers, we could simply say, with Howard Odum (1971, 115–116):

> A whole generation of citizens thought that the carrying capacity of the earth was proportional to the amount of land under cultivation and that higher efficiencies in using the energy of the sun had arrived. This is a sad hoax, for industrial man no longer eats potatoes made from solar energy, now he eats potatoes partly made of oil.

But this transformation has changed the global availability of food in several profound ways. In 1900 the gross global crop output (before storage and distribution losses) provided only a small margin above the average

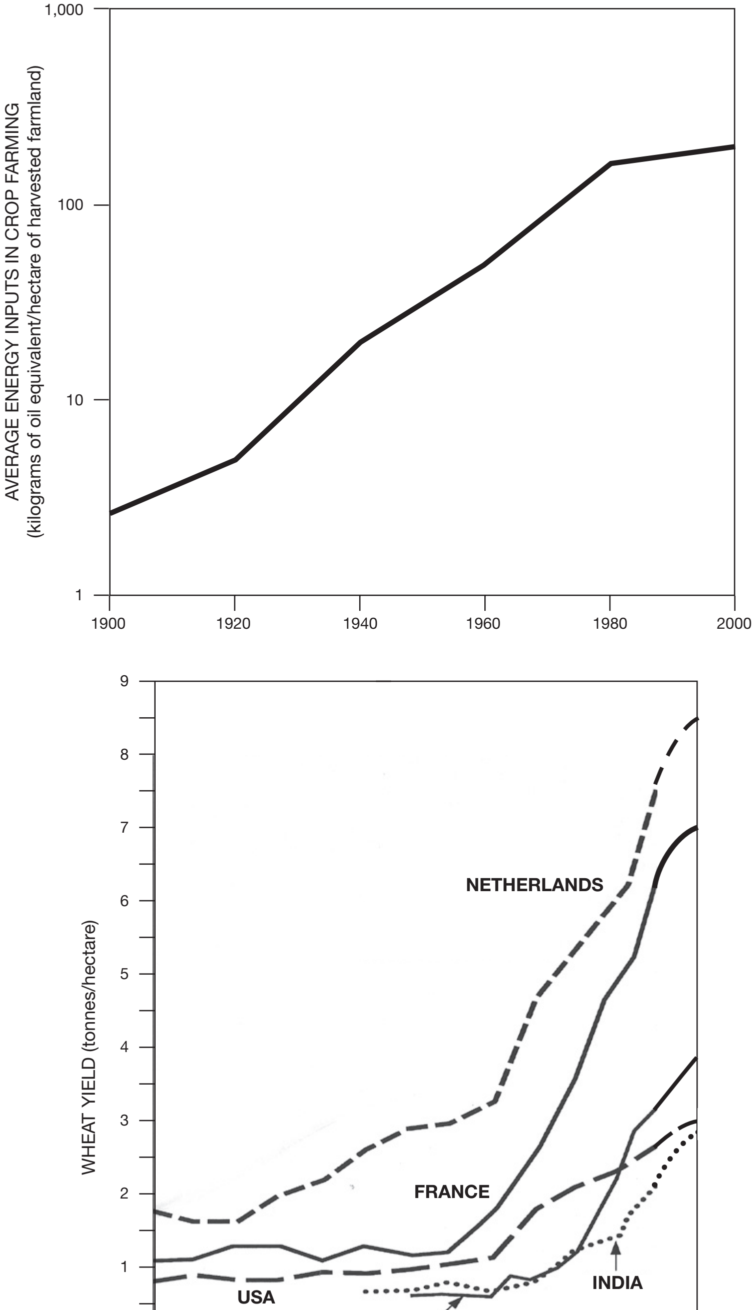


Figure 6.7

Total (direct and indirect) energy subsidies in modern farming (top), total harvests and rising wheat yields (bottom). Plotted from data in Smil (2008b), Palgrave Macmillan (2013), and FAO (2015a).

human food needs, which meant that a large share of humanity had only barely adequate or inadequate nutrition, and that the share of the harvest that could be used to feed animals was minimal. Greatly increased energy subsidies allowed the new staple cultivars (hybrid corn, introduced during the 1930s, and short-stalked wheat and rice cultivars, first adopted during the 1960s) to reach their full potential, resulting in rising yields of all staples and in an overall sixfold increase in harvested food energy (Smil 2000b, 2008).

At the beginning of the twenty-first century, the global harvest provided a daily supply averaging (for a population nearly four times as large as in 1900) about 2,800 kcal/capita, more than adequate if it were equitably accessible (Smil 2008a). The roughly 12% of the world's population that is still undernourished does not have enough to eat because of limited access to food, not because of its unavailability, and the food supply in affluent countries is now about 75% higher than the actual need, resulting in enormous food waste (30–40% of all food at the retail level) and high rates of overweight and obesity (Smil 2013a). Moreover, plenty of grain (50–60% in affluent countries) is fed to domestic animals. Chickens are the most efficient converters of feed (about three units of concentrate feed for a unit of meat); the feed:meat ratio for pork is about 9, and the production of grain-fed beef is most demanding, requiring up to 25 units of feed for a unit of meat.

This inferior ratio is also the function of the meat:live weight ratio: for chicken it is as high as 0.65 and for pork it is 0.53, but for large beef animals it is as low as 0.38 (Smil 2013d). But the energy loss in conversion to meat (and milk) has its nutritional rewards: the rising consumption of animal foods has brought high-protein diets to all rich nations (evident in taller statures) and has assured, on average, adequate nutrition even in most of the world's largest poor populous countries. Most notably, the energy content of China's average per capita diet is now, at about 3,000 kcal/day, about 10% ahead of the Japanese mean (FAO 2015a).

Industrialization

Critical ingredients of the industrialization process include a large number of interconnected changes (Blumer 1990), and this has been true at every scale of the unfolding process. By far the most important change on the factory floor was the introduction of electric motors powering individual machines, allowing precise and independent control by displacing generations of central drives transmitting the power of steam engines via leather belt and line shafts, but even this fundamental transformation would have had a limited impact if high-speed tools and better-quality steels had not been available to produce superior machines and finished components. As already noted, the intensifying international trade could not have happened without new, powerful prime movers, but their development in turn depended not only on advances in the technical design of machines but also on huge amounts of new liquid fuels being supplied, fuels produced by crude oil extraction and complex refining.

Similarly, the rising share of machine production concentrated in factories governed by hierarchical control required the location of workers near these establishments (hence new forms of urbanization) and the development of new skills and occupations (hence an unprecedented expansion of apprentice training and technical education). The utilization of a money economy and the mobility of labor and capital established new contractual relations that fostered the growth of migration and banking. Quests for mass output and low unit cost created new large markets whose functioning was predicated on reliable and inexpensive transportation and distribution.

And, contrary to common belief, the rising availability of coal-derived heat and mechanical power produced by steam engines was not necessary to initiate these complex industrialization changes. Cottage and workshop manufacturing, based on cheap countryside labor and serving national and even international markets, had been going on for generations before the beginning of coal-energized industrialization (Mendels 1972; Clarkson 1985; Hudson 1990). This proto-industrialization had a considerable presence not only in parts of Europe (Ulster, the Cotswolds, Picardy, Westphalia, Saxony, Silesia, and many more). Voluminous artisanal production for domestic and export markets had also been present in Ming and Qing China, in Tokugawa, Japan, and in parts of India.

A notable example is the carburization of wrought iron to produce Indian *wootz* steel, whose best-known transformation took the form of Damascene swords (Mushet 1804; Egerton 1896; Feuerbach 2006). Its production in some Indian regions (Lahore, Amritsar, Agra, Jaipur, Mysore, Malabar, Golconda) was on an almost industrial scale for exports to Persia and the Turkish Empire. The partially mechanized and relatively large-scale manufacturing of textiles based on water power was frequently the next step in the European transition from cottage production to centralized manufacturing. In a number of locations, industrial waterwheels and turbines competed successfully with steam engines for decades after the introduction of the new inanimate prime mover.

Nor was mass consumption a real novelty. We tend to think of materialism as a consequence of industrialization, but in parts of Western Europe, especially in the Netherlands and France, it was a major social force already during the fifteenth and sixteenth centuries (Mukerji 1981; Roche 2000). Similarly, in Tokugawa Japan (1603–1868), the richer inhabitants of cities, and especially of Edo, the country's capital, began to enjoy diversions ranging from buying illustrated books *(ehon)* to eating out (this is when sushi became popular), attending theater performances, and collecting colorful prints (*ukiyoe*) of landscapes and actors (Sheldon 1958; Nishiyama and Groemer 1997). The tastes and aspirations of increasing numbers of wealthier people provided an important cultural impetus to industrialization. They sought access to goods ranging from assortments of mundane cooking pots to exotic spices and fine textiles, from fascinating engraved maps to delicate tea services.

The term "industrial revolution" is as appealing and deeply entrenched as it is misleading. The process of industrialization was a matter of gradual, often uneven, advances. This was the case even in the regions that moved rather rapidly from domestic manufactures to concentrated large-scale production for distant markets. A spuriously accurate timing of these changes (Rostow 1965) ignores the complexity and the truly evolutionary nature of the whole process. Its English beginnings should go back at least to the late sixteenth century, but full development in Britain came only before 1850 (Clapham 1926; Ashton 1948). Even at that time traditional craftsmen greatly outnumbered machine-operating factory workers: the 1851 census showed the UK still had more shoemakers than coal miners, more blacksmiths than ironworkers (Cameron 1985).

To view the worldwide industrialization process largely as waves imitative of English developments (Landes 1969) is no less misleading. Even Belgium, whose advances resembled most closely the British progress, followed a distinct path. There was a much greater stress on metallurgy and a much lower importance of textiles. Critical national peculiarities resulted in far from uniform industrialization patterns. They included a French emphasis on water power, America's and Russia's longlasting reliance on wood, and Japan's tradition of meticulous craftsmanship. Coal and steam were initially not revolutionary inputs. Gradually they came to provide heat and mechanical power at an unprecedented level and with great reliability. Industrialization could then be broadened and accelerated at the same time, eventually becoming synonymous with an ever higher consumption of fossil energies.

Coal mining was not necessary for industrial expansion—but it was certainly critical for speeding it up. A comparison of Belgium and the Netherlands illustrates this effect. The highly urbanized Dutch society, equipped with excellent shipping and with relatively advanced commercial and financial capabilities, fell behind coal-rich, although otherwise poorer, Belgium, which became the most industrialized continental country in mid-nineteenth century Europe (Mokyr 1976). Other European regions whose coal-based economies took off early included the Rhine-Ruhr region, Bohemia and Moravia in the Habsburg Empire, and both Prussian and Austrian Silesia.

The pattern was repeated outside Western and Central Europe. In the United States, Pennsylvania, with its high-quality anthracites, and Ohio, with its excellent bituminous coal, emerged as the early leaders (Eavenson 1942). In pre–World War I Russia it was the discovery of the rich Donets coal deposits in Ukraine and the development of the Baku oil fields during the 1870s that ushered in the subsequent rapid industrial expansion (Falkus 1972). Japan's quest for modernity during the Meiji era was energized by coal from northern Kyushu. The country's first integrated modern iron and steel plant began its production only 48 years after the country's opening to the world, in 1901, with the blowing-in of the Higashida No. 1 blast furnace at Yawata Steel Works (the predecessor of Nippon Steel) in northern Kyushu (Yonekura 1994). India's largest commercial empire grew from J. Tata's blast furnace using Bihari coke in Jamshedpur beginning in 1911 (Tata Steel 2011).

Once energized by coal and steam power, traditional manufacturers could turn out larger volumes of good-quality products at lower prices. This achievement was a necessary precondition for mass consumption. The availability of an inexpensive and reliable supply of mechanical energy also allowed increasingly sophisticated machining. In turn, this led to more complex designs and greater specialization in the manufacture of parts, tools, and machines. New industries energized by coal, coke, and steam were set up to supply national and international markets with unprecedented speed. The making of high-pressure boilers and pipes started after 1810. The production of rails and railway locomotives and wagons rose rapidly after 1830, as did the making of water turbines and screw propellers after 1840. Iron hulls and submarine telegraph cables found large new markets after 1850, and commercial ways of making inexpensive steel—first in Bessemer converters after 1856, then in open-hearth (Siemens-Martin) furnaces during the 1860s (Bessemer 1905; Smil 2016)— found new large markets for finished products ranging from cutlery to rails and from plows to construction beams.

Rising fuel inputs and the replacement of tools by machines reduced human muscles to a marginal source of energy. Labor turned increasingly to supporting, controlling, and managing the productive process. The trend is well illustrated by an analysis of one and a half centuries of the England and Wales census and the Labour Force Survey (Stewart, De, and Cole 2015). In 1871 about 24% of all workers were in "muscle power" jobs (in agriculture, construction, and industry) and only about 1% were in "caring" professions (in health and teaching, child and home care, and welfare), but by 2011 caring jobs claimed 12% and muscle jobs only 8% of the labor force, and many of today's muscle jobs, such as cleaning and domestic service and routine factory line jobs, involve mostly mechanized tasks.

But even as the importance of human labor was declining, new systematic studies of individual tasks and complete factory processes demonstrated that labor productivity could be greatly increased by optimizing, rearranging, and standardizing muscular activities. Frederick Winslow Taylor (1856–1915) was the pioneer of such studies. Starting in 1880 he spent 26 years quantifying all key variables involved in steel cutting, reduced his findings to a simple set of slide-rule calculations, and drew general conclusions for efficiency management in *The Principles of Scientific Management* (Taylor 1911); a century later its lessons continue to guide some of the world's most successful makers of consumer products (box 6.3).

A radically new period of industrialization came when steam engines were eclipsed by electrification. Electricity is a superior form of energy, and not only in comparison with steam power. Only electricity combines instant and effortless access with the ability to serve very reliably every consuming sector except flying. The flip of a switch converts it into light, heat, motion, or chemical potential. Its easily adjustable flow allows a previously unsustainable precision, speed, and process control. Moreover, it is clean and silent at the point of consumption. And once proper wiring is in place, electricity can accommodate an almost infinite number of growing or changing uses—yet it requires no inventory.

These attributes made electrification of industries a truly revolutionary switch. After all, steam engines replacing waterwheels did not change the way of transmitting mechanical energy powering various industrial tasks. Consequently, this substitution did little to affect general factory layout. Space under factory ceilings remained crowded with mainline shafts linked to parallel countershafts transferring the motion by belts to individual machines (fig. 6.8). A prime mover's outage (whether caused by low water or by an engine failure) or a transmission failure (be it a line shaft crack or a slipped belt) disabled the whole setup. Such arrangements also generated large frictional losses and allowed only limited control of power at individual workplaces.

Box 6.3
From experiments with steel cutting to Japan's car exports

Frederick Winslow Taylor's main concern was with wasted labor, that is, with the inefficient use of energy—those "awkward, inefficient, or ill-directed movements of men" that "leave nothing visible or tangible behind them"—and argued for optimized physical exertion. Taylor's critics saw this as nothing but a stressful way of exploitation (Copley 1923; Kanigel 1997), but Taylor's effort was based on understanding the actual energetics of labor. He opposed excessive quotas (if the "man is overtired by his work, then the task has been wrongly set and this is as far as possible from the object of scientific management") and stressed that the combined knowledge of managers falls "far short of the combined knowledge and dexterity of workmen under them," and hence called for "the intimate cooperation of the management with the workmen" (Taylor 1911, 115).

Taylor's recommendations were initially rejected (Bethlehem Steel fired him in 1901), but his *Principles of Scientific Management* eventually became a key guide for global manufacturing. In particular, the global success of Japanese companies has been founded on a continuous effort to eliminate unproductive labor and excessive workloads, to eliminate an uneven pace of work, to encourage workers to participate in the production process by making suggestions for its improvement, and to minimize labor-management confrontation. Toyota's famous production system—an alliterative trio of *muda mura muri* (reducing non-value-adding activities, an uneven pace of production, and an excessive workload)—is nothing but pure Taylorism (Ohno 1988; Smil 2006).

The first electric motors powered shorter shafts for smaller groups of machines. After 1900 unit drives rapidly became the norm. Between 1899 and 1929 the total installed mechanical power in American manufacturing roughly quadrupled, while the capacities of industrial electrical motors grew nearly 60-fold and reached over 82% of the total available power, compared to less than 5% at the end of the nineteenth century (USBC 1954; Schurr et al. 1990). Afterward the share of electric power changed little: the substitution of steam- and direct water-powered drive by motors was practically complete just three decades after it began during the late 1890s. This efficient and reliable unit power supply did much more than remove the overhead clutter, with its inevitable noise and risk of accidents. The demise of the shaft drive freed up ceilings for the installation of superior illumination and ventilation and made possible a flexible plant design and easy

Figure 6.8
Interior of the main lathe workshop of the Stott Park Bobbin Mill in Finthswaite, Lakeside, Cumbria, showing the typical arrangement of overhead belts transmitting power from a large steam engine to individual machines. The mill produced wooden bobbins used by Lancashire's spinning and weaving industries (Corbis).

capacity expansion. The high efficiency of electric motors, combined with precise, flexible, and individual power control in a better working environment, brought much higher labor productivities.

Electrification also launched vast specialized industries. First came the manufacturing of light bulbs, dynamos, and transmission wires (after 1880) and steam and water turbines (after 1890). High-pressure boilers burning pulverized fuel were introduced after 1920; the construction of giant dams using large quantities of reinforced concrete started a decade later. The widespread installation of air pollution controls came after 1950, and the first nuclear power plants were commissioned before 1960. The rising demand for electricity also stimulated geophysical exploration, fuel extraction, and transportation. A great deal of fundamental research in material properties, control engineering, and automation was also necessary to produce better steels, other metals, and their alloys and to increase the reliability and extend the lifetime of expensive installations for extracting, transporting, and converting energies.

The availability of reliable and cheap electricity has transformed virtually every industrial activity. By far the most important effect on manufacturing was the widespread adoption of assembly lines (Nye 2013). Their classic, and now outdated, rigid Fordian variety was based on a moving conveyor introduced in 1913. The modern, flexible Japanese kind relies on just-in-time delivery of parts and on workers capable of doing a number of different tasks. The system, introduced in Toyota factories, combined elements of American practices with indigenous approaches and original ideas (Fujimoto 1999). The Toyota production system (*kaizen*) rested on continuous product improvement and dedication to the best achievable continuous quality control. Again, the fundamental commonality of all of these actions is minimizing energy waste.

The availability of inexpensive electricity has also created new metal-producing and electrochemical industries. Electricity allowed the large-scale smelting of aluminum by the electrolytic reduction of alumina (Al_2O_3) dissolved in an electrolyte, mainly cryolite (Na_3AlF_6). Starting in the 1930s electricity has been indispensable for the synthesis and shaping of an increasing variety of plastics and, most recently, for the introduction of a new class of composite materials, above all carbon fibers. The energy cost of these materials is about three times as high as that of aluminum, and their largest commercial use has been in replacing aluminum alloys in commercial aircraft construction: the latest Boeing 787 is about 80% composite by volume.

While new lightweight materials have been widely substituted for steel, steelmaking itself is increasingly done using electric arc furnaces, and new lighter but stronger steels have found many uses, particularly in the auto industry (Smil 2016). And before terminating this list, which could run for pages, I must stress that without electricity there could be no large-scale micromachining producing parts with exacting tolerances for such now ubiquitous applications as jet engines or medical diagnostic devices, and, of course, there would be neither accurate electronic controls nor the omnipresent computers and billions of telecommunication devices now in global use.

Although manufacturing's shares (as a percentage of the labor force or GDP) have been steadily declining in virtually all rich countries—in early 2015, those shares were just over 10% of workers and about 12% of the U.S. GDP (USDOL 2015)—industrialization continues, but its configuration has changed. Mass flows of energy and materials will remain at its foundation;

metals remain quintessential industrial materials; and iron, now used mostly in many kinds of steel, retains its dominance among metals. In 2014 steel production was nearly 20 times larger than the combined total output of the four leading nonferrous metals, aluminum, copper, zinc, and lead (USGS 2015). The smelting of iron ore in blast furnaces, followed by steelmaking in basic oxygen furnaces, and the use of recycled steel in electric arc furnaces dominate steel production. The massive growth of steel production would have been impossible without the much larger and more efficient blast furnaces (box 6.4, fig. 6.9).

Similarly, steelmaking techniques have become more efficient, not only because of reduced energy use but also because of the rising product yield (Takamatsu et al. 2014). The early Bessemer converters turned first

Box 6.4
Growth and mass and energy balances of blast furnaces

Few production structures with a medieval pedigree remain as important for the functioning of modern civilization do blast furnaces. As noted in chapter 5, Bell's 1840 redesign quintupled their internal volume, bringing it to 250 m^3. By 1880 the largest furnace surpassed 500 m^3; it reached 1,500 m^3 by 1950, and by 2015 the record inner volumes were between 5,500 and 6,000 m^3 (Smil 2016). The resulting increases in productivity brought the output of hot metal from 50 t/day in 1840 to more than 400 t/day by 1900. The 1,000 t/day mark was approached before World War II, and today's largest furnaces produce around 15,000 t/day, with the record rate at POSCO's Pohang 4 furnace, in South Korea, about 17,000 t/day.

The mass and energy flows needed to operate large blast furnaces and associated oxygen furnaces are prodigious (Geerdes, Toxopeus, and Van der Vliet 2009; Smil 2016). A blast furnace producing daily 10,000 t/day of iron and supplying an adjacent basic oxygen furnace will need 5.11 Mt of ore, 2.92 Mt of coal, 1.09 Mt of flux materials, and nearly 0.5 Mt of steel scrap. A large integrated steel mill thus receives every year nearly 10 Mt of materials. Modern furnaces now produce hot metal continuously for 15–20 years before their refractory brick interior and their carbon hearth are relined. These productivity gains have been accompanied by declines in specific coke consumption. In 1900 typical coke requirements were 1–1.5 t/t of hot metal, while by 2010 nationwide rates were about 370 kg/t in Japan and less than 340 kg/t in Germany (Lüngen 2013). The energy cost of coke-fueled iron smelting thus fell from about 275 GJ/t in 1750 to about 55 GJ/t in 1900, close to 30 GJ/t in 1950, and between 12 and 15 GJ/t by 2010.

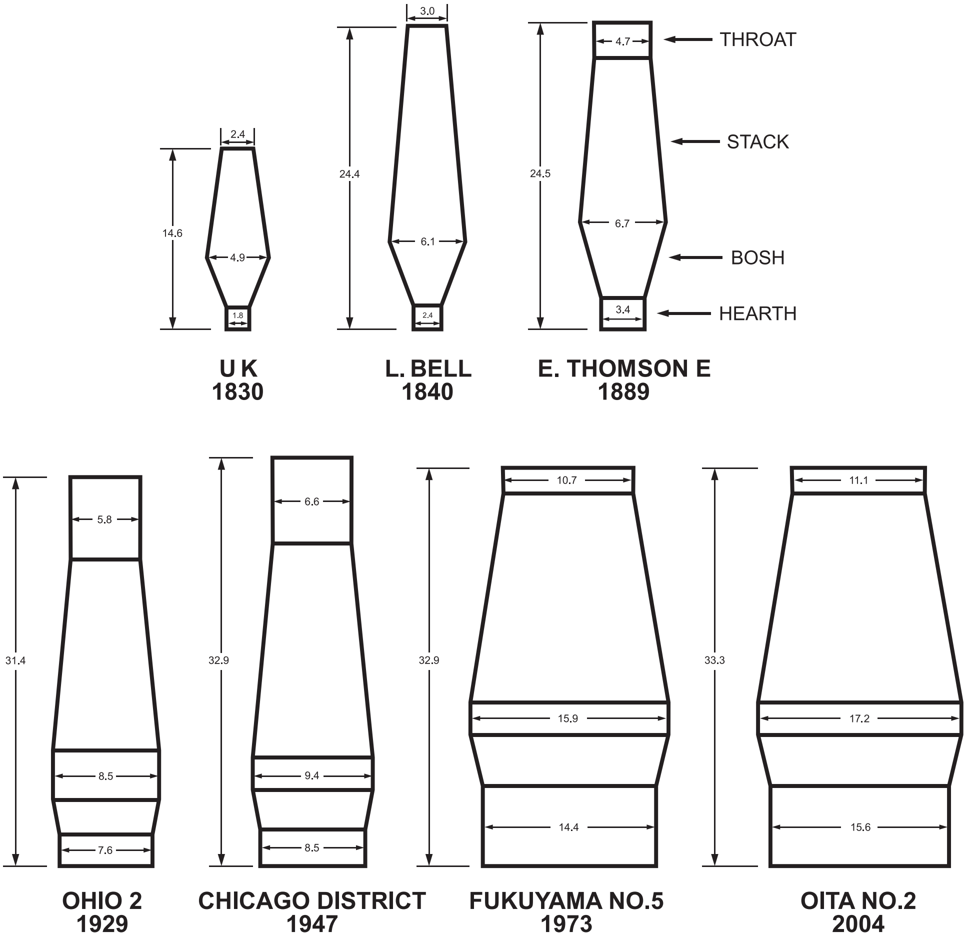


Figure 6.9

Changing designs of blast furnaces, 1830–2004. The principal trends have included taller and wider stacks, larger hearths, and lower and steeper boshes. The largest furnaces now produce more than 15,000 t of hot metal a day. Reproduced from Smil (2016).

less than 60% and later just above 70% of iron into steel. Open-hearth furnace eventually converted about 80%, and today's best basic oxygen furnaces, first introduced during the 1950s, yield as much as 95%, with electric arc furnaces converting up to 97%. And electric arc furnaces now consume less than 350 kWh/t kWh/t of steel, compared to more than 700 kWh/t in 1950; moreover, these gains have been accompanied by reduced emission rates: between 1960 and 2010 specific U.S. rates (per tonne of hot metal) fell by nearly 50% for CO_2 emissions and by 98% for dust emissions (Smil 2016). The energy cost of steel has been further lowered by continuous casting of the hot metal. This innovation supplanted the traditional production of ingots, which required reheating before further processing.

The resulting production increases have been large enough to translate into order-of-magnitude gains even in per capita terms: in 1850, before the beginning of modern steel production, fewer than 100,000 t of the metal were produced annually in artisanal ways, a mere 75 g/year/capita. In 1900, with 30 Mt, the global mean was 18 kg/capita; in the year 2000, with 850 Mt, the mean rose to140 kg/capita; and by 2015, with 1.65 Gt, it reached about 225 kg/capita, roughly 12 times the rate in 1900. My calculations show that in 2013 the worldwide production of iron and steel required at least 35 EJ of fuels and electricity, or less than 7% of the total of the world's primary energy supply, making it the world's largest energy-consuming industrial sector (Smil 2016). This compares to 23% for all other industries, 27% for transportation, and 36% for residential use and services. But if the sector's energy intensity had remained the same as in the 1960s, then the industry would have consumed at least 16% of the world's primary energy supply in 2015, an impressive illustration of continuing efficiency improvements.

By far the most important innovation in nonferrous metallurgy was the development of aluminum smelting. The element was isolated in 1824, but an economical process for its large-scale production was devised only in 1866. The independent inventions of Charles M. Hall in the United States and P. L. T. Héroult in France were based on electrolysis of aluminum oxide. The minimum energy needed to separate the metal is more than six times higher than that needed to smelt iron. Consequently, aluminum smelting advanced only slowly even after the beginning of large-scale electricity generation. During the 1880s specific electricity requirements were more than 50,000 kWh/t of aluminum, and subsequent steady improvements of the Hall-Héroult process lowered this rate by more than two-thirds by 1990 (Smil 2014b).

Aluminum's uses expanded first with advancing aviation. Metal bodies displaced wood and cloth during the late 1920s, and the demand rose sharply during World War II for the construction of fighters and bombers. Since 1945 aluminum and its alloys have become a substitute for steel wherever the design has required a combination of lightness and strength. These uses have ranged from automobiles to railway hopper cars to space vehicles, but this market is now also served by new lightweight steel alloys. And since the 1950s titanium has been replacing aluminum in high-temperature applications, above all in supersonic aircraft. Its production is at least three times as energy-intensive as aluminum's production (Smil 2014b).

Though the fundamental importance of mass-produced metals is often overlooked in a society preoccupied with the latest electronic advances, there is no doubt that modern manufacturing has been transformed by its continuing fusion with modern electronics, a union that has greatly enhanced available design options, introduced unprecedented precision controls and flexibility, and changed marketing, distribution, and performance monitoring. An international comparison showed that in the United States in 2005 services purchased by manufacturers from outside firms were 30% of the value added to finished goods, with similar shares (23–29%) in major EU economies, while in 2008 service-related occupations added up to a slight majority (53%) of all jobs in the U.S. manufacturing sector, to 44–50% in Germany, France, and the UK, and to 32% in Japan (Levinson 2012). And though many products do not look that different from their predecessors, they are actually very different hybrids (box 6.5).

Cars are just one prominent example of an industry that now finds research, design, marketing, and servicing no less important than the actual production of goods. Even if a specific embedded energy use (per vehicle, computer, or a production assembly) has increased (owing to the use of more energy-intensive materials, a larger mass, or better performance), remained the same, or declined, concerns other than a preoccupation with produced quantity have become very important, chief among them appearance, brand distinction, and quality considerations. This trend has major implications both for future energy use and for the structure of labor force, but not necessarily in any simple, unidirectional way (for more on this topic, see chapter 7).

Transportation

Several attributes apply to all forms of fossil-fueled, or electrified, transport. In contrast to traditional ways of moving people and goods they are much faster, often almost incredibly so: every year tens of millions of people now cross the Atlantic in 6–8 hours, though a century ago a crossing took nearly six days (Hugill 1993) and half a millennium ago the first crossing took five weeks. Transport conveyances are also incomparably more reliable: even the best coaches drawn by the strongest horse teams found it challenging to cross Alpine passes, succumbing to broken axles, crippled animals, and blinding storms; now hundreds of flights daily overfly the range and trains speed through deep tunnels. As for the expense, just before World War I the cost of a transatlantic crossing averaged $75 (Dupont, Keeling, and Weiss

Box 6.5
Cars as mechatronic machines

There is no better example of the fusion of mechanical and electronic components than a modern passenger car. In 1977 GM's Oldsmobile Toronado was the first production car with an electronic control unit (ECU) to govern spark timing. Four years later GM had about 50,000 lines of engine control software code in its domestic car line (Madden 2015). Now even inexpensive cars have up to 50 ECUs, and some premium brands (including the Mercedes-Benz S class) have up to 100 networked ECUs supported by software containing close to 100 million lines—compared to 5.7 million lines of software needed to operate the F-35, the U.S. Air Force's joint Strike Fighter, or 6.5 million lines for the Boeing 787, the latest model of the company's commercial jetliners (Charette 2009).

Car electronics are getting more complex, but comparing lines of code is a misleading choice. The main reason for bloated software in cars is to cover the excessive number of options and configurations offered with luxury models, including those for infotainment and navigation that have nothing to do with actual motoring; there is a great deal of re-used, auto-generated, and redundant code. Even so, electronics and software now represent up to 40% of the cost of premium vehicles: cars have been transformed from mechanical assemblies into mechatronic hybrids, and every addition of a useful control function—such as a lane-departure warning, automatic braking to a avoid rear-end collision, or advanced diagnostics—expands the software requirements and adds to the cost of a vehicle. The trend has been clear, but completely autonomous, self-driving vehicles are not coming as soon as many uncritical observers believe.

2012), or about $ 1,900 in 2015 monies. The return trip of nearly $4,000 in current dollars compares to about $1,000 for an average (undiscounted) London–New York flight.

While the early nineteenth century saw some important gains, in terms of both unit capacities and efficiencies and in the stationary harnessing of natural kinetic energies by water wheels and windmills, land transport, powered solely by animate muscles, had changed very little since the antiquity. For millennia, no mode of traveling on land was faster than riding a good horse. For centuries, no conveyance was less tiring than a well-sprung coach. By 1800 some roads had better hard tops, and many coaches were well sprung, but all of these were differences of degree, not of kind. Railways removed these constants in a matter of years. They not only shrank distances and redefined space, they did so with unprecedented comfort. Speed of a mile a minute (96 km/h) was first reached briefly on a scheduled English run in 1847; that was also the year of the greatest railway-building activity in the country, which laid a dense network of reliable links within just two generations (O'Brien 1983).

The large-scale construction of railways with trains pulled by increasingly powerful coal-fueled steam engines was accomplished in Europe and North America in less than 80 years: the 1820s were the decade of experimentation; by the 1890s the fastest trains traveled along some sections at more than 100 km/h. Very soon after their introduction passenger cars ceased to be merely carriages on rails and acquired heating and washrooms. For a higher price passengers also enjoyed good upholstery, fine meal services, and sleeping arrangements. Faster and more comfortable trains carried not only visitors and migrants to cities but also urbanites to the countryside. Thomas Cook offered railway holiday packages starting in in 1841. Commuter rail lines made the first great wave of suburbanization possible. Increasingly capacious freight trains brought bulky resources to distant industries and speedily distributed their products.

The total length of British railways was soon surpassed by American construction, which began in 1834 in Philadelphia. By 1860 the United States had 48,000 km of track, three times the UK total. By 1900 the difference was nearly tenfold. The first transcontinental link came in 1869, and by the end of the century there were four more such lines (Hubbard 1981). The Russian development also progressed fairly rapidly. Fewer than 2,000 km of track were laid by 1860, but the total rose to over 30,000 by 1890 and to nearly 70,000 km by 1913 (Falkus 1972). The transcontinental link across Siberia to Vladivostok, begun in 1891, was not fully completed until 1917. When the British withdrew from India in 1947 they left behind about 54,000 km of track (and 69,000 km in the whole subcontinent). No other mainland Asian country built a major railway network before World War II.

Since the end of World War II, competition from cars, buses, and planes reduced the relative importance of railways in most industrialized countries, but during the latter half of the twentieth century the Soviet Union, Brazil, Iraq, and Algeria were among the vigorous builders of new lines, and China was the Asian leader, with over 30,000 km of track added between 1950 and 1990. But the most successful innovation of the post–World War II period has been fast long-distance electrical train. The Japanese *shinkansen*, first run in 1964 between Tokyo and Osaka, reached a maximum of 250 km/h, and its latest trains (*nozomi*) go 300 km/h (Smil 2014a; fig. 6.10).

Figure 6.10
Shinkansen N700 Series at Kyoto Station in 2014, the 50th year of accident-free operation of Japan's rapid trains on the Tokaido line. Photograph by V. Smil.

French *trains a grand vitesse* (TGV) have been operating since 1983; the fastest scheduled trip is at nearly 280 km/h. Similarly rapid links now also exist in Spain (AVE), Italy (*Frecciarossa*) and Germany (Intercity), but China has become the new record-holder in the overall length of high-speed rail: in 2014 it had 16,000 km of dedicated track (Xinhua 2015). In contrast, America's solitary *Acela* (Boston-Washington, averaging just over 100 km/h) does not even qualify as a modern high-speed train.

If the count starts from the introduction of the first practical gasoline engines in the late 1880s, then the second transportation revolution on land, the progress of road vehicles powered by internal combustion engines, did not take less time. In the higher-income countries of Europe and North America it was twice interrupted by world wars. And while the United States had a high rate of car ownership already during the late 1920s, a comparable stage in Europe and Japan came only during the 1960s, and in China the age of mass car ownership began only in the year 2000, but, owing to the country's large population and rapid investment in new factories, China's car sales surpassed the U.S. total in 2010. By that time the world had about 870 million passenger cars and a total of more than one billion road vehicles (fig. 6.11).

The economic, social, and environmental changes brought by cars rank among the most profound transformations of the modern era (Ling 1990; Womack, Jones, and Roos 1991; Eckermann 2001; Maxton and Wormald 2004). In country after country (first in the United States during the mid-1920s), car making emerged as the leading industry in terms of product value. Cars have also become major commodities of international trade. Their exports from Germany (after 1960) and even more so from Japan (after 1970) have been benefiting those two economies for decades. Large segments of other industries—above all steel, rubber, glass, plastics, and oil refining—are dependent on making and driving cars. Highway building has involved massive state participation, leading to enormous cumulative capital investments. Hitler's *Autobahnen* of the 1930s preceded Eisenhower's system of interstates by a generation (starting in 1956, the total is now just above 77,000 km), and the latter system has been far surpassed by China's National Trunk Highway System, whose total length reached 112,000 km in 2015.

Certainly the most obvious car-generated impact has been the worldwide reordering of cities through the proliferation of freeways and parking spaces and the destruction of neighborhoods. Where space allows, there has been also a rapid increase in suburbanization (in North America also in exurbanization) and changes in location and forms of shopping and services. The social impacts have been even greater. Car ownership has been an important part of *embourgeoisement,* and some affordable designs that enabled this new mass ownership enjoyed amazing longevity (Siuru 1989). The first was Ford's Model T, whose price dropped as low as $265 in 1923 and whose production lasted 19 years (McCalley 1994). Other notable models were the Austin Seven, the Morris Minor, the Citroen 2CV, the Renault 4CV, the Fiat Topolino, and, the most popular of them all, Ferdinand Porsche's Hitler-inspired Volkswagen (box 6.6).

Freedom of personal travel has had enormous effects on both residential and professional mobility. These benefits have proved to be highly addictive. Boulding's (1974) analogy of a car as a mechanical steed turning its driver into a knight with an aristocrat's mobility, looking down at pedestrian peasants (and making it almost unthinkable to rejoin them), is hardly exaggerated. In 2010 there were only 1.25 people per motor vehicle (including trucks and buses) in the United States, and the rate was 1.7 in both Germany and Japan (World Bank 2015b). This widespread addiction to on-demand mobility makes it difficult to give up the habit: after a

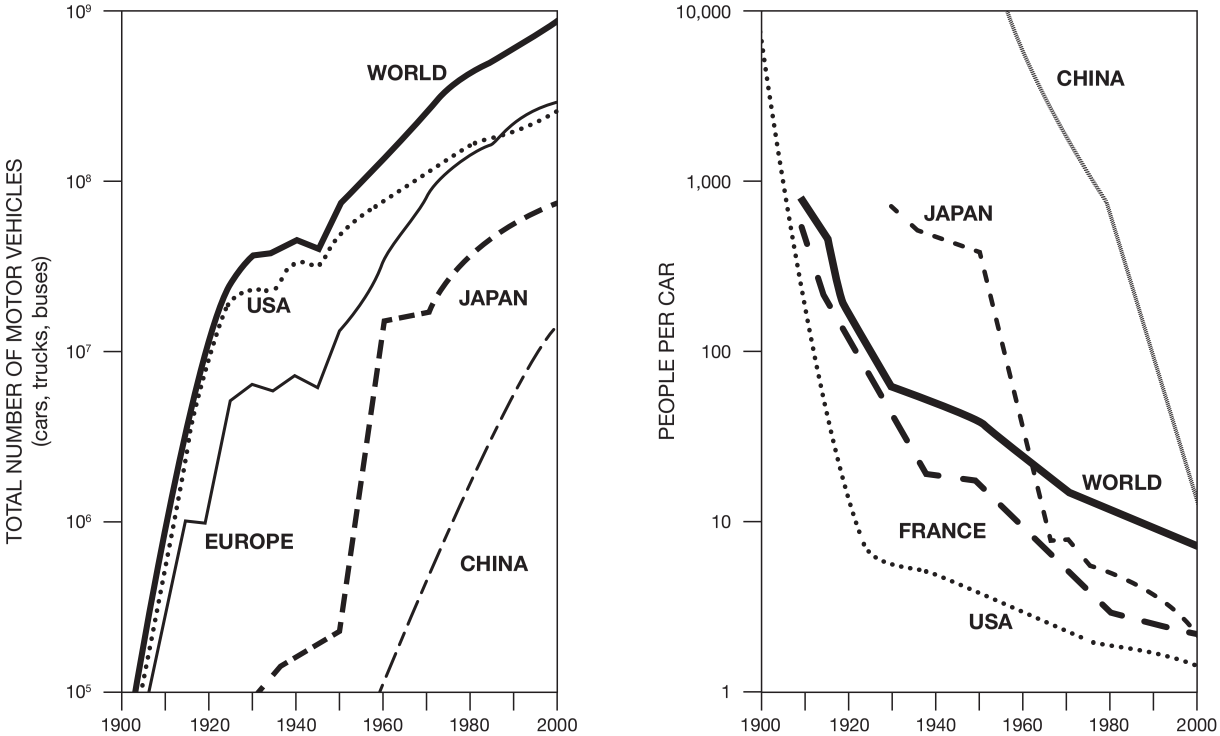


Figure 6.11
The worldwide total of road vehicles grew from about 10,000 in 1900 to more than one billion by 2010 (left). U.S. registrations were surpassed by the European total during the late 1980s, but the country still has the highest rate of ownership, about 1.25 people per vehicle in 2010 (right). Plotted from data in annual reports of the Motor Vehicle Manufacturers Association and World Bank (2015b).

Box 6.6
Volkswagen and other durable models

In terms of the aggregate production, size, and longevity (though with updated models), no car designed for the masses comes close to the one that Adolf Hitler decreed as the most suitable for his people (Nelson 1998; Patton 2004). In autumn 1933 Hitler set down the car's specifications—top speed 100 km/h, 7 L/km, capable of conveying two adults and three children, with air cooling, and at a cost below 1,000 RM—and Ferdinand Porsche (1875–1951) had the car, rather ugly and looking, at Hitler's insistence, like a beetle (*Käfer*), ready for production in 1938. War prevented any civilian production, and the Beetle's serial assembly began only in 1945, under the British Army command led by Major Ivan Hirst (1916–2000), who saved the damaged factory (Volkswagen AG 2013).

During the early years of West German *Wirtschaftswunder* (before mass ownership of Mercedeses, Audis, and BMWs), the car flooded German roads, and during the 1960s Volkswagen became the most popular import to the United States before it was displaced by Hondas and Toyotas. The production of the original Beetle stopped in Germany in 1977 but continued in Brazil until 1996 and in Mexico until 2003: the last car produced at the Puebla plant had number 21,529,464. The New Beetle, with an exterior redesigned by J. Mays and the engine in the front, was made between 1997 and 2011; since the model year 2012 the name of the latest design (A5) has reverted to Volkswagen Beetle.

The Renault 4CV, secretly designed during World War II, was the Beetle's French counterpart; more than one million cars were made between 1945 and 1961. The country's most famous basic car was the Citroen 2CV, made between 1940 and 1990: *deux cheveaux* marked just the number of cylinders; the engine actually had 29 hp (Siuru 1989). Fiat's little mouse, the Topolino, a two-seater with a wheelbase just short of 2 m, was made between 1936 and 1955, and the British Morris Minor was made between 1948 and 1971. All these models were eclipsed in popularity by Japanese designs: after relatively small exports during the 1960s and 1970s they became the global bestsellers during the 1980s.

recession-induced dip between 2009 and 2011, car sales in the United States reached near record levels of 16.5 million units by 2015.

We have gone to extraordinary lengths to preserve this privilege (and in North America we have made it easier by selling more than 90% of all vehicles on credit), and hence we cannot be surprised that Chinese and Indians want to emulate the North American experience. But like every addiction, this one exacts a high price. In 2015 the world had about 1.25 billion vehicles on the road, and in 2015 new passenger car sales reached about 73 million (Bank of Nova Scotia 2015), while traffic accidents cause annually nearly 1.3 million deaths and up to 50 million injuries (WHO 2015b), and automotive air pollution has been a key contributor to the worldwide phenomenon of seasonal (or semipermanent) photochemical smog in megacities on all continents (USEPA 2004). The life span of the average car now ranges from nearly 11 years in affluent countries to more than 15 years in low-income economies. Afterward, the steel (and copper and some rubber) is mostly recycled, but we have been willing to put up with enormous death, injury, and pollution costs.

Trucking has also had many profound socioeconomic consequences. Its first mass diffusion, in rural America after 1920, reduced the cost and sped up the movement of farm products to market. These benefits have been replicated first in Europe and Japan, and during the past two decades also in many Latin American and Asian countries. In rich countries, long-distance heavy trucking has become the backbone of food deliveries, as well as a key link in the distribution of industrial parts and manufactured goods, and its operation has benefited from the universal embrace of containers offloaded by cranes from oceangoing vessels directly onto flatbed trucks. In many rapidly growing economies trucking has obviated the construction of railways (Brazil being the best example) and opened up remote areas to commerce and development—but also to environmental destruction. In poor nations buses have been the leading means of long-distance passenger transport.

The first steamships crossed the North Atlantic no faster than the best contemporary sail ships with favorable winds. But already by the late 1840s the superiority of steam was clear, with the shortest crossing time cut to less than 10 days (fig. 6.12). By 1890 trips of less than six days were the norm, as were steel hulls. Steel did away with size restrictions: structural considerations limited the length of wooden hulls to about 100 m. Large ships of such famous lines as Cunard, Collins, or Hamburg-America became proud symbols of technical age. They were equipped with powerful engines and double-screw propellers, furnished with grand staterooms, and offered excellent service.

The opulence of these great liners contrasted with the crowding, smells, and tedium of steerage passages. By 1890 steamships carried more than half a million passengers a year to New York. By the late 1920s the total North Atlantic traffic surpassed one million passengers a year, and soon afterward the liners reached their maximum tonnages (fig. 6.12). But by 1957 airlines carried more people across the Atlantic than ships, and the introduction of regular jetliner service in the same year sealed the fate of long-distance passenger shipping: a decade later regularly scheduled transatlantic service came to an end. Commercial steamships got early boosts from the completion of the Suez Canal in 1869 and from the introduction of effective refrigeration during the 1880s. Its later growth was stimulated by the opening of the Panama Canal (1914), the deployment of large diesel engines (after 1920), and the transport of crude oil. Since the 1950s larger specialized

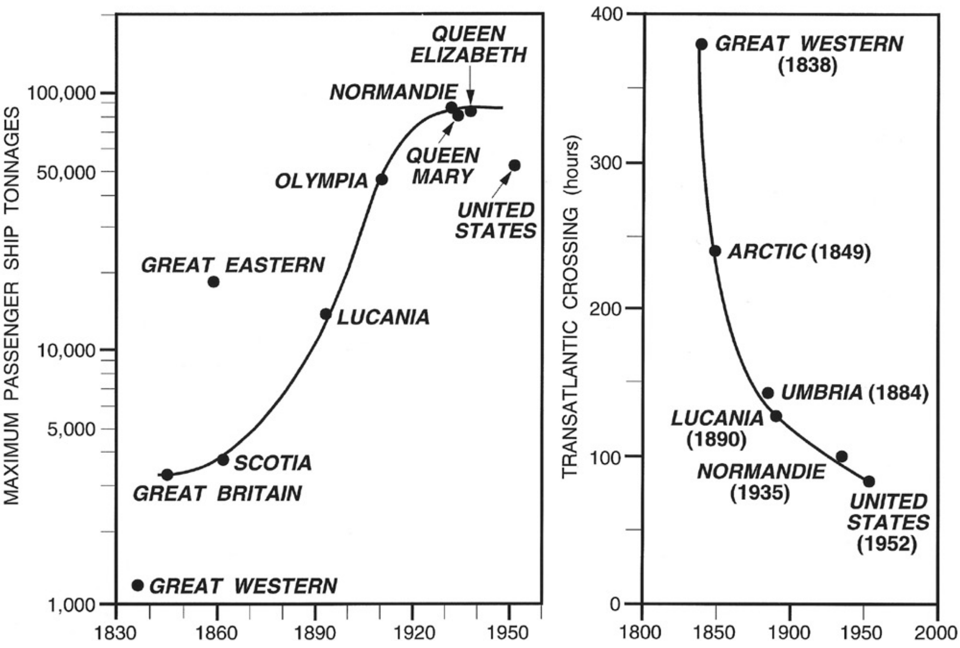


Figure 6.12
As the ships connecting Europe and North America grew in size (left) and became equipped with more powerful engines, the time needed to cross the Atlantic was cut from more than two weeks to just over three days (right). Plotted from data in Fry (1896), Croil (1898), and Stopford (2009).

ships have been needed to move not only oil but also widely traded bulky commodities (ores, lumber, grain, chemicals) and growing shipments of cars, machinery, and consumer goods.

Scheduled international air transport started with daily London-Paris flights in 1919 at speeds well below 200 km/h, and advanced to regular transoceanic links just before World War II: PanAm's Clipper reached Hong Kong from San Francisco after a six-day journey in March 1939 (fig. 6.13). The age of mass air travel came only with the introduction of jet aircraft in the late 1950s (British Comet, in service since 1952, was grounded in 1954 after three fatal disasters). The Boeing 707 (the first flight in 1957, in service since October 1958) was soon followed by the mid-range Boeing 727 (in regular service since February 1964 and produced until 1984) and the short- to mid-range Boeing 737. This smallest of all Boeing jetliners has become the bestselling airplane in history: by mid-2015 more than 8,600 planes had been delivered (compared to about 9,200 for all Airbus models). During the 1950s and 1960s McDonnell Douglas (DC-9, triple-engine DC-10), General Dynamics (Convair), Lockheed (Tristar), and Sud Aviation (Caravelle) introduced their own jetliners, but (leaving the Russian makers aside) by the century's end only the duopoly of the American Boeing and the European Airbus consortium was left (box 6.7).

The speed and range of these planes, the proliferation of airlines and flights, and the nearly universal linking of reservation systems have made it possible to travel among virtually all major cities of the planet in a single day (fig. 6.13). By the year 2000 the maximum range of wide-body jetliners reached 15,800 km, and in 2015 the longest scheduled flights (Dallas-Sydney and Johannesburg-Atlanta) lasted nearly 17 hours, while many cities are connected by frequent shuttle flights (in 2015 there were nearly 300 daily flights between Rio de Janeiro and São Paulo, nearly 200 between New York and Chicago). Moreover, the costs of flying have been steadily declining in real terms, in part because of lower fuel consumption. These achievements opened up new business opportunities as well as mass long-distance tourism to major cities and to subtropical and tropical beaches. They also opened up new possibilities for unprecedented migrant and refugee movements, for widespread drug smuggling, and for international terrorism involving aircraft hijacking.

Information and Communication

From their very conception, fossil-fueled societies have produced, stored, distributed, and used incomparably larger amounts of information than their predecessors. In East Asia and in early modern Europe, printing was an

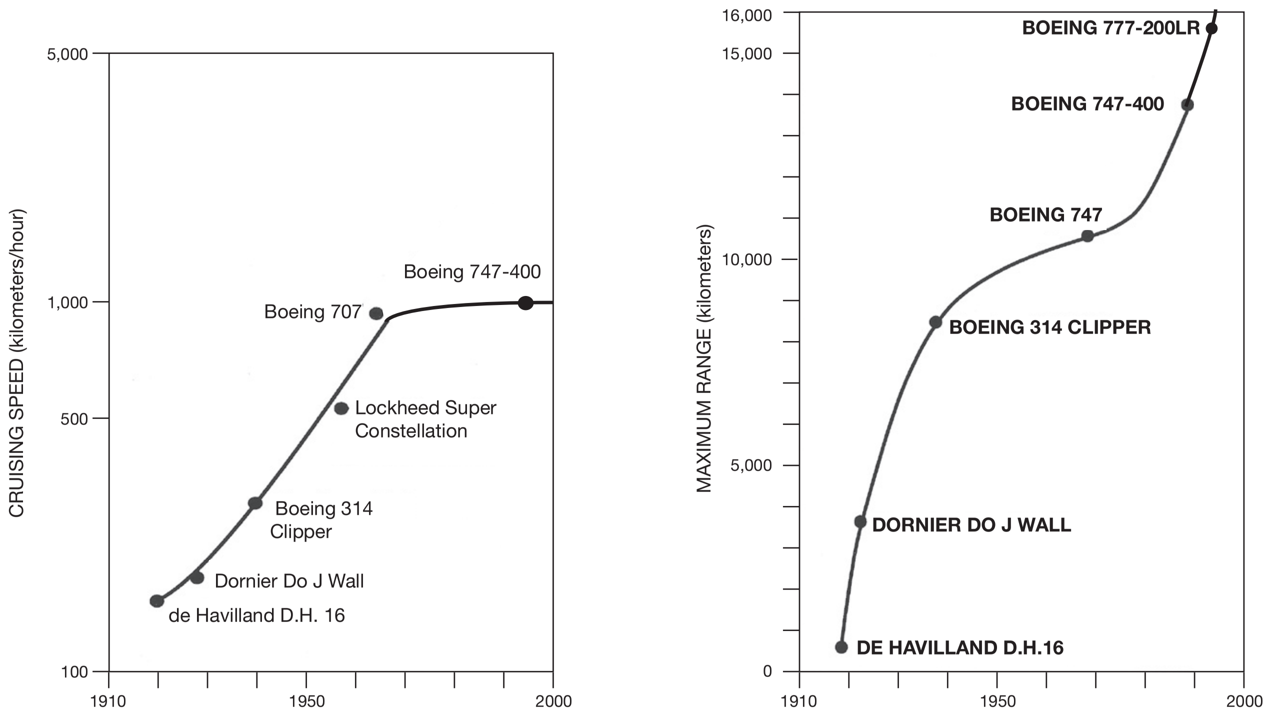


Figure 6.13

The first scheduled commercial flights (by the de Havilland D.H. 16 in 1919) averaged just over 150 km/h, and the plane had a maximum range of about 600 km (left). By the late 1950s the Boeing 707 could cruise at close to 1,000 km/h, and by the late 1990s the Boeing 777 could fly nonstop more than 15,000 km (right). The Concorde, flying at over twice the speed of sound, was a costly exception, not a precursor to a new generation of fast planes. Plotted from data in Taylor (1989) and Gunston (2002) and from technical specifications on the Boeing corporate website.

Box 6.7
Boeing and Airbus

Boeing is an old U.S. company—established by William E. Boeing (1881–1956) in 1916—and a maker of such iconic designs as the Boeing 314 Clipper and the 307 Stratoliner (both in 1938), the Boeing 707 (the first successful jetliner, in 1957), and the Boeing 747, the first wide-body plane, in 1969 (Boeing 2015). The company's latest innovation is the Boeing 787, an advanced design that uses lighter but stronger carbon fibers for 80% of the body, allowing 20% higher fuel efficiency than for the 767 (Boeing 2015). Airbus was set up in December 1970 with French and German participation, later joined by Spanish and British companies. Its first twinjet, the Airbus A300 (226 passengers), launched in October 1972, and its offerings expanded to a full range of planes, from the short-haul planes A319, 320, and 321 to the long-distance wide-body A340. In 2000 Airbus surpassed for the first time the number of planes sold by Boeing. Its greatest innovation has been the A380, a double-decker wide-body craft in service since 2007 with a maximum capacity of 853 passengers in a single class but, so far, ordered only in three-class configurations for 538 people (compared to 416 in three-class and 524 in two-class configuration for Boeing 747-400 planes).

The two companies have been in a close competition. Between 2001 and 2015 Boeing delivered 6,803 airplanes and Airbus produced 6,133 jetliners, and both companies have substantial multiyear order backlogs to supply the rising demand, particularly from Asia. Both companies have also made many cooperative agreements with aircraft and engine designers and with the suppliers of major airplane components in Europe, North America, and Asia, and both face a growing competition from below. The Canadian company Bombardier and the Brazilian Embraer have been enlarging their commuter jets: Bombardier's CRJ-900 seats 86, while Embraer's EMB-195 takes up to 122 passengers, and both of these companies, as well Russia's Sukhoi Superjet, the Commercial Aircraft Corporation of China, and Japan's Mitsubishi, are entering the lucrative market for narrow-body planes that is now served by the Boeing 737 and the Airbus A319/320.

established commercial activity for hundreds of years before the introduction of fossil fuels, but hand typesetting was laborious and print runs were limited by the slowness of hand-operated wooden screw presses. Iron frames sped up the work, but even advanced designs of Gutenberg's printing press could make no more than 240 impressions per hour (Johnson 1973). But even the first press powered by a steam engine—designed by Friedrich Koenig and Andreas Friedrich Bauer and sold to the *Times* in 1814—could do 1,100 impressions per hour. By 1827 that figure was 5,000, and the first rotary presses of the 1840s managed 8,000 impressions per hour; two decades later the rate was up to 25,000 (Kaufer and Carley 1993).

Mass editions of inexpensive newspaper became a quotidian reality, with news traveling faster thanks to telegraph (commercially for the first time in 1838) and less than two generations later telephone (1876), and before the century's end two new information-communication techniques had become commercial: sound recordings and replays and film. Except for printing, all these techniques were developed during the high-energy age based on fossil fuels. Except for photography and the early phonographs, none of them could function without electricity. And except for printed matter, now in retreat as many e-formats are taking its place, all these techniques have been expanding their user base and acquiring new modes of information capture, storage, recording, viewing, and sharing in the instantaneously interconnected world.

Inexpensive, reliable, and truly global telecommunication became possible only with electricity. The first century of its development was dominated by messages transmitted by wires. Decades of experiments in various countries ended with the first practical telegraph, demonstrated by William Cooke and Charles Wheatstone in 1837 (Bowers 2001). Its success depended on a reliable source of electricity, which was provided by Alessandro Volta's battery, designed in 1800. The adoption of the coding system of Samuel Morse in 1838 and the rapid extension of land lines in conjunction with railways were the most notable early developments. Undersea links (across the English Channel in 1851, across the Atlantic in 1866) and a wealth of technical innovations (including some of Edison's early inventions) combined to make the telegraph global within just two generations. By 1900 multiplex wires with automatic coding carried millions of words every day. The messages ranged from personal to diplomatic codes, and included reams of stock market quotations and business orders.

The telephone, patented by Alexander Graham Bell in 1876 just hours ahead of Elisha Grey's independent filing (Hounshell 1981), had an even faster acceptance in local and regional service (Mercer 2006). Reliable and cheap long-distance links were introduced rather slowly. The first trans-American link came only in 1915, and the transatlantic telephone cable was laid only in 1956. To be sure, radio-telephone links were available from the late 1920s, but they were neither cheap nor reliable. Great telephone monopolies provided affordable and reliable service, but they were not great innovators: the classic black rotary-dial telephone was introduced in the early 1920s and remained the only choice for the next four decades: the first electronic touch-tone phones appeared in the United States only in 1963.

Techniques for the storage, reproduction, and transmission of sound and pictures were developing concurrently with advances in telephony. Thomas Edison's 1877 phonograph was a simple hand-operated machine, as was Emile Berliner's (1851–1929) more complex gramophone in 1888 (Gronow and Saunio 1999). Electric record players took over only during the 1920s. Image making advanced rather slowly from its French beginnings, most notably in the work of J. N. Niepce and L. J. M. Daguerre during the 1820s and 1830s (Newhall 1982; Rosenblum 1997). Kodak's first inexpensive box camera came out in 1888, and developments sped up after 1890 with breakthroughs in cinematography: the first public short movies by the Lumière brothers were projected in 1895. Sound movies came in the late 1920s (the first feature film was *The Jazz Singer*, in 1927), the first color feature (after years of short color movies) came in 1935, and the invention of xerography by Chester Carlson (1906–1968) came two years later (Owen 2004).

The quest for wireless transmission started with Heinrich Hertz's (1857–1894) generation of electromagnetic waves in 1887, anticipated by James Clerk Maxwell's (1831–1879) formulation of the theory of electromagnetic radiation (Maxwell 1865; fig. 6.14). Subsequent practical progress was fast. By 1899 Guglielmo Marconi's (1874–1937) signals had crossed the English Channel, two years later the Atlantic (Hong 2001). In 1897 Ferdinand Braun (1850–1918) invented the cathode ray tube, the device that made possible both television cameras and receivers. In 1906 Lee de Forest (1873–1961) built the first triode, whose indispensability for broadcasting, long-distance telephony, and computers ended only with the invention of the transistor.

Regular radio broadcasts started in 1920. BBC offered the first scheduled television service in 1936, and RCA followed suit in 1939 (Huurdeman 2003). Mechanical calculators—starting with prescient designs by Charles Babbage and Edward Scheutz after 1820 (Lindgren 1990; Swade 1991) and culminating in the establishment of IBM in 1911—were finally left behind with the development of the first electronic computers during World War II. But these machines—the British Mark, the U.S. Harvard Mark 1, and the ENIAC—were unique, dedicated, massive (room-sized, to accommodate thousands of glass vacuum tubes) devices with no immediate commercial prospects.

This impressive concatenation of greatly improved and entirely new communication and information techniques and services was entirely

Figure 6.14
Engraved portrait of James Clerk Maxwell, based on a photograph by Fergus (Corbis). Maxwell's formulation of the theory of electromagnetism opened the way to the still unfolding exploits of modern wireless electronics that have brought inexpensive instant communication and global connectivity: the e-world of the twenty-first century rests on Maxwell's insights.

overshadowed by post–World War II developments. Their shared foundation was the rise of solid state electronics, which began with the American invention of the transistor, a miniature solid-state semiconductor device, the equivalent of a vacuum tube that can amplify and switch electronic signals. Julius Edgar Lilienfeld filed his patent for the field-effect transistor in Canada in 1925 and a year later in the United States (Lilienfeld 1930); the patent application clearly outlines the way to control and amplify the flow of current between the two terminals of a conducting solid.

But Lilienfeld did not attempt to build any device, and the first experimental success, by two Bell Labs researchers, Walter Brattain and John Bardeen, on December 16, 1947, used a germanium crystal (Bardeen and Brattain 1950). But as the Bell System Memorial site now admits, "It's perfectly clear that Bell Labs didn't invent the transistor, they re-invented it" while failing to acknowledge a great deal of pioneering research and design done since the very first decade of the twentieth century (Bell System Memorial 2011). In any case, it was not the crude point-contact device used by Brattain and Bardeen but more useful junction field-effect transistor patented in 1951 by William Shockley (1910–1989) that transformed electronic computing. In the same year Gordon K. Teal and Ernest Buehler succeeded in making larger silicon crystals and mastering improved methods for crystal pulling and silicon doping (Shockley 1964; Smil 2006).

A very important theoretical advance was made in 1948 when Claude Shannon opened the way to quantitative appraisals of the energy cost of communication (Shannon 1948). Despite the impressive progress made during the intervening years (a three orders of magnitude increase in carrying simultaneous conversation by a single cable, now no thicker than a human hair), Shannon's theoretical limits indicated that the performance could be improved by several orders of magnitude. But there was no immediate post–World War II rush to commercialize electronic computing, and Remington Rand's first UNIVAC (Universal Automatic Computer, an outgrowth of the Eckert-Mauchly ENIAC) was sold to the U.S. Census Bureau only in 1951.

The calculating speed of the new programmable machines started to rise exponentially as transistors supplanted vacuum tubes. Business use of computers in the United States finally took off only during the late 1950s, with Fairchild Semiconductor, Texas Instruments (which marketed the first silicon transistor in 1954), and IBM as the most accomplished developers of hardware and software (Ceruzzi 2003; Lécuyer and Brock 2010). In 1958–1959 Jack S. Kilby (1923–2005) at Texas Instruments and Robert Noyce (1927–1990) at Fairchild Semiconductor independently invented miniaturized circuits integrated into the body of semiconductor material (Noyce 1961; Kilby 1964). Noyce's design of a planar transistor opened the new era of solid-state electronics (box 6.8).

The U.S. military was the first customer for integrated circuits. In 1965, when the number of transistors on a microchip had doubled to 64 from 32 in the previous year, Gordon Moore predicted that this doubling would continue (Moore 1965). In 1975 he relaxed the pace to a doubling every two years (Moore 1975), and this rule, now commonly known as Moore's law, has held ever since (fig. 6.15). The world's first microprocessor-controlled commercial product was a programmable calculator by Busicom, a small Japanese company; its four-chip set was designed by the just

Box 6.8
Invention of integrated circuits

When working as the director of research at Fairchild Semiconductors in Santa Clara, California, Robert Noyce wrote in his lab notebook that

> it would be desirable to make multiple devices on a single piece of silicon, in order to be able to make interconnections between devices as part of the manufacturing process, and thus reduce size, weight, etc. as well as cost per active element. (Reid 2001, 13)

Noyce's 1959 patent application for a "semiconductor device-and-lead structure" showed a planar integrated circuit. It specified

> dished junctions extending to the surface of a body of extrinsic semiconductor, an insulating surface layer consisting essentially of oxide of the same semiconductor extending across the junctions, and leads in the form of vacuum-deposited or otherwise formed metal strips extending over and adherent to the insulating oxide layer for making electrical connections to and between various regions of the semiconductor body without shorting the junctions. (Noyce 1961, 1)

Noyce's patent (U.S. 2,981,877) was granted in April 1961, Kilby's (U.S. 3,138,743) only in July 1964, and lengthy interference proceeding, litigation, and appeals were settled only in 1971 when the Supreme Court ruled in Noyce's favor. By that time that was an immaterial victory because back in the summer of 1966 the two companies had agreed to share their production licenses and require other fabricators to make separate arrangements with both of them. In principle, Kilby's and Noyce's ideas were identical, but Noyce died of a heart attack in 1990, whereas Kilby lived long enough to share a Nobel Prize for Physics in the year 2000 "for his part in the invention of the integrated circuit."

established Intel in 1969–1970 (Augarten 1984). Busicom sold only a few large calculator models using the MCS-4 chip set before it went bankrupt in 1974; fortuitously, Intel had had the foresight to buy back the rights for the processor before that happened, and it released the world's first universal microprocessor—the 3 mm × 4 mm Intel 4004 containing 2,250 metal-oxide semiconductor transistors and priced at $200—in November 1971. With 60,000 operations per second, it was the functional equivalent of the room-sized ENIAC of 1945 (Intel 2015).

The universal deployment of these increasingly powerful microprocessors in conjunction with increasingly capacious memory devices has affected every sector of modern manufacturing, transportation, services, and communication, and the spectacular growth of these capabilities has been accompanied by steadily declining costs and improving reliability (Williams 1997; Ceruzzi 2003; Smil 2013c; Intel 2015). Microchips have

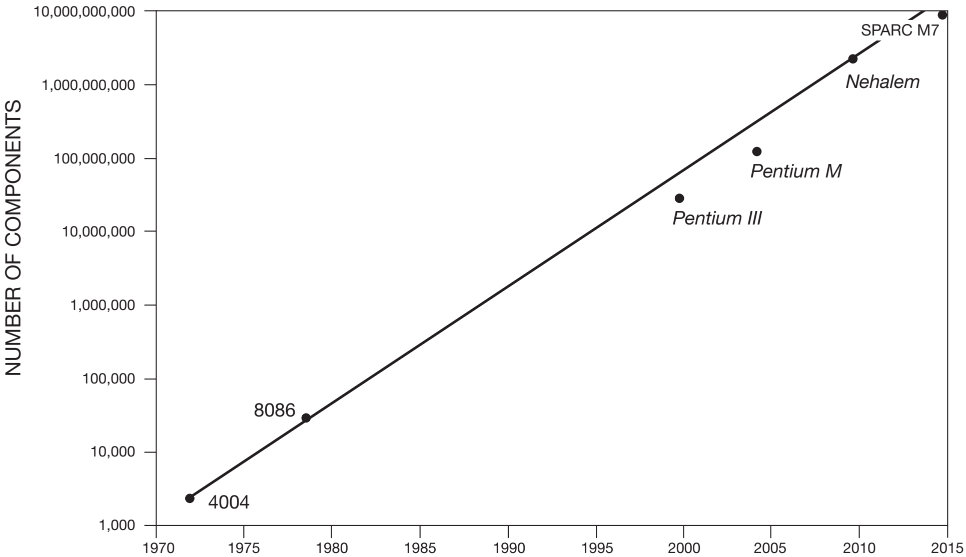


Figure 6.15
Moore's law in operation. The first commercially available microchip (Intel 4004) had 2,250 metal-oxide semiconductor transistors, while the latest designs have more than ten billion components, an increase of six orders of magnitude. Plotted from data in Smil (2006) and Intel (2015).

become the most ubiquitous complex artifacts of modern civilization: more than 200 billion of them are produced every year, and the devices can be found in products ranging from mundane household items and appliances (thermostats, ovens, furnaces, and in every electronic gadget) to the automated fabrication of complex assemblies, including the design and making of microprocessors themselves. They govern the timing of fuel ignition in automotive engines, optimize the operation of jetliner turbines, and guide rockets to place satellites on their predetermined paths.

But the most personalized impact of microprocessors has been through the mass ownership of portable electronic devices, above all cellular phones. This development was preceded by the rise of personal computers, by the surprisingly long development of the Internet, and by a period of relatively slow adoption of mobile phones. Xerox Palo Alto Research Center (PARC) invented personal computing during the 1970s by combining the processing power of microchips with a mouse, a graphical user interface, icons and pop-up menus, laser printing, text editing, spell checking, and access to file servers and printers with point-and-click actions (Smil 2006; fig. 6.16). Without these advances Steven Wozniak and Steven Jobs could not have

Figure 6.16
Utilitarian but revolutionary: the Xerox Alto desk computer, released in 1973, was the first nearly complete embodiment of the basic features characteristic of all later PCs (Wikimedia photograph).

introduced the first successful commercial PC, the Apple II with color graphics, in 1977 (Moritz 1984). IBM's PC was released in 1981, and in the United States ownership of PCs rose from two million units in 1983 to nearly 54 million units by 1990 (Stross 1996). Lighter, portable machines, laptops and tablets, matured only during the late 1990s, and Apple's iPad was introduced in 2010.

Communication using computers was first proposed in 1962 by J.C.R. Licklider, the first director of the Pentagon's Advanced Research Project Agency, and began in 1969 with ARPANET, limited to just four sites, at the Stanford Research Institute, UCLA, UCSB, and the University of Utah. In 1972 Ray Tomlinson of BBN Technologies designed programs for sending messages to other computers and chose the @ sign as the locator symbol for email addresses (Tomlinson 2002). In 1983 ARPANET converted a protocol that made it possible to communicate across a system of networks, and by 1989, when it ended its operation, it had more than 100,000 hosts. A year later Tim Berners-Lee created the hypertext-based World Wide Web at Geneva's CERN in order to organize online scientific information (Abbate 1999). The early web was not easy to navigate, but that changed rapidly with the introduction of efficient browsers, starting with Netscape in 1993.

The first major electronic advance in telephony was the possibility of inexpensive intercontinental calls, thanks to automatic dialing via geostationary satellites. This innovation resulted from a combination of microelectronic advances and powerful rocket launchers during the 1960s, and as the underlying costs declined, calls became cheaper. But the first radical change in telephony came only with the introduction of mobile phones (cell phones): first demonstrated in 1973, an expensive paid service with bulky Motorola sets was available in the United States in 1983, but ownership began to rise rapidly (with Japan and the EU ahead of the United States) only during the late 1990s. Global cell phone sales surpassed 100 million units in 1997, the year that Ericsson introduced the first smart phone.

Cell phone sales reached the billion mark by 2009, and by the end of 2015 there were 7.9 billion devices in use and the total annual shipments of mobile devices, including tablets, notebooks, and netbooks, had reached nearly 2.2 billion units, among them 1.88 billion cell phones (Gartner 2015; mobiForge 2015). This impressive and rapidly changing system of communication, entertainment, monitoring and data devices, and software requires a significant amount of energy to be embodied in highly energy-intensive electronic devices and is utterly dependent on an incessant and highly reliable electricity supply to energize the requisite infrastructures, ranging from data centers to cell towers (box 6.9).

Of special note is the enormous progress that has been made since the 1960s in designing and deploying a wide range of diagnostic, measuring, and remote sensing techniques. These advances yielded a previously unimaginable wealth of information. X-rays, discovered by W. K. Roentgen (1845–1923) in 1895, were the only such option in 1900. By 2015 these techniques ranged from ultrasound (used both in medical diagnoses and in engineering) to high-resolution imaging (MRI, CT), and from radar (developed on the eve of World War II, and now an indispensable tool in transportation and weather monitoring) to a wide range of satellite-based sensors acquiring data in various bands of the electromagnetic spectrum and

Box 6.9
Energy embodied in mobile phones and cars

Even a compact car weighs 10,000 times as much as a smart phone (1.4 t vs. 140 g), and hence it embodies considerably more energy. But the energy difference is far smaller than the four orders magnitude mass disparity, and aggregate accounts make for a surprising comparison. A cell phone embodies about 1 GJ of energy, whereas a typical passenger car now requires about 100 GJ to produce, only 100 times as much energy as goes into a cell phone. In 2015 worldwide sales of mobile phones came very close to 2 billion units, and hence their production consumed about 2 EJ (an equivalent of about 48 million metric tons of crude oil). About 72 million cars were sold worldwide in 2015, and their production embodied roughly 7.2 EJ—or only less than four times the total for mobile phones.

Mobile phones have very short life spans, on the average just two years, and their production now embodies globally about 1 EJ per average year of use. Passenger cars last on the average at least a decade, and their production embodies globally about 0.72 EJ per year of use—30% less than the making of mobile phones! This means that even if these approximate aggregates err in opposite directions (in reality, cars embodying more and mobile phone less energy), the two totals would still be not only of the same order of magnitude but surprisingly close. Operating energy costs are, of course, vastly different. A smart phone consumes annually just 4 kWh of electricity, less than 30 MJ during its two years of service, or just 3% of its embodied energy cost. In contrast, during its lifetime a compact car will consume four to five times as much energy (as gasoline or diesel fuel) as its embodied content. But the costs of electrifying the world's information and communications networks is rising: it claimed nearly 5% of worldwide electricity generation in 2012 and will approach 10% by 2020 (Lannoo 2013).

enabling much improved weather forecasting and natural resource management.

Economic Growth

To talk about energy *and* the economy is a tautology: every economic activity is fundamentally nothing but a conversion of one kind of energy to another, and monies are just a convenient (and often rather unrepresentative) proxy for valuing the energy flows. Not surprisingly, Frederick Soddy, a Nobelian physicist approaching the discipline from this perspective, argued that "the flow of energy should be the primary concern of economics" (Soddy 1933, 56). At the same time, energy flow is a poor measure of intellectual activity: education certainly embodies a great deal of energy expended on its infrastructures and employees, but brilliant ideas (which are by no means directly related to the intensity of schooling) do not require large increases of the brain's metabolic rate.

This obvious fact explains much of the recent decoupling of GDP growth from overall energy demand: we impute much higher monetary values to the nonphysical endeavors that now constitute the largest share of the economic product. In any case, energy has been of marginal concern in modern economic studies; only ecological economists have seen it as their primary focus (Ayres, Ayres, and Warr 2003; Stern 2010). And the public concern about energy and the economy has been disproportionately focused on prices in general, and on the prices of crude oil, the world's most important traded commodity, in particular.

In the West it was OPEC's two rounds of oil price increases during the 1970s—both the source of Middle Eastern consumption excesses and a threat to the region's stability—that became a particular object of critique, blamed for economic dislocations and social turmoil. But OPEC's price rise had a salutary (and long overdue) effect on the efficiency with which the countries importing OPEC oil consumed refined fuels. In 1973, after four decades of slow deterioration, the average specific fuel consumption of new American passenger cars was higher than in the early 1930s, 17.7 L/100 km versus 14.8 L/100 km, or, in American usage, 13.4 mpg versus 16 mpg (Smil 2006)—a rare example of a modern energy conversion becoming less efficient.

Higher oil prices forced the reversal, and between 1973 and 1987 the average fuel demand of new cars on the North American market was cut in half as the CAFE (Corporate Automobile Fuel Efficiency) standard fell to 8.6 L/100 km (27.5 mpg). Unfortunately, the post-1985 fall in oil prices first stopped and then even reversed (with more SUVs and pickups) this efficiency progress, and return to rationality came only in 2005. And OPEC's price rise had a beneficial effect for the global economy as it significantly reduced its average oil intensity (amount of oil used per unit of GDP). Power plants stopped burning liquid fuels; iron makers replaced injections of fuel oil to blast furnaces by powdered coal; jet engines became more efficient; and many industrial processes converted to natural gas. The results have been quite impressive. By 1985 the U.S. economy needed 37% less oil to produce a dollar of GDP than it did in 1970; by the year 2000 its oil intensity was 53% lower; and by 2014 it required 62% less crude oil to create a dollar of GDP than it did in 1970 (Smil 2015c).

And (a curiously neglected fact) Western governments have been making more money from oil than OPEC. In 2014, taxes in G7 countries accounted for about 47% of the price of a liter of oil, compared to about 39% going to the producers, with the respective national shares at 60/30 in the UK, 52/34 in Germany, and 15/61 in the United States (OPEC 2015). Moreover, to ensure a secure supply, many governments (including those of market economies) have engaged in a great deal of industry regulation, while governments in many oil-producing countries have been buying political support with heavy subsidies of energy prices (GSI 2015). Saudi subsidies claimed more than 20% of all government expenditures in 2010, and China's coal subsidies have resulted in prices fixed even below the production cost.

Growth—its origins, rate, and persistence—has been the leading concern of modern economic inquiries (Kuznets 1971; Rostow 1971; Barro 1997; Galor 2005), and hence the links between energy consumption and the increase gross economic product (either gross domestic product, GDP, for individual economies, or GWP, gross world product, for studying global trends) have received a great deal of attention (Stern 2004, 2010; World Economic Forum 2012; Ayres 2014). Traditional preindustrial economies were either largely stationary or managed to grow by a few percent per decade, and average per capita energy consumption advanced at an even slower pace: there is no shortage of testimonies from the early decades of the nineteenth century showing that the living conditions of some impoverished groups were not very different from those that had prevailed even two or three or four centuries before.

In contrast, the fossil-fueled economies have seen unprecedented rates of growth, though modified by the cyclical nature of economic expansion (van Duijn 1983; ECRI 2015) and interrupted by major internal or international conflicts. Industrializing societies of the nineteenth century saw their economies growing by 20–60% in a decade. Such growth rates meant that the output of the British economy in 1900 was nearly ten times larger than in 1800. America's GDP doubled in just 20 years, between 1880 and 1900. Japanese output during the Meiji era (1868–1912) rose 2.5 times. Economic growth during the first half of the twentieth century was affected by two world wars and the great economic crisis of the 1930s, but there had never been a period of such rapid and widespread growth of output and prosperity as between 1950 and 1973.

The steady pre-1970 decline in real crude oil prices was a critical ingredient of this unprecedented expansion. American per capita GDP, already the world's highest, rose by 60%. The West German rate more than tripled, and the Japanese rate more than sextupled. A number of the poor populous countries of Asia and Latin America also entered a phase of vigorous economic growth. OPEC's first round of oil price increases (1973–1974) temporarily stopped this growth. The second round of oil price increases, in 1979, was caused by the overthrow of the Iranian monarchy and the ascent of fundamentalist ayatollahs to power. The global economic slowdown of the early 1980s was accompanied by record inflation and high unemployment, but during the 1990s stabilized low oil prices supported another period of growth, which ended only in 2008 with the world's worst post–World War II recession, followed by a weak recovery.

Ayres, Ayres, and Warr (2003) identified the declining price of useful work as the growth engine of the U.S. economy during the twentieth century, useful work being the product of exergy (the maximum work possible in an ideal energy conversion process) and conversion efficiency. Once the historical data of economic output are normalized (with GDP values expressed in constant, inflation-adjusted monies and with the national products used to calculate GWP given in terms of purchasing power parity rather than by using official exchange rates), impressively strong long-term correlations between economic growth and energy use emerge on both global and national levels.

Between 1900 and 2000 the use of all primary energy (after subtracting processing losses and nonfuel uses of fossil fuels) rose nearly eightfold, from 44 to 382 EJ, and the GWP increased more than 18 times, from about $2 trillion to nearly $37 trillion in constant 1990 monies (Smil 2010a; Maddison Project 2013), implying an elasticity of less than 0.5. High correlations of the two variables can be found for a single country over time, but the elasticities differ: during the twentieth century, the Japanese GDP increased 52-fold and total energy use rose 50-fold (an elasticity very close to 1.0), while the multiples for the United States were, respectively, nearly 10-fold and 25-fold (an elasticity of less than 0.4), and for China nearly 13-fold and 20-fold (an elasticity of 0.6).

The expected closeness of the link between the two variables is further confirmed by very high correlations (>0.9) between averages of per capita GDP and energy supply when the set includes all the world's countries. This is clearly one of the unusually high correlations in the normally unruly realm of socioeconomic affairs, but the effect weakens considerably once we examine more homogeneous groups of countries: to become rich requires a substantial increase in energy use, but the relative energy consumption increase among affluent societies, whether measured per GDP unit or per capita, varies widely, producing very low correlations.

For example, Italy and South Korea have a very similar per capita GDP—adjusted for purchasing power, it was about $35,000 in 2014—but South Korea's per capita energy use is nearly 90% higher than Italy's. Conversely, Germany and Japan have a nearly identical annual energy consumption, about 170 GJ/capita, but in 2014 Germany's GDP was nearly 25% higher (IMF 2015; USEIA 2015d). And the rise in absolute energy consumption required to produce higher economic outputs hides an important relative decline. High-income, high-energy mature economies have a significantly lower energy intensity (energy per unit of GDP) than they had during earlier stages of their development (box 6.10, fig. 6.17).

The most important lesson to be drawn from looking at long-term trends of per capita energy use and economic growth is that respectable rates of the latter can be achieved with progressively lower use of the former. In the United States a continuing if slow population growth has brought further increases in the absolute consumption of fuels and electricity, but the average per capita use of primary energy has been flat (with only minor fluctuations) for three decades, since the mid-1980s, yet the real GDP (in chained 2009 dollars) per capita gained nearly 57%, growing from $32,218 in 1985 to $50,456 in 2014 (FRED 2015). Similarly, in both France and Japan (where population is now declining) per capita primary energy use has stabilized since the mid-1990s—yet in the following two decades the average per capita GDP increased, respectively, by about 20% and 10%.

But these outcomes must be interpreted with caution as those periods of relative energy-GDP decoupling coincided with extensive offshoring of U.S., European, and Japanese energy-intensive heavy industries and manufacturing to Asia in general and to China in particular: it would be premature to conclude that the recent experience of those three major economies is a harbinger of a widespread decoupling trend. And mainly because of China's enormous growth in pre-2014 energy demand (achieving a nearly 4.5-fold increase since 1990), the global primary energy supply had to rise nearly 60% in order to produce a 2.8-fold rise in GWP during the 25 years after 1990 (an elasticity of 0.56). Moreover, declines in electricity intensity have been much slower than the declines in overall energy intensity. Between 1990 and 2015 the global drop was just short of 20% (compared to >40% for all energy), and the U.S. decline was also 20%, but rapidly modernizing China saw no decline between 1990 and 2015.

The primary energy (and electricity) intensity of global economic growth have been declining, but, because of the size of the world economy and the continuing population growth in Asia and Africa, the coming decades will repeat, though in a modified way, the past experience as large quantities of

Box 6.10
Declining energy intensity of economic growth

Historical statistics show a steady decline in the British energy intensity following the rapid rise brought by the adoption of steam engines and railways between 1830 and1850 (Humphrey and Stanislaw 1979). Canadian and U.S. intensities followed the declining British trend with a lag of 60–70 years. The U.S. rate peaked before 1920, the Chinese maximum was reached during the late 1970s, and India's energy intensity began to decline only in the twenty-first century (Smil 2003). Between 1955 and 1973 the U.S. energy intensity was flat (fluctuating just ±2%), while the real GDP grew 2.5-fold, but then it resumed its decline, and by 2010 US it was 45% below the 1980 level.

In contrast, the Japanese energy intensity was rising until 1970, but between 1980 and 2010 it declined by 25% (USEIA 2015d), and the Chinese decline has been particularly large, almost 75% between 1980 and 2013 (China Energy Group 2014), as much a reflection of exceedingly low efficiencies of early post-Mao China as of the modernization advances since 1980. On the other hand, India, still in an earlier stage of economic development, saw only a 7% drop between 1980 and 2010. These declines stem from a combination of several factors: the declining importance of energy-intensive capital inputs that characterize earlier stages of economic development, heavily focused on basic infrastructures; improved conversion efficiencies of combustion and electricity use; and the rising shares of the service sector (retail, education, banking), where adding value requires less energy per unit of GDP than in extractive industries or manufacturing.

Major differences in the national energy intensities of otherwise similarly accomplished economies are also explained by the composition of primary energy use (somebody must produce energy-intensive metals), the efficiency of final conversions (hydroelectricity is always superior to coal), climate, and the size of the territory (Smil 2003). With the United States at 100, the relative rates in 2011 were about 60 in Japan and Germany, 70 in Sweden, 150 in Canada, and 340 in China. Interestingly, Kaufmann (1992) showed that most of the post-1950 decline in energy intensity in affluent economies resulted from shifts in the kind of energies used and the type of dominant goods and services rather than from technical advances.

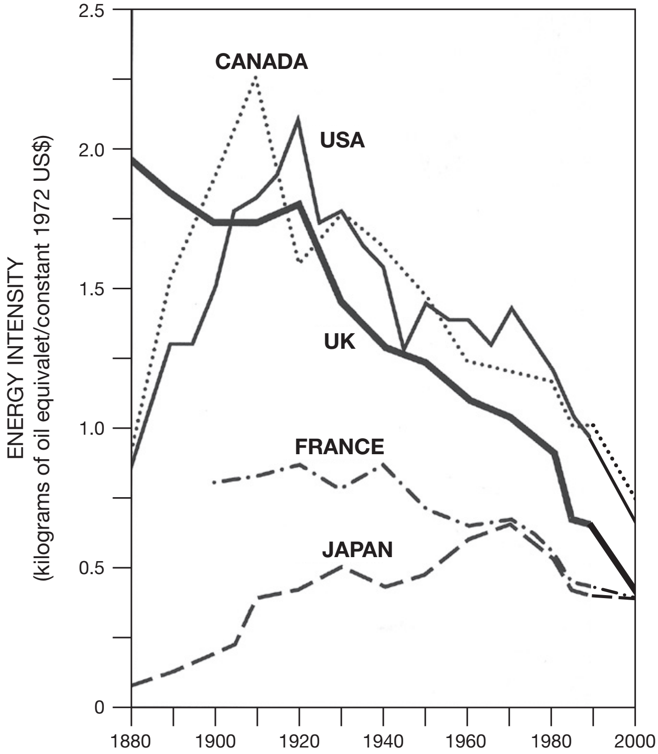


Figure 6.17

A declining energy intensity of the GDP has been a universal feature of maturing economies. Based on data in Smil 2003 and USEIA 2015d.

fuels and large additions of electricity-generating capacities will be required to energize economic growth in modernizing countries. Obviously, both the initiation and the maintenance of strong economic growth are matters of complex, interdependent inputs. They require technical improvements and responsive institutional arrangements, most notably sound banking and legal systems. Appropriate government policies, good educational systems, and a high level of competitiveness are also essential. But if today's low-income countries are to move from poverty to an incipient affluence (replicating China's post-1990 economic trajectory), then none of those factors could make a difference without the rising consumption of fuels and electricity: a decoupling of economic growth and energy consumption during early stages of modern economic development would defy the laws of thermodynamics.

Consequences and Concerns

The negative consequences of high energy use by modern societies range from obvious physical manifestations to gradual changes whose undesirable outcomes become apparent only after many generations. In the first category is an abundant food supply fostering indefensibly high food waste and contributing to unprecedented rates of overweight (a body mass index between 25 and 30) and obesity (body mass index >30). This trend toward heavier bodies is further reinforced by reduced energy expenditures, by more sedentary lifestyles resulting from mass replacement of muscle exertion by machines, and by the ubiquitous use of cars even for short trips that used to be made on foot. By 2012, 69% of the U.S. population was overweight or obese, up from 33% during the 1950s (CDC 2015), a clear proof that those conditions have been acquired through the combination of overeating and reduced physical activity.

The United States is hardly the only country with increasing shares of overweight and obese population (the rates are even higher in Saudi Arabia, and some of the fastest increases in excess weight are now found among Chinese children), but the trend is not (yet?) global: many European populations and most of the sub-Saharan Africa populations still have appropriate body masses. In any case, my intent is not to focus only on the negative impacts of intensive use of energy. Every one of the five fundamental global consequences of modern energy use I will examine has brought many welcome improvements along with effects whose worrisome impacts can be seen on scales ranging from local to global.

Continuing urbanization—since 2007, more than half of humanity has been living in cities—has been a major source of innovation. It has improved the physical quality of life and offers unprecedented opportunities for education and cultural exploits even as it has caused harmful levels of air and water pollution, led to excessive crowding, and created appalling living conditions for the poorest urban residents. High-energy societies enjoy a much higher standard of living than their traditional predecessors, and these gains have led to expectations of continued improvements: but because of persevering (and often deep) economic inequalities, these benefits have been unevenly distributed; moreover, there is no guarantee that further gains, requiring further deficit spending, will continue as populations age.

Energy prices, trade in fuels and electricity, and the security of energy supplies have become important political factors in both energy-importing and energy-exporting countries; in particular, periods of high and low oil prices have had major consequences for economies heavily dependent on hydrocarbon exports. The increased destructiveness of weapons and the increased risks of a nuclear conflict with truly global environmental and economic consequences have been accompanied by a widespread recognition of the futility of thermonuclear war and by steps to reduce the possibilities of such conflicts. And the massive combustion of fossil fuels has brought many negative environmental impacts, above all the risk of rapid global warming, and it will be very challenging to mitigate this threat.

Urbanization

Cities, even large cities, have a long history (Mumford 1961; Chandler 1987). Rome of the first century CE housed more than half a million people. Harun ar-Rashid's early ninth-century Baghdad had 700,000 people, and contemporary Changan (the capital of the Tang dynasty) had about 800,000 inhabitants. A thousand years later Beijing, the capital of the Qing dynasty, topped one million, and in 1800 there were about 50 cities above 100,000. But even in Europe no more than 10% of people lived in cities in 1800. The subsequent rapid increases in both the population of the world's largest cities and the overall shares of city dwellers would have been impossible without fossil fuels. Traditional societies could support only a small number of large cities because their energies had to come from croplands and woodlands that were at least 50 times and commonly about 100 times larger than the size of the settlement itself (box 6.11).

Modern cities use fuels with much higher efficiency, but their high concentrations of housing, factories, and transport push their power density to 15 W/m^2 in sprawling, warm-climate places and, in industrial cities in colder climates, up to 150 W/m^2 of their area. However, both coals and crude oils supplying these needs are extracted with power densities ranging usually between 1,000 and 10,000 W/m^2 (Smil 2015b). This means that an industrial city needs to rely on a coalfield or oil field whose size is no more than one-seventh and as little as 1/1,000th of its built-up area, and on new powerful prime movers to transport fuels from their basically punctiform places of extraction to urban users. While traditional cities had to be supported by the concentration of diffuse energy flows harvested over large areas, modern cities are supplied by the diffusion of fossil energies extracted in concentrated fashion from relatively small areas.

As far as food is concerned, a modern city of 500,000 people consuming daily 11 MJ/capita (with one-third coming from animal foods requiring, on the average, four times their energy value in feed) needs only about 70,000 ha to grow the crops, even when their mean yield would be just 4 t/ha. This

Box 6.11
Power densities of traditional urban energy supply and use

With average per capita food intakes of about 9 MJ/day originating, as preindustrial diets usually did, overwhelmingly (90%) from plant foods, and with typical grain yields of just 750 kg/ha, a traditional city of 500,000 people would have needed about 150,000 ha of cropland. In a colder climate, annual fuel (wood and charcoal) needs would have been about 2 t/capita. If supplied on a sustainable basis from forests or from fuelwood groves with annual yields of 10 t/ha, around 100,000 ha would have been needed to fuel the city. A densely populated city of that size occupied as little as 2,500 ha and had to rely on an area about 100 times its size for its food and fuel.

In terms of average power densities, this example implies about 25 W/m^2 for total energy consumption and 0.25 W/m^2 for the supply. The actual range of power densities was fairly large. Depending on their food intakes, cooking and heating practices, energy requirements for small manufactures, and combustion efficiencies, the total energy consumption of preindustrial cities was between 5 and 30 W/m^2 of their area. The sustainable production of fuel from nearby forests and woodlots yielded anywhere between 0.1 and 1 W/m^2. Consequently, cities had to rely on cropped and wooded areas 50–150 times larger than their own size—and the absence of powerful and inexpensive prime movers limited the capacity to transport food and fuel from distant regions, putting pressure on the plant resources of the surrounding areas (Smil 2015b).

would be less than half the total in the traditional city example, and fossil fuels and electricity also make large-scale, long-distance food imports affordable. And only electricity and liquid transportation fuels have made it possible to pump drinking water, remove and treat sewage and garbage, and meet the transportation and communication needs of megacities (cities with more than 10 million people). All modern cities are creations of fossil energy flows converted with high power densities, but megacities make exceptionally high claims: a survey by Kennedy and co-workers (2015) concluded that in 2011 the world's 27 megacities (with less than 7% of the global population) consumed 9% of all electricity and 10% of all gasoline.

The rise of fossil-fueled (initially just coal-fueled) cities was rapid. In 1800 only one of the world's ten largest cities, London (number two), was in a country whose energy use was dominated by coal. A century later nine out of ten were in that category: London, New York, Paris, Berlin, Chicago, Vienna, Saint Petersburg, Philadelphia, and Manchester, while Tokyo was the capital of a country where biomass fuels provided still about half of all primary energy (Smil 2010a). The worldwide share of urban population in 1900 was only about 15%—but it was far higher in the world's three largest coal producers. The rate was over 70% in UK, approaching 50% in Germany, and nearly 40% in the United States. The subsequent continuation of urban growth has also brought a remarkable increase in the total number of very large cities. By 2015 nearly 550 urban agglomerations surpassed one million inhabitants compared to 13 in 1900 and only two, Beijing and Greater London, in 1800 (City Population 2015).

Fossil fuels also energized the push and pull forces of migration: urban growth has been driven by the push of agricultural mechanization and by the pull of industrialization. Urbanization and industrialization are not, of course, synonymous, but the two processes have been closely tied by many mutually amplifying links. Most notably, technical innovation in Europe and North America had overwhelmingly urban origins, and cities continue to be the fonts of innovation (Bairoch 1988; Wolfe and Bramwell 2008). Bettencourt and West (2010) concluded that as the population of a city doubles. economic productivity goes up by an average of 130%, with both total and per capita productivity rising, and Pan and co-workers (2013) attributed this result largely to "superlinear scaling" as increases in urban population density give residents greater opportunity for face-to-face interaction.

The massive shift of urban jobs into service sectors is largely a post–World War II development. By 2015 these transfers had brought urban populations above 75% of the total not only in nearly all Western nations but also in Brazil and Mexico (respectively about 90% and 80%). Only in many African and Asian countries do urban shares of the population remain below 50%, with India at 35% and Nigeria at 47%, but China is at 55%. China's relatively low figure has been heavily influenced by decades of tightly controlled migration in Maoist China, with rapid urbanization beginning only in the 1990s. The economic, environmental, and social effects of these great human translocations have been among the most avidly studied phenomena of modern history. The misery, deprivation, filth, and disease common in rapidly growing nineteenth-century cities spawned a particularly vast literature. These writings ranged from primarily descriptive (Kay 1832) to largely indignant (Engels 1845) and from a series of parliamentary hearings to bestselling novels (Dickens 1854; Gaskell 1855).

Similar realities—minus the threat of most contagious diseases, now eliminated by inoculation—can be seen today in many Asian, African, or Latin American cities. But people are still moving in. Now as earlier, they are often leaving conditions that, on balance, were even worse, a fact commonly neglected both by the original reformist writings and by subsequent debates about the disadvantages of urbanization. Now as earlier, one must weigh the dismal state of urban environments—aesthetic affronts, air and water pollution, noise, crowding, dismal living conditions in slums—against their often no less objectionable rural counterparts.

Common rural environmental burdens include very high concentrations of indoor air pollutants (particularly fine particulate matter) from unvented biomass combustion, inadequate heating in colder climates, unsafe water supplies, poor personal hygiene, dilapidated, overcrowded housing, and minimal or no opportunities to see the children properly educated. Moreover, the drudgery of field labor in the open is seldom preferable even to unskilled industrial work in a factory. In general, typical factory tasks require lower energy expenditures than does common farm work, and in a surprisingly short time after the beginning of mass urban industrial employment the duration of factory work became reasonably regulated.

Later came progressively higher wages, in combination with such benefits as health insurance and pension plans. Together with better educational opportunities these changes led to appreciable improvements in typical standards of living. This led eventually to the emergence of a substantial urban middle class in all largely laissez-faire economies. The appeal of this great, although now certainly tarnished, Western accomplishment is felt strongly throughout the industrializing world. And it was undoubtedly an important factor in the demise of Communist regimes, which proved slow to deliver similar benefits. And there is no doubt about the consequence of urbanization for energy consumption; living in cities requires substantial increases in the per capita provision of energy even in the absence of heavy industries or large ports: the fossil fuels and electricity required to sustain a person who moved to one of Asia's new growing cities can be easily an order of magnitude higher than the meager amounts of biomass fuels used in the village of her birth to cook and (if need be) to heat a room.

Quality of Life

Rising energy consumption has been exerting usually gradual (but in some instances, as in post-1990 China, fairly abrupt) and largely desirable effects on the average quality of life—a term broader than standard of living as it also encompasses such key intangible variables as education and personal freedoms. During the decades of rapid post–World War II economic growth, many previously poor countries moved to the intermediate energy consumption category as their inhabitants improved their overall quality of life (though often at the price of concomitant environmental degradation), but the distribution of global energy use remains extremely skewed. In 1950 only about 250 million people, or one-tenth of the global population, living in the world's most affluent economies consumed more than 2 t of oil equivalent (84 GJ) a year per capita—yet they claimed 60% of the world's primary energy (excluding traditional biomass). By the year 2000, such populations numbered nearly a quarter of all mankind and claimed nearly three quarters of all fossil fuels and electricity. In contrast, the poorest quarter of humanity used less than 5% of all commercial energies (fig. 6.18).

By 2015, thanks to China's rapid economic growth, the share of the global population consuming more than 2 t of oil equivalents jumped to 40%, the greatest equalization advance in history. Stunning as they are, these averages do not capture the real differences in the average quality of life because poor countries devote a much smaller share of their total energy

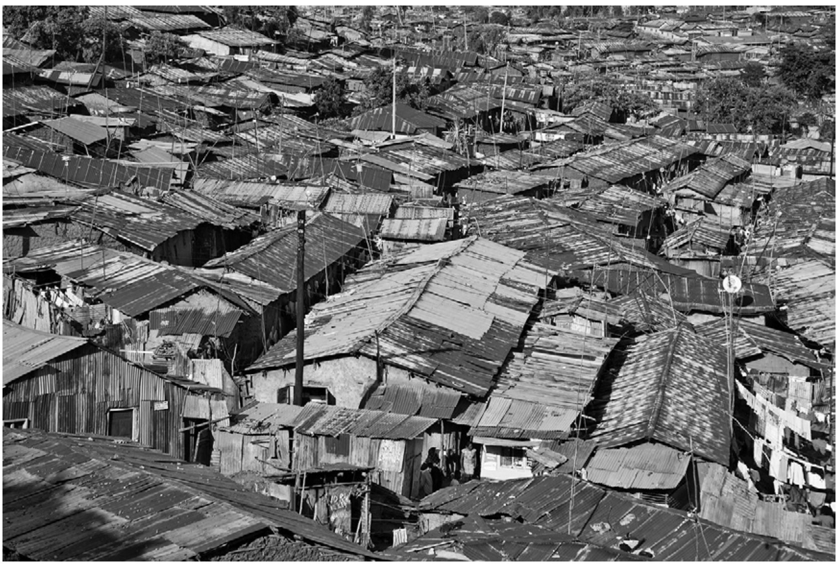


Figure 6.18
Kibera, one of Nairobi's largest slums (Corbis). Kenya's per capita use of modern energies averages about 20 GJ/year, but slum dwellers in Africa and Asia consume as little as 5 GJ/year, or less than 2% of the U.S. mean.

consumption to private household and transportation uses and convert those energies with lower efficiency. The real difference in typical direct per capita energy use among the richest and the poorest quarters of the mankind is thus closer to being 40-fold rather than "just" 20-fold. This enormous disparity is one of the few main reasons for the chronic gap in economic achievements and in the prevailing quality of life. In turn, these inequalities are a major source of persistent global political instability.

Those countries that have made it into the intermediate consumption category have gone through similar stages of improvements but at a very different pace: what took the early industrializers of Western Europe two or even three generations has been recently accomplished in South Korea and China in a single generation of compressed development (an advantage of determined late-starters). In the early stages of economic growth, these benefits are rather limited because fuels and electricity are overwhelmingly channeled into building up an industrial base. The slowly increasing acquisition of household and personal goods and better basic diets have been the first signs of improvement, starting in the cities and gradually diffusing to the countryside.

Among the first gains are a greater variety and better quality of basic cookware, dishes, and utensils; more, and usually more colorful, pieces of clothing; better shoes; better personal hygiene (more frequent washing and laundering); purchases of additional pieces of furniture; purchases of small gifts for special occasions; and pictures (starting with cheap reproductions) on walls. In North America and Europe of the early twentieth century possession of an increasing range of electrical appliances came during the next stage of *embourgeoisement,* but the low cost of new electric (air conditioning, microwave ovens, TVs) and electronic appliances and devices (above all mobile phones) means that in many Asian and in some African countries, families acquired them even before they owned other better household items.

The next stage sees further improvements in the variety and quality of the food supply and better health care, and the progress starts spilling into the countryside. The educational level of urban populations begins to rise, and there are increasing signs of incipient affluence, including car ownership, new house comforts, and travel abroad for people in higher income groups. Again, some of these gains have been recently conflated or inverted, particularly in Asia. Eventually comes the stage of mass consumption with its many physical comforts and frequent ostentatious displays. Longer periods of schooling, high personal mobility, and growing expenditures on leisure and health are part of this change.

Correlations of this sequence with average per capita energy consumption have been unmistakable, but what is usually compared—average per capita consumption calculated by aggregating a nation's primary energy supply and dividing it by the population total—is not the best variable. The per capita average consumption of the total primary energy supply tells us nothing either about the consumption breakdown (the military may claim a disproportionately large amount, as it did in the USSR and as it does in North Korea and Pakistan) or about the typical (or average) efficiency of energy conversions (higher, and hence delivering more final services per unit of gross energy, in Japan than in India). Better insights might come from comparing average rates of residential energy consumption, but that tack, too, is hardly perfect: fuels and electricity consumed by households will count, but considerable indirect energy inputs (required to build houses or to manufacture cars, household appliances, electronics, and furniture) are excluded.

Keeping this mind, and also realizing that national peculiarities (from climatic to economic singularities) preclude any simple classification, the relationship between energy use and quality of life can be divided into three basic categories. No country whose annual primary commercial energy consumption (leaving aside traditional biofuels) averages less than 5 GJ/capita (that is, about 120 kg of oil equivalent) can guarantee even basic necessities to all its inhabitants. In 2010 Ethiopia was still well below that minimum, Bangladesh barely above it; China was there before 1950, as were large parts of Western Europe before 1800.

As the rate of commercial energy use approaches 1 t of oil equivalent (42 GJ), industrialization advances, incomes rise, and the quality of life improves noticeably. China of the 1980s, Japan of the 1930s and again of the 1950s, and Western Europe and the United States between 1870 and 1890 are all examples of this stage of development. Incipient affluence requires, even with fairly efficient energy use, at least 2 t of oil equivalent (84 GJ) per capita per year. France made it during the 1960s, Japan during the 1970s. China reached that level by 2012, but its rate is not fully comparable with the Western rates because too much of its energy use still goes into industry (almost 30% in 2013), too little for private discretionary energy use (IEA 2015a).

But both the French and the Chinese gains illustrate the speed of recent changes. The French census of 1954 revealed the striking deficiencies in housing: less than 60% of households had running water, only 25% had an indoor toilet, and only 10% had a bathroom and central heating (Prost 1991). By the mid-1970s refrigerators were in almost 90% of households, toilets in 75%, 70% had bathrooms, and about 60% enjoyed central heating and washing machines. By 1990 all these possessions became virtually universal, and 75% of all families also owned a car, compared to fewer than 30% in 1960. Such growing affluence had to be reflected in a rising use of energy. Between 1950 and 1960 the average French per capita energy consumption rose by about 25%, but between 1960 and 1974 it soared by over 80%; and while between 1950 and 1990 the per capita supply of all fuels more than doubled, gasoline consumption rose nearly sixfold and electricity use went up more than eightfold (Smil 2003).

Even faster advances have taken place in China. In 1980, when the economic reforms started (four years after Mao Zedong's death), per capita energy consumption averaged about 19 GJ; by 2000 it was nearly 35 GJ; in 2010, after quadrupling in three decades, it was roughly 75 GJ; and in 2015 it was just above 90 GJ (Smil 1976; China Energy Group 2015), a level comparable to the Spanish mean during the early 1980s. Moreover, disproportionate shares of these gains have been used in construction. Nothing indicates this better than this fact: while the US consumption of cement added up to about 4.5 Gt during the entire twentieth century, China emplaced more of it (4.9 Gt) in its new construction projects in just the three years of 2008–2010 (Smil 2014b). No wonder that the country now has the world's largest modern networks of high-speed railways and interprovincial freeways.

No other form of energy has had a more wide-ranging impact on rising quality of life than the provision of affordable electricity: on the personal level the effects have been pervasive and life-spanning (premature babies are kept in incubators, vaccines to inoculate them are kept in refrigerators, dangerous illness are diagnosed by noninvasive techniques in time to be treated, the critically ill are hooked up to electronic monitors). But one of electricity's most consequential social impacts has been to transform many chores of household work and hence to disproportionately benefit women. This change has been, even in the Western world, fairly recent.

For generations, a rising energy consumption made little difference for everyday household work. Indeed, it could make it worse. As the standards of hygiene and social expectations rose with better education, women's work in Western countries often got harder. No matter if it was washing, cooking, and cleaning in cramped English apartments (Spring-Rice 1939) or doing daily chores in American farmhouses, women's work was still exceedingly hard during the 1930s. Electricity was the eventual liberator. Regardless of the availability of other energy forms, it was only the introduction of electricity that did away with exhausting and often dangerous labor (Caro 1982; box 6.12).

Many electric appliances were available already by 1900: during the 1890s General Electric was selling electric irons, fans, and an immersion water heater coil that could boil a pint of water in 12 minutes (Electricity Council 1973). The high cost of these appliances, limited house wiring, and slow progress in rural electrification delayed their widespread adoption, both in Europe and in North America, until the 1930s. Refrigeration has been a more important innovation than gas or electric cooking (Pentzer 1966). The first home refrigerators were marketed by Kelvinator Company in 1914. American ownership rose sharply only during the 1940s, and refrigerators became common in Europe only after 1960. Their importance has increased with the growing reliance on fast food. Refrigeration now accounts for up to 10% of all electricity used in the households of rich nations.

Electricity's conquest of household services continues to bring further time and labor savings in rich countries. Self-cleaning ovens, food processors and microwave cooking (developed in 1945, but introduced in small household models only during the late 1960s) have become common throughout the rich world. Refrigerator, washing machine, and microwave

Box 6.12
Importance of electricity for easing housework

The liberating effects of electricity are unforgettably illustrated in Robert Caro's (1982) first volume of Lyndon Johnson's biography. As Caro points out, it was not the shortage of energy that made life in Texas Hill County so hard (households had plenty of wood and kerosene) but the absence of electricity. In a moving, almost physically painful, account Caro describes the drudgery, and danger, of ironing with heavy wedges of metal heated on wood stoves, the endless pumping and carrying of water for cooking, washing, and animals, the grinding of feed, and sawing wood. These burdens, which fell largely on women, were much harder than the typical labor requirements in poor countries as the Hill County farmers of the 1930s strove to maintain a much higher standard of life and run much larger farming operations than peasants in Asia or Latin America. For example, the water needs for a family of five came to nearly 300 t/year, and to supply them required an equivalent of more than 60 eight-hour days and walking about 2,500 km. Not surprisingly, nothing could have been as revolutionary in the life of these people as the extension of transmission lines.

ownership has also approached saturation levels among better-off segments of Asian and Latin American populations, and they also have a high ownership of air conditioning units. Patented first by Willis Carrier (1876–1950) in 1902, air conditioning was limited for decades to industrial applications. The first units scaled down for household use came during the 1950s in the United States, and their widespread adoption opened up the American Sun Belt to mass migration from northern states and increased the appeal of subtropical and tropical tourist destinations (Basile 2014). Household air conditioners are now also used widely in urban areas of hot-weather countries, most of them being single-room wall units (fig. 6.19).

Modern societies have elevated economic growth, and hence rising energy use, to the level of unquestioned desiderata, implicitly assuming

Figure 6.19
A Shanghai high-rise apartment building with air conditioners for virtually every room (Corbis).

that using more will always have its rewards. But economic growth and rising energy use should be seen only as the means of securing a better quality of life, a concept that includes not only the satisfaction of basic physical needs (health, nutrition) but also the development of the human intellect (ranging from basic education to individual freedoms). Such an inherently multidimensional concept cannot be contracted into a single representative indicator, but it turns out that a few variables serve as its sensitive markers.

Infant mortality (deaths/1,000 live births) and life expectancy at birth are two obvious and unambiguous indicators of the physical quality of life. Infant mortality is an excellent proxy for conditions ranging from disposable income and quality of housing to the adequacy of nutrition, level of education, and a state's investment in health care: very few babies die in countries where families live in good housing and where well-educated parents (themselves well nourished) feed them properly and have access to medical care. And, naturally, life expectancy quantifies the long-term effects of these critical factors. Education and literacy data are not as revealing: enrollment ratios tell us about access but not about quality, and detailed achievement studies (such as the OECD's Programme for International Student Assessment, or PISA) are not available for most countries. Another option is to use the UNDP's Human Development Index (HDI), which combines life expectancy at birth, adult literacy, combined educational enrolment, and per capita GDP.

Comparing these measures with average energy use leads to some important conclusions. Some societies have been able to secure adequate diets, basic health care and schooling, and a decent quality of life with an annual energy use as low as 40–50 GJ/capita. Relatively low infant mortalities, below 20/1,000 newborns; relatively high female life expectancies, above 75 years; and an HDI above 0.8 could be achieved with 60–65 GJ/capita, while the world's top rates (infant mortality below 10/1,000 newborns, female life expectancies above 80, HDI > 0.9) require at least 110 GJ/capita. There is no discernible improvement in fundamental quality of life above that level.

Energy use is thus related to quality of life in a fairly linear manner only during the lower stages of development (going from quality of life in Niger to quality of life in Malaysia). Plotted values show distinct inflections of the best-fit lines at between 50 and 70 GJ/capita, followed by diminishing returns, topped by a plateau above (depending on the studied quality-of-life variable) 100–120 GJ/capita (fig. 6.20). This means that the effect of energy consumption on improving quality of life—measured by variables that

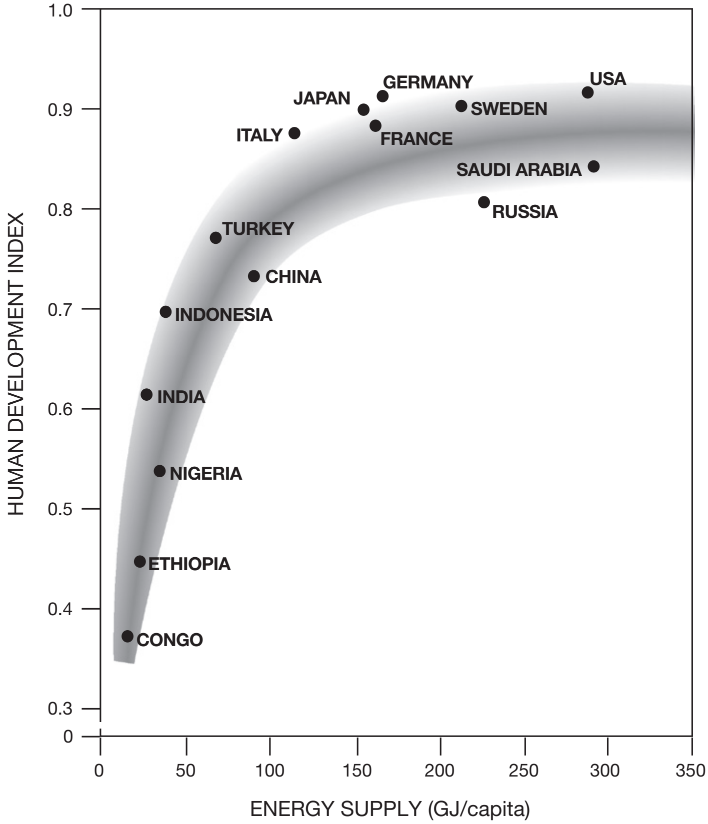


Figure 6.20
Average per capita energy consumption and the human development index in 2010. Plotted from data in UNDP (2015) and World Bank (2015a).

truly matter, not by the ownership of yachts—reaches a saturation level well below the rates of energy use prevailing in affluent countries, with the leading EU economies and Japan at about 150 GJ/capita, Australia at 230 GJ/capita, the United States at 300 GJ/capita, and Canada at about 385 GJ/capita in 2015 (BP 2015). Additional increases in discretionary energy use go into ostentatious housing (as average family size has declined, the average size of U.S. houses has more than doubled since the 1950s), the ownership of multiple expensive vehicles, and frequent flying.

More remarkably, America's high energy use has been accompanied by quality-of-life indicators that are inferior not only when compared with the performance of leading EU countries or Japan (whose energy use is only half the U.S. rate) but when compared with the performance of many countries with intermediate energy use. In 2013 the United States, with 6.6 of every 1,000 live-born babies dying in the first year of life, ranked 31st worldwide, below not only France (3.8), Germany (3.5), and Japan (2.6) but also more than twice as high as Greece's infant mortality (CDC 2015). Even worse, in 2013 America's life expectancy ranked 36th worldwide, with an average of 79.8 years for both sexes, which was hardly better than in Castro's Cuba (79.4) and behind the life expectancy of Greece, Portugal, and South Korea (WHO 2015a).

The educational achievements of students in OECD countries are regularly assessed by PISA,, and the latest results show America's 15-year-old ranking just below that of Russia, Slovakia, and Spain and far lower than that German, Canadian, or Japanese teenagers (PISA 2015). In science, U.S. children were just below the mean OECD score (497 vs. 501); in reading they were barely above the mean (498 vs. 496), and far behind all populous affluent Western nations. PISA, much like any such study, has its weaknesses, but the large differences in relative rankings are clear: there is not the slightest indication that America's high energy use has any beneficial effect on the country's educational achievements.

Political Implications

The dependence of modern societies on incessant, reliable, and inexpensive supplies of fossil fuels and electricity (delivered at required, and now invariably massive, rates) has generated a multitude of political concerns and responses, domestic and foreign. Perhaps the most universal concern is the concentration of decision-making power resulting from higher levels of integration, be it in government, business, or the military. As Adams (1975, 120–121) noted, when "more energetic processes and forms enter a society, control over them becomes disproportionately concentrated in the hands of a few, so that fewer independent decisions are responsible for greater releases of energy."

But far greater perils arise when these concentrated controls become superconcentrated in a single individual who decides to use them in an aggressive and destructive manner. Their misdirection can result in enormous human suffering, the prodigious waste of labor and resources, damage to the environment, and the destruction of a cultural heritage. Examples of such excessive concentration of control to unleash destructive forces have been a recurrent phenomenon in history; if measured only in human casualties, then the decisions made by the Spanish kings of the sixteenth century, by Napoleon Bonaparte (1769–1821), by Kaiser Wilhelm II

(1859–1941), or by Adolf Hitler (1889–1945) resulted in millions of deaths. The Spanish *conquista* of the Americas had eventually led, directly (through battle death and enslavement) and indirectly (through infectious diseases and famines) to the deaths of tens of millions (López 2014); Napoleon's serial aggression cost at least 2.5 million and as many as 5 million lives (Gates 2011); Prussian aggression was the proximate cause of more than 17 million World War I deaths; and the total death toll in World War II, military and civilian, approached 50 million (War Chronicle 2015).

But the unchallenged decisions of the two Communist dictators who could convert their manias into terrible realities with greater flows of fossil fuels and electricity are the unsurpassed epitomes of the perils of concentrated control. In 1953, the year of Stalin's death, the USSR's energy use was more than 25 times the total in 1921 when the country emerged from its civil war (Clarke and Dubravko 1983). Yet the generalissimo's paranoia led to the deaths of tens of millions in massive purges, the resettlement of entire populations (Crimean Tatars, Volga Germans, Chechens), the gulag empire, and the economic prostration of the world's potentially richest nation; the total number of deaths will be never tallied accurately, but it is at least on the order of 15–20 million (Conquest 2007).

Similarly, on Mao Zedong's death in 1976, China's energy production was more than 20 times the 1949 total (Smil 1988). But the Great Helmsman's delusions brought successive waves of death in the Great Leap Forward, followed by the worst famine in human history—between 1959 and 1961 more than 30 million Chinese died (Yang 2012)—and then the destruction of the Cultural Revolution. Again, no accurate total will be ever known, but the total of 1949–1976 deaths could be close to 50 million (Dikötter 2010). And while the probability of the ultimate threat—a thermonuclear war between great powers—has been reduced, thanks to the reduction of American and Russian warhead arsenals, its possibility remains, and the decision to launch would be made, on either side, by a very small group of people.

There has been no better example of the global political and economic consequences of concentrated controls of energy flows than the decisions made by the Organization of Petroleum Exporting Countries since 1973. Given the importance of crude oil in modern economies and the dominance of the global export market by a few Middle Eastern countries, it is inevitable that any decisions of a few individuals, particularly those in Saudi Arabia whose enormous oil-production capacity dominated OPEC's directions, will have profound consequences for global prosperity. OPEC's dissatisfaction with low royalties and the ensuing quintupling of world oil prices in 1973–1974, and their further near quadrupling in 1979–1980, ushered in, and then further deepened, a period of worldwide economic dislocation marked by high inflation and significantly reduced economic growth (Smil 1987; Yergin 2008).

In response, all major Western importers and Japan set up emergency energy-sharing agreements coordinated by the International Energy Agency, mandated the establishment of strategic petroleum reserves (some countries had also promoted closer bilateral ties with OPEC nations), and subsidized the quest for domestic fuel self-sufficiency by promoting alternative sources of energy. France's development of nuclear electricity and Japan's energy conservation effort have been especially remarkable and effective. But China's rapid economic rise—the country became a net oil importer in 1994—and the declining output of traditional oil fields, whether in Alaska or in the North Sea, were key reasons behind another rise in world oil price to a record level of about $145/bbl in July 2008, a run-up that ended only with the economic crisis in the fall of 2008 and with the oil price barely above $30/bbl in December 2008.

As the economies recovered and the Chinese demand continued to rise, oil prices rose once again above $100/bbl in July 2014, but then the falling demand and rising supply (mainly owing to the reemergence of the United States as the world's largest producer, thanks to rapid increases in shale oil production through hydraulic fracturing) brought a deep reversal. But this time there was a key difference: in order to protect the country's global market share, Saudi leaders decided to keep producing at maximum output rather than, as in the past, cutting output and propping up the price. Once again, the decisions made by a few men has worldwide consequences for the political stability of countries heavily dependent on oil exports, as well as for all major non-OPEC oil producers, including the United States and Canada.

Falling oil prices have once again brought expectations of OPEC's near demise—but the peculiarities of the highly uneven distribution of crude oil reserves (a leading strategic concern of the twentieth century that has not lost its importance in the twenty-first century) remain in favor of the Middle Eastern producers. The Persian Gulf basin is an unparalleled singularity: it has 12 of the world's 15 largest oil fields, and in 2015 it contained about 65% of the world's reserves of liquid oil (BP 2015). These riches explain the lasting interest in the region's stability. This desire is immensely complicated by the near chronic disarray of the area, which is made up of artificial states separated by arbitrary borders cutting across ancient ethnic group and containing complex religious enmities.

Notable post–World War II outside involvements in the region started with the Soviet attempt to take over northern Iran (1945–1946). Americans had landed twice in Lebanon, in 1958 and 1982, when their resolve was broken by a single terrorist bombing of the Beirut barracks in 1983 (Hammel 1985). Western countries armed Iran heavily (before 1979, during the last decade of Shah Reza Pahlevi's reign) and also Saudi Arabia, and the Soviets did the same with Egypt, Syria, and Iraq. The Western tilt (weapons, intelligence, and credit) benefited Iraq during the Iraq-Iran War (1980–1988). The pattern of intervention culminated in the Desert Shield and Desert Storm Operations of 1990–1991, a massive response by the U.S.-led, UN-sanctioned alliance assembled to reverse the Iraqi invasion of Kuwait (CMI 2010).

By that move, Iraq doubled the oil reserves under its control, pushing them to about 20% of the global total. The Iraqi advance seriously threatened the nearby Saudi oil fields, and perhaps even the very existence of the monarchy, which controlled one quarter of the world's oil reserves. But after the swift defeat, Saddam Hussein remained in power, and after the events of 9/11, fears of further aggression (misplaced, as proven later when no weapons of mass destruction were found in Iraq) led to the U.S. occupation of Iraq in March 2003, which was followed by years of internal violence and the eventual loss of part of the country to the so-called Islamic State. But later in this chapter I will argue, agreeing with Lesser (1991), that resource-related objectives, seemingly so paramount in the Middle Eastern conflicts, have historically been determined by broader strategic aims, not vice versa. And the failure of the Arab OPEC nations to turn oil into a political weapon (enacting an oil embargo against the United States and the Netherlands in the wake of the October 1973 Arab-Israeli Yom Kippur War) was not the first instance of using the energy supply to carry an ideological message.

The symbolic power of electric light was exploited by such diverse actors as large U.S. companies and Germany's Nazi party. American industrialists displayed the power of light for the first time during the 1894 Columbian Exposition in Chicago, and then by flooding the downtowns of large cities with "White Ways" (Nye 1992). The Nazis used walls of light to awe the participants at massed party rallies of the 1930s (Speer 1970). Electrification became the embodiment of such disparate political ideals as Lenin's quest for a Communist state form and Franklin Roosevelt's New Deal. Lenin summarized his goal in a terse slogan, "Communism equals the Soviet power plus electrification," and the Soviet preference for building giant hydroelectric projects was kept alive after the USSR's demise in post-Mao China.

Roosevelt used federal involvement in building dams and electrifying the countryside as a means of economic recovery, some of it in the country's most backward regions (Lilienthal 1944).

Weapons and Wars

Weapons production has become a leading industrial activity, one now heavily supported by advanced research, and all major economies have also become large-scale exporters of armaments. Only a fraction of these expenditures could be justified by real security needs, and waste and misallocation of investments and skilled manpower—most notably the development of weapons irrelevant for new forms of warfare (massed tank warfare is hardly the best response to jihadi terrorism)—have marked the history of modern weapon procurement. Not surprisingly, many technical advances brought about by new fuels and new prime movers were rapidly adapted for destructive uses. First they increased the power and the effectiveness of existing techniques. Later they made it possible to design new classes of weapons capable of unprecedented reach, speed, and destruction.

The culmination of these efforts came with the construction of enormous nuclear arsenals and with the deployment of intercontinental ballistic missiles capable of reaching any target on Earth. Accelerated destructiveness of modern weapons is well illustrated by contrasting the typical mid-nineteenth-century and the mid-twentieth-century weapons with their predecessors half a century before. The two principal classes of weapons used during the American Civil War (1861–1865), infantry muskets and 12-pound guns (both muzzle-loading with smooth bores), would have been quite familiar to the veterans of the Napoleonic Wars (Mitchell 1931). In contrast, among the weapons that dominated the battlefields of World War II—tanks, fighter and bomber planes, aircraft carriers, and submarines—only the last ones existed, and just in early experimental stages, during the 1890s. A revealing way to illustrate the energetic dimension of these developments is to compare the actual kinetic and explosive power of commonly used weapons.

As the basis for the first kind of comparison, it is useful to recall (as shown in chapter 4) that the kinetic energy of the two most common handheld weapons of the preindustrial era, arrows (shot from bows) and swords, was merely on the order of 10^1 J (mostly between 15 and 75 J), and that an arrow released from a heavy crossbow may hit a target with 100 J of kinetic energy. In contrast, bullets shot from the muzzles of muskets and rifles would have kinetic energies on the order of 10^3 J (10 to 100 times higher), while the shells fired from modern guns (including those mounted on tanks) rate at 10^6 J. The calculations for half a dozen specific weapons are shown in box 6.13: the values for gun shells are only the kinetic energies of projectiles and exclude the energies of the explosives they may or may not carry.

Rockets and missiles, propelled by solid or liquid fuels, cause most of their damage by targeted explosion of their warheads, not by their kinetic energy, but when the first (unguided) World War II German V-1 missiles failed to explode, the kinetic energy of their impact was 15–18 MJ. And the most famous recent example using an object with high kinetic energy to inflict extraordinary damage was the steering of large Boeing aircraft (767 and 757) into the World Trade Center skyscrapers by jihadi hijackers on September 11, 2001. The towers were actually designed to absorb a jetliner impact, but only of a slow-flying (80 m/s) Boeing 707 that might get lost on its approach to Newark, La Guardia, or JFK airport. The Boeing 767–200 is only about 15% heavier than was the 707, but the plane hit the tower at no less than 200 m/s, and hence its kinetic energy was more than seven times higher (about 3.5 GJ vs. roughly 480 MJ).

Even so, the structures were not brought down by the impact as the airplanes acted much as bullets hitting a massive tree: they could not push the massive structure, but they penetrated it by first destroying exterior columns. Karim and Fatt (2005) showed that 46% of the initial kinetic energy of the aircraft was used to damage the exterior columns, and that they would not have been destroyed if they had had a minimum thickness of 20 mm. The collapse of the towers was thus caused by the burning of fuel (more than 50 t of kerosene, or 2 TJ) and the building's interior combustibles, which caused thermal weakening of structural steel and nonuniform heating of the long floor joists, which precipitated the staggered floor

Box 6.13

Kinetic energy of projectiles propelled by explosives

Weapon	Projectile	Kinetic energy (J)
Civil war musket	Bullet	1×10^3
Assault rifle (M16)	Bullet	2×10^3
Eighteenth-century cannon	Iron ball	300×10^3
World War I artillery gun	Shrapnel shell	1×10^6
World War II heavy AA gun	High-explosive shell	6×10^6
M1A1 Abrams tank	Depleted U shell	6×10^6

collapse and led to the free-fall speed, as the towers fell in only about 10 s (Eagar and Musso 2001).

The explosive power of modern weapons began to rise with the invention of compounds more powerful than gunpowder: they, too, are self-oxidizing, but their high detonation velocities create a shock wave. This new class of chemicals was prepared by the nitration of such organic compounds as cellulose, glycerine, phenol, and toluene (Urbanski 1967). Ascanio Sobrero prepared nitroglycerin in 1846 and J. F. E. Schultze introduced nitrocellulose in 1865, but nitroglycerin's practical use was made possible only by Alfred Nobel's two inventions: mixing the compound with diatomaceous earth (an inert porous substance) to create dynamite and the introduction of a practical detonator, the Nobel igniter (Fant 2014).

Depending on the composition, gunpowder's detonation velocity can be only a few hundred m/s, while dynamite's is up to 6,800 m/s. Trinitrotoluene (TNT) was synthesized by Joseph Wilbrand in 1863 and was used as an explosive (detonation velocity of 6,700 m/s) by the end of the nineteenth century, while the most powerful prenuclear explosive, cyclonite (cyclotrimethylenetrinitramine or RDX, Royal Demolition eXplosive, detonation velocity of 8,800 m/s), was first made by Hans Henning in 1899. These explosives have been used ever since in gun shells, mines, torpedoes, and bombs, and in recent decades also strapped to the bodies of suicide bombers. But many terrorist attacks using car and truck bombs have been made just with a mixture of a common fertilizer (ammonium nitrate) and fuel oil: ANFO consists 94% of NH_4NO_3 (as an oxidizing agent) and 6% of fuel oil, both readily available ingredients whose effect is the result of the mass of the explosive used, not of any extraordinary detonation velocity (box 6.14).

The combination of better propellants and high-quality steels increased the range of field and naval guns from less than 2 km during the 1860s to

Box 6.14
Kinetic energy of explosive devices

Explosive device	Explosive	Kinetic energy (J)
Hand grenade	TNT	2×10^6
Suicide bomber with a belt	RDX	100×10^6
World War II gun shrapnel	TNT	600×10^6
Truck bomb (500 kg)	ANFO	2×10^9

over 30 km by 1900. The combination of long-range guns, heavy armor, and steam turbines for naval propulsion made it possible to build new heavy battleships: the HMS *Dreadnought*, launched in 1906, was their prototype (Blyth, Lambert, and Ruger 2011). The ship was powered by steam turbines (introduced by the Royal Navy in 1898), as were all of the largest passenger ships of the pre–World War I years, starting in 1907 with the *Mauretania* and *Lusitania*, and as are today the U.S. nuclear aircraft carriers of the Nimitz class (Smil 2005). Other notable pre–World War I destructive innovations included machine guns, submarines, and the first prototypes of military planes. The horrible trench stalemates of World War I were sustained by the massive deployment of heavy field guns, machine guns, and mortar launchers. Neither poisonous gases (first used in 1915) nor the first extensive use of fighter planes and tanks (in 1916, heavily only in 1918) could break the hold of that massive firepower deployed in frontal attacks (Bishop 2014).

The interwar years saw the rapid development of tanks and fighter and bomber planes. All-metal bodies replaced the early wood-canvas-wire construction, and the first purpose-built aircraft carriers came in 1922 (Polmar 2006). These weapons launched the aggression of World War II. Early German successes were largely a matter of rapid tank-led penetrations, and Japan's surprise attack on Pearl Harbor on December 7, 1941, could be made only with long-range fighters (the Mitsubishi A6M2 Zero, range of 1,867 km) and bombers (the Aichi 3A2, range of 1,407 km, and the Nakajima B5N2, range of 1,093 km) launched from a large carrier force (Hoyt 2000; National Geographic Society 2001; Smith 2015).

The same classes of weapons were essential in defeating the Axis powers. First it was a combination of excellent fighter planes (Supermarine Spitfires and Hawker Hurricanes) and radar during the Battle of Britain in August and September 1940 (Collier 1962; Hough and Richards 2007). Then came America's effective use of carrier planes (starting with the pivotal Battle of Midway in 1942), and the crushing Soviet tank superiority (model T-42) during the Red Army's westward thrust. The postwar arms race began during the war with the development of jet propulsion, the firing of German ballistic missiles (the V-2 was first used in 1944), and the explosion of the first nuclear bombs, the Trinity, New Mexico, test on July 11, the Hiroshima bombing on August 6, 1945, and the Nagasaki bombing three days later. The total energy released by these first nuclear bombs was orders of magnitude above that of any previous explosive weapons—but also orders of magnitude below those of subsequent hydrogen bomb designs.

The first modern field gun, the French *canon 75 mm modèle 1897*, fired shells filled with nearly 700 g of picric acid, whose explosive energy reached 2.6 MJ (Benoît 1996). Perhaps the best-known gun of World War II was the German anti-aircraft FlaK *(Flugzeugabwehrkanone*) 18, whose variant was also used in Tiger tanks (Hogg 1997); it fired shrapnel shells whose explosive energy was 4 MJ. But the most powerful explosives of World War II were the massive bombs that were dropped on cities. The most powerful bomb carried by the Flying Fortress (Boeing B-17) had an explosive energy of 3.8 GJ. But the greatest damage was done by dropping incendiary bombs on Tokyo on March 9–10, 1945 (box 6.15, fig. 6.21).

The Hiroshima bomb released 63 TJ of energy, about half of it as the blast and 35% as thermal radiation (Malik 1985). These two effects caused a large number of instant deaths, while ionizing radiation caused both instant and delayed casualties. The bomb exploded at 8:15 a.m. on August 7, 1945, about 580 m above ground; the temperature at the point of explosion was several million degrees Centigrade, compared to 5,000°C for conventional

Box 6.15
Firebombing of Tokyo, March 9–10, 1945

The raid, the largest of its kind ever, involved 334 B-29 bombers that offloaded their bombs at low (about 600–750 m) altitude (Caidin 1960; Hoyt 2000). Most of these were large, 230 kg cluster bombs, each releasing 39 M-69 incendiary bombs filled with napalm, a mixture of polystyrene, benzene, and gasoline (Mushrush et al. 2000); simple 45 kg jelled-gasoline and phosphorus bombs were also used. About 1,500 t of incendiary compounds were dropped on the city, and their total energy content (assuming an average density of napalm of 42.8 GJ/t) amounted to about 60 TJ, nearly as much as the power of the Hiroshima bomb.

But energy released by burning napalm was only a small fraction of the total released by the city's incinerated wooden buildings. According to the Tokyo Metropolitan Police Department, the fire destroyed 286,358 buildings and structures (U.S. Strategic Bombing Survey 1947), and conservative assumptions (250,000 wooden buildings, just 4 t of wood per building, 18 GJ/t of dry lumber) result in in some 18 PJ of energy released from the combustion of the city's wooden housing, two orders of magnitude (300 times) larger than the energy of incendiary bombs. The destroyed area amounted to about 4,100 ha, and at least 100,000 people died. For comparison, the totally destroyed area in Hiroshima was about 800 ha, and the best estimate of immediate deaths was 66,000.

Figure 6.21
Aftermath of Tokyo bombing of March 1945 (Corbis).

explosives. The fireball expanded to its maximum size of 250 m in one second, the highest blast velocity at the hypocenter was 440 m/s, and the maximum pressure reached was 3.5 kg/cm^2 (Committee for the Compilation of Materials 1991). The Nagasaki bomb released about 92 TJ.

These weapons appear minuscule compared to the most powerful thermonuclear bomb, tested by the USSR over Novaya Zemlya on October 30, 1961: the *tsar bomba* released 209 PJ of energy (Khalturin et al. 2005). Less than 15 months later Nikita Khrushchev revealed that Soviet scientists had built a bomb that was twice as powerful. Comparisons of explosive powers are usually done not in joules but in units of TNT equivalents (1 t TNT = 4.184 GJ): the Hiroshima bomb was equivalent to 15 kt TNT, the *tsar bomba* to 50 Mt TNT. Typical warheads on intercontinental missiles have a power of between 100 kt and 1 Mt, but up to 10 of them can be carried by such missiles as the U.S. submarine-launched Poseidon or the Russian SS-11. To emphasize the magnitudes of energy release I do not use scientific notation (exponents) in the staggering ladder of maximum destructivity of explosive weapons (box 6.16).

Box 6.16
Maximum energy of explosive weapons

Year	Weapon	Energy (J)
1900	Picrite-filled shell from French 75 mm modèle 1897 gun	2,600,000
1940	Amatol/TNT-filled shrapnel from German 88 mm FlaK	4,000,000
1944	The largest bomb carried by the Boeing B-17	3,800,000,000
1945	Hiroshima bomb	63,000,000,000,000
1945	Nagasaki bomb	92,400,000,000,000
1961	Soviet *tsar bomba* tested in 1961	209,000,000,000,000,000

The two nuclear superpowers eventually amassed about 5,000 strategic nuclear warheads (and an arsenal of more than 15,000 other nuclear warheads on shorter-range missiles) with an aggregate destructive energy of about 20 EJ. This was irrational overkill. As Victor Weisskopf (1983, 25) noted, "Nuclear weapons are not weapons of war. The only purpose they can possibly have is to deter their use by the other side, and for that purpose far fewer are good enough." And yet this excess actually served the West well as a mighty deterrent that prevented an obviously unwinnable global thermonuclear war.

But the development of nuclear bombs imposed a significant drain on national treasuries as it required enormous investment and very large amounts of energy, mostly for separating the fissile isotope of uranium (Kesaris 1977; WNA 2015a). Gaseous diffusion required about 9 GJ/SWU (separative work unit), but modern gas centrifuge plants need only 180 MJ/SWU, and with 227 SWU needed to produce a kilogram of weapons-grade uranium, the latter rate works to or about 41 GJ/kg. And the triad of means amassed to deliver nuclear warheads—long-distance bombers, intercontinental ballistic missiles, and nuclear submarines—also consisted of prime movers (jet and rocket engines) and structures whose production and operation were highly energy-intensive.

The production of conventional weapons also requires energy-intensive materials, and their deployment is energized by secondary fossil fuels (gasoline, kerosene, diesel fuel) and the electricity used to power the machines that carry them and to equip and to provision the soldiers who operate them. While ordinary steel could be made from iron ore and pig iron with as little as 20 MJ/kg, specialty steels used in heavy armored equipment require 40–50 MJ/kg, and the use of depleted uranium (for armor-piercing shells and enhanced armor protection) is even more energy-intensive. Aluminum and titanium (and their alloys), the principal materials used to build modern aircraft, embody respectively between 170 and 250 MJ/kg (aluminum) and 450 MJ/kg (titanium), while lighter and stronger composite fibers require typically between 100 and 150 MJ/kg.

Such powerful modern war machines are obviously designed for optimized combat performance, not for minimized energy consumption, and they are extraordinarily energy-intensive. For example, America's 60 t M1/A1 Abrams main battle tank is powered by a 1.1 MW AGT-1500 Honeywell gas turbine and consumes (depending on mission, terrain, and weather) 400–800 L/100 km (*Army Technology* 2015). By comparison, a large Mercedes S600 needs about 15 L/100 km and a Honda Civic sips just 8 L/100 km. And flying at supersonic speeds (up to 1.6–1.8 Mach) such highly maneuverable combat aircraft as the F-16, Lockheed's Fighting Falcon, and F/A-18, the McDonnell Douglas Hornet, needs so much aviation fuel that their extended missions are possible only with in-flight refueling from large tanker planes, such as the KC-10, the KC-135, and the Boeing 767.

Another feature of modern warfare that demands high energy inputs is the use of weapons in massive configurations. The most concentrated tank attack during 1918 involved almost 600 machines (at that time relatively light models), but nearly 8,000 tanks, 11,000 airplanes, and more than 50,000 guns and rocket launchers were deployed by the Red Army during its final assault on Berlin in April 1945 (Ziemke 1968). As an example of the intensity of modern airfare, during the Gulf War (Operation Desert Storm, January–April 1991) and the months preceding it (Operation Desert Shield, August 1990–January 1991), some 1,300 aircraft flew more than 116,000 sorties (Gulflink 1991).

Yet another phenomenon that has greatly contributed to overall energy costs has been is the necessity to ramp up the mass production of military equipment in very short periods of time. The two world wars offer the best examples. In August 1914 Britain had only 154 military airplanes, but four years later the country's aircraft factories were employing about 350,000 people and producing 30,000 airplanes a year (Taylor 1989). When the United States declared war on Germany in April 1917 it had fewer than 300 second-rate planes, none able to carry machine guns or bombs, but three months later the U.S. Congress approved an unprecedented appropriation of $640 million (almost U.S. $12 billion in 2015 monies) to build 22,500

Liberty engines for new fighters (Dempsey 2015). And the American industrial acceleration during World War II was even more impressive.

During the last quarter of 1940 only 514 planes were delivered to the U.S. Army Air Force. In 1941 the total reached 8,723, in 1942 it was 26,448, in 1943 the total surpassed 45,000, and 1944 saw American factories complete 51,547 new planes (Holley 1964). America's aircraft production was the largest manufacturing sector of the wartime economy: it employed two million workers, required nearly a quarter of all wartime expenditures, and produced in total 295,959 airplanes, compared to 117,479 British, 111,784 German, and 68,057 Japanese airplanes (Army Air Forces 1945; Yenne 2006). Ultimately, Allied victories were to the result of their superiority in harnessing destructive energy. By 1944 the United States, the USSR, the UK, and Canada were making three times as much combat munitions as Germany and Japan (Goldsmith 1946). The increasing destructiveness of weapons and the more concentrated delivery of explosives can be illustrated by comparing both discrete events and overall conflict casualties (box 6.17).

Box 6.17
Casualties of modern wars

Combat casualties during the Battle of the Somme (July–November 19, 1916) totaled 1.043 million. Those during the Battle of Stalingrad (August 23, 1942–February 2, 1943) surpassed 2.1 million (Beevor 1998). Battle death rates—expressed as fatalities per 1,000 men of armed forces fielded at the beginning of a conflict—were below 200 during the first two modern wars involving major powers (the Crimean War of 1853–1856 and the Franco-Prussian War of 1870–1871); they surpassed 1,500 during World War I and 2,000 during World War II, and were above 4,000 for Russia (Singer and Small 1972). Germany lost about 27,000 combatants per million people during World War I but more than 44,000 during World War II.

The civilian casualties of modern warfare grew even faster. During World War II they reached about 40 million, more than 70% of the 55 million total casualties. The bombing of large cities produced huge losses within days or just hours (Kloss 1963; Levine 1992). The total German bombing casualties reached nearly 600,000 dead and almost 900,000 wounded. About 100,000 people died during nighttime raids by B-29 bombers, which leveled about 83 km^2 of Japan's four principal cities between March 10 and 20, 1945. The effects of the firebombing of Tokyo and of the nuclear attack on Hiroshima have already been described (see box 6.15).

Calculating the energy cost of major armed conflicts requires important arbitrary delimitations of what should be included in such totals. After all, societies in mortal danger do not operate two separate civilian and military sectors, for wartime economic mobilization affects nearly all activities. Available summations put the total U.S. cost of major twentieth-century conflicts at about $334 billion for World War I, $4.1 trillion for World War II, and $748 billion for the Vietnam War (1964–1972), all expressed in constant 2011 dollars (Daggett 2010). Expressing those costs in current monies and multiplying those totals by adjusted averages of prevailing energy intensities of the country's GDP would amount to defensible approximations of the minimum energy costs of those conflicts.

Adjustments are required because the wartime industrial production and transportation consumed more energy per unit of their output than did the average unit of GDP. As approximations, I chose the respective multiples of 1.5, 2, and 3 for the three conflicts. As a result, participation in World War I required about 15% of the total U.S. energy consumption in 1917 and 1918, and it averaged about 40% during World War II, but it was no more than 4% for the years of the Vietnam War. Peak shares were obviously higher, ranging from 54% for the United States in 1944 to 76% for the USSR in 1942 and a similar share for Germany in 1943.

There is no obvious correlation between overall energy use and success in waging modern acts of aggression (or preventing them). The clearest case for a positive correlation between energy outlay and a fairly swift victory is the U.S. mobilization for World War II, energized by a 46% increase in the total use of primary energy between 1939 and 1944. But in conventional sense America was even more dominant during the Vietnam War—the amount of explosives used was three times as much as all bombs dropped by the U.S. Air Force during World War II on Germany and Japan, and the United States had state-of-the art jet fighters, bombers, helicopters, aircraft carriers, and defoliants—yet it could not, for a variety of political and strategic reasons, translate that dominance into another victory.

And, of course, the absence of any correlation between energies expended and results achieved is most obviously illustrated by terrorist attacks. Completely reversing the Cold War paradigm, in which weapons were extremely expensive to produce and carefully guarded by states, terrorists use weapons that are cheap and widely available. A few hundred kilograms of ANFO (ammonium nitrate/fuel oil) for a truck bomb, a few tens of kilograms for a car bomb, or just a few kilograms of high explosives (often spiked with metal bits) fastened to the bodies of suicide bombers suffice to cause scores or even hundreds of deaths (in 1983 two truck bombs killed 307 people, mostly U.S. servicemen, in their Beirut barracks) and many more injuries, and to terrorize the targeted population.

The 19 hijackers of 9/11 had no weapons other than a few box cutters, and the entire operation, including flight lessons, cost less than $500,000 (bin Laden 2004, 3)—while even the narrowest estimate of the monetary burden (New York City's comptroller report, issued a year after the attack) put the city's direct cost at $95 billion, including about $22 billion to replace the buildings and infrastructure and $17 billion in lost wages (Thompson 2002). A national perspective evaluating lost GDP, the decline of stock values, losses by the airline and tourist industries, higher insurance and shipping rates, and increased security and defense spending put the cost at more than $500 billion (Looney 2002). Adding even a partial cost of the subsequent invasion and occupation of Iraq would raise the total well above a trillion dollars, and as the experience since the attack has shown, there is no easy military solution, as both the classic powerful weapons and the latest smart machines are of a limited use against fanatically motivated individuals or groups willing to die in suicide attacks.

There is no doubt that the mutually assured destruction (MAD) concept has been the main reason why the two nuclear superpowers have not fought a thermonuclear war, but at the same time, the magnitude of the nuclear stockpiles amassed by the two adversaries, and hence their embedded energy cost, has gone far beyond any rationally defensible deterrent level. Every step in developing, deploying, safeguarding, and maintaining nuclear warheads and their carriers (intercontinental bombers and ballistic missiles, nuclear-powered submarines) is energy-intensive. An order-of-magnitude estimate is that at least 5% of all U.S. and Soviet commercial energy that was consumed between 1950 and 1990 was claimed by developing and amassing these weapons and the means of their delivery (Smil 2004).

But even if the burden was twice as high, it might be argued the cost has been acceptable compared to the toll of a thermonuclear exchange that would have, even in the case of a limited exchange, resulted in tens of millions of casualties from the direct effects of blast, fire, and ionizing radiation (Solomon and Marston 1986). A thermonuclear exchange between the United States and the USSR limited to targeting strategic facilities would have caused at least 27 million and up to 59 million deaths during the late 1980s (von Hippel et al. 1988). Such a prospect has acted as a very powerful deterrent from launching, and after the 1960s even seriously contemplating, the first attack.

Unfortunately, the cost attributable to nuclear weapons would not cease even with their instant abolishment: their disarming and expensive safeguarding and the cleanup of contaminated production sites would continue for decades to come, and the estimated U.S. costs of these operations have been rising. It would be even more costly to clean up the more severely contaminated nuclear weapons sites in the countries of the former USSR. Fortunately, the costs of nuclear warhead decommissioning can be much reduced by reusing the recovered fissile material for electricity generation (WNA 2014).

Highly enriched uranium (HEU, containing at least 20% and up to 90% U-235) is blended down with depleted uranium (mostly U-238), natural uranium (0.7% U-235), or partially enriched uranium to produce low-enriched uranium (<5% U-235) used for power reactors. According to a 1993 agreement between the United States and Russia (megatons for megawatts), Russia converted 500 t of HEU from its warheads and strategic stockpiles (equivalent to around 20,000 nuclear bombs) to reactor-ready fuel (averaging about 4.4% U-235) and sold it to power the U.S. civilian reactors.

I cannot leave this section on energy and war without making a few remarks about energy as the casus belli. The belief in this link has been all too common, its latest iteration being the U.S. invasion of Iraq in 2003, done, we are assured, to get at Iraqi oil. And for historians, the most often cited example of the link is the Japanese attack on the United States in December 1941. Roosevelt's administration first abrogated the 1911 Treaty of Commerce and Navigation (in January 1940), then it stopped licensing for exports of aviation gasoline and machine tools (in July 1940) and followed that by a ban on exporting scrap iron and steel (in September 1940). That, according to a still far from abandoned Japanese justification, led the country with little choice but to attack the United States in order to have a free hand to assault Southeast Asia with its Sumatran and Burmese oil fields.

But Pearl Harbor was preceded by nearly a decade of expansive Japanese militarism, beginning with the 1933 conquest of Manchuria and escalating with the 1937 attack on China: Japan could have had continued access to U.S. oil if it had abandoned its aggressive China policy (Ienaga 1978). Not surprisingly, Marius Jansen, one of the leading historians of modern Japan, wrote about the peculiarly self-inflicted nature of the entire confrontation with the United States (Jansen 2000). And who would claim that Hitler's serial aggression—against Czechoslovakia (in 1938 and 1939), Poland (1939), Western Europe (beginning in 1939), and the USSR (1941)—and his genocidal war against Jews were motivated by a quest for energy resources?

Neither were there any energy-related motives for the Korean War (started on Stalin's orders), for the conflict in Vietnam (the French fighting the Communist guerrillas until 1954, the United States between 1964 and 1972), the Soviet occupation of Afghanistan (1979–1989), the U.S. war against the Taliban (launched in October 2001)—or for late twentieth-century cross-border conflicts (China-India, several rounds between India-Pakistani, Eritrea-Ethiopia, and many more) and civil wars (Angola, Uganda, Sri Lanka, Colombia). And while Nigeria's war with the secessionist Biafra (1967–1970) and Sudan's endless civil war (now transformed into the Sudan–South Sudan conflict and tribal warfare within South Sudan) had a clear oil component, both stemmed primarily from religious and ethnic enmities, and the Sudanese conflict began in 1956, decades before any oil discoveries.

Finally, we are left with the two wars in which oil has been widely seen as the real cause. The Iraqi invasion of Kuwait in August 1990 doubled the conventional crude oil reserves under Saddam Hussein's control and threatened the nearby Saudi giant oil fields (Safania, Zuluf, Marjan, and Manifa, on- and offshore just south of Kuwait) and the survival of the monarchy. But there was more at stake than oil, including the Iraqi quest for nuclear and other nonconventional weapons (in 1990 nobody doubted that) and the risks of another Arab-Israeli war (the Iraqi missile attacks on Israel were designed to provoke such a conflict). And if control of oil resources was the primary objective of the 1991 Gulf War, why was the victorious army ordered to stop its uncheckable progress, and why it did not occupy at least Iraq's richest southern oil fields?

What have been the results of the 2003 U.S. invasion of Iraq? American imports of Iraqi oil had actually peaked in 2001 when Saddam Hussein was still in control, at about 41 Mt, after the invasion they kept on declining steadily, and in 2015 they totaled less than 12 Mt, not even 3% of all U.S. imports (USEIA 2016b)—and those, of course, have been diminishing steadily as hydraulic fracturing has made the country once again the world's largest producer of crude oil and natural gas liquids (BP 2016). The verdict is simple: the United States does not need Iraqi oil, East Asia has been its largest buyer—so did the United States go into Iraq to secure Chinese oil supplies? Even the case seen by many as a clear-cut demonstration of energy-driven war is anything but! The conclusion is clear: broader strategic aims, whether well justified or misplaced, and not a quest for resources have led America into its post–World War II conflicts.

Environmental Changes

The provision and the use of fossil fuels and electricity are the largest causes of anthropogenic pollution of the atmosphere and greenhouse gas emissions and are leading contributors to water pollution and land use changes. The combustion of all fossil fuels entails, of course, rapid oxidation of their carbon, which produces increasing emissions of CO_2, while methane (CH_4), a more potent greenhouse gas, is released during the production and transportation of natural gas; small volumes of nitrous oxide (N_2O) are also released from fossil fuel combustion. The combustion of coal used to be a large source of particulate matter and sulfur and nitrogen oxides (SO_x and NO_x), but the stationary emissions of these gases are now largely controlled by electrostatic precipitators, desulfurization, and NO_x-removal processes (Smil 2008a). Even so, emissions from coal combustion continue to have significant health impacts (Lockwood 2012).

Water pollution arises mainly from accidental oil spills (from pipelines, rail cars, barges and tankers, refineries) and acid mine drainage. Major land use changes are caused by surface coal mining, by reservoirs created by major hydroelectric dams, by right-of-way corridors for high-voltage transmission lines, by building extensive storage, refining, and distribution facilities for liquid fuels, and, most recently, by construction of large wind and solar farms. Indirectly, fuels and electricity are responsible for many more pollution flows and ecosystemic degradations. The most notable ones arise from industrial production (above all from ferrous metallurgy and chemical syntheses), agricultural chemicals, urbanization, and transportation. These impacts have been increasing in both extent and intensity and affecting the environment on scales ranging from local to regional. Their costs have been forcing all major economies to devote growing attention to environmental management.

By the 1960s one of these degradations, acid deposition in Central and Western Europe and in eastern North America, created mostly by emissions of SO_x and NO_x from large coal-fired power plants but also by automotive emissions, reached semicontinental scale and until the mid-1980s was widely seen as the most pressing environmental problem facing affluent countries (Smil 1985, 1997). A combination of actions—the switch to low-sulfur coal and to sulfur-free natural gas in electricity generation, the use of cleaner gasoline and diesel and more efficient car engines, and the installation of flue gas desulfurization at major pollution sources—not only arrested the acidification process but by 1990 had reversed it, and precipitation in Europe and North America became less acid (Smil 1997). But the problem has reoccurred since 1990 in East Asia following China's large post-1980 increase in coal combustion.

The partial destruction of the ozone layer above Antarctica and the surrounding ocean briefly assumed the top spot among environmental concerns associated with energy use. The possibility of reduced concentrations of stratospheric ozone protecting the planet from excessive ultraviolet radiation was accurately foreseen in 1974, and the phenomenon was first measured above Antarctica in 1985 (Rowland 1989). Ozone loss has been caused largely by releases of chlorofluorocarbons (CFCs, used mostly as refrigerants), but an effective international treaty, the Montreal Protocol, signed in 1987, and a switch to less harmful compounds soon eased the worries (Andersen and Sarma 2002).

The threat to stratospheric ozone was only the first of several new concerns about the global consequences of environmental change (Turner et al. 1990; McNeill 2001; Freedman 2014). Prominent worries have ranged from the loss of global biodiversity to plastic accumulation in the oceans, but one global environmental concern has been paramount since the late 1980s: anthropogenic emissions of greenhouses gases causing relatively rapid climate change, above all tropospheric warming and ocean acidification and sea-level rise. The behavior of greenhouse gases and their likely warming effect were fairly well understood by the end of the nineteenth century (Smil 1997). The leading anthropogenic contributor is CO_2, the end-product of the efficient combustion of all fossil and biomass fuels, and the destruction of forests (above all in wet tropics) and grasslands has been the second most important source of CO_2 emissions (IPCC 2015).

Since 1850, when it was just 54 Mt C (multiply by 3.667 to convert to CO_2), the global anthropogenic generation of CO_2 has been rising exponentially with the increasing consumption of fossil fuels: as already noted, by 1900 it had risen to 534 Mt C and in 2010 it surpassed 9 Gt C (Boden and Andres 2015). In 1957 Hans Suess and Roger Revelle concluded that

> human beings are now carrying out a large scale geophysical experiment of a kind that could not have happened in the past nor be reproduced in the future. Within a few centuries we are returning to the atmosphere and oceans the concentrated organic carbon stored in sedimentary rocks over hundreds of millions of years. (Revelle and Suess 1957, 19)

The first systematic measurements of rising background CO_2 levels, organized by Charles Keeling (1928–2005), began in 1958 near the summit of Mauna Loa in Hawaii and at the South Pole (Keeling 1998). Mauna Loa concentrations have been used as the global marker of rising tropospheric

CO_2: they averaged almost 316 ppm in 1959, surpassed 350 ppm in 1988, and were 398.55 ppm in 2014 (NOAA 2015; fig. 6.22). Other greenhouse gases are emitted by human activities in much smaller volumes than CO_2, but because their molecules absorb relatively more of the outgoing infrared radiation (methane 86 times as much over 20 years, nitrous oxides 268 times as much as CO_2), their combined contribution now accounts for about 35% of anthropogenic radiative forcing (box 6.18).

The consensus position is that, to avoid the worst consequences of global warming, the average temperature rise should be limited to less than 2°C, but this would require immediate and substantial curtailment of fossil fuel combustion and a rapid transition to noncarbon sources of energy—not an impossible but a highly unlikely development, given the dominance of fossil fuel in the global energy system and the enormous energy requirements of low-income societies: some of those large new needs can come from renewable electricity generation, but there is no affordable, mass-scale alternative available for transportation fuels, feedstocks (ammonia, plastics), or iron ore smelting.

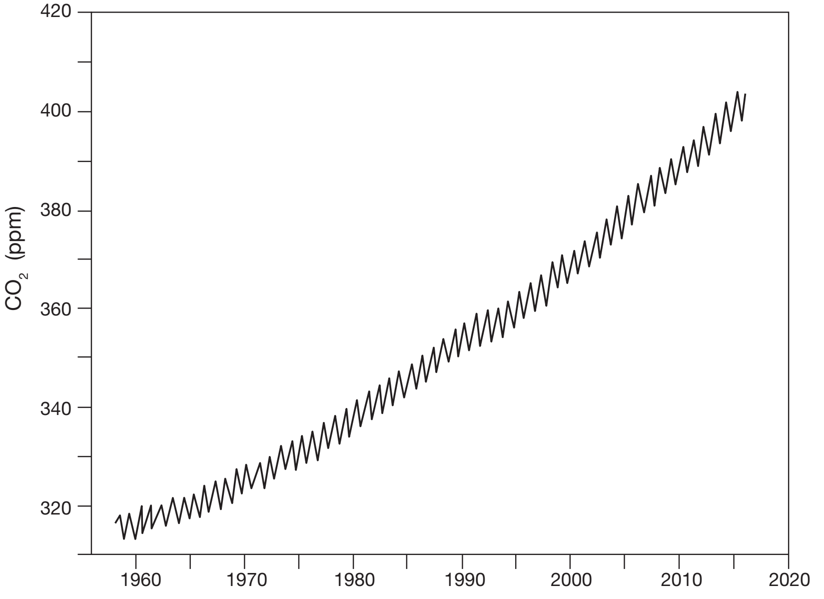


Figure 6.22
Atmospheric CO_2 measured at Mauna Loa Observatory in Hawaii (NOAA 2015).

Box 6.18
Greenhouse gases and rising tropospheric temperature

In 2014 the global rate of anthropogenic radiative forcing (capacity of greenhouse gases to affect the planet's energy balance) reached 2.936 W/m^2, with CO_2 contributing 65% (Butler and Montzka 2015). As for the sources, fossil fuels account for more than 60%, land use changes (mainly deforestation) for 10%, and methane emissions (mainly from livestock) for about 20%. The globally averaged surface temperature increase (combined data for ocean and land) shows a linear rise of 0.85°C (0.65–1.06°C) between 1880 and 2012 (IPCC 2015). Uncertainties regarding the future level of global emissions and the complexity of atmospheric, hydrospheric, and biospheric processes and interactions governing the global carbon cycle make it impossible to construct reliable models forecasting temperature and sea-level rises for the year 2100. The latest consensus assessment shows that (depending largely on future emission rates) by the end of the twenty-first century (2081–2100), the average global temperature will be at least 0.3–1.7°C higher than during 1986–2005, but that it may rise by as much as 2.6–4.8°C (IPCC 2015).

In any case, the Arctic region will continue warming more rapidly. Obviously, the lower rates would make it easier to adapt, while the highest likely increases would pose many serious problems. The multitude of changes attributable to global warming range from new precipitation patterns, coastal flooding, and shifts in ecosystem boundaries to the spread of warm-climate, vector-borne diseases. Changes in plant productivity, loss of near-shore real estate, sectoral unemployment, and large-scale migration from affected regions would be the key economic consequences. There is no easy technical fix (such as CO_2 capture from the air or CO_2 storage underground, both required to handle, at an affordable cost, more than 10 Gt CO_2/year in order to be effective) for anthropogenic greenhouse gas emissions. The only potentially successful approach to deal with these changes is through unprecedented international cooperation. Unintentionally, this worrisome challenge also offers a fundamental motivation for a new departure in managing human affairs.